Geometry of Quantum States

An Introduction to Quantum Entanglement

Ingemar Bengtsson and
Karol Życzkowski

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521814515

GEOMETRY OF QUANTUM STATES

An Introduction to Quantum Entanglement

GEOMETRY OF QUANTUM STATES

An Introduction to Quantum Entanglement

INGEMAR BENGTSSON AND KAROL ŻYCZKOWSKI

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521814515

First published in print format 2006

ISBN-13 978-0-511-19174-9 eBook (NetLibrary)
ISBN-10 0-511-19174-x eBook (NetLibrary)

ISBN-13 978-0-521-81451-5 hardback
ISBN-10 0-521-81451-0 hardback

ISBN-13 978-0-521-89140-0
ISBN-10 0-521-89140-x

Contents

Preface

The geometry of quantum states is a highly interesting subject in itself, but it is also relevant in view of possible applications in the rapidly developing fields of quantum information and quantum computing.

But what is it? In physics words like 'states' and 'system' are often used. Skipping lightly past the question of what these words mean – it will be made clear by practice – it is natural to ask for the properties of the space of all possible states of a given system. The simplest state space occurs in computer science: a 'bit' has a space of states that consists simply of two points, representing on and off. In probability theory the state space of a bit is really a line segment, since the bit may be 'on' with some probability between zero and one. In general the state spaces used in probability theory are 'convex hulls' of a discrete or continuous set of points. The geometry of these simple state spaces is surprisingly subtle – especially since different ways of distinguishing probability distributions give rise to different notions of distance, each with their own distinct operational meaning. There is an old idea saying that a geometry can be understood once it is understood what linear transformations are acting on it, and we will see that this is true here as well.

The state spaces of classical mechanics are – at least from the point of view that we adopt – just instances of the state spaces of classical probability theory, with the added requirement that the sample spaces (whose 'convex hull' we study) are large enough, and structured enough, so that the transformations acting on them include canonical transformations generated by Hamiltonian functions.

In quantum theory the distinction between probability theory and mechanics goes away. The simplest quantum state space is these days known as a 'qubit'. There are many physical realizations of a qubit, from silver atoms of spin $1/2$ (assuming that we agree to measure only their spin) to the qubits that are literally designed in today's laboratories. As a state space a qubit is a three-dimensional ball; each diameter of the ball is the state space of some classical bit, and there are so many bits that their sample spaces conspire to form a space – namely the surface of the

ball – large enough to carry the canonical transformations that are characteristic of mechanics. Hence the word quantum mechanics.

It is not particularly difficult to understand a three-dimensional ball, or to see how this description emerges from the usual description of a qubit in terms of a complex two-dimensional Hilbert space. In this case we can take the word *geometry* literally: there will exist a one-to-one correspondence between pure states of the qubit and the points of the surface of the Earth. Moreover, at least as far as the surface is concerned, its geometry has a statistical meaning when transcribed to the qubit (although we will see some strange things happening in the interior).

As the dimension of the Hilbert space goes up, the geometry of the state spaces becomes very intricate, and qualitatively new features arise – such as the subtle way in which composite quantum systems are represented. Our purpose is to describe this geometry. We believe it is worth doing. Quantum state spaces are more wonderful than classical state spaces, and in the end composite systems of qubits may turn out to have more practical applications than the bits themselves already have.

A few words about the contents of our book. As a glance at the table of contents will show, there are 15 chapters, culminating in a long chapter on 'entanglement'. Along the way, we take our time to explore many curious byways of geometry. We expect that you – the reader – are familiar with the principles of quantum mechanics at the advanced undergraduate level. We do not really expect more than that, and should you be unfamiliar with quantum mechanics we hope that you will find some sections of the book profitable anyway. You may start reading any chapter: if you find it incomprehensible we hope that the cross-references and the index will enable you to see what parts of the earlier chapters may be helpful to you. In the unlikely event that you are not even interested in quantum mechanics, you may perhaps enjoy our explanations of some of the geometrical ideas that we come across.

Of course there are limits to how independent the chapters can be of each other. Convex set theory (Chapter 1) pervades all statistical theories, and hence all our chapters. The ideas behind the classical Shannon entropy and the Fisher–Rao geometry (Chapter 2) must be brought in to explain quantum mechanical entropies (Chapter 12) and quantum statistical geometry (Chapters 9 and 13). Sometimes we have to assume a little extra knowledge on the part of the reader, but since no chapter in our book assumes that all the previous chapters have been understood, this should not pose any great difficulties.

We have made a special effort to illustrate the geometry of quantum mechanics. This is not always easy, since the spaces that we encounter more often than not have a dimension higher than three. We have simply done the best we could. To facilitate self-study each chapter concludes with problems for the reader, while some additional geometrical exercises are presented in Appendix 3.

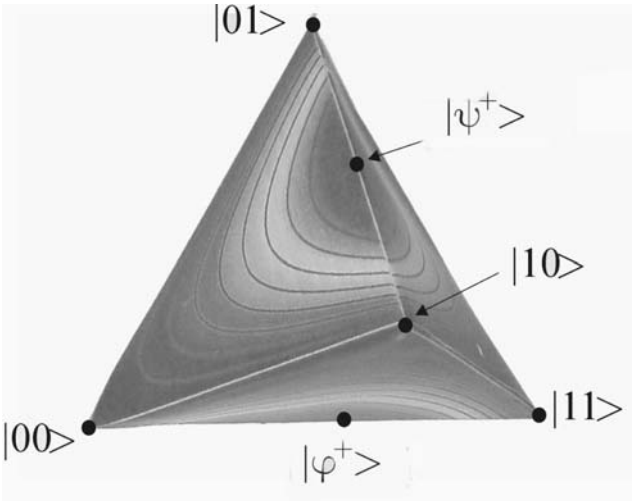


Figure 0.1. Black and white version of the cover picture which shows the entropy of entanglement for a 3-D cross section of the 6-D manifold of pure states of two qubits. The hotter the colour, the more entangled the state. For more information study Sections 15.2 and 15.3 and look at Figures 15.1 and 15.2.

We limit ourselves to finite-dimensional state spaces. We do this for two reasons. One of them is that it simplifies the story very much, and the other is that finite-dimensional systems are of great independent interest in real experiments.

The entire book may be considered as an introduction to quantum entanglement. This very non-classical feature provides a key resource for several modern applications of quantum mechanics including quantum cryptography, quantum computing and quantum communication. We hope that our book may be useful for graduate and postgraduate students of physics. It is written first of all for readers who do not read the mathematical literature everyday, but we hope that students of mathematics and of the information sciences will find it useful as well, since they also may wish to learn about quantum entanglement.

We have been working on the book for about five years. Throughout this time we enjoyed the support of Stockholm University, the Jagiellonian University in Kraków, and the Center for Theoretical Physics of the Polish Academy of Sciences in Warsaw. The book was completed at Waterloo during our stay at the Perimeter Institute for Theoretical Physics. The motto at its main entrance – ΑΣΠΟΥΔΑΣΤΟΣ ΠΕΡΙ ΓΕΩΜΕΤΡΙΑΣ ΜΗΔΕΙΣ ΕΙΣΙΤΩ[1] – proved to be a lucky omen indeed, and we are pleased to thank the Institute for creating optimal working conditions

[1] Let no one uninterested in geometry enter here.

for us, and to thank all the knowledgable colleagues working there for their help, advice and support. We also thank the *International Journal of Modern Physics A* for permission to reproduce a number of figures.

We are grateful to Erik Aurell for his commitment to Polish–Swedish collaboration; without him the book would never have been started. It is a pleasure to thank our colleagues with whom we have worked on related projects: Johan Brännlund, Åsa Ericsson, Sven Gnutzmann, Marek Kuś, Florian Mintert, Magdalena Sinołęcka, Hans-Jürgen Sommers and Wojciech Słomczyński. We are grateful to them and to many others who helped us to improve the manuscript. If it never reached perfection, it was our fault, not theirs. Let us mention some of the others: Robert Alicki, Anders Bengtsson, Iwo Białynicki-Birula, Rafał Demkowicz-Dobrzański, Johan Grundberg, Sören Holst, Göran Lindblad and Marcin Musz. We have also enjoyed constructive interactions with Matthias Christandl, Jens Eisert, Peter Harremoës, Michał, Paweł and Ryszard Horodeccy, Vivien Kendon, Artur Łoziński, Christian Schaffner, Paul Slater and William Wootters.

Five other people provided indispensable support: Martha and Jonas in Stockholm, and Jolanta, Jaś and Marysia in Kraków.

Ingemar Bengtsson Karol Życzkowski
Waterloo
12 March 2005

1

Convexity, colours and statistics

> What picture does one see, looking at a physical theory from a distance, so that the details disappear? Since quantum mechanics is a statistical theory, the most universal picture which remains after the details are forgotten is that of a convex set.
>
> *Bogdan Mielnik*

1.1 Convex sets

Our object is to understand the geometry of the set of all possible states of a quantum system that can occur in nature. This is a very general question, especially since we are not trying to define 'state' or 'system' very precisely. Indeed we will not even discuss whether the state is a property of a thing, or of the preparation of a thing, or of a belief about a thing. Nevertheless we can ask what kind of restrictions are needed on a set if it is going to serve as a space of states in the first place. There is a restriction that arises naturally both in quantum mechanics and in classical statistics: the set must be a *convex set*. The idea is that a convex set is a set such that one can form 'mixtures' of any pair of points in the set. This is, as we will see, how probability enters (although we are not trying to define 'probability' either).

From a geometrical point of view a *mixture* of two states can be defined as a point on the segment of the straight line between the two points that represent the states that we want to mix. We insist that given two points belonging to the set of states, the straight line segment between them must belong to the set too. This is certainly not true of any set. But before we can see how this idea restricts the set of states we must have a definition of 'straight lines' available. One way to proceed is to regard a convex set as a special kind of subset of a flat Euclidean space $\mathbf{E}^n$. Actually we can get by with somewhat less. It is enough to regard a convex set as a subset of an affine space. An *affine space* is just like a vector space, except that no special choice of origin is assumed. The *straight line* through the two points $\mathbf{x}_1$

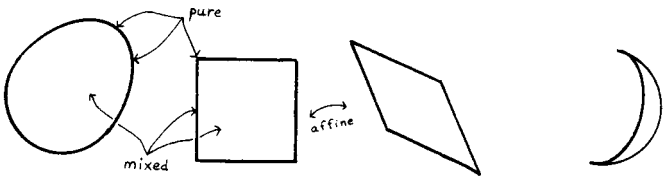


Figure 1.1. Three convex sets in two dimensions, two of which are affine transformations of each other. The new moon is not convex. An observer in Singapore will find the new moon tilted but still not convex, since convexity is preserved by rotations.

and $\mathbf{x}_2$ is defined as the set of points

$$\mathbf{x} = \mu_1 \mathbf{x}_1 + \mu_2 \mathbf{x}_2 \,, \qquad \mu_1 + \mu_2 = 1 \,. \tag{1.1}$$

If we choose a particular point $\mathbf{x}_0$ to serve as the origin, we see that this is a one parameter family of vectors $\mathbf{x} - \mathbf{x}_0$ in the plane spanned by the vectors $\mathbf{x}_1 - \mathbf{x}_0$ and $\mathbf{x}_2 - \mathbf{x}_0$. Taking three different points instead of two in Eq. (1.1) we define a *plane*, provided the three points do not belong to a single line. A k-dimensional plane is obtained by taking $k + 1$ generic points, where $k < n$. An $(n - 1)$-dimensional plane is known as a *hyperplane*. For $k = n$ we describe the entire space $\mathbf{E}^n$. In this way we may introduce *barycentric coordinates* into an n-dimensional affine space. We select $n + 1$ points $\mathbf{x}_i$, so that an arbitrary point $\mathbf{x}$ can be written as

$$\mathbf{x} = \mu_0 \mathbf{x}_0 + \mu_1 \mathbf{x}_1 + \cdots + \mu_n \mathbf{x}_n \,, \qquad \mu_0 + \mu_1 + \cdots + \mu_n = 1 \,. \tag{1.2}$$

The requirement that the barycentric coordinates μ_i add up to one ensures that they are uniquely defined by the point $\mathbf{x}$. (It also means that the barycentric coordinates are not coordinates in the ordinary sense of the word, but if we solve for μ_0 in terms of the others then the remaining independent set is a set of n ordinary coordinates for the n-dimensional space.) An *affine map* is a transformation that takes lines to lines and preserves the relative length of line segments lying on parallel lines. In equations an affine map is a combination of a linear transformation described by a matrix $\mathbf{A}$ with a translation along a constant vector $\mathbf{b}$, so $\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{A}$ is an invertible matrix.

By definition a subset S of an affine space is a *convex set* if for any pair of points $\mathbf{x}_1$ and $\mathbf{x}_2$ belonging to the set it is true that the *mixture* $\mathbf{x}$ also belongs to the set, where

$$\mathbf{x} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 \,, \qquad \lambda_1 + \lambda_2 = 1 \,, \qquad \lambda_1,\ \lambda_2 \geq 0 \,. \tag{1.3}$$

Here λ_1 and λ_2 are barycentric coordinates on the line through the given pair of points; the extra requirement that they be positive restricts $\mathbf{x}$ to belong to the segment of the line lying between the pair of points.

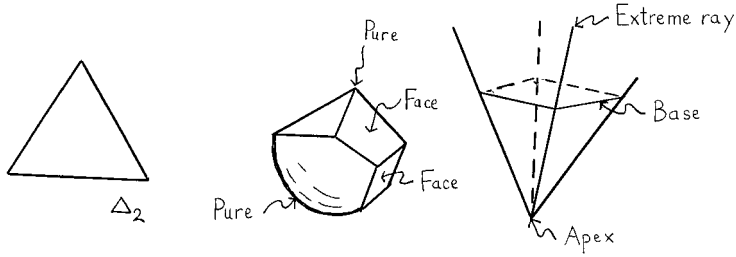


Figure 1.2. The convex sets we will consider are either convex bodies (like the simplex on the left or the more involved example in the centre) or convex cones with compact bases (an example is shown on the right).

It is natural to use an affine space as the 'container' for the convex sets since convexity properties are preserved by general affine transformations. On the other hand it does no harm to introduce a flat metric on the affine space, turning it into an Euclidean space. There may be no special significance attached to this notion of distance, but it helps in visualizing what is going on. From now on, we will assume that our convex sets sit in Euclidean space, whenever it is convenient to do so.

Intuitively a convex set is a set such that one can always see the entire set from whatever point in the set one happens to be sitting at. They can come in a variety of interesting shapes. We will need a few definitions. First, given any subset of the affine space we define the *convex hull* of this subset as the smallest convex set that contains the set. The convex hull of a finite set of points is called a *convex polytope*. If we start with $p+1$ points that are not confined to any $(p-1)$-dimensional subspace then the convex polytope is called a *p-simplex*. The p-simplex consists of all points of the form

$$\mathbf{x} = \lambda_0\mathbf{x}_0 + \lambda_1\mathbf{x}_1 + \cdots + \lambda_p\mathbf{x}_p\,, \quad \lambda_0 + \lambda_1 + \cdots + \lambda_p = 1\,, \quad \lambda_i \geq 0\,. \tag{1.4}$$

(The barycentric coordinates are all non-negative.) The *dimension* of a convex set is the largest number n such that the set contains an n-simplex. In discussing a convex set of dimension n we usually assume that the underlying affine space also has dimension n, to ensure that the convex set possesses interior points (in the sense of point set topology). A closed and bounded convex set that has an interior is known as a *convex body*.

The intersection of a convex set with some lower dimensional subspace of the affine space is again a convex set. Given an n-dimensional convex set S there is also a natural way to increase its dimension by one: choose a point $\mathbf{y}$ not belonging to the n-dimensional affine subspace containing S. Form the union of all the *rays* (in this chapter a ray means a half line), starting from $\mathbf{y}$ and passing through S. The result is called a *convex cone* and $\mathbf{y}$ is called its *apex*, while S is its *base*. A ray is

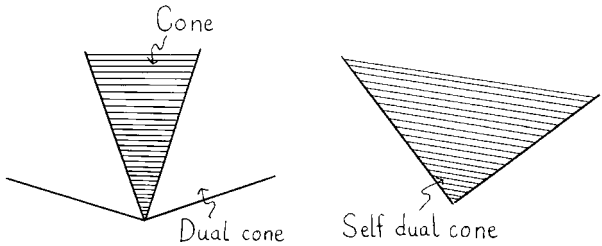


Figure 1.3. Left: a convex cone and its dual, both regarded as belonging to Euclidean 2-space. Right: a self dual cone, for which the dual cone coincides with the original. For an application of this construction see Figure 11.6.

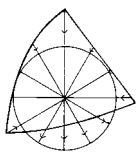


Figure 1.4. A convex body is homeomorphic to a sphere.

in fact a one-dimensional convex cone. A more interesting example is obtained by first choosing a p-simplex and then interpreting the points of the simplex as vectors starting from an origin O not lying in the simplex. Then the $(p+1)$-dimensional set of points

$$\mathbf{x} = \lambda_0 \mathbf{x}_0 + \lambda_1 \mathbf{x}_1 + \cdots + \lambda_p \mathbf{x}_p \,, \qquad \lambda_i \geq 0 \tag{1.5}$$

is a convex cone. Convex cones have many appealing properties, including an inbuilt partial order among its points: $\mathbf{x} \leq \mathbf{y}$ if and only if $\mathbf{y} - \mathbf{x}$ belongs to the cone. Linear maps to $\mathbb{R}$ that take positive values on vectors belonging to a convex cone form a dual convex cone in the dual vector space. Since we are in the Euclidean vector space $\mathbf{E}^n$, we can identify the dual vector space with $\mathbf{E}^n$ itself. If the two cones agree the convex cone is said to be *self dual*. One self dual convex cone that will appear now and again is the *positive orthant* or *hyperoctant* of $\mathbf{E}^n$, defined as the set of all points whose Cartesian coordinates are non-negative. We use the notation $\mathbf{x} \geq 0$ to denote the fact that $\mathbf{x}$ belongs to the positive orthant.

From a purely topological point of view all convex bodies are equivalent to an n-dimensional ball. To see this choose any point $\mathbf{x}_0$ in the interior and then for every point in the boundary draw a ray starting from $\mathbf{x}_0$ and passing through the boundary point (as in Figure 1.4). It is clear that we can make a continuous transformation of the convex body into a ball with radius one and its centre at $\mathbf{x}_0$ by moving the points of the container space along the rays.

Convex bodies and convex cones with compact bases are the only convex sets that we will consider. Convex bodies always contain some special points that cannot be obtained as mixtures of other points: whereas a half space does not! These points are called *extreme points* by mathematicians and *pure* points by physicists (actually, originally by Weyl), while non-pure points are called *mixed*. In a convex cone the rays from the apex through the pure points of the base are called *extreme rays*; a point $\mathbf{x}$ lies on an extreme ray if and only if $\mathbf{y} \leq \mathbf{x} \Rightarrow \mathbf{y} = \lambda\mathbf{x}$ with λ between zero and one. A subset F of a convex set that is stable under mixing and purification is called a *face* of the convex set. This phrase means that if

$$\mathbf{x} = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \,, \qquad 0 \leq \lambda \leq 1 \tag{1.6}$$

then $\mathbf{x}$ lies in F if and only if $\mathbf{x}_1$ and $\mathbf{x}_2$ lie in F. A face of dimension k is a k-face. A 0-face is an extremal point, and an $(n-1)$-face is also known as a *facet*. It is interesting to observe that the set of all faces on a convex body form a partially ordered set; we say that $F_1 \leq F_2$ if the face F_1 is contained in the face F_2. It is a partially ordered set of the special kind known as a *lattice*, which means that a given pair of faces always has a greatest lower bound (perhaps the empty set) and a lowest greater bound (perhaps the convex body itself).

To stem the tide of definitions let us quote two theorems that have an 'obvious' ring to them when they are stated abstractly but which are surprisingly useful in practice:

Theorem 1.1 (Minkowski's) *Any convex body is the convex hull of its pure points.*

Theorem 1.2 (Carathéodory's) *If X is a subset of $\mathbb{R}^n$ then any point in the convex hull of X can be expressed as a convex combination of at most $n+1$ points in X.*

Thus any point $\mathbf{x}$ of a convex body S may be expressed as a *convex combination* of pure points:

$$\mathbf{x} = \sum_{i=1}^{p} \lambda_i \mathbf{x}_i, \qquad \lambda_i \geq 0 \,, \qquad p \leq n+1 \,, \qquad \sum_i \lambda_i = 1 \,. \tag{1.7}$$

This equation is quite different from Eq. (1.2) that defined the barycentric coordinates of $\mathbf{x}$ in terms of a fixed set of points $\mathbf{x}_i$, because – with the restriction that all the coefficients be non-negative – it may be impossible to find a finite set of $\mathbf{x}_i$ so that every $\mathbf{x}$ in the set can be written in this new form. An obvious example is a circular disc. Given $\mathbf{x}$ one can always find a finite set of pure points $\mathbf{x}_i$ so that the equation holds, but that is a different thing.

It is evident that the pure points always lie in the boundary of the convex set, but the boundary often contains mixed points as well. The simplex enjoys a very special property, which is that any point in the simplex can be written as a mixture of pure

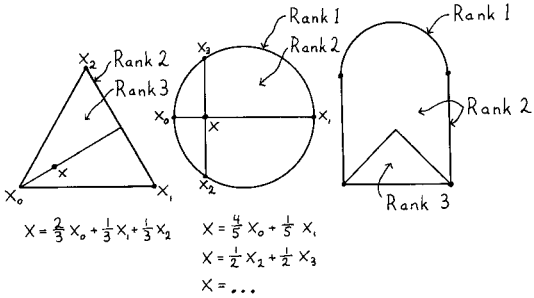


Figure 1.5. In a simplex a point can be written as a mixture in one and only one way. In general the rank of a point is the minimal number of pure points needed in the mixture; the rank may change in the interior of the set as shown in the rightmost example.

points in one and only one way (as in Figure 1.5). This is because for the simplex the coefficients in Eq. (1.7) are barycentric coordinates and the result follows from the uniqueness of the barycentric coordinates of a point. No other convex set has this property. The *rank* of a point $\mathbf{x}$ is the minimal number p needed in the convex combination (Eq. (1.7)). By definition the pure points have rank one. In a simplex the edges have rank two, the faces have rank three, and so on, while all the points in the interior have maximal rank. From Eq. (1.7) we see that the maximal rank of any point in a convex body in $\mathbb{R}^n$ does not exceed $n + 1$. In a ball all interior points have rank two and all points on the boundary are pure, regardless of the dimension of the ball. It is not hard to find examples of convex sets where the rank changes as we move around in the interior of the set (see Figure 1.5).

The simplex has another quite special property, namely that its lattice of faces is *self dual*. We observe that the number of k-faces in an n-dimensional simplex is

$$\binom{n+1}{k+1} = \binom{n+1}{n-k}. \tag{1.8}$$

Hence the set of $(n - k - 1)$-dimensional faces can be put in one-to-one correspondence with the set of k-faces. In particular, the pure points ($k = 0$) can be put in one-to-one correspondence with the set of facets (by definition, the $(n - 1)$-dimensional faces). For this, and other, reasons its lattice of subspaces will have some exceptional properties, turning it into what is technically known as a *Boolean* lattice.[1]

There is a useful dual description of convex sets in terms of supporting hyperplanes. A *support hyperplane* of S is a hyperplane that intersects the set and is such

[1] Because it is related to what George Boole thought were the laws of thought; see Varadarajan's book on quantum logic (Varadarajan, 1985).

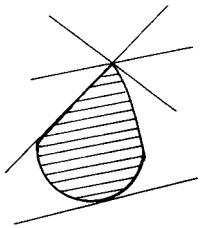


Figure 1.6. Support hyperplanes of a convex set.

that the entire set lies in one of the closed half spaces formed by the hyperplane (see Figure 1.6). Hence a support hyperplane just touches the boundary of S, and one can prove that there is a support hyperplane passing through every point of the boundary of a convex body. By definition a *regular point* is a point on the boundary that lies on only one support hyperplane, a *regular support hyperplane* meets the set at only one point, and the entire convex set is regular if all its boundary points as well as all its support hyperplanes are regular. So a ball is regular, while a convex polytope or a convex cone is not – indeed all the support hyperplanes of a convex cone pass through its apex. Convex polytopes arise as the intersection of a finite number of closed half spaces in $\mathbb{R}^n$, and any pure point of a convex polytope saturates n of the inequalities that define the half spaces; again a statement with an 'obvious' ring that is useful in practice.

In a flat Euclidean space a linear function to the real numbers takes the form $\mathbf{x} \to \mathbf{a} \cdot \mathbf{x}$, where $\mathbf{a}$ is some constant vector. Geometrically, this defines a family of parallel hyperplanes. The following theorem is important:

Theorem 1.3 (Hahn–Banach separation) *Given a convex body and a point* $\mathbf{x_0}$ *that does not belong to it, then one can find a linear function* f *that takes positive values for all points belonging to the convex body, while* $f(\mathbf{x_0}) < 0$.

This is again almost obvious if one thinks in terms of hyperplanes.[2]

We will find much use for the concept of *convex functions*. A real function $f(\mathbf{x})$ defined on a closed convex subset X of $\mathbb{R}^n$ is called *convex*, if for any $\mathbf{x}, \mathbf{y} \in X$ and $\lambda \in [0, 1]$ it satisfies

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) \,. \tag{1.9}$$

The name refers to the fact that the *epigraph* of a convex function, that is the region lying above the curve $f(\mathbf{x})$ in the graph, is convex. Applying the inequality $k - 1$

[2] To readers who wish to learn more about convex sets – or who wish to see proofs of the various assertions that we have left unproved – we recommend Eggleston (1958).

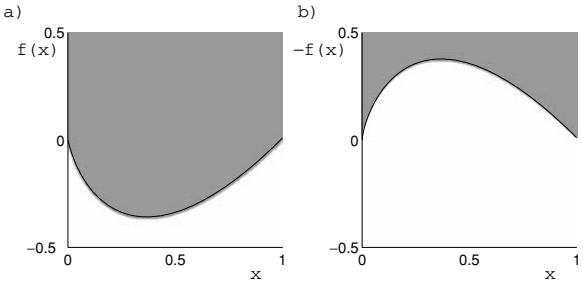


Figure 1.7. (a) the convex function $f(x) = x \ln x$, and (b) the concave function $g(x) = -x \ln x$. The names stem from the shaded epigraphs of the functions which are convex and concave, respectively.

times we see that

$$f\Big(\sum_{j=1}^{k} \lambda_j \mathbf{x}_j\Big) \;\le\; \sum_{j=1}^{k} \lambda_j f(\mathbf{x}_j), \tag{1.10}$$

where $\mathbf{x}_j \in X$ and the non-negative weights sum to unity, $\sum_{j=1}^{k} \lambda_j = 1$. If a function f from $\mathbb{R}$ to $\mathbb{R}$ is differentiable, it is convex if and only if

$$f(y) - f(x) \geq (y - x) f'(x)\,. \tag{1.11}$$

If f is twice differentiable, it is convex if and only if its second derivative is non-negative. For a function of several variables to be convex, the matrix of second derivatives must be positive definite. In practice, this is a very useful criterion. A function f is called *concave* if $-f$ is convex.

One of the main advantages of convex functions is that it is (comparatively) easy to study their minima and maxima. A minimum of a convex function is always a global minimum, and it is attained on some convex subset of the domain of definition X. If X is not only convex but also compact, then the global maximum sits at an extreme point of X.

1.2 High-dimensional geometry

In quantum mechanics the spaces we encounter are often of very high dimension; even if the dimension of Hilbert space is small, the dimension of the space of density matrices will be high. Our intuition, on the other hand, is based on two- and three-dimensional spaces, and frequently leads us astray. We can improve ourselves by asking some simple questions about convex bodies in flat space. We choose to look at balls, cubes and simplices for this purpose. A flat metric is assumed. Our

questions will concern the *inspheres* and *outspheres* of these bodies (defined as the largest inscribed sphere and the smallest circumscribed sphere, respectively). For any convex body the outsphere is uniquely defined, while the insphere is not – one can show that the upper bound on the radius of inscribed spheres is always attained by some sphere, but there may be several of those.

Let us begin with the surface of a ball, namely the n-dimensional *sphere* $\mathbf{S}^n$. In equations, a sphere of radius r is given by the set

$$X_0^2 + X_1^2 + \cdots + X_n^2 = r^2 \tag{1.12}$$

in an $(n+1)$-dimensional flat space $\mathbf{E}^{n+1}$. A sphere of radius one is denoted $\mathbf{S}^n$. The sphere can be parametrized by the angles $\phi, \theta_1, \ldots, \theta_{n-1}$ according to

$$\begin{cases} X_0 = r\cos\phi \sin\theta_1 \sin\theta_2 \ \ldots \ \sin\theta_{n-1} \\ X_1 = r\sin\phi \sin\theta_1 \sin\theta_2 \ \ldots \ \sin\theta_{n-1} \\ X_2 = \quad r\cos\theta_1 \sin\theta_2 \ \ldots \ \sin\theta_{n-1} \\ \ \vdots \quad \vdots \qquad\qquad \vdots \\ X_n = \qquad r\cos\theta_{n-1} \end{cases} \quad \begin{array}{l} 0 < \theta_i < \pi \\ 0 \le \phi < 2\pi \end{array} . \tag{1.13}$$

The volume element $\mathrm{d}A$ on the unit sphere then becomes

$$\mathrm{d}A = \mathrm{d}\phi \mathrm{d}\theta_1 \ \ldots \ \mathrm{d}\theta_{n-1} \sin\theta_1 \sin^2\theta_2 \ \ldots \ \sin^{n-1}\theta_{n-1} . \tag{1.14}$$

We want to compute the volume $\mathrm{vol}(\mathbf{S}^n)$ of the n-sphere, that is to say its 'hyperarea' – meaning that $\mathrm{vol}(\mathbf{S}^2)$ is measured in square metres, $\mathrm{vol}(\mathbf{S}^3)$ in cubic metres, and so on. A clever trick simplifies the calculation: consider the well-known Gaussian integral

$$I = \int \mathrm{e}^{-X_0^2 - X_1^2 - \cdots - X_n^2} \, \mathrm{d}X_0 \mathrm{d}X_1 \ \ldots \ \mathrm{d}X_n = (\sqrt{\pi})^{n+1} . \tag{1.15}$$

Using the spherical polar coordinates introduced above our integral splits into two, one of which is related to the integral representation of the Euler Gamma function, $\Gamma(x) = \int_0^\infty \mathrm{e}^{-t} t^{x-1} \, \mathrm{d}t$, and the other is the one we want to do:

$$I = \int_0^\infty \mathrm{d}r \int_{\mathbf{S}^n} \mathrm{d}A \mathrm{e}^{-r^2} r^n = \frac{1}{2}\Gamma\left(\frac{n+1}{2}\right) \mathrm{vol}(\mathbf{S}^n) . \tag{1.16}$$

We do not have to do the integral over the angles. We simply compare these results and obtain (recalling the properties of the Gamma function)

$$\mathrm{vol}(\mathbf{S}^n) = 2\frac{\pi^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})} = \begin{cases} \frac{2(2\pi)^p}{(2p-1)!!} & \text{if } n = 2p \\ \\ \frac{(2\pi)^{p+1}}{(2p)!!} & \text{if } n = 2p+1 \end{cases} , \tag{1.17}$$

where double factorial is the product of every other number, $5!! = 5 \cdot 3 \cdot 1$ and $6!! = 6 \cdot 4 \cdot 2$. An alarming thing happens as the dimension grows. For large x we can approximate the Gamma function using Stirling's formula

$$\Gamma(x) = \sqrt{2\pi}\mathrm{e}^{-x}x^{x-\frac{1}{2}}\left(1 + \frac{1}{12x} + o(\frac{1}{x^2})\right) . \tag{1.18}$$

Hence for large n we obtain

$$\mathrm{vol}(\mathbf{S}^n) \sim \sqrt{2}\left(\frac{2\pi\mathrm{e}}{n}\right)^{\frac{n}{2}} . \tag{1.19}$$

This is small if n is large! In fact the 'biggest' unit sphere – in the sense that it has the largest hyperarea – is $\mathbf{S}^6$, which has

$$\mathrm{vol}(\mathbf{S}^6) = \frac{16}{15}\pi^3 \approx 33.1 . \tag{1.20}$$

Incidentally Stirling's formula gives 31.6, which is already rather good. We hasten to add that $\mathrm{vol}(\mathbf{S}^2)$ is measured in square metres and $\mathrm{vol}(\mathbf{S}^6)$ in $(\mathrm{metre})^6$, so that the direct comparison makes no sense.

There is another funny thing to be noticed. If we compute the volume of the n-sphere without any clever tricks, simply by integrating the volume element $\mathrm{d}A$ using angular coordinates, then we find that

$$\begin{aligned}\mathrm{vol}(\mathbf{S}^n) &= 2\pi \int_0^\pi \mathrm{d}\theta \sin\theta \int_0^\pi \mathrm{d}\theta \sin^2\theta \ \ldots \ \int_0^\pi \mathrm{d}\theta \sin^{n-1}\theta \\ &= \mathrm{vol}(\mathbf{S}^{n-1}) \int_0^\pi \mathrm{d}\theta \sin^{n-1}\theta .\end{aligned} \tag{1.21}$$

As n grows the integrand of the final integral has an increasingly sharp peak close to the equator $\theta = \pi/2$. Hence we conclude that when n is high most of the hyperarea of the sphere is confined to a 'band' close to the equator. What about the volume of an n-dimensional unit ball $\mathbf{B}^n$? By definition it has unit radius and its boundary is $\mathbf{S}^{n-1}$. Its volume, using the radial integral $\int_0^1 r^{n-1}\,\mathrm{d}r = 1/n$ and the fact that $\Gamma(x+1) = x\Gamma(x)$, is

$$\mathrm{vol}(\mathbf{B}^n) = \frac{\mathrm{vol}(\mathbf{S}^{n-1})}{n} = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \sim \frac{1}{\sqrt{2\pi}}\left(\frac{2\pi\mathrm{e}}{n}\right)^{\frac{n}{2}} . \tag{1.22}$$

Again, as the dimension grows the denominator grows faster than the numerator and therefore the volume of a unit ball is small when the dimension is high. We can turn this around if we like: a ball of unit volume has a large radius if the dimension is high. Indeed since the volume is proportional to r^n, where r is the radius, it follows

that the radius of a ball of unit volume grows like $\sqrt{n}$ when Stirling's formula applies.

The fraction of the volume of a unit ball that lies inside a radius r is r^n. We assume $r < 1$, so this is a quickly shrinking fraction as n grows. The curious conclusion of this is that when the dimension is high almost all of the volume of a ball lies very close to its surface. In fact this is a crucial observation in statistical mechanics. It is also the key property of n-dimensional geometry: when n is large the 'amount of space' available grows very fast with distance from the origin.

In some ways it is easier to see what is going on if we consider hypercubes $\boxtimes_n$ rather than balls. Take a cube of unit volume. In n dimensions it has 2^n corners, and the longest straight line that we can draw inside the hypercube connects two opposite corners. It has length $L = \sqrt{1^2 + \cdots + 1^2} = \sqrt{n}$. Or expressed in another way, a straight line of any length fits into a hypercube of unit volume if the dimension is large enough. The reason why the longest line segment fitting into the cube is large is clearly that we normalized the volume to one. If we normalize $L = 1$ instead we find that the volume goes to zero like $(1/\sqrt{n})^n$. Concerning the insphere (the largest inscribed sphere, with *inradius* r_n) and the outsphere (the smallest circumscribed sphere, with *outradius* R_n), we observe that

$$R_n = \frac{\sqrt{n}}{2} = \sqrt{n} r_n \ . \tag{1.23}$$

The ratio between the two grows with the dimension, $\zeta_n \equiv R_n/r_n = \sqrt{n}$. Incidentally, the somewhat odd statement that the volume of a sphere goes to zero when the dimension n goes to infinity can now be interpreted: since $\mathrm{vol}(\boxtimes_n) = 1$ the real statement is that $\mathrm{vol}(\mathbf{S}^n)/\mathrm{vol}(\boxtimes_n)$ goes to zero when n goes to infinity.

Now we turn to simplices, whose properties will be of some importance later on. We concentrate on *regular simplices* Δ_n, for which the distance between any pair of corners is one. For $n = 1$ this is the unit interval, for $n = 2$ a regular triangle, for $n = 3$ a regular tetrahedron, and so on. Again we are interested in the volume, the radius r_n of the insphere, and the radius R_n of the outsphere. We will also compute χ_n, the angle between the lines from the 'centre of mass' to a pair of corners. For a triangle it is $\arccos(-1/2) = 2\pi/3 = 120^\circ$, but it drops to $\arccos(-1/3) \approx 110^\circ$ for the tetrahedron. A practical way to go about all this is to think of Δ_n as a (part of) a cone having Δ_{n-1} as its base. It is then not difficult to show that

$$R_n = n r_n = \sqrt{\frac{n}{2(n+1)}} \quad \text{and} \quad r_n = \sqrt{\frac{1}{2(n+1)n}} \ , \tag{1.24}$$

so their ratio grows linearly, $\zeta = R_n/r_n = n$. The volume of a cone is $V = Bh/n$, where B is the area of the base, h is the height of the cone and n is the dimension.

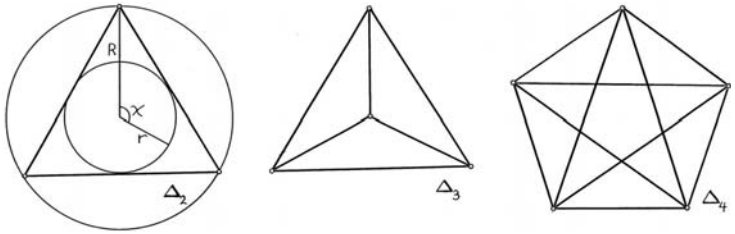


Figure 1.8. Regular simplices in two, three and four dimensions. For Δ_2 we also show the insphere, the outsphere, and the angle discussed in the text.

For the simplex we obtain

$$\mathrm{vol}(\Delta_n) \;=\; \frac{1}{n!}\sqrt{\frac{n+1}{2^n}}\,. \tag{1.25}$$

We can check that the ratio of the volume of the largest inscribed sphere to the volume of the simplex goes to zero. Hence most of the volume of the simplex sits in its corners, as expected. The angle χ_n subtended by an edge as viewed from the centre is given by

$$\sin\frac{\chi_n}{2} = \frac{1}{2R_n} = \sqrt{\frac{n+1}{2n}} \quad\Leftrightarrow\quad \cos\chi_n = -\frac{1}{n}\,. \tag{1.26}$$

When n is large we see that χ_n tends to a right angle. This is as it should be. The corners sit on the outsphere, and for large n almost all the volume of the circumsphere lies close to the equator – hence, if we pick one corner and let it play the role of the north pole, all the other corners are likely to lie close to the equator. Finally it is interesting to observe that it is known for convex bodies in general that the radius of the circumsphere is bounded by

$$R_n \;\le\; L\sqrt{\frac{n}{2(n+1)}}\,, \tag{1.27}$$

where L is the length of the longest line segment contained in the body. The regular simplex saturates this bound.

The effects of increasing dimension are clearly seen if we look at the ratio between surface (hyper) area and volume for bodies of various shapes. Rather than fixing the scale, let us study the dimensionless quantities $\zeta_n = R_n/r_n$ and $\eta(X) \equiv R\,\mathrm{vol}(\partial X)/\mathrm{vol}(X)$, where X is the body, ∂X its boundary, and R its outradius. For n-balls we get

$$\eta_n(\mathbf{B}^n) = R\,\frac{\mathrm{vol}(\partial\mathbf{B}^n)}{\mathrm{vol}(\mathbf{B}^n)} = R\,\frac{\mathrm{vol}(\mathbf{S}^{n-1})}{\mathrm{vol}(\mathbf{B}^n)} = \frac{Rn}{R} = n\,. \tag{1.28}$$

Next consider a hypercube of edge length L. Its boundary consists of $2n$ facets, that are themselves hypercubes of dimension $n-1$. This gives

$$\eta_n(\boxtimes_n) = R\,\frac{\mathrm{vol}(\partial\boxtimes_n)}{\mathrm{vol}(\boxtimes_n)} = \frac{\sqrt{n}L}{2}\,\frac{2n\,\mathrm{vol}(\boxtimes_{n-1})}{\mathrm{vol}(\boxtimes_n)} = \frac{n^{3/2}L}{L} = n^{3/2}\,. \tag{1.29}$$

A regular simplex of edge length L has a boundary consisting of $n+1$ regular simplices of dimension $n-1$. We obtain the ratio

$$\eta_n(\Delta_n) = R\,\frac{\mathrm{vol}(\partial\Delta_n)}{\mathrm{vol}(\Delta_n)} = L\sqrt{\frac{n}{2(n+1)}}\,\frac{(n+1)\mathrm{vol}(\Delta_{n-1})}{\mathrm{vol}(\Delta_n)} = n^2\,. \tag{1.30}$$

In this case the ratio η_n grows quadratically with n, reflecting the fact that simplices have sharper corners than those of the cube.

The reader may know about the five regular Platonic solids in three dimensions. When $n > 4$ there are only three kinds of regular solids, namely the simplex, the hypercube, and the *cross-polytope*. The latter is the generalization to arbitrary dimension of the octahedron. It is dual to the cube; while the cube has 2^n corners and $2n$ facets, the cross-polytope has $2n$ corners and 2^n facets. The two polytopes have the same values of ζ_n and η_n.

These results are collected in Table 14.2. We observe that $\eta_n = n\zeta_n$ for all these bodies. There is a reason for this. When Archimedes computed volumes, he did so by breaking them up into cones and using the formula $V = Bh/n$, where V is the volume of the cone and B is the area of its base. Then we get

$$\eta_n = R\,\frac{\sum_{\text{cones}} B}{\left(\sum_{\text{cones}} B\right) h/n} = \frac{nR}{h}\,. \tag{1.31}$$

If the height h of the cones is equal to the inradius of the body, the result follows.[3]

1.3 Colour theory

How do convex sets arise? An instructive example occurs in colour theory, and more particularly in the psychophysical theory of colour. (This means that we will leave out the interesting questions about how our brain actually processes the visual information until it becomes a percept.) In a way tradition suggests that colour theory should be studied before quantum mechanics, because this is what Schrödinger was doing before inventing his wave equation.[4] The object of our attention is *colour space*, whose points are the colours. Naturally one might worry

[3] Consult Ball (1997) for more information on the subject of this section. For a discussion of rotations in higher dimensions consult Section 8.3.

[4] Schrödinger (1926b) wrote a splendid review of the subject. Readers who want a more recent discussion may enjoy the book by Williamson and Cummins (1983).

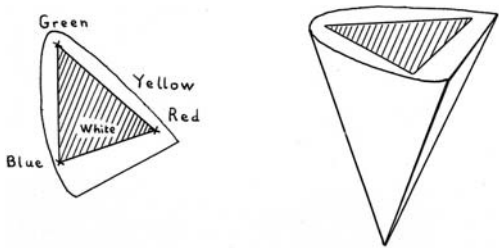


Figure 1.9. Left: the chromaticity diagram, and the part of it that can be obtained by mixing red, green and blue. Right: when the total illumination is taken into account, colour space becomes a convex cone.

that the space of colours may differ from person to person but in fact it does not. The perception of colour is remarkably universal for human beings (colour-blind persons not included). What has been done experimentally is to shine mixtures of light of different colours on white screens; say that three reference colours consisting of red, green and blue light are chosen. Then what one finds is that by adjusting the mixture of these colours the observer will be unable to distinguish the resulting mixture from a given colour C. To simplify matters, suppose that the overall brightness has been normalized in some way, then a colour C is a point on a two-dimensional *chromaticity diagram*. Its position is determined by the equation

$$C = \lambda_0 R + \lambda_1 G + \lambda_2 B \ . \tag{1.32}$$

The barycentric coordinates λ_i will naturally take positive values only in this experiment. This means that we only get colours inside the triangle spanned by the reference colours R, G and B. Note that the 'feeling of redness' does not enter into the experiment at all.

But colour space is not a simplex, as designers of TV screens learn to their chagrin. There will always be colours C' that cannot be reproduced as a mixture of three given reference colours. To get out of this difficulty one shines a certain amount of red (say) on the sample to be measured. If the result is indistinguishable from some mixture of G and B then C' is determined by the equation

$$C' + \lambda_0 R = \lambda_1 G + \lambda_2 B \ . \tag{1.33}$$

If not, repeat with R replaced by G or B. If necessary, move one more colour to the left-hand side. The empirical finding is that all colours can be assigned a position on the chromaticity diagram in this way. If we take the overall intensity into account we find that the full colour space is a three-dimensional convex cone with the chromaticity diagram as its base and complete darkness as its apex (of course this is to the extent that we ignore the fact that very intense light will cause

the eyes to boil rather than make them see a colour). The pure colours are those that cannot be obtained as a mixture of different colours; they form the curved part of the boundary. The boundary also has a planar part made of purple.

How can we begin to explain all this? We know that light can be characterized by its spectral distribution, which is some positive function I of the wave length λ. It is therefore immediately apparent that the space of spectral distributions is a convex cone, and in fact an infinite-dimensional convex cone since a general spectral distribution $I(\lambda)$ can be defined as a convex combination

$$I(\lambda) = \int \mathrm{d}\lambda' I(\lambda')\delta(\lambda - \lambda') \,, \qquad I(\lambda') \geq 0 \,. \tag{1.34}$$

The delta functions are the pure states. But colour space is only three-dimensional. The reason is that the eye will assign the same colour to many different spectral distributions. A given colour corresponds to an equivalence class of spectral distributions, and the dimension of colour space will be given by the dimension of the space of equivalence classes. Let us denote the equivalence classes by $[I(\lambda)]$, and the space of equivalence classes as colour space. Since we know that colours can be mixed to produce a third quite definite colour, the equivalence classes must be formed in such a way that the equation

$$[I(\lambda)] = [I_1(\lambda)] + [I_2(\lambda)] \tag{1.35}$$

is well defined. The point here is that whatever representatives of $[I_1(\lambda)]$ and $[I_2(\lambda)]$ we choose we always obtain a spectral distribution belonging to the same equivalence class $[I(\lambda)]$. We would like to understand how this can be so.

In order to proceed it will be necessary to have an idea about how the eye detects light (especially so since the perception of sound is known to work in a quite different way). It is reasonable – and indeed true – to expect that there are chemical substances in the eye with different sensitivities. Suppose for the sake of the argument that there are three such 'detectors'. Each has an adsorption curve $A_i(\lambda)$. These curves are allowed to overlap; in fact they do. Given a spectral distribution each detector then gives an output

$$c_i = \int \mathrm{d}\lambda \; I(\lambda)A_i(\lambda) \,. \tag{1.36}$$

Our three detectors will give us only three real numbers to parametrize the space of colours. Equation (1.35) can now be derived. According to this theory, colour space will inherit the property of being a convex cone from the space of spectral distributions. The pure states will be those equivalence classes that contain the pure spectral distributions. On the other hand the dimension of colour space will be determined by the number of detectors, and not by the nature of the pure states.

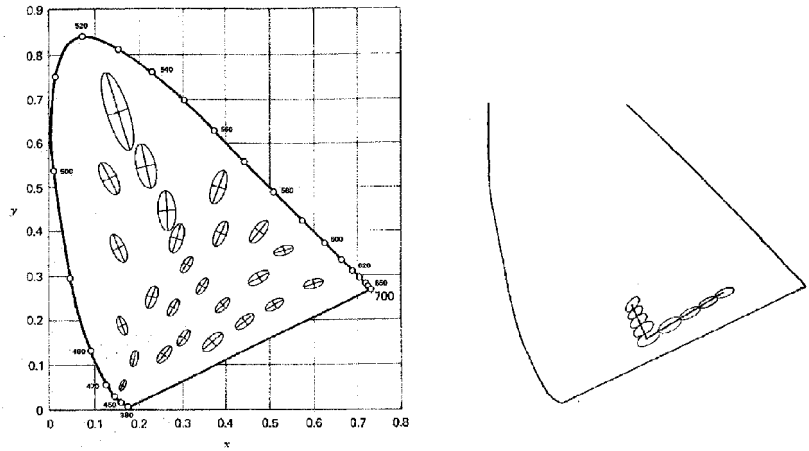


Figure 1.10. To the left, we see the MacAdam ellipses, taken from MacAdam, *Journal of the Optical Society of America* 32, p. 247 (1942). They show the points where the colour is just distinguishable from the colour at the centre of the ellipse. Their size is exaggerated by a factor of ten. To the right, we see how these ellipses can be used to define the length of curves on the chromaticity diagram – the two curves shown have the same length.

This is where colour-blind persons come in; they are missing one or two detectors and their experiences can be predicted by the theory. By the way, frogs apparently enjoy a colour space of four dimensions while lions make do with one.

Like any convex set, colour space is a subset of an affine space and the convex structure does not single out any natural metric. Nevertheless colour space does have a natural metric. The idea is to draw surfaces around every point in colour space, determined by the requirement that colours situated on the surfaces are just distinguishable from the colour at the original point by an observer. In the chromaticity diagram the resulting curves are known as *MacAdam ellipses*. We can now introduce a metric on the chromaticity diagram which ensures that the MacAdam ellipses are circles of a standard size. This metric is called the *colour metric*, and it is curved. The distance between two colours as measured by the colour metric is a measure of how easy it is to distinguish the given colours. On the other hand this natural metric has nothing to do with the convex structure per se.

Let us be careful about the logic that underlies the colour metric. The colour metric is defined so that the MacAdam ellipses are circles of radius ϵ, say. Evidently we would like to consider the limit when ϵ goes to zero (by introducing increasingly sensitive observers), but unfortunately there is no experimental justification for this here. We can go on to define the length of a curve in colour space as the smallest

number of MacAdam ellipses that is needed to completely cover the curve. This gives us a natural notion of distance between any two points in colour space since there will be a curve between them of shortest length (and it will be uniquely defined, at least if the distance is not too large). Such a curve is called a *geodesic*. The *geodesic distance* between two points is then the length of the geodesic that connects them. This is how distances are defined in Riemannian geometry, but it is worthwhile to observe that only the 'local' distance as defined by the metric has a clear operational significance here. There are many lessons from colour theory that are of interest in quantum mechanics, not least that the structure of the convex set is determined by the nature of the detectors.

1.4 What is 'distance'?

In colour space distances are used to quantify distinguishability. Although our use of distances will mostly be in a similar vein, they have many other uses too – for instance, to prove convergence for iterative algorithms. But what are they? Though we expect the reader to have a share of inborn intuition about the nature of geometry, a few indications of how this can be made more precise are in order. Let us begin by defining a *distance* $D(\mathbf{x}, \mathbf{y})$ between two points in a vector space (or more generally, in an affine space). This is a function of the two points that obeys three axioms:

(1) The distance between two points is a non-negative number $D(\mathbf{x}, \mathbf{y})$ that equals zero if and only if the points coincide.
(2) It is symmetric in the sense that $D(\mathbf{x}, \mathbf{y}) = D(\mathbf{y}, \mathbf{x})$.
(3) It satisfies the triangle inequality $D(\mathbf{x}, \mathbf{y}) \leq D(\mathbf{x}, \mathbf{z}) + D(\mathbf{z}, \mathbf{y})$.

Actually both axiom (2) and axiom (3) can be relaxed – we will see what can be done without them in Section 2.3 – but as is often the case it is even more interesting to try to restrict the definition further, and this is the direction that we are heading in now. We want a notion of distance that meshes naturally with convex sets, and for this purpose we add a fourth axiom:

(4) It obeys $D(\lambda\mathbf{x}, \lambda\mathbf{y}) = \lambda D(\mathbf{x}, \mathbf{y})$ for non-negative numbers λ.

A distance function obeying this property is known as a *Minkowski distance*. Two important consequences follow, neither of them difficult to prove. First, any convex combination of two vectors becomes a metric straight line in the sense that

$$\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \quad \Rightarrow \quad D(\mathbf{x}, \mathbf{y}) = D(\mathbf{x}, \mathbf{z}) + D(\mathbf{z}, \mathbf{y}) \,, \quad 0 \leq \lambda \leq 1 \,. \tag{1.37}$$

Second, if we define a unit ball with respect to a Minkowski distance we find that such a ball is always a convex set.

Let us discuss the last point in a little more detail. A Minkowski metric is naturally defined in terms of a *norm* on a vector space, that is a real valued function $||\mathbf{x}||$ that obeys

$$\begin{aligned} &\text{i)} \quad ||\mathbf{x}|| \geq 0 \ , \text{ and } ||\mathbf{x}|| = 0 \ \Leftrightarrow \ \mathbf{x} = 0 \ . \\ &\text{ii)} \quad ||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \ . \\ &\text{iii)} \quad ||\lambda \mathbf{x}|| = |\lambda| \ ||\mathbf{x}|| \ , \quad \lambda \in \mathbf{R} \ . \end{aligned} \tag{1.38}$$

The distance between two points $\mathbf{x}$ and $\mathbf{y}$ is now defined as $D(\mathbf{x}, \mathbf{y}) \equiv ||\mathbf{x} - \mathbf{y}||$, and indeed it has the properties (1)–(4). The unit ball is the set of vectors $\mathbf{x}$ such that $||\mathbf{x}|| \leq 1$, and it is easy to see that

$$||\mathbf{x}|| \ , \ ||\mathbf{y}|| \leq 1 \quad \Rightarrow \quad ||\lambda \mathbf{x} + (1 - \lambda)\mathbf{y}|| \leq 1 \ . \tag{1.39}$$

So the unit ball is convex. In fact the story can be turned around at this point – any centrally symmetric convex body can serve as the unit ball for a norm, and hence it defines a distance. (A centrally symmetric convex body K has the property that, for some choice of origin, $\mathbf{x} \in K \ \Rightarrow \ -\mathbf{x} \in K$.) Thus the opinion that balls are round is revealed as an unfounded prejudice. It may be helpful to recall that water droplets are spherical because they minimize their surface energy. If we want to understand the growth of crystals in the same terms, we must use a notion of distance that takes into account that the surface energy depends on direction.

We need a set of norms to play with, so we define the l_p*-norm* of a vector by

$$||\mathbf{x}||_p \equiv \left(|x_1|^p + |x_2|^p + \ \cdots \ + |x_n|^p\right)^{\frac{1}{p}} \ , \quad p \geq 1 \ . \tag{1.40}$$

In the limit we obtain the *Chebyshev norm* $||\mathbf{x}||_\infty = \max_i x_i$. The proof of the triangle inequality is non-trivial and uses *Hölder's inequality*

$$\sum_{i=1}^{N} |x_i \, y_i| \leq ||\mathbf{x}||_p ||\mathbf{y}||_q \ , \quad \frac{1}{p} + \frac{1}{q} = 1 \ , \tag{1.41}$$

where $p, q \geq 1$. For $p = 2$ this is the *Cauchy–Schwarz inequality*. If $p < 1$ there is no Hölder inequality, and the triangle inequality fails. We can easily draw a picture (namely Figure 1.11) of the unit balls $\mathbf{B}_p$ for a few values of p, and we see that they interpolate beween a hypercube (for $p \to \infty$) and a cross-polytope (for $p = 1$), and that they fail to be convex for $p < 1$. We also see that in general these balls are not invariant under rotations, as expected because the components of the vector in a special basis were used in the definition. The topology induced by the l_p-norms is the same, regardless of p. The corresponding distances $D_p(\mathbf{x}, \mathbf{y}) \equiv ||\mathbf{x} - \mathbf{y}||_p$ are known as the l_p*-distances*.

Depending on circumstances, different choices of p may be particularly relevant. The case $p = 1$ is relevant if motion is confined to a rectangular grid (say, if you are

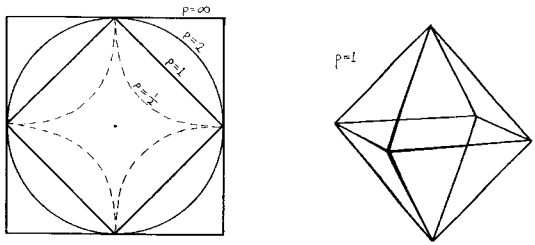


Figure 1.11. Left: points at distance 1 from the origin, using the l_1-norm for the vectors (the inner square), the l_2-norm (the circle) and the l_∞-norm (the outer square). The $l_{\frac{1}{2}}$-case is shown dashed – the corresponding ball is not convex because the triangle inequality fails, so it is not a norm. Right: in three dimensions one obtains, respectively, an octahedron, a sphere and a cube. We illustrate the $p = 1$ case.

a taxi driver on Manhattan). As we will see (in Section 13.1) it is also of particular relevance to us. It has the slightly awkward property that the shortest path between two points is not uniquely defined. Taxi drivers know this, but may not be aware of the fact that it happens only because the unit ball is a polytope, that is it is convex but not strictly convex. The l_1-distance goes under many names: *taxi cab*, *Kolmogorov*, or *variational distance*.

The case $p = 2$ is consistent with Pythagoras' theorem and is the most useful choice in everyday life; it was singled out for special attention by Riemann when he made the foundations for differential geometry. Indeed we used a $p = 2$ norm when we defined the colour metric at the end of Section 1.3. The idea is that once we have some coordinates to describe colour space then the MacAdam ellipse surrounding a point is given by a quadratic form in the coordinates. The interesting thing – that did not escape Riemann – is the ease with which this 'infinitesimal' notion of distance can be converted into the notion of geodesic distance between arbitrary points. (A similar generalization based on other l_p-distances exists and is called *Finslerian geometry*, as opposed to the *Riemannian geometry* based on $p = 2$.)

Riemann began by defining what we now call differentiable manifolds of arbitrary dimension;[5] for our purposes here let us just say that this is something that locally looks like $\mathbb{R}^n$ in the sense that it has open sets, continuous functions and differentiable functions; one can set up a one-to-one correspondence between the

[5] Riemann lectured on the hypotheses which lie at the foundations of geometry in 1854, in order to be admitted as a Dozent at Göttingen. As Riemann says, only two instances of continuous manifolds were known from everyday life at the time: the space of locations of physical objects, and the space of colours. In spite of this he gave an essentially complete sketch of the foundations of modern geometry. For a more detailed account see (for instance) Murray and Rice (1993). A very readable, albeit old-fashioned, account is by our Founding Father: Schrödinger (1950). For beginners the definitions in this section can become bewildering; if so our advice is to ignore them, and look at some examples of curved spaces first.

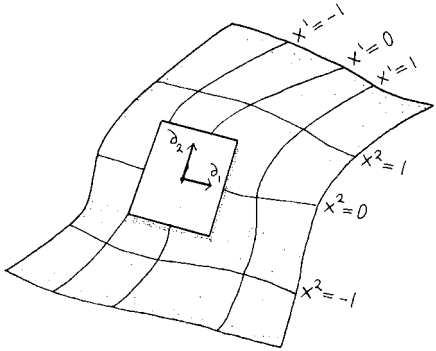


Figure 1.12. The tangent space at the origin of some coordinate system. Note that there is a tangent space at every point.

points in some open set and n numbers θ^i, called *coordinates*, that belong to some open set in $\mathbb{R}^n$. There exists a *tangent space* $\mathbf{T}_q$ at every point q in the manifold; intuitively we can think of the manifold as some curved surface in space and of a tangent space as a flat plane touching the surface at some point. By definition the tangent space $\mathbf{T}_q$ is the vector space whose elements are *tangent vectors* at q, and a tangent vector at a point of a differentiable manifold is defined as the tangent vector of a smooth curve passing through the point. Intuitively, it is a little arrow sitting at the point. Formally, it is a *contravariant* vector (with index upstairs). Each tangent vector V^i gives rise to a directional derivative $\sum_i V^i \partial_i$ acting on the functions on the space; in differential geometry it has therefore become customary to think of a tangent vector as a derivative operator. In particular we can take the derivatives in the directions of the coordinate lines, and any directional derivative can be expressed as a linear combination of these. Therefore, given any coordinate system θ^i, the derivatives ∂_i with respect to the coordinates form a basis for the tangent space – not necessarily the most convenient basis one can think of, but one that certainly exists. To sum up, a tangent vector is written as

$$\mathbf{V} = \sum_i V^i \partial_i \,, \tag{1.42}$$

where $\mathbf{V}$ is the vector itself and V^i are the components of the vector in the coordinate basis spanned by the basis vectors ∂_i.

It is perhaps as well to emphasize that the tangent space $\mathbf{T}_q$ at a point q bears no a-priori relation to the tangent space $\mathbf{T}_{q'}$ at a different point q', so that tangent vectors at different points cannot be compared unless additional structure is introduced. Such an additional structure is known as 'parallel transport' or 'covariant derivatives', and will be discussed in Section 3.2.

At every point q of the manifold there is also a *cotangent space* $\mathbf{T}_q^*$, the vector space of linear maps from $\mathbf{T}_q$ to the real numbers. Its elements are called *covariant vectors*. Given a coordinate basis for $\mathbf{T}_q$ there is a natural basis for the cotangent space consisting of n covariant vectors $\mathrm{d}\theta^i$ defined by

$$\mathrm{d}\theta^i(\partial_j) = \delta^i_j \ , \tag{1.43}$$

with the Kronecker delta appearing on the right-hand side. The tangent vector ∂_i points in the coordinate direction, while $\mathrm{d}\theta^i$ gives the level curves of the coordinate function. A general element of the cotangent space is also known as a *one-form*. It can be expanded as $U = U_i \mathrm{d}\theta^i$, so that covariant vectors have indices downstairs. The linear map of a tangent vector $\mathbf{V}$ is given by

$$U(\mathbf{V}) = U_i \mathrm{d}\theta^i(V^j \partial_j) = U_i V^j \mathrm{d}\theta^i(\partial_j) = U_i V^i \ . \tag{1.44}$$

From now on the *Einstein summation convention* is in force, which means that if an index appears twice in the same term then summation over that index is implied. A natural next step is to introduce a scalar product in the tangent space, and indeed in every tangent space. (One at each point of the manifold.) We can do this by specifying the scalar products of the basis vectors ∂_i. When this is done we have in fact defined a *Riemannian metric tensor* on the manifold, whose components in the coordinate basis are given by

$$g_{ij} = \langle \partial_i, \partial_j \rangle \ . \tag{1.45}$$

It is understood that this has been done at every point q, so the components of the metric tensor are really functions of the coordinates. The metric g_{ij} is assumed to have an inverse g^{ij}. Once we have the metric it can be used to raise and lower indices in a standard way ($V_i = g_{ij} V^j$). Otherwise expressed it provides a canonical isomorphism between the tangent and cotangent spaces.

Riemann went on to show that one can always define coordinates on the manifold in such a way that the metric at any given point is diagonal and has vanishing first derivatives there. In effect – provided that the metric tensor is a positive definite matrix, which we assume – the metric gives a 2-norm on the tangent space at that special point. Riemann also showed that in general it is not possible to find coordinates so that the metric takes this form everywhere; the obstruction that may make this impossible is measured by a quantity called the *Riemann curvature tensor*. It is a linear function of the second derivatives of the metric (and will make its appearance in Section 3.2). The space is said to be flat if and only if the Riemann tensor vanishes, which is if and only if coordinates can be found so that the metric takes the same diagonal form everywhere. The 2-norm was singled out by Riemann precisely because his grandiose generalization of geometry to the case of arbitrary differentiable manifolds works much better if $p = 2$.

With a metric tensor at hand we can define the length of an arbitrary curve $x^i = x^i(t)$ in the manifold as the integral

$$\int \mathrm{d}s = \int \sqrt{g_{ij} \frac{\mathrm{d}x^i}{\mathrm{d}t} \frac{\mathrm{d}x^j}{\mathrm{d}t}}\, \mathrm{d}t \tag{1.46}$$

along the curve. The shortest curve between two points is called a geodesic, and we are in a position to define the geodesic distance between the two points just as we did at the end of Section 1.3. The geodesic distance obeys the axioms that we laid down for distance functions, so in this sense the metric tensor defines a distance. Moreover, at least as long as two points are reasonably close, the shortest path between them is unique.

One of the hallmarks of differential geometry is the ease with which the tensor formalism handles coordinate changes. Suppose we change to new coordinates $x^{i'} = x^{i'}(x)$. Provided that these functions are invertible the new coordinates are just as good as the old ones. More generally, the functions may be invertible only for some values of the original coordinates, in which case we have a pair of partially overlapping *coordinate patches*. It is elementary that

$$\partial_{i'} = \frac{\partial x^j}{\partial x^{i'}} \partial_j \ . \tag{1.47}$$

Since the vector $\mathbf{V}$ itself is not affected by the coordinate change – which is after all just some equivalent new description – Eq. (1.42) implies that its components must change according to

$$V^{i'} \partial_{i'} = V^i \partial_i \quad \Rightarrow \quad V^{i'}(x') = \frac{\partial x^{i'}}{\partial x^j} V^j(x) \ . \tag{1.48}$$

In the same way we can derive how the components of the metric change when the coordinate system changes, using the fact that the scalar product of two vectors is a scalar quantity that does not depend on the coordinates:

$$g_{i'j'} U^{i'} V^{j'} = g_{ij} U^i V^j \quad \Rightarrow \quad g_{i'j'} = \frac{\partial x^k}{\partial x^{i'}} \frac{\partial x^l}{\partial x^{j'}} g_{kl} \ . \tag{1.49}$$

We see that the components of a tensor, in some basis, depend on that particular and arbitrary basis. This is why they are often regarded with feelings bordering on contempt by professionals, who insist on using 'coordinate free methods' and think that 'coordinate systems do not matter'. But in practice few things are more useful than a well-chosen coordinate system. And the tensor formalism is tailor made to construct scalar quantities invariant under coordinate changes.

In particular the formalism provides invariant measures that can be used to define lengths, areas, volumes, and so on, in a way that is independent of the choice of coordinate system. This is because the square root of the determinant of the metric

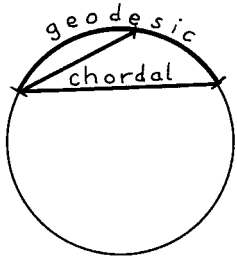


Figure 1.13. Here is how to measure the geodesic and the chordal distances between two points on the sphere. When the points are close these distances are also close; they are consistent with the same metric.

tensor, $\sqrt{g}$, transforms in a special way under coordinate transformations:

$$\sqrt{g'}(x') = \left(\det \frac{\partial x'}{\partial x}\right)^{-1} \sqrt{g}(x) \,. \tag{1.50}$$

The integral of a scalar function $f'(x') = f(x)$, over some manifold $\mathbf{M}$, then behaves as

$$I = \int_{\mathbf{M}} f'(x')\sqrt{g'}(x')\, \mathrm{d}^n x' = \int_{\mathbf{M}} f(x)\sqrt{g}(x)\, \mathrm{d}^n x \tag{1.51}$$

– the transformation of $\sqrt{g}$ compensates for the transformation of $\mathrm{d}^n x$, so that the measure $\sqrt{g}\mathrm{d}^n x$ is invariant. A submanifold can always be locally defined via equations of the general form $x = x(x')$, where x' are intrinsic coordinates on the submanifold and x are coordinates on the *embedding space* in which it sits. In this way Eq. (1.49) can be used to define an *induced metric* on the submanifold, and hence an invariant measure as well. Equation (1.46) is in fact an example of this construction – and it is good to know that the geodesic distance between two points is independent of the coordinate system.

Since this is not a textbook on differential geometry we leave these matters here, except that we want to draw attention to some possible ambiguities. First there is an ambiguity of notation. The metric is often presented in terms of the squared *line element*,

$$\mathrm{d}s^2 = g_{ij}\mathrm{d}x^i \mathrm{d}x^j \,. \tag{1.52}$$

The ambiguity is this: in modern notation $\mathrm{d}x^i$ denotes a basis vector in cotangent space, and $\mathrm{d}s^2$ is a linear operator acting on the tensor product $\mathbf{T} \otimes \mathbf{T}$. There is also an old-fashioned way of reading the formula, which regards $\mathrm{d}s^2$ as the length squared of that tangent vector whose components (at the point with coordinates x) are $\mathrm{d}x^i$. A modern mathematician would be appalled by this, rewrite it as $g_x(\mathrm{d}s, \mathrm{d}s)$,

and change the label ds for the tangent vector to, say, A. But a liberal reader will be able to read Eq. (1.52) in both ways. The old-fashioned notation has the advantage that we can regard ds as the distance between two 'nearby' points given by the coordinates x and $x + \mathrm{d}x$; their distance is equal to ds plus terms of higher order in the coordinate differences. We then see that there are ambiguities present in the notion of distance too. To take the sphere as an example, we can define a distance function by means of geodesic distance. But we can also define the distance between two points as the length of a chord connecting the two points, and the latter definition is consistent with our axioms for distance functions. Moreover both definitions are consistent with the metric, in the sense that the distances between two nearby points will agree to lowest order. However, in this book we will usually regard it as understood that once we have a metric we are going to use the geodesic distance to measure the distance between two arbitrary points.

1.5 Probability and statistics

The reader has probably surmised that our interest in convex sets has to do with their use in statistics. It is not our intention to explain the notion of probability, not even to the extent that we tried to explain colour. We are quite happy with the Kolmogorov axioms, that define probability as a suitably normalized positive measure on some set Ω. If the set of points is finite, this is simply a finite set of positive numbers adding up to one. Now there are many viewpoints on what the meaning of it all may be, in terms of frequencies, propensities and degrees of reasonable beliefs. We do not have to take a position on these matters here because the geometry of probability distributions is invariant under changes of interpretation.[6] We do need to fix some terminology however, and will proceed to do so.

Consider an experiment that can yield N possible outcomes, or in mathematical terms a *random variable* X that can take N possible values x_i belonging to a *sample space* Ω, which in this case is a discrete set of points. The probabilities for the respective outcomes are

$$P(X = x_i) = p^i \,. \tag{1.53}$$

For many purposes the actual outcomes can be ignored. The interest centres on the probability distribution $P(X)$ considered as the set of N real numbers p^i such that

$$p^i \geq 0 \,, \qquad \sum_{i=1}^{N} p^i = 1 \,. \tag{1.54}$$

[6] The reader may consult the book by von Mises (1957) for one position, and the book by Jaynes (2003) for another. Jaynes regards probability as quantifying the degree to which a proposition is plausible, and finds that $\sqrt{p_i}$ has a status equally fundamental as that of p_i.

(We will sometimes be a little ambiguous about whether the index should be up or down – although it should be upstairs according to the rules of differential geometry.) Now look at the space of all possible probability distributions for the given random variable. This is a simplex with the p^i playing the role of barycentric coordinates; a convex set of the simplest possible kind. The pure states are those for which the outcome is certain, so that one of the p^i is equal to one. The pure states sit at the corners of the simplex and hence they form a zero-dimensional subset of its boundary. In fact the space of pure states is isomorphic to the sample space. As long as we keep to the case of a finite number of outcomes – the *multinomial probability distribution* as it is known in probability theory – nothing could be simpler.

Except that, as a subset of an n-dimensional vector space, an n-dimensional simplex is a bit awkward to describe using Cartesian coordinates. Frequently it is more convenient to regard it as a subset of an $N = (n + 1)$-dimensional vector space instead, and use the unrestricted p^i to label the axes. Then we can use the l_p-norms to define distances. The meaning of this will be discussed in Chapter 2; meanwhile we observe that the probability simplex lies somewhat askew in the vector space, and we find it convenient to adjust the definition a little. From now on we set

$$D_p(P, Q) \equiv ||P - Q||_p \equiv \left(\frac{1}{2} \sum_{i=1}^{N} |p_i - q_i|^p \right)^{\frac{1}{p}} , \qquad 1 \leq p . \tag{1.55}$$

The extra factor of $1/2$ ensures that the edge lengths of the simplex equal 1, and also has the pleasant consequence that all the l_p-distances agree when $N = 2$. However, it is a little tricky to see what the l_p-balls look like inside the probability simplex. The case $p = 1$, which is actually important to us, is illustrated in Figure 1.14; we are looking at the intersection of a cross-polytope with the probability simplex. The result is a convex body with $N(N - 1)$ corners. For $N = 2$ it is a hexagon, for $N = 3$ a cuboctahedron, and so on.

The l_1-distance has the interesting property that probability distributions with orthogonal *support* – meaning that the product $p_i q_i$ vanishes for each value of i – are at maximal distance from each other. One can use this observation to show, without too much effort, that the ratio of the radii of the in- and outspheres for the l_1-ball obeys

$$\frac{r_{\text{in}}}{R_{\text{out}}} = \sqrt{\frac{2}{N}} \text{ if } N \text{ is even} , \qquad \frac{r_{\text{in}}}{R_{\text{out}}} = \sqrt{\frac{2N}{N^2 - 1}} \text{ if } N \text{ is odd} . \tag{1.56}$$

Hence, although some corners have been 'chopped off', the body is only marginally more spherical than is the cross-polytope. Another way to say the same thing is that, with our normalization, $||\mathbf{p}||_1 \leq ||\mathbf{p}||_2 \leq R_{\text{out}} ||\mathbf{p}||_1 / r_{\text{in}}$.

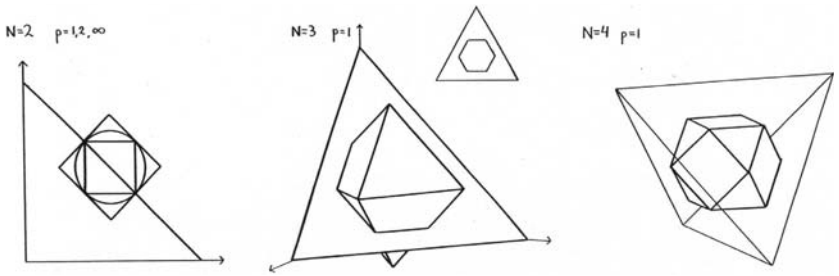


Figure 1.14. For $N = 2$ we show why all the l_p-distances agree when the definition (Eq. 1.55) is used. For $N = 3$ the l_1-distance gives hexagonal 'spheres', arising as the intersection of the simplex with an octahedron. For $N = 4$ the same construction gives an Archimedean solid known as the cuboctahedron.

We end with some further definitions, that will put more strain on the notation. Suppose we have two random variables X and Y with N and M outcomes and described by the distributions P_1 and P_2, respectively. Then there is a *joint probability distribution* P_{12} of the joint probabilities,

$$P_{12}(X = x_i, Y = y_j) = p_{12}^{ij} \,. \tag{1.57}$$

This is a set of NM non-negative numbers summing to one. Note that it is not implied that $p_{12}^{ij} = p_1^i p_2^j$; if this does happen the two random variables are said to be *independent*, otherwise they are *correlated*. More generally K random variables are said to be independent if

$$p_{12\ldots K}^{ij\ldots k} = p_1^i \, p_2^j \, \ldots \, p_K^k \,, \tag{1.58}$$

and we may write this schematically as $P_{12\ldots K} = P_1 P_2 \ldots P_K$. A *marginal distribution* is obtained by summing over all possible outcomes for those random variables that we are not interested in. Thus a first order distribution, describing a single random variable, can be obtained as a marginal of a second order distribution, describing two random variables jointly, by

$$p_1^i = \sum_j p_{12}^{ij} \,. \tag{1.59}$$

There are also special probability distributions that deserve special names. Thus the *uniform* distribution for a random variable with N outcomes is denoted by $Q_{(N)}$ and the distributions where one outcome is certain are collectively denoted by $Q_{(1)}$. The notation can be extended to include

$$Q_{(M)} = (\frac{1}{M}, \frac{1}{M}, \ldots, \frac{1}{M}, 0, \ldots, 0) \,, \tag{1.60}$$

with $M \leq N$ and possibly with the components permuted.

With these preliminaries out of the way, we will devote Chapter 2 to the study of the convex sets that arise in classical statistics, and the geometries that can be defined on them – in itself, a preparation for the quantum case.

Problems

Problem 1.1 Helly's theorem states that if we have $N \geq n + 1$ convex sets in $\mathbb{R}^n$ and if for every $n + 1$ of these convex sets we find that they share a point, then there is a point that belongs to all of the N convex sets. Show that this statement is false if the sets are not assumed to be convex.

Problem 1.2 Compute the inradius and the outradius of a simplex, that is prove Eq. (1.24).

2

Geometry of probability distributions

> Some people hate the very name of statistics, but I find them full of beauty and interest.
>
> *Sir Francis Galton*

In quantum mechanics one often encounters sets of non-negative numbers that sum to unity, having a more or less direct interpretation as probabilities. This includes the squared moduli of the coefficients when a pure state is expanded in an orthonormal basis, the eigenvalues of density matrices, and more. Continuous distributions also play a role, even when the Hilbert space is finite dimensional. From a purely mathematical point of view a probability distribution is simply a *measure* on a sample space, constrained so that the total measure is one. Whatever the point of view one takes on this, the space of states will turn into a convex set when we allow probabilistic mixtures of its pure states. In classical mechanics the sample space is phase space, which is typically a continuous space. This leads to technical complications but the space of states in classical mechanics does share a major simplifying feature with the discrete case, namely that every state can be expressed as a mixture of pure states in a unique way. This was not so in the case of colour space, nor will it be true for the convex set of all states in quantum mechanics.

2.1 Majorization and partial order

Our first aim is to find ways of describing probability distributions; we want to be able to tell when a probability distribution is 'more chaotic' or 'more uniform' than another. One way of doing this is provided by the theory of *majorization*.[1] We will

[1] This is a large research area in linear algebra. Major landmarks include the books by Hardy, Littlewood and Pólya (1929), Marshall and Olkin (1979), and Alberti and Uhlmann (1982). See also Ando (1989); all unproved assertions in this section can be found there.

regard a probability distribution as a vector $\vec{x}$ belonging to the positive hyperoctant in $\mathbb{R}^N$, and normalized so that the sum of its components is unity.

The set of all normalized vectors forms an $(N-1)$-dimensional simplex Δ_{N-1}. We are interested in transformations that take probability distributions into each other, that is transformations that preserve both positivity and the l_1-norm of positive vectors.

Now consider two positive vectors, $\vec{x}$ and $\vec{y}$. We order their components in decreasing order, $x_1 \geq x_2 \geq \cdots \geq x_N$. When this has been done we may write $x_i^{\downarrow}$. We say that $\vec{x}$ is *majorized* by $\vec{y}$, written

$$\vec{x} \prec \vec{y} \quad \text{if and only if} \quad \begin{cases} \text{(i): } \sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow} \quad \text{for} \quad k = 1, \ldots, N \\ \text{(ii): } \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \ . \end{cases} \tag{2.1}$$

We assume that all our vectors are normalized in such a way that their components sum to unity, so condition (ii) is automatic. It is evident that $\vec{x} \prec \vec{x}$ (majorization is reflexive) and that $\vec{x} \prec \vec{y}$ and $\vec{y} \prec \vec{z}$ implies $\vec{x} \prec \vec{z}$ (majorization is transitive) but it is not true that $\vec{x} \prec \vec{y}$ and $\vec{y} \prec \vec{x}$ implies $\vec{x} = \vec{y}$, because one of these vectors may be obtained by a permutation of the components of the other. But if we arrange the components of all vectors in decreasing order then indeed $\vec{x} \prec \vec{y}$ and $\vec{y} \prec \vec{x}$ does imply $\vec{x} = \vec{y}$; majorization does provide a *partial order* on such vectors. The ordering is only partial because given two vectors it may happen that none of them majorize the other. Moreover there is a *smallest element*. Indeed, for every vector $\vec{x}$ it is true that

$$\vec{x}_{(N)} \equiv (1/N, 1/N, \ldots, 1/N) \prec \vec{x} \prec (1, 0, \ldots, 0) \equiv \vec{x}_{(1)} \ . \tag{2.2}$$

Note also that

$$\vec{x}_1 \prec \vec{y} \quad \text{and} \quad \vec{x}_2 \prec \vec{y} \quad \Rightarrow \quad (a\vec{x}_1 + (1-a)\vec{x}_2) \prec \vec{y} \tag{2.3}$$

for any real $a \in [0, 1]$. Hence the set of vectors majorized by a given vector is a convex set. In fact this set is the convex hull of all vectors that can be obtained by permuting the components of the given vector.

Vaguely speaking it is clear that majorization captures the idea that one vector may be more 'uniform' or 'mixed' than another, as seen in Figure 2.1. We can display all positive vectors of unit l_1-norm as a probability simplex; for $N = 3$ the convex set of all vectors that are majorized by a given vector is easily recognized (Figure 2.2). For special choices of the majorizing vector we get an equilateral triangle or a regular tetrahedron; for $N = 4$ a number of Platonic and Archimedean solids appear in this way (an Archimedean solid has regular but not equal faces (Cromwell, 1997)). See Figure 2.3.

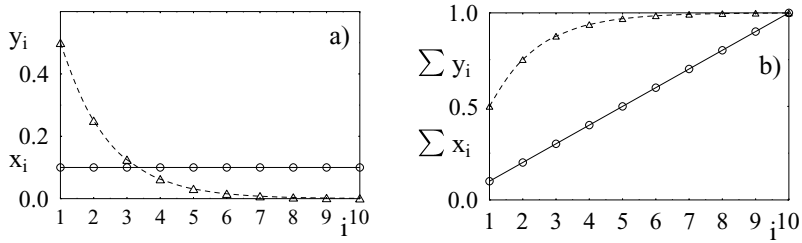


Figure 2.1. Idea of majorization: in panel (a) the vector $\vec{x} = \{x_1, \ldots, x_{10}\}$ ($\circ$) is majorized by $\vec{y} = \{y_1, \ldots, y_{10}\}$ ($\triangle$). In panel (b) we plot the distribution functions and show that Eq. (2.1) is obeyed.

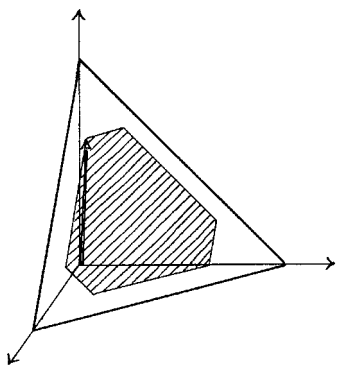


Figure 2.2. The probability simplex for $N = 3$ and the shaded convex set that is formed by all vectors that are majorized by a given vector; its pure points are obtained by permuting the components of the given vector.

Many processes in physics occur in the direction of the majorization arrow (because the passage of time tends to make things more uniform). Economists are also concerned with majorization. When Robin Hood robs the rich and helps the poor he aims for an income distribution that is majorized by the original one (provided that he acts like an isometry with respect to the l_1-norm, that is that he does not keep anything in his own pocket). We will need some information about such processes and we begin by identifying a suitable class of transformations. A *stochastic* matrix is a matrix B with N rows, whose matrix elements obey

$$\begin{aligned} &\text{(i):}\ \ B_{ij} \geq 0 \\ &\text{(ii):}\ \ \textstyle\sum_{i=1}^{N} B_{ij} = 1\ . \end{aligned} \tag{2.4}$$

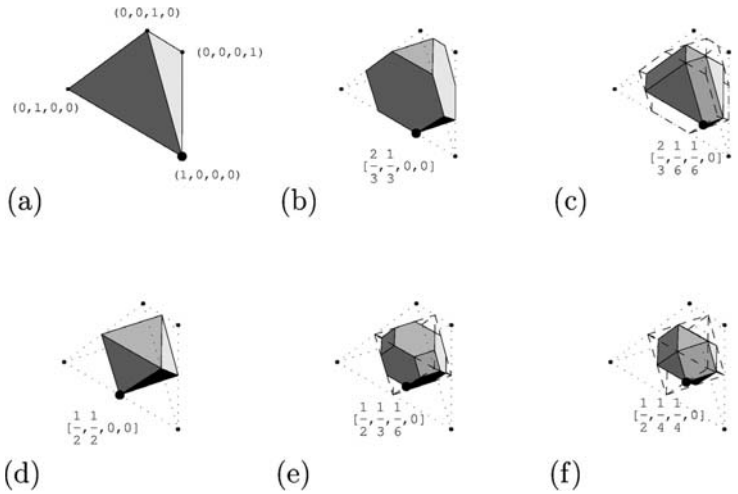


Figure 2.3. Panel (a) shows the probability simplex for $N = 4$. The set of vectors majorized by a given vector gives rise to the convex bodies shown in (b)–(f); these bodies include an octahedron (d), a truncated octahedron (e), and a cuboctahedron (f).

A *bistochastic* or *doubly stochastic* matrix is a square stochastic matrix obeying the additional condition[2]

$$\text{(iii)}: \sum_{j=1}^{N} B_{ij} = 1 \ . \tag{2.5}$$

Condition (i) means that B preserves positivity. Condition (ii) says that the sum of all the elements in a given column equals one, and it means that B preserves the l_1-norm when acting on positive vectors, or in general that B preserves the sum $\sum_i x_i$ of all the components of the vector. Condition (iii) means that B is *unital*, that is it leaves the 'smallest element' $\vec{x}_{(N)}$ invariant. Hence it causes some kind of contraction of the probability simplex towards its centre, and the classical result by Hardy et al. (1929) does not come as a complete surprise:

Lemma 2.1 (Hardy, Littlewood and Pólya's (HLP)) *$\vec{x} \prec \vec{y}$ if and only if there exists a bistochastic matrix B such that $\vec{x} = B\vec{y}$.*

For a proof see Problem 2.4. The product of two bistochastic matrices is again bistochastic; they are closed under multiplication but they do not form a group.

[2] Bistochastic matrices were first studied by Schur (1923).

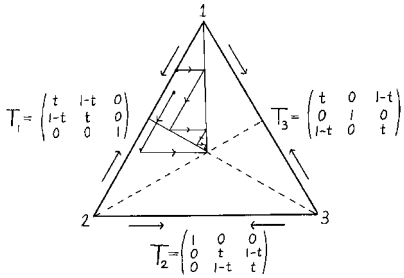


Figure 2.4. How T-transforms, and sequences of T-transforms, act on the probability simplex. The distribution $(3/4, 1/4, 0)$ is transformed to the uniform ensemble with an infinite sequence of T-transforms, while for the distribution $(14, 7, 3)/24$ we use a finite sequence (T_2T_1).

A general 2×2 bistochastic matrix is of the form

$$T = \begin{bmatrix} t & 1-t \\ 1-t & t \end{bmatrix}, \qquad t \in [0, 1]\,. \tag{2.6}$$

In higher dimensions there will be many bistochastic matrices that connect two given vectors. Of a particularly simple kind are *T-transforms* ('T' as in transfer), that is matrices that act non-trivially only on two components of a vector. It is geometrically evident from Figure 2.4 that if $\vec{x} \prec \vec{y}$ then it is always possible to find a sequence of not more than $N-1$ T-transforms such that $\vec{x} = T_{N-1}T_{N-2} \ldots T_1\vec{y}$. (Robin Hood can achieve his aim using T-transforms to transfer income.) On the other hand (except for the 2×2 case) there exist bistochastic matrices that cannot be written as sequences of T-transforms at all.

A matrix B is called *unistochastic* if there exists a unitary matrix U such that $B_{ij} = |U_{ij}|^2$. (No sum is involved – the equality concerns individual matrix elements.) In the special case that there exists an orthogonal matrix O such that $B_{ij} = (O_{ij})^2$ the matrix B is called *orthostochastic*. Due to the unitarity condition every unistochastic matrix is bistochastic, but the converse does not hold, except when $N = 2$. On the other hand we have the following (Horn, 1954):

Lemma 2.2 (Horn's) *$\vec{x} \prec \vec{y}$ if and only if there exists an orthostochastic matrix B such that $\vec{x} = B\vec{y}$.*

There is an easy to follow algorithm for how to construct such an orthostochastic matrix (Bhatia, 1997), which may be written as a product of $(N-1)$ T-transforms acting in different subspaces. In general, however, a product of an arbitrary number of T-transforms needs not be unistochastic (Poon and Tsing, 1987).

A theme that will recur is to think of a set of transformations as a space in its own right. The space of linear maps of a vector space to itself is a linear space of

its own in a natural way; to superpose two linear maps we define

$$(a_1T_1 + a_2T_2)\vec{x} \equiv a_1T_1\vec{x} + a_2T_2\vec{x} \; . \tag{2.7}$$

Given this linear structure it is easy to see that the set of bistochastic matrices forms a convex set and in fact a convex polytope. Of the equalities in Eq. (2.4) and Eq. (2.5) only $2N - 1$ are independent, so the dimension is $N^2 - 2N + 1 = (N - 1)^2$. We also see that permutation matrices (having only one non-zero entry in each row and column) are pure points of this set. The converse holds:

Theorem 2.1 (Birkhoff's) *The set of $N \times N$ bistochastic matrices is a convex polytope whose pure points are the $N!$ permutation matrices.*

To see this note that Eq. (2.4) and Eq. (2.5) define the set of bistochastic matrices as the intersection of a finite number of closed half spaces in $\mathbb{R}^{(N-1)^2}$. (An equality counts as the intersection of a pair of closed half spaces.) According to Section 1.1 the pure points must saturate $(N - 1)^2 = N^2 - 2N + 1$ of the inequalities in condition (i). Hence at most $2N - 1$ matrix elements can be non-zero; therefore at least one row (and by implication one column) contains one unit and all other entries zero. Effectively we have reduced the dimension one step, and we can now use induction to show that the only non-vanishing matrix elements for a pure point in the set of bistochastic matrices must equal 1, which means that it is a permutation matrix. Note also that using Carathéodory's theorem (again from Section 1.1) we see that every $N \times N$ bistochastic matrix can be written as a convex combination of $(N - 1)^2$ permutation matrices, and it is worth adding that there exist easy-to-follow algorithms for how to actually do this.

Functions which preserve the majorization order are called *Schur convex*;

$$\vec{x} \prec \vec{y} \quad \text{implies} \quad f(\vec{x}) \leq f(\vec{y}). \tag{2.8}$$

If $\vec{x} \prec \vec{y}$ implies $f(\vec{x}) \geq f(\vec{y})$ the function is called *Schur concave*. Clearly $-f(\vec{x})$ is Schur concave if $f(\vec{x})$ is Schur convex, and conversely. The key theorem here is:

Theorem 2.2 (Schur's) *A differentiable function $F(x_1, \ldots, x_N)$ is Schur convex if and only if F is permutation invariant and if, for all $\vec{x}$,*

$$(x_1 - x_2)\left(\frac{\partial F}{\partial x_1} - \frac{\partial F}{\partial x_2}\right) \geq 0 \; . \tag{2.9}$$

Permutation invariance is needed because permutation matrices are (the only) bistochastic matrices that have bistochastic inverses. The full proof is not difficult when one uses T-transforms (Ando, 1989). Using Schur's theorem, and assuming

that $\vec{x}$ belongs to the positive orthant, we can easily write down a supply of Schur convex functions. Indeed any function of the form

$$F(\vec{x}) = \sum_{i=1}^{N} f(x_i) \tag{2.10}$$

is Schur convex, provided that $f(x)$ is a convex function on $\mathbb{R}$ (in the sense of Section 1.1). In particular, the l_p-norm of a vector is Schur convex. Schur concave functions include the *elementary symmetric functions*

$$s_2(\vec{x}) = \sum_{i<j} x_i x_j \ , \qquad s_3(\vec{x}) = \sum_{i<j<k} x_i x_j x_k, \tag{2.11}$$

and so on up to $s_N(\vec{x}) = \prod_i x_i$.

This concludes our tour of the majorization order and the transformations that go with it. Since we laid so much stress on bistochastic matrices, let us end with an interesting theorem that applies to stochastic matrices in general. We quote it in an abbreviated form (see Bhatia (1997) for more):

Theorem 2.3 (Frobenius–Perron's) *An irreducible $N \times N$ matrix whose matrix elements are non-negative has a real, positive and simple eigenvalue and a corresponding positive eigenvector. The modulus of any other eigenvalue never exceeds this.*

For a stochastic matrix the leading eigenvalue is 1 and all others lie in the unit disc of the complex plane – all stochastic matrices have an invariant eigenvector, although it need not sit in the centre of the probability simplex.

Stochastic maps are also called *Markov maps*. A *Markov chain* is a sequence of Markov maps and may be used to provide a discrete time evolution of probability distributions. It is not easy to justify why time evolution should be given by linear maps, but Markov maps are useful in a wide range of physical problems. In classical mechanics it is frequently assumed that time evolution is governed by a Hamiltonian. This forces the space of pure states to be infinite dimensional (unless one adopts finite number fields (Wootters, 1987)). Markovian evolution does not require this, so we will see a lot of Markov maps in this book. Another context in which stochastic matrices appear is the process of *coarse graining* (or *randomization*, as it is known in the statistical literature): suppose that we do not distinguish between two of three outcomes described by the probability distribution $\vec{p}$. Then we may map (p_0, p_1, p_2) to $(q_0, q_1) \equiv (p_0, p_1 + p_2)$ and this map is effected by a stochastic matrix. In this book we will be much concerned with functions that are monotonely increasing, or decreasing, under stochastic maps.

2.2 Shannon entropy

Let P be a probability distribution for a finite number N of possible outcomes, that is we have a vector $\vec{p}$ whose N components obey $p_i \geq 0$ and $\sum_i p_i = 1$. We ask for functions of the variables p_i that can tell us something interesting about the distribution. Perhaps surprisingly there is a single choice of such a function that (at least arguably) is more interesting than any other. This is the *Shannon entropy*

$$S(P) = -k \sum_{i=1}^{N} p_i \ln p_i \ , \tag{2.12}$$

where k is a positive number that we usually set equal to 1. Note that the entropy is associated to a definite random variable X and can be written as $S(X)$, but the only property of the random variable that matters is its probabilitity distribution P. The reason why it is called entropy was explained by Shannon in a disarming way:

> My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have an advantage." [3]

Entropy is a concept that has evolved a long way from its thermodynamic origins – and as so often happens a return to the origin may be difficult, but this will not concern us much in this book. The Shannon entropy can be interpreted as a measure of the uncertainty about the outcome of an experiment that is known to occur according to the probability distribution P, or as the amount of information needed to specify the outcome that actually occurs, or very precisely as the expected length of the communication needed for this specification. It takes the value zero if and only if one outcome is certain (that is for a pure state, when one component of $\vec{p}$ equals one) and its maximum value $k \ln N$ when all outcomes are equally likely (the uniform distribution). In this section only we set $k = 1/\ln 2$, which in effect means that we use logarithms to the base 2 in the definition of S:

$$S(P) = -\sum_{i=1}^{N} p_i \log_2 p_i \ . \tag{2.13}$$

With this choice the entropy is said to be measured in units of *bits*. If we have $N = 2^a$ possible outcomes we see that the maximum value of S is $\log_2 N = a$ bits,

[3] Quoted by Tribus and McIrvine (1971); Shannon's original work (Shannon, 1948) is available in book form as Shannon and Weaver (1949). A version of Eq. (2.12), with all the p_i equal to $1/W$, is engraved on Boltzmann's tombstone.

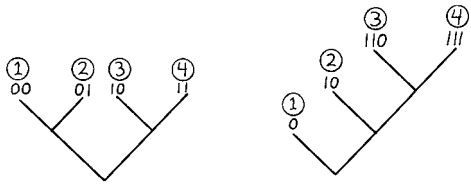


Figure 2.5. 'Code trees' for four possible outcomes. The one on the right gives a shorter expected length if $p_1 > p_3 + p_4$.

which is the length of the string of binary digits one can use to label the outcomes. If the outcomes occur with unequal probabilities it will pay, on the average, to label the more likely outcomes with shorter strings, and indeed the entropy goes down.

The context in which the Shannon entropy has an absolutely precise interpretation along these lines is when we have a *source* that produces outcomes of an infinite sequence of *independent and identically distributed* random variables. Suppose we want to code the outcomes of a set of such *i.i.d.* events as a string of binary digits (zeros and ones). We define a *code* as a map of a sequence of outcomes into strings of binary numbers. The coding has to be done in such a way that a given string has an unambiguous interpretation as a sequence of outcomes. Some examples will make this clear: if there are four possible outcomes then we can code them as 00, 01, 10 and 11, respectively. These are *code words*. Without ambiguity, the string 0100010100 then means (01, 00, 01, 01, 00). Moreover it is clear that the length of string needed to code one outcome is always equal to 2 bits. But other codes may do better, in the sense that the average number of bits needed to encode an outcome is less than 2; we are interested in codes that minimize the *expected length*

$$L = \sum_i p_i l_i \ , \tag{2.14}$$

where l_i is the length of the individual code words (in bits). In particular, suppose that $p_1 \geq p_2 \geq p_3 \geq p_4$, and label the first outcome as 0, the second as 10, the third as 110 and the fourth as 111. Then the string we had above is replaced by the shorter string 10010100, and again this can be broken down in only one way, as (10, 0, 10, 10, 0). That the new string is shorter was expected because we used a short code word to encode the most likely outcome, and we can read our string in an unambiguous way because the code has the *prefix property*: no code word is the prefix of any other. If we use both 0 and 1 as code words the prefix property is lost and we cannot code for more than two outcomes. See Figure 2.5 for the structure of the codes we use.

We are now faced with the problem of finding an *optimal code*, given the probability distribution. The expected length L_* of the code words used in an optimal code obeys $L_* \leq L$, where L is the expected length for an arbitrary code. We will

not describe the construction of optimal codes here, but once it is admitted that such a code can be found then we can understand the statement given in the next theorem:

Theorem 2.4 (Shannon's noiseless coding theorem) *Given a source distribution P, let L_* be the expected length of a code word used in an optimal code. Then*

$$S(P) \leq L_* \leq S(P) + 1 \, . \tag{2.15}$$

For the proof of this interesting theorem – and how to find an optimal code – we must refer to the literature (Cover and Thomas, 1991). Our point is that the noiseless coding theorem provides a precise statement about the sense in which the Shannon entropy is a measure of information. (It may seem as if L_* would be a better measure. It is not for two reasons. First, we do not have a closed expression for L_*. Second, as we will see below, when many code words are being sent the expected length per code word can be made equal to S.)

The Shannon entropy has many appealing properties. Let us make a list:

- **Positivity**. Clearly $S(P) \geq 0$ for all discrete probability distributions.
- **Continuity**. $S(P)$ is a continuous function of the distribution.
- **Expansibility**. We adopt the convention that $0 \ln 0 = 0$. Then it will be true that $S(p_1, \ldots, p_N) = S(p_1, \ldots, p_N, 0)$.
- **Concavity**. It is a concave function in the sense of convex set theory; mixing of probability distributions increases the entropy.
- **Schur concavity**. As explained in Section 2.1 the Shannon entropy is Schur concave, so it tells us something about the majorization order that we imposed on the set of all probability distributions.
- **Additivity.** If we have a joint probability distribution P_{12} for two random variables and if they are independent, so that the joint probabilities are products of the individual probabilities, then

  $$S(P_{12}) = S(P_1) + S(P_2). \tag{2.16}$$

  In words, the information needed to describe two independent random variables is the sum of the information needed to describe them separately.[4]
- **Subadditivity**. If the two random variables are not independent then

  $$S(P_{12}) \leq S(P_1) + S(P_2) \, , \tag{2.17}$$

  with equality if and only if the random variables are independent. Any correlation between them means that once we know the result of the first trial the amount of information needed to specify the outcome of the second decreases.
- **The recursion property**. Suppose that we coarse grain our description in the sense that we do not distinguish between all the outcomes. Then we are dealing with a new probability

[4] In the statistical mechanics community additivity is usually referred to as *extensitivity*.

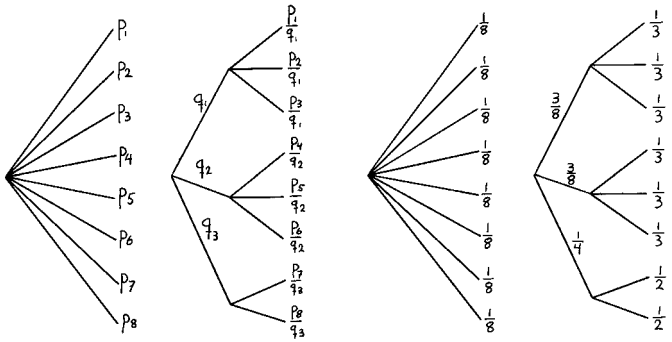


Figure 2.6. The recursion property illustrated; to the right it is used to determine $S(\frac{3}{8}, \frac{3}{8}, \frac{1}{4})$ in terms of the Shannon entropy for uniform distributions.

distribution Q with components

$$q_1 = \sum_{i=1}^{k_1} p_i \ , \qquad q_2 = \sum_{i=k_1+1}^{k_2} p_i \ , \qquad \dots \ , \qquad q_r = \sum_{i=N-k_r+1}^{N} p_i \ , \tag{2.18}$$

for some partition $N = k_1 + k_2 + \cdots + k_r$. It is easy to show that

$$S(P) = S(Q) + q_1 S\left(\frac{p_1}{q_1}, \dots, \frac{p_{k_1}}{q_1}\right) + \cdots + q_r S\left(\frac{p_{k_N-k_r+1}}{q_r}, \dots, \frac{p_N}{q_r}\right) \ . \tag{2.19}$$

This is the recursion property and it is illustrated in Figure 2.6; it tells us how a choice can be broken down into successive choices.

The recursion property can be used as an axiom that singles out the Shannon entropy as unique.[5] Apart from the recursion property we assume that the entropy is a positive and continuous function of P, so that it is enough to consider rational values of $p_i = m_i/M$, where $M = \sum_i m_i$. We also define

$$A(N) = S(1/N, 1/N, \dots, 1/N) \ . \tag{2.20}$$

Then we can start with the equiprobable case of M outcomes and, using Eq. (2.19), obtain $S(P)$ in an intermediate step:

$$A(M) = S(P) + \sum_{i=1}^{N} p_i A(m_i) \ . \tag{2.21}$$

In the special case that $M = Nm$ and all the m_i are equal to m we get

$$A(Nm) = A(N) + A(m) \ . \tag{2.22}$$

[5] Increasingly elegant sets of axioms for the Shannon entropy were given by Shannon and Weaver (1949), Khinchin (1957) and Faddejew (1957). Here we only give the key points of the argument.

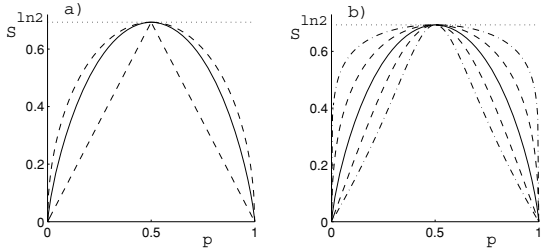


Figure 2.7. The Shannon entropy for $N = 2$ (solid curve). In panel (a) we show the simple power bounds (Eq. 2.24), and in panel (b) some bounds provided by the Rényi entropies from Section 2.7, upper bounds with $q = 1/5$ (dotted) and $1/2$ (dashed) and lower bounds with $q = 2$ (dashed, concave) and 5 (not concave).

The unique solution[6] of this equation is $A(N) = k \ln N$ and k must be positive if the entropy is positive. Shannon's formula for the entropy then results by solving Eq. (2.21) for $S(P)$.

The argument behind Shannon's coding theorem can also be used to define the Shannon entropy, provided that we rely on additivity as well. Let the source produce the outcomes of $\mathcal{N}$ independent events. These can be regarded as a single outcome for a random variable described by a joint probability distribution and must be provided by a code word. Let $L_*^{\mathcal{N}}$ be the expected length for the code word used by an optimal code. Because the Shannon entropy is additive for independent distributions we get

$$\mathcal{N} S(P) \leq L_*^{\mathcal{N}} \leq \mathcal{N} S(P) + 1 \,. \tag{2.23}$$

If we divide through by $\mathcal{N}$ we get the expected length per outcome of the original random variable – and the point is that this converges to $S(P)$, exactly. By following this line of reasoning we can give a formulation of the noiseless coding theorem that is more in line with that used later (in Section 12.2). But here we confine ourselves to the observation that, in this specific sense, the Shannon entropy is exactly what we want a measure of information to be!

Now what does the function look like? First we switch back to using the natural logarithm in the definition (we set $k = 1$), then we take a look at the Shannon entropy for $N = 2$. Figure 2.7 also shows that

$$2 \ln 2 \, \min\{x, 1 - x\} \leq S(x) \leq 2 \ln 2 \sqrt{x(1 - x)} \,. \tag{2.24}$$

Much sharper power (and logarithmic) bounds have been provided by Topsøe (2001). Bounds valid for $N = 2$ only are more interesting than one might think, as we will see when we prove the Pinsker inequality in Section 13.1.

[6] This is the tricky point: see Rényi (1961) for a comparatively simple proof.

2.3 Relative entropy

The Shannon entropy is a function that tells us something about a probability distribution. We now consider two probability distributions P and Q, and we look for a measure of how different they are – a distance between them, or at least a *distinguishability measure*. Information theory offers a favoured candidate: the *relative entropy*

$$S(P||Q) = \sum_{i=1}^{N} p_i \ln \frac{p_i}{q_i} \,. \tag{2.25}$$

It is also known as the *Kullback–Leibler entropy* or the *information divergence*.[7] In terms of lengths of messages, the relative entropy $S(P||Q)$ measures how much the expected length of a code word grows, if the coding is optimal but made under the erroneous assumption that the random variable is described by the distribution Q. Relative entropy is not a distance in the sense of Section 1.4 since it is not symmetric under interchange of P and Q. Nevertheless it does provide a measure of how different the two distributions are from each other. We should be clear about the operational context in which a given distinguishability measure is supposed to work; here we assume that we have a source of independent and identically distributed random variables, and we try to estimate the probability distribution by observing the frequency distribution of the various outcomes. The *Law of Large Numbers* guarantees that in the limit when the number $\mathcal{N}$ of samplings goes to infinity the probability to obtain any given frequency distribution not equal to the true probability distribution vanishes. It turns out that the relative entropy governs the rate at which this probability vanishes as $\mathcal{N}$ increases.

A typical situation that one might want to consider is that of an experiment with N outcomes, described by the probability distribution Q. Let the experiment be repeated $\mathcal{N}$ times. We are interested in the probability $\mathcal{P}$ that a frequency distribution P (different from Q) will be observed. To simplify matters, assume $N = 2$. For instance, we may be flipping a biased coin described by the (unknown) probability distribution $Q = (q, 1-q)$. With Jakob Bernoulli, we ask for the probability $\mathcal{P}$ that the two outcomes occur with frequencies $P = (m/\mathcal{N}, 1 - m/\mathcal{N})$. This is a matter of counting how many strings of outcomes there are with each given frequency, and the exact answer is

$$\mathcal{P}\left(\frac{m}{\mathcal{N}}\right) = \binom{\mathcal{N}}{m} q^m (1-q)^{\mathcal{N}-m} \,. \tag{2.26}$$

[7] The relative entropy was introduced by Kullback and Leibler (1951), and even earlier than that by Jeffreys in 1939 (Jeffreys, 1961). Sanov's theorem (Sanov, 1957) appeared in 1957. In mathematical statistics an asymmetric distance measure is referred to as a divergence.

For large $\mathcal{N}$ we can approximate this using Stirling's formula, in the simplified form $\ln \mathcal{N}! \approx \mathcal{N} \ln \mathcal{N} - \mathcal{N}$. For $\mathcal{N} \approx 100$ it is accurate within a per cent or so. The result is

$$\ln\left(\mathcal{P}(m/\mathcal{N})\right) \approx -\mathcal{N}\left[\frac{m}{\mathcal{N}}\left(\ln\frac{m}{\mathcal{N}} - q\right) + (1 - \frac{m}{\mathcal{N}})\left(\ln(1 - \frac{m}{\mathcal{N}}) - \ln(1-q)\right)\right] \tag{2.27}$$

and we recognize the relative entropy on the right-hand side. Indeed the probability to obtain the frequency distribution P is

$$\mathcal{P}(P) \approx \mathrm{e}^{-\mathcal{N}S(P||Q)} . \tag{2.28}$$

This simple exercise should be enough to advertise the fact that relative entropy means something.

There are more precise ways of formulating this conclusion. We have the following theorem:

Theorem 2.5 (Sanov's) *Let an experiment with N outcomes described by the probability distribution Q be repeated $\mathcal{N}$ times, and let E be a set of probability distributions for N outcomes such that E is the closure of its interior. Then, for $\mathcal{N}$ large, the probability $\mathcal{P}$ that a frequency distribution belonging to E will be obtained is*

$$\mathcal{P}(E) \sim \mathrm{e}^{-\mathcal{N}S(P_*||Q)} , \tag{2.29}$$

where P_* is that distribution in E that has the smallest value of $S(P||Q)$.

That E is the closure of its interior means that it is a 'nice' set, without isolated points, 'spikes', and so on. An important ingredient in the proof is that the number of frequency distributions that may be realized is bounded from above by $(\mathcal{N}+1)^N$, since any one of the N outcomes can occur at most $\mathcal{N}+1$ times. On the other hand the number of possible ordered strings of outcomes grows exponentially with $\mathcal{N}$. But this is all we have to say about the proof; our main concern is the geometrical significance of relative entropy, not to prove the theorems of information theory.[8]

The relative entropy is always a positive quantity. In fact

$$S(P||Q) = \sum_{i=1}^{N} p_i \ln\frac{p_i}{q_i} \;\geq\; \frac{1}{2}\sum_{i=1}^{N}(p_i - q_i)^2 = D_2^2(P, Q) . \tag{2.30}$$

To prove this, note that any smooth function f obeys

$$f(x) = f(y) + (x-y)f'(y) + \frac{1}{2}(x-y)^2 f''(\xi) , \qquad \xi \in (x, y) . \tag{2.31}$$

[8] All unproved assertions in this section are proved in the book by Cover and Thomas (1991). A number of further results can be found in their chapter 10.

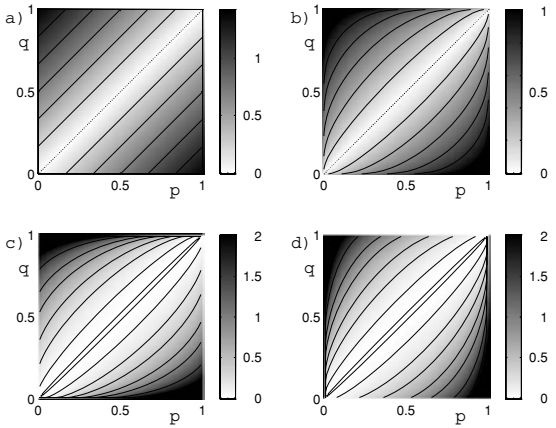


Figure 2.8. Measures of distinguishability between probability distributions $P = (p, 1-p)$ and $Q = (q, 1-q)$: (a) Euclidean distance, (b) Bhattacharyya distance (see Section 2.5), (c) and (d) relative entropies $S(P||Q)$ and $S(Q||P)$.

Now set $f(x) = -x \ln x$, in which case $f''(x) \leq -1$ when $0 \leq x \leq 1$. A little rearrangement proves that the inequality (2.30) holds term by term, so that this is not a very sharp bound.

The relative entropy is asymmetric. That this is desirable can be seen if we consider two coins, one fair and one with heads on both sides. Start flipping one of them to find out which is which. A moment's thought shows that if the coin we picked is the fair one, this will most likely be quickly discovered, whereas if we picked the other coin we can never be completely sure – after all the fair coin can give a sequence of a thousand heads, too. If statistical distance is a measure of how difficult it will be to distinguish two probability distributions in a repeated experiment then this argument shows that such a measure must have some asymmetry built into it. Indeed, using notation from Eq. (1.60),

$$S(Q_{(N)}||Q_{(1)}) = \infty \quad \text{and} \quad S(Q_{(1)}||Q_{(N)}) = \ln N \; . \tag{2.32}$$

The asymmetry is pronounced. For further intuition about this, consult Section 13.1. The asymmetry can also be studied in Figures 2.8(c) and (d) for $N = 2$ and in Figure 2.9 for $N = 3$.

There is a kind of 'Pythagorean theorem' that supports the identification of relative entropy as a (directed) distance squared:

Theorem 2.6 ('Pythagorean') *The distribution P lies in a convex set E and Q lies outside E. Choose a distribution P_* on the boundary of the convex set such that*

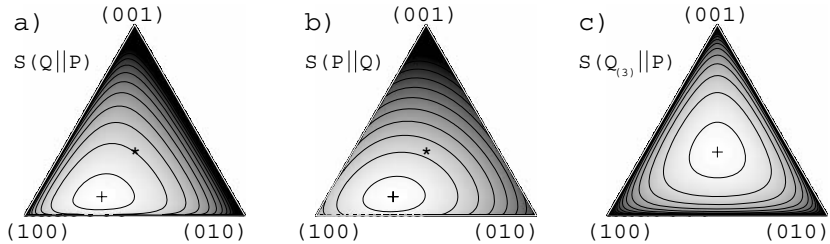


Figure 2.9. Contour lines of the relative entropy $S(Q||P)$ (a) and $S(P||Q)$ for $Q = (0.6, 0.4, 0.1)$ (+) as a function of P. Panel (c) shows an analogous picture for $S(Q_{(3)}||P)$, while the contours of $S(P||Q_{(3)})$ consisting of points of the same Shannon entropy are shown in Figure 2.14(c).

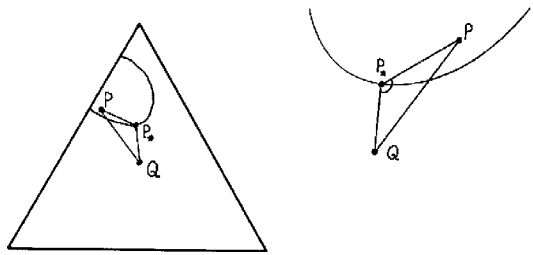


Figure 2.10. A Pythagorean property of relative entropy (left), and the same property in Euclidean geometry where $D^2_{PQ} \geq D^2_{PP_*} + D^2_{P_*Q}$. The angle is obtuse because the set to which P belongs is convex (right).

$S(P_*||Q)$ assumes its minimum value for P_* belonging to E and fixed Q. Then

$$S(P||Q) \geq S(P||P_*) + S(P_*||Q)\,. \tag{2.33}$$

To prove this, consider distributions on the straight line $P_\lambda = \lambda P + (1-\lambda)P_*$, which lies inside E because E is convex. A minor calculation shows that

$$\lambda = 0 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}\lambda} S(P_\lambda||Q) = S(P||Q) - S(P||P_*) - S(P_*||Q)\,. \tag{2.34}$$

But the derivative cannot be negative at $\lambda = 0$, by the assumption we made about P^*, and the result follows. It is called a Pythagorean theorem because, under the same conditions, if we draw Euclidean straight lines between the points and measure the Euclidean angle at P_* then that angle is necessarily obtuse once P_* has been chosen to minimize the Euclidean distance from the convex set to Q. If D_{PQ} denotes the Euclidean distance between the points then Pythagoras' theorem in its usual formulation states that $D^2_{PQ} \geq D^2_{PP_*} + D^2_{P_*Q}$. In this sense the relative entropy really behaves like a distance squared.

The convexity properties of the relative entropy are interesting. It is *jointly convex*. This means that when $\lambda \in [0, 1]$ it obeys

$$S(\lambda P_1 + (1-\lambda)P_2 || \lambda Q_1 + (1-\lambda)Q_2) \leq \lambda S(P_1||Q_1) + (1-\lambda) S(P_2||Q_2) \,. \tag{2.35}$$

Convexity in each argument separately is an easy consequence (set $Q_1 = Q_2$). Moreover relative entropy behaves in a characteristic way under Markov maps. Let T be a stochastic map (that is, a matrix that takes a probability vector to another probability vector). Then we have the following property:

- **Monotonicity under stochastic maps**.

$$S(TP||TQ) \;\leq\; S(P||Q) \,. \tag{2.36}$$

In words, the relative entropy between two probability distributions always decreases in a Markov chain, so that the distinguishability of two distributions decreases with time. It is really this property that makes relative entropy such a useful concept; functions of pairs of probability distributions that have this property are known as *monotone* functions. The increase of entropy follows if T is also bistochastic, that is to say that it leaves the uniform probability distribution invariant. The uniform distribution is denoted $Q_{(N)}$. Clearly

$$S(P||Q_{(N)}) = \sum_i p_i \ln (N p_i) = \ln N - S(P) \,. \tag{2.37}$$

Since $T Q_{(N)} = Q_{(N)}$ the decrease of the relative entropy implies that the Shannon entropy of the distribution P is increasing with time; we are getting closer to the uniform distribution. (And we have found an interesting way to show that $S(P)$ is a Schur concave function.)

The entropy increase can also be bounded from above. To do so we first define the entropy of a stochastic matrix T with respect to some fixed probability distribution P as

$$S_P(T) \equiv \sum_{i=1}^{N} p_i S(\vec{T}_i) \,; \qquad \vec{T}_i = (T_{1i}, T_{2i}, \ldots, T_{Ni}) \,. \tag{2.38}$$

(Thus $\vec{T}_i$ are the column vectors of the matrix, which are probability vectors because T is stochastic.) One can prove that (Słomczyński, 2002)

$$S_P(T) \;\leq\; S(TP) \;\leq\; S_P(T) + S(P) \,. \tag{2.39}$$

Again, this is a result that holds for all stochastic matrices.

2.4 Continuous distributions and measures

The transition from discrete to continuous probability distributions appears at first sight to be trivial – just replace the sum with an integral and we arrive at the *Boltzmann entropy*

$$S_B = -\int_{-\infty}^{\infty} \mathrm{d}x\ p(x) \ln p(x)\ . \tag{2.40}$$

This is a natural generalization of Shannon's entropy to continuous probability distributions, but depending on the behaviour of the distribution function $p(x)$ it may be ill-defined, and it is certainly not bounded from below in general, since $p(x)$ need not be bounded from above. To see this let $p(x)$ be a step function taking the value t^{-1} for $x \in [0, t]$ and zero elsewhere. The entropy S_B is then equal to $\ln t$ and goes to negative infinity as $t \to 0$. In particular a pure classical state corresponds to a delta function distribution and has $S_B = -\infty$. In a way this is as it should be – it takes an infinite amount of information to specify such a state exactly.

The next problem is that there is no particular reason why the entropy should be defined in just this way. It is always a delicate matter to take the continuum limit of a discrete sum. In fact the formula as written will change if we change the coordinate that we are using to label the points on the real line. In the continuous case random variables are scalars, but the probability density transforms like a density. What this means is that

$$\langle A \rangle = \int \mathrm{d}x\ p(x)A(x) = \int \mathrm{d}x'\ p'(x')A'(x')\ , \tag{2.41}$$

where $A'(x') = A(x)$. Hence $p'(x') = J^{-1}p(x)$, where J is the Jacobian of the coordinate transformation. This means that the logarithm in our definition of the Boltzmann entropy will misbehave under coordinate transformations. There is a simple solution to this problem, consisting in the observation that the relative entropy with respect to some *prior* probability density $m(x)$ does behave itself. In equations,

$$S_B = -\int_{-\infty}^{\infty} \mathrm{d}x\ p(x) \ln \frac{p(x)}{m(x)} \tag{2.42}$$

is well behaved. The point is that the quotient of two densities transforms like a scalar under coordinate changes. Now, depending on how we take the continuum limit of the Shannon entropy, different densities $m(x)$ will arise. Finding an appropriate $m(x)$ may be tricky, but this should not deter us from using continuous distributions when necessary. We cannot avoid them in this book, because continuous distributions occur also in finite-dimensional quantum mechanics. In most cases we will simply use the 'obvious' translation from sums to integrals along the lines of Eq. (2.40),

but it should be kept in mind that this really implies a special choice of prior density $m(x)$.

At this point we begin to see why mathematicians regard probability theory as a branch of measure theory. We will refer to expressions such as $\mathrm{d}\mu(x) \equiv p(x)\mathrm{d}x$ as *measures*. Mathematicians tend to use a slightly more sophisticated terminology at this point. Given a suitable subset A of the space coordinatized by x, they define the measure as a real valued function μ such that

$$\mu(A) \equiv \int_A \mathrm{d}\mu(x) \,. \tag{2.43}$$

The measure measures the volume of the subset. This terminology can be very helpful, but we will not use it much.

When the sample space is continuous the space of probability distributions is infinite dimensional. But in many situations we may be interested in a finite-dimensional subset. For instance, we may be interested in the two-dimensional submanifold of *normal distributions*

$$p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \,, \tag{2.44}$$

with the mean μ and the standard deviation σ serving as coordinates in the submanifold. In general a finite-dimensional submanifold of the space of probability distributions is defined by the function $p(x;\theta^1,\ldots,\theta^n)$, where θ^a are coordinates in the submanifold. When we think of this function as a function of θ^a it is known as the *likelihood function*. This is a setup encountered in the theory of *statistical inference*, that is to say the art of guessing what probability distribution governs an experiment, given only the results of a finite number of samplings, and perhaps some kind of prior knowledge. (The word 'likelihood' is used because, to people of resolutely frequentist persuasions, there can be no probability distribution for θ^a.) The likelihood function is a hypothesis about the form that the probability distribution takes. This hypothesis is to be tested, and we want to design experiments that allow us to make a 'best guess' for the values of the parameters θ^a. As a first step we want to design a statistical geometry that is helpful for this kind of question.

A somewhat idiosyncratic notation is used in the statistical literature to represent tangent spaces. Consider the *log-likelihood function*

$$l(x;\theta) = \ln p(x;\theta) \,. \tag{2.45}$$

Here x denotes coordinates on the sample space Ω and θ are coordinates on a subspace of probability distributions. The idea is to describe tangent space through the natural isomorphism between its basis vectors ∂_a (used in Section 1.4) and the *score vectors* l_a, which are the derivatives with respect to the coordinates θ^a of the

log-likelihood function. That is to say

$$\partial_a \equiv \frac{\partial}{\partial \theta^a} \quad \leftrightarrow \quad l_a \equiv \frac{\partial l}{\partial \theta^a} \,. \tag{2.46}$$

For this to make any sense we must assume that the score vectors form a set of n linearly independent functions. (It is not supposed to be self-evident why one should use precisely the log-likelihood function here, but it is an interesting function to consider – the Shannon entropy is nothing but its expectation value.) The expectation values of the score vectors vanish. To see this recall that the expectation value is a sum over the sample space which we denote schematically as an integral:

$$\langle l_a \rangle \equiv \int_\Omega \mathrm{d}x \; p(x;\theta) l_a(x;\theta) = \int_\Omega \mathrm{d}x \; \partial_a p(x;\theta) = \frac{\partial}{\partial \theta^a} \int_\Omega \mathrm{d}x \; p(x;\theta) = 0 \,. \tag{2.47}$$

It is assumed that the interchange of the order of integration with respect to x and differentiation with respect to θ is allowed. A general tangent vector $A(x) = A^a l_a(x)$ is a linear combination of score vectors; as such it is a random variable with expectation value zero. These at first sight peculiar definitions actually achieve something: sets of probability distributions are turned into manifolds, and random variables are turned into tangent vectors. In the next section we will see that scalar products between tangent vectors have a statistical meaning, too.

2.5 Statistical geometry and the Fisher–Rao metric

One way of introducing a metric on the space of probability distribution is to consider the situation where a large number $\mathcal{N}$ of samplings are made from a given probability distribution P, at first assumed to be discrete. This case is sufficiently simple so that we can do all the details. We ask if we can find out what P actually is, given only the result of the $\mathcal{N}$ samplings. We considered this problem in Section 2.3; when there are only two possible outcomes the probability to obtain the frequencies $(f_1, f_2) = (m/\mathcal{N}, 1 - m/\mathcal{N})$ is

$$\mathcal{P}\left(\frac{m}{\mathcal{N}}\right) = \binom{\mathcal{N}}{m} p^m (1-p)^{\mathcal{N}-m} \,. \tag{2.48}$$

We give the answer again because this time we denote the true probability distribution by $P = (p, 1-p)$. When $\mathcal{N}$ is large, we can use Stirling's formula to approximate the result; in textbooks on probability it is usually given as

$$\mathcal{P}\left(\frac{m}{\mathcal{N}}\right) = \frac{1}{\sqrt{2\pi \mathcal{N} p(1-p)}} \, \mathrm{e}^{-\frac{\mathcal{N}}{2} \frac{(\frac{m}{\mathcal{N}} - p)^2}{p(1-p)}} = \frac{1}{\sqrt{2\pi \mathcal{N} p_1 p_2}} \, \mathrm{e}^{-\frac{\mathcal{N}}{2} \sum_{i=1}^2 \frac{(f_i - p_i)^2}{p_i}} \,. \tag{2.49}$$

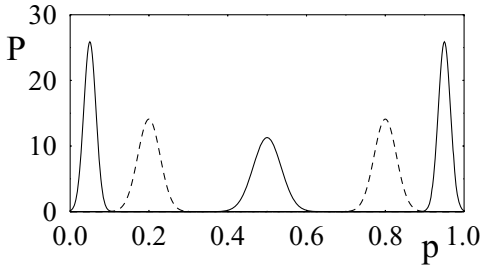


Figure 2.11. Bernoulli's result: the shapes of the Gaussians will determine how easily points can be distinguished.

This is consistent with Eq. (2.28) when the observed frequencies are close to P. The error that we have committed is smaller than $k/\mathcal{N}$, where k is some real number independent of $\mathcal{N}$; we assume that $\mathcal{N}$ is large enough so that this does not matter. Figure 2.11 shows that the Gaussians are strongly peaked when the outcomes of the experiment are nearly certain, while they are much broader in the middle of the probability simplex. Statistical fluctuations are small when we are close to pure states; in some sense the visibility goes up there.

The result is easily generalized to the case of N outcomes; we obtain the normal distribution

$$\mathcal{P}(F) \propto \mathrm{e}^{-\frac{\mathcal{N}}{2}\sum_{i=1}^{N}\frac{(f^i-p^i)^2}{p^i}} \,. \tag{2.50}$$

We raised all indices, because we will be doing differential geometry soon! Our question is: given only the frequencies f^i, do we dare to claim that the probabilities take the values $p^i = f^i$, rather than some other values q^i? This is clearly a matter of taste. We may choose some number ϵ and decide that we dare to do it if the probability vector $\vec{q}$ lies outside an ellipsoid centred at the point $\vec{p}$ in the probability simplex, consisting of all probability vectors of the form $\vec{p} + \mathrm{d}\vec{p}$, where

$$\sum_{i=1}^{N} \frac{\mathrm{d}p^i \mathrm{d}p^i}{p^i} \leq \epsilon^2 \,. \tag{2.51}$$

The analogy to the MacAdam ellipses of colour theory should be clear – and if ϵ is small we are doing differential geometry already. The vector $\mathrm{d}\vec{p}$ is a tangent vector, and we introduce a metric tensor g_{ij} through the equation

$$\mathrm{d}s^2 = \sum_{i,j} g_{ij}\mathrm{d}p^i\mathrm{d}p^j = \frac{1}{4}\sum_{i=1}^{N}\frac{\mathrm{d}p^i\mathrm{d}p^i}{p^i} \quad \Leftrightarrow \quad g_{ij} = \frac{1}{4}\frac{\delta_{ij}}{p^i} \,. \tag{2.52}$$

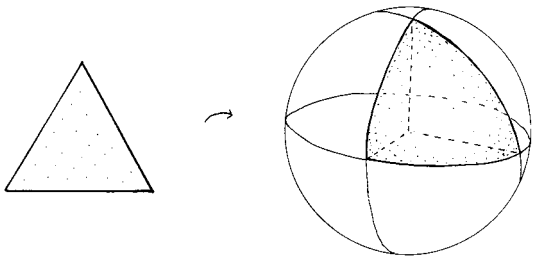


Figure 2.12. The convex (flat) and statistical (round) geometry of a simplex when $N = 3$.

(For clarity, we do not use Einstein's summation convention here.) This Riemannian metric is known as the *Fisher–Rao metric*, or as the *Fisher information matrix*.[9]

What is it? To see this, we will try to deform the probability simplex into some curved space, in such a way that all the little ellipsoids become spheres of equal size. A rubber model of the two-dimensional case would be instructive to have, but since we do not have one let us do it with equations. We introduce new coordinates X^i , all of them obeying $X^i \geq 0$, through

$$X^i = \sqrt{p^i} \quad \Rightarrow \quad \mathrm{d}X^i = \frac{\mathrm{d}p^i}{2\sqrt{p^i}} \,. \tag{2.53}$$

Then the Fisher–Rao metric takes the very simple form

$$\mathrm{d}s^2 = \sum_{i=1}^{N} \mathrm{d}X^i \mathrm{d}X^i \,. \tag{2.54}$$

All the little error ellipsoids have become spheres. If we remember the constraint

$$\sum_{i=1}^{N} p^i = \sum_{i=1}^{N} X^i X^i = 1 \,, \tag{2.55}$$

then we see that the geometry is that of a unit sphere. Because of the restricted coordinate ranges we are in fact confined to the positive hyperoctant of $\mathbf{S}^{N-1}$. Figure 2.12 illustrates the transformation that we made.

Section 3.1 will provide more details about spheres if this is needed. But we can right away write down the geodesic distance D_{Bhatt} between two arbitary probability distributions P and Q, now regarded as the length of the great circle between them. This is simply the angle between two vectors, with components $X^i = \sqrt{p^i}$ and

[9] The factor of 1/4 is needed only for agreement with conventions that we use later on. The Fisher–Rao metric was introduced by Rao (1945), based on earlier work by Fisher (1925) and Bhattacharyya (1943).

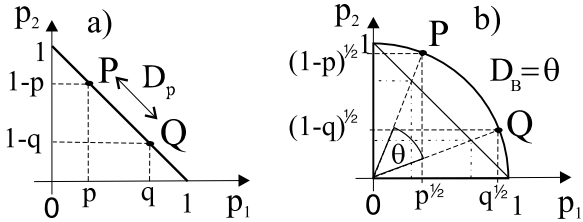


Figure 2.13. For $N = 2$ the space of discrete probabilities reduces to (a) a straight line segment along which all the l_p distances agree, or (b) a quarter of a circle equipped with the Bhattacharyya distance, measured by the angle θ.

$Y^i = \sqrt{q^i}$. Hence

$$\cos D_{\text{Bhatt}} = \sum_{i=1}^{N} \sqrt{p^i q^i} = B(P, Q) \,. \tag{2.56}$$

In statistics this distance is known as the *Bhattacharyya distance*. The right-hand side of the equation is known as the *Bhattacharyya coefficient*, and it looks intriguingly similar to the scalar product in quantum mechanics. Its square is known as the *classical fidelity*. Alternatively we can define the chordal distance D_{H} as the distance between the points in the flat embedding space where the round hyperoctant sits, that is

$$D_{\text{H}} = \left(\sum_{i=1}^{N} \left(\sqrt{p^i} - \sqrt{q^i} \right)^2 \right)^{\frac{1}{2}} \,. \tag{2.57}$$

In statistics this is known as the *Hellinger distance*. It is clearly a monotone function of the Bhattacharyya distance.

In writing down these distances we have allowed the geometry to run ahead of the statistics. The infinitesimal Fisher–Rao distance does have a clear cut operational significance, in terms of the statistical distinguishability of nearby probability distributions, in the limit of a large number of samplings. But this is not really the case for the finite Bhattacharyya and Hellinger distances, at least not as far as we know. The Kullback–Leibler relative entropy is better in this respect, and should be connected to the Fisher–Rao metric in some way. Let us assume that two probability distributions are close, and let us expand the relative entropy (Eq. 2.25) to lowest order. We find

$$S(P||P + \mathrm{d}P) = \sum_i p^i \ln \frac{p^i}{p^i + \mathrm{d}p^i} \approx \frac{1}{2} \sum_i \frac{\mathrm{d}p^i \mathrm{d}p^i}{p^i} \,. \tag{2.58}$$

Expanding $S(P + \mathrm{d}P||P)$ gives the same result – the asymmetry present in the relative entropy does not make itself felt to lowest order. So the calculation simply confirms the feeling that developed in Section 2.3, namely that relative entropy should be thought of as a distance squared. Infinitesimally, this distance is precisely that defined by the Fisher–Rao metric. Note that we can also regard the Fisher–Rao metric as the Hessian matrix of the Shannon entropy:

$$g_{ij} = -\frac{1}{4}\partial_i \partial_j S(p) = \frac{1}{4}\partial_i \partial_j \sum_{k=1}^{N} p^k \ln p^k \,. \tag{2.59}$$

The fact that Shannon entropy is concave implies that the Fisher–Rao metric is positive definite. In Section 3.2 we will tell the story backwards and explain that relative entropy exists precisely because there exists a 'potential' whose second derivatives form the components of the Riemannian metric.

As we made clear in Section 1.4 there are many ways to introduce the concept of distance on a given space, and we must ask in what sense, if any, our concept of statistical distance is unique? There is a clear cut answer to this question. We will consider maps between two probability distributions P and Q of the form $Q = TP$, where T is a stochastic matrix. We define a *monotone* function as a function of pairs of probability distributions such that

$$f(TP, TQ) \;\leq\; f(P, Q) \,. \tag{2.60}$$

Then there is a comfortable result:

Theorem 2.7 (Čencov's) *For multinomial distributions the Fisher–Rao metric is (up to normalization) the only metric whose geodesic distance is a monotone function.*

In this desirable sense the Fisher–Rao metric is unique. It is not hard to see that the Fisher–Rao metric has the desired property. We must show that the length of an arbitrary tangent vector $\mathrm{d}p^i$ at the point p^i decreases under a stochastic map, that is to say that

$$||T\,\mathrm{d}p||^2 \leq ||\mathrm{d}p||^2 \quad \Leftrightarrow \quad \sum_{i=1}^{N} \frac{(T\,\mathrm{d}p)^i (T\,\mathrm{d}p)^i}{(Tp)^i} \;\leq\; \sum_{i=1}^{N} \frac{\mathrm{d}p^i \mathrm{d}p^i}{p^i} \,. \tag{2.61}$$

To do so, we first prove that

$$\left(\sum_j T_{ij}\mathrm{d}p_j\right)^2 \;\leq\; \left(\sum_j T_{ij}\frac{\mathrm{d}p_j \mathrm{d}p_j}{p_j}\right)\left(\sum_j T_{ij}p_j\right) \,. \tag{2.62}$$

Looking closely at this expression (take square roots!) we see that it is simply the Cauchy–Schwarz inequality in disguise. Dividing through by the rightmost factor

and summing over i we find – precisely because $\sum_i T_{ij} = 1$, that is because the map is stochastic – that Eq. (2.61) follows.[10]

It often happens that one is interested in some submanifold of the probability simplex, coordinatized by a set of coordinates θ^a. The Fisher–Rao metric will induce a metric on such submanifolds, given by

$$g_{ab} = \sum_{i,j} \frac{\partial p^i}{\partial \theta^a} \frac{\partial p^j}{\partial \theta^b} g_{ij} = \frac{1}{4} \sum_i \frac{\partial_a p^i \partial_b p^i}{p^i} . \tag{2.63}$$

As long as the submanifold – the *statistical model*, as it may be referred to here – is finite dimensional, this equation is easily generalized to the case of continuous probability distributions $p(x)$, where the probability simplex itself is infinite dimensional:

$$g_{ab} = \frac{1}{4} \int_\Omega \mathrm{d}x \, \frac{\partial_a p \partial_b p}{p} . \tag{2.64}$$

This metric is unaffected by reparametrizations of the sample space, that is to say by changes to new coordinates $x'(x)$ – and indeed obviously so since the derivatives are with respect to θ.

The odd-looking notation using score vectors comes into its own here. Using them we can rewrite Eq. (2.64) as

$$g_{ab} = \frac{1}{4} \int_\Omega \mathrm{d}x \; p l_a l_b \equiv \frac{1}{4} \langle l_a l_b \rangle . \tag{2.65}$$

The components of the metric tensor are the scalar products of the basis vectors in tangent space. A general tangent vector is, in the language we now use, a random variable with vanishing mean, and the scalar product of two arbitrary tangent vectors becomes

$$\frac{1}{4} \langle A(x) B(x) \rangle = \frac{1}{4} \int_\Omega \mathrm{d}x \; p(x) A(x) B(x) . \tag{2.66}$$

Except for the annoying factor of $1/4$, this is the *covariance* of the two random variables.

In order to get used to these things let us compute the Fisher–Rao metric on the two-dimensional family of normal distributions, with the mean μ and variance σ as coordinates. See Eq. (2.44). We have two score vectors

$$l_\mu = \partial_\mu \ln p(x; \mu, \sigma) = \frac{x - \mu}{\sigma^2} \quad l_\sigma = \partial_\sigma \ln p(x; \mu, \sigma) = \frac{(x - \mu)^2}{\sigma^3} - \frac{1}{\sigma} . \tag{2.67}$$

[10] We do not prove uniqueness here, but in Section 13.1 we show that the flat metric is not monotone. Unfortunately Čencov's proof is difficult (Čencov, 1982); for an accessible proof of his key point see Campbell (1986). By the way, Čencov will also appear in our book under the name of Chentsov.

Taking the appropriate expectation values we find (after a short calculation) that the statistical metric on the normal family is

$$\mathrm{d}s^2 = \frac{1}{4\sigma^2}(\mathrm{d}\mu^2 + 2\mathrm{d}\sigma^2)\,. \tag{2.68}$$

This is a famous metric of constant negative curvature, known as the *Poincaré metric* on the upper half plane. The μ-axis itself consists of the pure states, and points there are infinitely far away from any point in the interior of the upper half plane; if the outcome is certain (so that the standard deviation σ is zero) then it is trivial to distinguish this distribution from any other one.

As our parting shot in this section, consider a statistical inference problem where we want to find the values of the parameters θ^a from samplings of some random variable. A random variable ξ^a is said to be an *unbiased estimator* of the parameter θ^a if $\langle \xi^a \rangle = \theta^a$. So, there are no systematic errors in the experiment. Still, there is a limit on how well the unbiased estimators can perform their task:

Theorem 2.8 (Cramér–Rao's) *The variance of an unbiased estimator obeys*

$$\langle \xi^a \xi^b \rangle - \langle \xi^a \rangle \langle \xi^b \rangle \geq \frac{1}{4} g^{ab}\,. \tag{2.69}$$

The inequality means that the left-hand side minus the right-hand side is a positive semi-definite matrix.[11]

2.6 Classical ensembles

Let us define an *ensemble* as a set equipped with a probability measure. Choosing a probability simplex as our set, we obtain an ensemble of probability distributions, or states, characterized by a distribution $P(\vec{p})$. The physical situation may be that of a machine producing an unlimited set of independent and identically distributed copies of a physical system each characterized by the same unknown probability distribution $\vec{p}$, in which case $P(\vec{p})$ tells us something about the machine, or about our knowledge of the machine.

We seem to be on the verge of breaking our promise not to commit ourselves to any particular interpretation of probability, since we appear to imply that the original probability $\vec{p}$ is an inherent property of the individual physical systems that are being produced. We have not introduced any agent whose degree of reasonable belief can reasonably be said to be quantified by $\vec{p}$. Closer inspection reveals that the subjective interpretation of probability can be saved. To do so one observes that there is really only one thing that we are ignorant of, namely the full sequence of

[11] We have left much unsaid here. Amari (1985) contains a highly recommended book length discussion of statistical geometry. We also recommend the review by Ingarden (1981).

observations. But beyond that we make use of some prior assumptions that specify exactly what we mean by 'independent and identically distributed copies'. Then *de Finetti's theorem* tells us that this situation can be treated as if it was described by the distribution $P(\vec{p})$ on the probability simplex. This is a subtle point, and we refer to the literature for a precise account of our ignorance.[12]

We have now explained one setting in which a distribution $P(\vec{p})$ arises. Others can be imagined. From the subjective point of view, the prior should summarize whatever information we have before the data have been analysed, and it should be systematically updated as the observations are coming in. The prior should obey one important requirement: it should lead to manageable calculations in practice. What we want to do here is much simpler. We ask what we should expect to observe if the state of a physical system is picked 'at random'. A *random state* is a state picked from the probability simplex according to some measure, and that measure is now to be chosen in a way that captures our idea of randomness. Hence we ask for the *uniform prior* corresponding to complete ignorance. But we are using an alarmingly vague language. What is complete ignorance?

Presumably, the uniform prior is such that the system is equally likely to be found anywhere on the simplex, but there are at least two natural candidates for what this should mean. We could say that the simplex is flat. Then we use the measure

$$\mathrm{d}P_\Delta = (N-1)!\,\delta\left(\sum_{i=1}^{N} p_i - 1\right)\mathrm{d}p_1\mathrm{d}p_2\ldots\mathrm{d}p_N\ . \tag{2.70}$$

(At the cost of a slight inconsistency, we have lowered the index on p_i again. We find this more congenial!) This probability measure is correctly normalized, because we get one when we integrate over the positive cone:

$$\begin{aligned}\int_{\mathbb{R}_+^N}\mathrm{d}P_\Delta &= (N-1)!\int_0^1\mathrm{d}p_1\int_0^{1-p_1}\mathrm{d}p_2\ldots\int_0^{1-p_1-\cdots-p_{N-2}}\mathrm{d}p_{N-1}\\ &= \int_0^1\mathrm{d}p_1\int_0^1\mathrm{d}p_2\ldots\int_0^1\mathrm{d}p_{N-1} = 1\ .\end{aligned} \tag{2.71}$$

The volume of the simplex is $1/(N-1)!$ times the volume of the parallelepiped spanned by its edges.

But if we use the Fisher–Rao rather than the flat metric, then the simplex is round rather than flat, and the appropriate density to integrate is proportional to the square root of the determinant of the Fisher–Rao metric. The measure becomes

$$\mathrm{d}P_{\mathrm{FR}} = \frac{\Gamma\left(\frac{N}{2}\right)}{\pi^{N/2}}\,\delta\left(\sum_{i=1}^{N} p_i - 1\right)\frac{\mathrm{d}p_1\mathrm{d}p_2\ldots\mathrm{d}p_N}{\sqrt{p_1p_2\ldots p_N}}\ . \tag{2.72}$$

[12] For an engaging account both backwards (with references to earlier literature) and forwards (to quantum information theory) see Caves, Fuchs and Schack (2001b). But note that from now on we will use whatever language we find convenient, rather than get involved in issues like this.

To check the normalization we first change coordinates to $X^i = \sqrt{p_i}$, and then use the fact that we already know the volume of a round hyperoctant – namely $1/2^N$ times the volume of $\mathbf{S}^{N-1}$, given in Eq. (1.17). This choice of prior is known as *Jeffreys' prior* and is generally agreed to be the best choice. Jeffreys arrived at his prior through the observation that for continuous sample spaces the corresponding expression has the desirable property of being invariant under reparametrizations of the sample space. Note the implication: somehow Jeffreys is claiming that 'most' of the probabilities that we encounter in the world around us are likely to be close to one or zero.

Further choices are possible (and are sometimes made in mathematical statistics). Thus we have the *Dirichlet distribution*

$$\mathrm{d}P_s \;\propto\; \delta\left(\sum_{i=1}^{N} p_i - 1\right)(p_1 p_2 \dots p_N)^{s-1}\,\mathrm{d}p_1 \mathrm{d}p_2 \dots \mathrm{d}p_N \;, \tag{2.73}$$

which includes the flat and round measures for $s = 1$ and $s = 1/2$, respectively. To find a simple way to generate probability vectors according to these distributions study Problem 2.5.

2.7 Generalized entropies

The Shannon entropy is arguably the most interesting function of $\vec{p}$ that one can find. But in many situations the only property that one really requires is that the function be Schur concave, that is consistent with the majorization order. For this reason we are willing to call any Schur concave function a *generalized entropy*. Similarly a *generalized relative entropy* must be monotone under stochastic maps. Equation (2.10), when adjusted with a sign, provides us with a liberal supply of Schur concave functions. To obtain a supply of generalized monotone entropies, let g be a convex function defined on $(0, \infty)$ such that $g(1) = 0$. Then the expression

$$S_g(P||Q) = \sum_i p_i\, g(q_i/p_i) \tag{2.74}$$

is monotone under stochastic maps. The choice $g(t) = -\ln t$ gives the Kullback–Leibler relative entropy that we have already studied. All of these generalized relative entropies are consistent with the Fisher–Rao metric, in the sense that

$$S_g(P||P + \mathrm{d}P) = \sum_i p_i g\left(1 + \frac{\mathrm{d}p_i}{p_i}\right) \approx \frac{g''(1)}{2} \sum_i \frac{\mathrm{d}p_i \mathrm{d}p_i}{p_i} \;. \tag{2.75}$$

As long as we stay with classical probability theory the Fisher–Rao geometry itself remains unique.

We will study generalized entropies in some detail. We define the *moments* of a probability distribution as

$$f_q(\vec{p}) \;=\; \sum_{i=1}^{N} p_i^q \;. \tag{2.76}$$

They are Schur concave for $q \leq 1$, and Schur convex for $q \geq 1$. The set of moments $f_q(\vec{p})$ for $q = 2, \ldots, N$ determines the vector $\vec{p}$ up to a permutation of its components, just as knowing the traces, $\mathrm{Tr}A^k, k = 1, \ldots, N$, of a Hermitian matrix A one can find its spectrum. The analogy is exact: in Chapter 12 we will think of classical probability distributions as diagonal matrices, and face the problem of generalizing the definitions of the present chapter to general Hermitian matrices. Instead of the Shannon entropy, we will have the von Neumann entropy that one can calculate from the spectrum of the matrix. But it is much easier to raise a matrix A to some integer power, than to diagonalize it. Therefore it is easier to compute the moments than to compute the von Neumann entropy.

When $q = 2$ the moment f_2 is also known as the *purity* (because it vanishes if and only if the state is pure) or as the *index of coincidence* (because it gives the probability of getting identical outcomes from two independent and equally distributed events). The *linear entropy* is defined as $S_L = 1 - f_2$ and the *participation number* as $R = 1/f_2$ (it varies between zero and N and is interpreted as the 'effective number of events' that contribute to it).

In order to bring the moments closer to the Shannon entropy we can define the *Havrda–Charvát entropies* [13] as

$$S_q^{HC}(P) \;\equiv\; \frac{1}{1-q}\Big[\sum_{i=1}^{N} p_i^q - 1\Big] \,. \tag{2.77}$$

We now get the Shannon entropy in the limit $q \to 1$. These entropies are Schur concave but not recursive. They are not additive for independent random variables; when $P_{12}^{ij} = p_1^i \, p_2^j$ we have

$$S_q^{HC}(P_{12}) = S_q^{HC}(P_1) + S_q^{HC}(P_2) + (1-q) S_q^{HC}(P_1) S_q^{HC}(P_2) \,. \tag{2.78}$$

To ensure additivity a logarithm can be used in the definition. A one parameter family of Schur concave and additive entropies are the *Rényi entropies* [14]

$$S_q(P) \;\equiv\; \frac{1}{1-q} \ln\Big[\sum_{i=1}^{N} p_i^q\Big] \,. \tag{2.79}$$

[13] They were first studied by at Havrda and Charvát (1967); in statistical physics they go under the name of *Tsallis entropies*. For a review of their uses, see Tsallis (2002) or Kapur (1994).

[14] Rényi (1961) introduced them as examples of functions that are additive, and obey all of Khinchin's axioms for an entropy except the recursion property.

Table 2.1. *Properties of generalized entropies*

Entropy	Shannon	Rényi	Havrda–Charvát
Formula	$-\sum_{i=1}^{N} p_i \ln p_i$	$\frac{1}{1-q} \ln \left(\sum_{i=1}^{N} p_i^q \right)$	$\frac{1}{1-q} \left(\sum_{i=1}^{N} p_i^q - 1 \right)$
Recursivity	yes	no	no
Additivity	yes	yes	no
Concavity	yes	for $0 < q \leq 1$	for $q > 0$

We assume that $q \geq 0$. An added advantage of the Rényi entropies is that they are normalized in a uniform way, in the sense that they vanish for pure states and attain their maximal value $\ln N$ at the uniform distribution.

We summarize some properties of generalized entropies in Table 2.1. There we see a possible disadvantage of the Rényi entropies for $q > 1$, which is that we cannot guarantee that they are concave in the ordinary sense. In fact concavity is lost for $q > q_* > 1$, where q_* is N dependent.[15]

Special cases of the Rényi entropies include $q = 0$, which is the logarithm of the number of non-zero components of the distribution and is known as the *Hartley entropy*.[16] When $q \to 1$, we have the Shannon entropy (sometimes denoted S_1), and when $q \to \infty$ the *Chebyshev entropy* $S_\infty = -\ln p_{\max}$, a function of the largest component $p_{\max}$. Figure 2.14 shows some iso-entropy curves in the $N = 3$ probability simplex; equivalently we see curves of constant $f_q(\vec{p})$. The special cases $q = 1/2$ and $q = 2$ are of interest because their iso-entropy curves form circles, with respect to the Bhattacharyya and Euclidean distances, respectively. For $q = 20$ we are already rather close to the limiting case $q \to \infty$, for which we would see the intersection of the simplex with a cube centred at the origin in the space where the vector $\vec{p}$ lives – compare the discussion of l_p-norms in Chapter 1. For $q = 1/5$ the maximum is already rather flat. This resembles the limiting case S_0, for which the entropy reflects the number of events which may occur: it vanishes at the corners of the triangle, is equal to $\ln 2$ at its sides and equals $\ln 3$ for any point inside it.

For any given probability vector P the Rényi entropy is a continuous, non-increasing function of its parameter,

$$S_t(P) \;\leq\; S_q(P) \quad \text{for any} \quad t > q \;. \tag{2.80}$$

To show this, introduce the auxiliary probability vector $r_i \equiv p_i^q / \sum_i p_i^q$. Observe that the derivative $\partial S_q / \partial q$ may be written as $-S(P||R)/(1-q)^2$. Since the relative

[15] Peter Harremoës has informed us of the bound $q_* \leq 1 + \ln(4)/\ln(N-1)$.
[16] The idea of measuring information regardless of its contents originated with Hartley (1928).

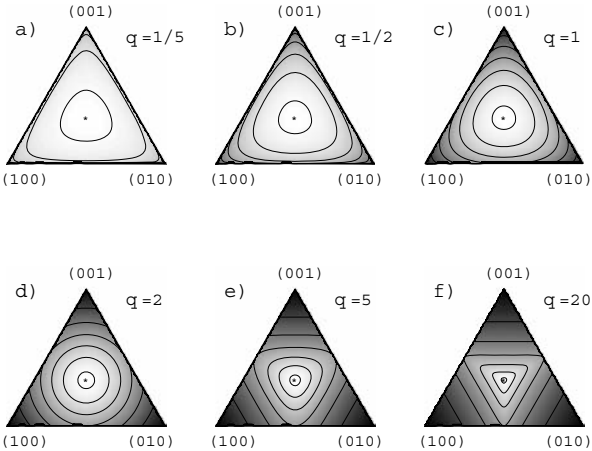


Figure 2.14. The Rényi entropy is constant along the curves plotted for (a) $q = 1/5$; (b) $q = 1/2$; (c) $q = 1$; (d) $q = 2$; (e) $q = 5$ and (f) $q = 20$.

entropy $S(P||R)$ is non–negative, the Rényi entropy S_q is a non-increasing function of q. In Figure 2.7(b) we show how this fact can be used to bound the Shannon entropy S_1.

In a similar way one proves (Beck and Schlögl, 1993) analogous inequalities valid for $q \geq 0$:

$$\frac{\mathrm{d}}{\mathrm{d}q}\left[\frac{q-1}{q}S_q\right] \geq 0\,, \qquad \frac{\mathrm{d}^2}{\mathrm{d}q^2}\left[(1-q)S_q\right] \geq 0\,. \tag{2.81}$$

The first inequality reflects the fact that the l_q-norm is a non–increasing function. It allows one to obtain useful bounds on Rényi entropies,

$$\frac{q-1}{q}S_q(P) \leq \frac{s-1}{s}S_s(P) \quad \text{for} \quad \text{any} \quad q \leq s\,. \tag{2.82}$$

Due to the second inequality the function $(1-q)S_q$ is convex. However, this does not imply that the Rényi entropy itself is a convex function of q;[17] it is non-convex for probability vectors P with one element dominating.

The Rényi entropies are correlated. For $N = 3$ we can see this if we superpose the various panels of Figure 2.14. Consider the superposition of the iso-entropy curves for $q = 1$ and 5. Compared with the circle for $q = 2$ the isoentropy curves for $q < 2$ and $q > 2$ are deformed (with a three-fold symmetry) in the opposite way: together they form a kind of David's star with rounded corners. Thus if we

[17] As erroneously claimed in Życzkowski (2003).

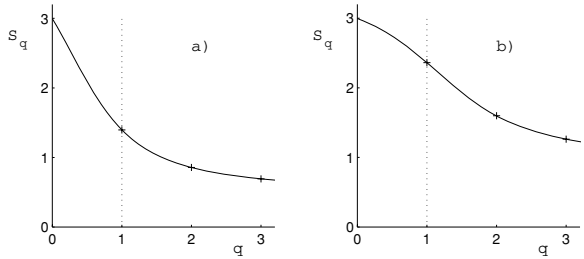


Figure 2.15. Rényi entropies S_q for $N = 20$ probability vectors: (a) convex function for a power–law distribution, $p_j \sim j^{-2}$; (b) non-convex function for $P = \frac{1}{100}(43, 3, \ldots, 3)$.

move along a circle of constant S_2 in the direction of decreasing S_5 the Shannon entropy S_1 increases, and conversely.

The problem, what values the entropy S_q may admit, provided S_t is given, has been solved by Harremoës and Topsøe (2001). They proved a simple but not very sharp upper bound on S_1 by S_2, valid for any distribution $P \in \mathbb{R}^N$

$$S_2(P) \leq S_1(P) \leq \ln N + 1/N - \exp\bigl(-S_2(P)\bigr) \,. \tag{2.83}$$

The lower bound provided by a special case of Eq. (2.80) is not tight. Optimal bounds are obtained[18] by studying both entropies along families of interpolating probability distributions

$$Q_{(k,l)}(a) \equiv a\,Q_{(k)} + (1-a)Q_{(l)} \quad \text{with} \quad a \in [0,1] \,. \tag{2.84}$$

For instance, the upper bound for S_q as a function of S_t with $t > q$ can be derived from the distribution $Q_{(1,N)}(a)$. For any value of a we compute S_t, invert this relation to obtain $a(S_t)$ and arrive at the desired bound by plotting $S_q[a(S_t)]$. In this way we may bound the Shannon entropy by a function of S_2,

$$S_1(P) \leq (1-N)\frac{1-a}{N}\ln\frac{1-a}{N} - \frac{1+a(N-1)}{N}\ln\frac{1+a(N-1)}{N} \,, \tag{2.85}$$

where $a = [(N\exp[-S_2(P)]/(N-1)]^{1/2}$. This bound is shown in Figure 2.16(c) and (d) for $N = 3$ and $N = 5$, respectively. Interestingly, the set M_N of possible distributions plotted in the plane S_q versus S_t is not convex. All Rényi entropies agree at the distributions $Q_{(k)}$, $k = 1, \ldots, N$. These points located at the diagonal, $S_q = S_t$, belong to the lower bound. It consists of $N-1$ arcs derived from interpolating distributions $Q_{(k,k+1)}$ with $k = 1, \ldots, N-1$. As shown Figure 2.16 the set M_N

[18] This result (Harremoës and Topsøe, 2001) was later generalized (Berry and Sanders, 2003) for other entropy functions.

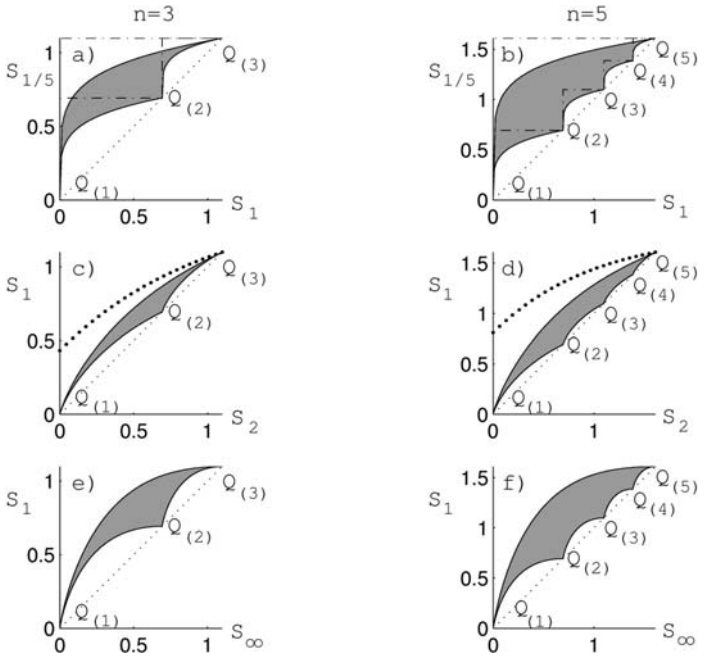


Figure 2.16. The set M_N of all possible discrete distributions for $N = 3$ and $N = 5$ in the Rényi entropies plane S_q and S_t: $S_{1/5}$ and S_1 (a and b); S_1 and S_2 (c and d), and S_1 and S_∞ (e and f). Thin dotted lines in each panel stand for the lower bounds (Eq. (2.80)), dashed-dotted lines in panels (a) and (b) represent bounds between S_0 and S_1, while bold dotted curves in panel (c) and (d) are the upper bounds (Eq. (2.83)).

resembles a medusa[19] with N arms. Its actual width and shape depends on the parameters t and q (Życzkowski, 2003).

Problems

Problem 2.1 The difference $S_{\rm str} \equiv S_1 - S_2$, called *structural entropy*, is useful to characterize the non–homogeneity of a probability vector (Pipek and Varga, 1992). Plot $S_{\rm str}$ for $N = 3$, and find its maximal value.

Problem 2.2 Let $\vec{x}$, $\vec{y}$ and $\vec{z}$ be positive vectors with components in decreasing order and such that $\vec{z} \prec \vec{y}$. Prove that $\vec{x} \cdot \vec{z} \leq \vec{x} \cdot \vec{y}$.

[19] Polish or Swedish readers will know that a medusa is a (kind of) jellyfish.

Problem 2.3 (a) Show that any bistochastic matrix B written as a product of two T-transforms is orthostochastic. (b) Show that the product of $(N-1)$ T-transforms of size N acting in different subspaces forms an orthostochastic matrix.

Problem 2.4 Prove the Hardy–Littlewood–Pólya lemma.

Problem 2.5 Take N independent real (complex) random numbers z_i generated according to the real (complex) normal distribution. Define normalized probability vector P, where $p_i \equiv |z_i|^2 / \sum_{i=1}^{N} |z_i|^2$ with $i = 1, \ldots, N$. What is its distribution on the probability simplex Δ_{N-1}?

Problem 2.6 To see an analogy between the Shannon and the Havrda–Charvát entropies prove that (Abe, 1997)

$$S(P) \equiv -\sum_i p_i \ln p_i = -\Big[\frac{\mathrm{d}}{\mathrm{d}x}\Big(\sum_i p_i^x\Big)\Big]\Big|_{x=1} \tag{2.86}$$

$$S_q^{HC}(P) \equiv \frac{1}{1-q}\Big(\sum_i p_i^q - 1\Big) = -\Big[D_q\Big(\sum_i p_i^x\Big)\Big]\Big|_{x=1}, \tag{2.87}$$

where the 'multiplicative' *Jackson* q-derivative reads

$$D_q\big(f(x)\big) \equiv \frac{f(qx) - f(x)}{qx - x}, \quad \text{so that} \quad \lim_{q\to 1} D_q(f(x)) = \frac{\mathrm{d}f(x)}{\mathrm{d}x}. \tag{2.88}$$

Problem 2.7 For what values of q are the Rényi entropies concave when $N = 2$?

3

Much ado about spheres

> He who undertakes to deal with questions of natural sciences without the help of geometry is attempting the infeasible.
>
> *Galileo Galilei*

In this chapter we will study spheres, mostly two- and three-dimensional spheres, for two reasons: because spheres are important, and because they serve as a vehicle for introducing many geometric concepts (such as symplectic, complex and Kähler spaces, fibre bundles and group manifolds) that we will need later on. It may look like a long detour, but it leads into the heart of quantum mechanics.

3.1 Spheres

We expect that the reader knows how to define a round n-dimensional sphere through an embedding in a flat $(N = n + 1)$-dimensional space. Using Cartesian coordinates, the n-sphere $\mathbf{S}^n$ is the surface

$$X \cdot X = \sum_{I=0}^{n} X^I X^I = (X^0)^2 + (X^1)^2 + \cdots + (X^n)^2 = 1 \,, \tag{3.1}$$

where we gave a certain standard size (unit radius) to our sphere and also introduced the standard scalar product in $\mathbb{R}^N$. The Cartesian coordinates $(X^0, X^1, \ldots, X^n) = (X^0, X^i) = X^I$, where $1 \leq i \leq n$ and $0 \leq I \leq n$, are known as *embedding coordinates*. They are not intrinsic to the sphere but useful anyway.

Our first task is to introduce a few more intrinsic coordinate systems on $\mathbf{S}^n$, in addition to the polar angles used in Section 1.2. Eventually this should lead to the insight that coordinates are not important, only the underlying space itself counts. We will use a set of coordinate systems that are obtained by projecting the sphere from a point on the axis between the north and south poles to a plane parallel to the equatorial plane. Our first choice is perpendicular projection to the equatorial plane,

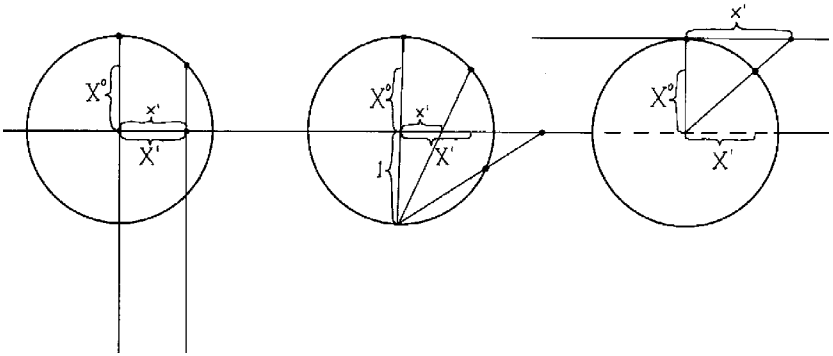


Figure 3.1. Three coordinate system that we use. To the left, orthographic projection from infinity of the northern hemisphere into the equatorial plane. In the middle, stereographic projection from the south pole of the entire sphere (except the south pole itself) onto the equatorial plane. To the right, gnomonic projection from the centre of the northern hemisphere onto the tangent plane at the north pole.

known as *orthographic projection* among mapmakers. The point of projection is infinitely far away. We set

$$X^i = x^i \qquad X^0 = \sqrt{1 - r^2}\,, \qquad r^2 \equiv \sum_{i=1}^{n} x^i x^i < 1\,. \tag{3.2}$$

This coordinate patch covers the region where $X^0 > 1$; we need several coordinate patches of this kind to cover the entire sphere. The metric when expressed in these coordinates is

$$\mathrm{d}s^2 = \mathrm{d}X^0 \mathrm{d}X^0 + \sum_{i=1}^{n} \mathrm{d}X^i \mathrm{d}X^i = \frac{1}{1 - r^2}\left[(1 - r^2)\mathrm{d}x \cdot \mathrm{d}x + (x \cdot \mathrm{d}x)^2\right]\,, \tag{3.3}$$

where

$$x \cdot \mathrm{d}x \equiv \sum_{i=1}^{n} x^i \mathrm{d}x^i \qquad \text{and} \qquad \mathrm{d}x \cdot \mathrm{d}x \equiv \sum_{i=1}^{n} \mathrm{d}x^i \mathrm{d}x^i\,. \tag{3.4}$$

An attractive feature of this coordinate system is that, as a short calculation shows, the measure defined by the metric becomes simply $\sqrt{g} = 1/X^0$.

An alternative choice of intrinsic coordinates – perhaps the most useful one – is given by *stereographic projection* from the south pole to the equatorial plane, so that

$$\frac{x^i}{X^i} = \frac{1}{1 + X^0} \qquad \Leftrightarrow \qquad X^i = \frac{2x^i}{1 + r^2} \qquad X^0 = \frac{1 - r^2}{1 + r^2}\,. \tag{3.5}$$

A minor calculation shows that the metric now becomes manifestly *conformally flat*, that is to say that it is given by a conformal factor Ω^2 times a flat metric:

$$\mathrm{d}s^2 = \Omega^2 \delta_{ij} \mathrm{d}x^i \mathrm{d}x^j = \frac{4}{(1+r^2)^2}\, \mathrm{d}x \cdot \mathrm{d}x \ . \tag{3.6}$$

This coordinate patch covers the region $X^0 > -1$, that is to say the entire sphere except the south pole itself. To cover the entire sphere we need at least one more coordinate patch, say the one that is obtained by stereographic projection from the north pole. In the particular case of $\mathbf{S}^2$ one may collect the two stereographic coordinates into one complex coordinate z; the relation between this coordinate and the familiar polar angles is

$$z = x^1 + \mathrm{i}x^2 = \tan\frac{\theta}{2} \mathrm{e}^{\mathrm{i}\phi} \ . \tag{3.7}$$

We will use this formula quite frequently.

A third choice is *gnomonic* or *central projection*. (The reader may want to know that the gnomon being referred to is the vertical rod on a primitive sundial.) We now project one half of the sphere from its centre to the tangent plane touching the north pole. In equations

$$x^i = \frac{X^i}{X^0} \quad \Leftrightarrow \quad X^i = \frac{x^i}{\sqrt{1+r^2}} \qquad X^0 = \frac{1}{\sqrt{1+r^2}} \ . \tag{3.8}$$

This leads to the metric

$$\mathrm{d}s^2 = \frac{1}{(1+r^2)^2} \left[(1+r^2) \mathrm{d}x \cdot \mathrm{d}x - (x \cdot \mathrm{d}x)^2 \right] \ . \tag{3.9}$$

One hemisphere only is covered by gnomonic coordinates. (The formalism presented in Section 1.4 can be used to transform between the three coordinate systems that we presented, but it was easier to derive each from scratch.)

All coordinate systems have their special advantages. Let us sing the praise of stereographic coordinates right away. The topology of coordinate space is $\mathbb{R}^n$, and when stereographic coordinates are used the sphere has only one further point not covered by these coordinates, so the topology of $\mathbf{S}^n$ is the topology of $\mathbb{R}^n$ with one extra point attached 'at infinity'. The conformal factor ensures that the round metric is smooth at the added point, so that 'infinity' in coordinate space lies at finite distance on the sphere. One advantage of these coordinates is that all angles come out correctly if we draw a picture in a flat coordinate space, although distances far from the origin are badly distorted. We say that the map between the sphere and the coordinate space is *conformal*, that is it preserves angles. The stereographic picture makes it easy to visualize $\mathbf{S}^3$, which is conformally mapped to ordinary flat space

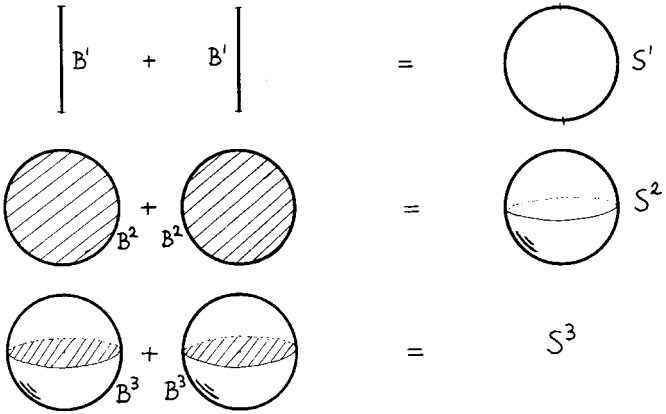


Figure 3.2. A circle is the sum of two intervals, a 2-sphere is the sum of two discs glued together along the boundaries, and the 3-sphere is the sum of two balls again with the boundaries identified. In the latter case the gluing cannot be done in three dimensions. See Appendix 3 for a different picture in the same vein.

in such a way that the north pole is at the origin, the equator is the unit sphere, and the south pole is at infinity. With a little training one can get used to this picture, and learn to disregard the way in which it distorts distances. If the reader prefers a compact picture of the 3-sphere this is easily provided: use the stereographic projection from the south pole to draw a picture of the northern hemisphere only. This gives the picture of a solid ball whose surface is the equator of the 3-sphere. Then project from the north pole to get a picture of the southern hemisphere. The net result is a picture consisting of two solid balls whose surfaces must be mentally identified with each other.

When we encounter a new space, we first ask what symmetries it has, and what its geodesics are. Here the embedding coordinates are very useful. An infinitesimal *isometry* (a transformation that preserves distances) is described by a *Killing vector field* pointing in the direction that points are transformed. We ask for the flow lines of the isometry. A sphere has exactly $n(n+1)/2$ linearly independent Killing vectors at each point, namely

$$J_{IJ} = X_I \partial_J - X_J \partial_I \ . \tag{3.10}$$

(Here we used the trick from Section 1.4 to represent a tangent vector as a differential operator.) On the 2-sphere the flow lines are always circles at constant distance from a pair of antipodal *fixed points* where the flow vanishes. The situation gets more interesting on the 3-sphere, as we will see.

A geodesic is the shortest curve between any pair of nearby points on itself. On the sphere a geodesic is a great circle, that is the intersection of the sphere

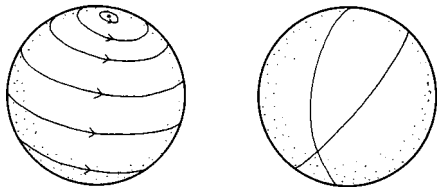


Figure 3.3. Killing flows and geodesics on the 2-sphere.

with a two-dimensional plane through the origin in the embedding space. Such a curve is 'obviously' as straight as it can be, given that it must be confined to the sphere. (Incidentally this means that gnomonic coordinates are useful, because the geodesics will appear as straight lines in coordinate space.) The geodesics can also be obtained as the solutions of the Euler–Lagrange equations coming from the constrained Lagrangian

$$L = \frac{1}{2}\dot{X}\cdot\dot{X} + \Lambda(X\cdot X - 1)\,, \tag{3.11}$$

where Λ is a Lagrange multiplier and the overdot denotes differentiation with respect to the *affine parameter* along the curve. We rescale the affine parameter so that $\dot{X}\cdot\dot{X} = 1$, and then the general solution for a geodesic takes the form

$$X^I(\tau) = k^I\cos\tau + l^I\sin\tau\,, \qquad k\cdot k = l\cdot l = 1\,,\ \ k\cdot l = 0\,. \tag{3.12}$$

The vectors k^I and l^I span a plane through the origin in $\mathbb{R}^N$. Since $X^I(0) = k^I$ and $\dot{X}^I(0) = l^I$, the conditions on these vectors say that we start on the sphere, with unit velocity, in a direction tangent to the sphere. The entire curve is determined by these data.

Let us now choose two points along the geodesic, with different values of the affine parameter τ, and compute

$$X(\tau_1)\cdot X(\tau_2) = \cos(\tau_1 - \tau_2)\,. \tag{3.13}$$

With the normalization of the affine parameter that we are using $|\tau_1 - \tau_2|$ is precisely the length of the curve between the two points, so we get the useful formula

$$\cos d = X(\tau_1)\cdot X(\tau_2)\,, \tag{3.14}$$

where d is the geodesic distance between the two points. It is equal to the angle between the unit vectors $X^I(\tau_1)$ and $X^I(\tau_2)$ – and we encountered this formula before in Eq. (2.56).

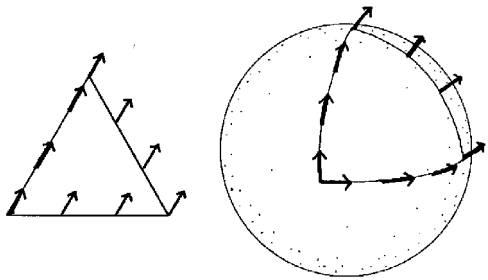


Figure 3.4. A paradox? On the left we parallel transport a vector around the edge of the flat probability simplex. On the right the same simplex is regarded as a round octant, and the obvious (Levi–Civita) notion of parallel transport gives a different result. It is the same space but two different affine connections!

3.2 Parallel transport and statistical geometry

Let us focus on the positive octant (or hyperoctant) of the sphere. In the previous chapter its points were shown to be in one-to-one correspondence to the set of probability distributions over a finite sample space, and this set was sometimes thought of as round (equipped with the Fisher metric) and sometimes as flat (with convex mixtures represented as straight lines). How can we reconcile these two ways of looking at the octant? To answer this question we will play a little formal game with connections and curvatures.[1]

Curvature is not a very immediate property of a space. It has to do with how one can compare vectors sitting at different points with each other, and we must begin with a definite prescription for how to *parallel transport* vectors from one point to another. For this we require a *connection* and a *covariant derivative*, such as the *Levi–Civita connection* that is defined (using the metric) in Appendix A1.2. Then we can transport a vector V^i along any given curve with tangent vector X^i by solving the ordinary differential equation $X^j \nabla_j V^i = 0$ along the curve. But there is no guarantee that the vector will return to itself if it is transported around a closed curve. Indeed it will not if the *curvature tensor* is non-zero. It must also be noted that the prescription for parallel transport can be changed by changing the connection. Assume that in addition to the metric tensor g_{ij} we are given a totally symmetric *skewness tensor* T_{ijk}. Then we can construct the one-parameter family

[1] In this section we assume that the reader knows some Riemannian geometry. Readers who have forgotten this can refresh their memory with Appendix A1.2. Readers who never knew about it may consult, say, Murray and Rice (1993) or Schrödinger (1950) – or take comfort in the fact that Riemannian geometry is used in only a few sections of our book.

of affine connections

$$\overset{(\alpha)}{\Gamma}_{ijk} = \Gamma_{ijk} + \frac{\alpha}{2} T_{ijk} \,. \tag{3.15}$$

Here Γ_{ijk} is the ordinary Levi–Civita connection (with one index lowered using the metric). Since T_{ijk} transforms like a tensor all α-connections transform as connections should. The covariant derivative and the curvature tensor will be given by Eqs. (A1.5) and (A1.10), respectively, but with the Levi–Civita connection replaced by the new α-connection. We can also define α-geodesics, using Eq. (A1.8) but with the new connection. This is *affine differential geometry*; a subject that at first appears somewhat odd, because there are 'straight lines' (geodesics) and distances along them (given by the affine parameter of the geodesics), but these distances do not fit together in the way they would do if they were consistent with a metric tensor.[2]

In statistical geometry the metric tensor has a potential, that is to say that there is a convex function Φ such that the Fisher–Rao metric is

$$g_{ij}(p) \;=\; \frac{\partial^2 \Phi}{\partial p^i \partial p^j} \;\equiv\; \partial_i \partial_j \Phi(p) \,. \tag{3.16}$$

In Eq. (2.59) this function is given as minus the Shannon entropy, but for the moment we want to play a game and keep things general. Actually the definition is rather strange from the point of view of differential geometry, because it uses ordinary rather than covariant derivatives. The equation will therefore be valid only with respect to some preferred *affine coordinates* that we call p^i here, anticipating their meaning; we use an affine connection defined by the requirement that it vanishes in this special coordinate system. But if we have done one strange thing we can do another. Therefore we can define a totally symmetric third rank tensor using the same preferred affine coordinate system, namely

$$T_{ijk}(p) = \partial_i \partial_j \partial_k \Phi(p) \,. \tag{3.17}$$

Now we can start the game.

Using the definitions in Appendix A1.2 it is easy to see that

$$\Gamma_{ijk} = \frac{1}{2} \partial_i \partial_j \partial_k \Phi(p) \,. \tag{3.18}$$

But we also have an α-connection, namely

$$\overset{(\alpha)}{\Gamma}_{ijk} = \frac{1+\alpha}{2} \partial_i \partial_j \partial_k \Phi(p) = (1+\alpha)\Gamma_{ijk} \,. \tag{3.19}$$

[2] The statistical geometry of α-connections is due to Čencov (1982) and Amari (1985).

A small calculation is now enough to relate the α-curvature to the usual one:

$$\overset{(\alpha)}{R}_{ijkl} = (1-\alpha^2) R_{ijkl} \,. \tag{3.20}$$

Equation (3.19) says that the α-connection vanishes when $\alpha = -1$, so that the space is (-1)-flat. Our preferred coordinate system is preferred in the sense that a line that looks straight in this coordinate system is indeed a geodesic with respect to the (-1)-connection. The surprise is that Eq. (3.20) shows that the space is also $(+1)$-flat, even though this is not obvious just by looking at the $(+1)$-connection in this coordinate system.

We therefore start looking for a coordinate system in which the $(+1)$-connection vanishes. We try the functions

$$\eta^i = \frac{\partial \Phi}{\partial p^i} \,. \tag{3.21}$$

Then we define a new function $\Psi(\eta)$ through a *Legendre transformation*, a trick familiar from thermodynamics:

$$\Psi(\eta) + \Phi(p) - \sum_i p^i \eta^i = 0 \,. \tag{3.22}$$

Although we express the new function as a function of its 'natural' coordinates η^i, it is first and foremost a function of the points in our space. By definition

$$\mathrm{d}\Psi = \sum_i p^i \mathrm{d}\eta^i \qquad \Rightarrow \qquad p^i = \frac{\partial \Psi}{\partial \eta^i} \,. \tag{3.23}$$

Our coordinate transformation is an honest one in the sense that it can be inverted to give the functions $p^i = p^i(\eta)$. Now let us see what the tensors g_{ij} and T_{ijk} look like in the new coordinate system. An exercise in the use of the chain rule shows that

$$g_{ij}(\eta) = \frac{\partial p^k}{\partial \eta^i} \frac{\partial p^l}{\partial \eta^j} g_{kl}\big(p(\eta)\big) = \frac{\partial p^j}{\partial \eta^i} = \frac{\partial^2 \Psi}{\partial \eta^i \partial \eta^j} \equiv \partial_i \partial_j \Psi(\eta) \,. \tag{3.24}$$

For T_{ijk} a slightly more involved exercise shows that

$$T_{ijk}(\eta) \;=\; -\partial_i \partial_j \partial_k \Psi(\eta) \,. \tag{3.25}$$

To show this, first derive the matrix equation

$$\sum_{k=1}^{n} g_{ik}(\eta)\, g_{kj}(p) \;=\; \delta_{ij} \,. \tag{3.26}$$

The sign in Eq. (3.25) is crucial since it implies that $\overset{(+1)}{\Gamma}_{ijk} = 0$; in the new coordinate system the space is indeed manifestly (+1)-flat.

We now have two different notions of affine straight lines. Using the (−1)-connection they are

$$\dot{p}^j \overset{(-1)}{\nabla}_j \dot{p}^i = \ddot{p}^i = 0 \quad \Rightarrow \quad p^i(t) = p_0^i + t p^i \,. \tag{3.27}$$

Using the (+1)-connection, and working in the coordinate system that comes naturally with it, we get instead

$$\dot{\eta}^j \overset{(+1)}{\nabla}_j \dot{\eta}^i = \ddot{\eta}^i = 0 \quad \Rightarrow \quad \eta^i(t) = \eta_0^i + t \eta^i \,. \tag{3.28}$$

We will see presently what this means.

There is a final manoeuvre that we can do. We go back to Eq. (3.22), look at it, and realize that it can be modified so that it defines a function of pairs of points P and P' on our space, labelled by the coordinates p and η', respectively. This is the function

$$S(P||P') = \Phi\big(p(P)\big) + \Psi\big(\eta'(P')\big) - \sum_i p^i(P)\, \eta'^i(P') \,. \tag{3.29}$$

It vanishes when the two points coincide, and since this is an extremum the function is always positive. It is an asymmetric function of its arguments. To lowest non-trivial order the asymmetry is given by the skewness tensor; indeed

$$S(p||p + \mathrm{d}p) = \sum_{i,j} g_{ij}\, \mathrm{d}p^i \mathrm{d}p^j - \sum_{i,j,k} T_{ijk}\, \mathrm{d}p^i \mathrm{d}p^j \mathrm{d}p^k + \cdots \tag{3.30}$$

$$S(p + \mathrm{d}p||p) = \sum_{i,j} g_{ij}\, \mathrm{d}p^i \mathrm{d}p^j + \sum_{i,j,k} T_{ijk}\, \mathrm{d}p^i \mathrm{d}p^j \mathrm{d}p^k + \cdots \tag{3.31}$$

(We are of course entitled to use the same coordinates for both arguments, as we just did, provided we do the appropriate coordinate transformations.)

To bring life to this construction it remains to seed it with some interesting function Φ and see what interpretation we can give to the objects that come with it. We already have one interesting choice, namely minus the Shannon entropy. First we play the game on the positive orthant $\mathbb{R}_+^N$, that is to say we use the index i ranging from 1 to N, and assume that all the $p^i > 0$ but leave them otherwise unconstrained. Our input is

$$\Phi(p) \;=\; \sum_{i=1}^{N} p^i \ln p^i \,. \tag{3.32}$$

Our output becomes

$$\mathrm{d}s^2 = \sum_{i=1}^{N} \frac{\mathrm{d}p^i \mathrm{d}p^i}{p^i} = \sum_{i=1}^{N} \mathrm{d}x^i \mathrm{d}x^i \; ; \qquad 4p^i = (x^i)^2 \; , \tag{3.33}$$

$$\eta^i = \ln p^i + 1 \; , \tag{3.34}$$

$$\Psi(\eta) = \sum_{i=1}^{N} p^i \; , \tag{3.35}$$

$$S(p||p') = \sum_{i=1}^{N} p^i \ln \frac{p^i}{p'^i} + \sum_{i=1}^{N} (p'^i - p^i) \; , \tag{3.36}$$

$$(+1)\text{-geodesic} : \qquad p^i(t) = p^i(0) \mathrm{e}^{t[\ln p^i(1) - \ln p^i(0)]} \; . \tag{3.37}$$

In Eq. (3.33) the coordinate transformation $p^i = 4(x^i)^2$ shows that the Fisher–Rao metric on $\mathbb{R}^N_+$ is flat. To describe the situation on the probability simplex we impose the constraint

$$\sum_{i=1}^{N} p^i = 1 \qquad \Leftrightarrow \qquad \sum_{i=1}^{N} x^i x^i = 4 \; . \tag{3.38}$$

Taking this into account we see that the Fisher–Rao metric on the probability simplex is the metric on the positive octant of a round sphere with radius 2. For maximum elegance, we have chosen a different normalization of the metric, compared with what we used in Section 2.5. The other entries in the list have some definite statistical meaning, too. We are familiar with the relative entropy $S(p||p')$ from Section 2.3. The geodesics defined by the (-1)-connection, that is to say the lines that look straight in our original coordinate system, are convex mixtures of probability distributions. The (-1)-connection is therefore known as the *mixture connection* and its geodesics are (one-dimensional) *mixture families*. The space is flat with respect to the mixture connection, but (in a different way) also with respect to the $(+1)$-connection. The coordinates in which this connection vanishes are the η^i. The $(+1)$-geodesics, that is the lines that look straight when we use the coordinates η^i, are known as (one-dimensional) *exponential families* and the $(+1)$-connection as the *exponential connection*. Exponential families are important in the theory of statistical inference; it is particularly easy to pin down (using samplings) precisely where you are along a given curve in the space of probability distributions, if that curve is an exponential family. We notice one interesting thing: it looks reasonable to define the mean value of $p^i(0)$ and $p^i(1)$ as $p^i(1/2)$, where the parameter is the affine parameter along a geodesic. If our geodesic is a mixture family, this is the arithmetic mean, while it will be the geometric mean $p^i(1/2) = \sqrt{p^i(0)p^i(1)}$ if the geodesic is an exponential family.

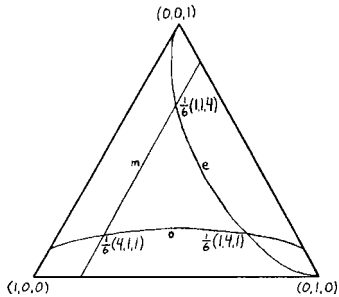


Figure 3.5. Here we show three different kinds of geodesics – mixture (m), exponential (e) and metric (0) – on the simplex; since the mixture coordinates p^i are used only the mixture geodesic appears straight in the picture.

Since we have shown that there are three different kinds of straight lines on the probability simplex – mixture families, exponential families, and geodesics with respect to the round metric – we should also show what they look like. Figure 3.5 is intended to make this clear. The probability simplex is *complete* with respect to the exponential connection, meaning that the affine parameter along the exponential geodesics goes from minus to plus infinity – whereas the mixture and metric geodesics cross its boundary at some finite value of the affine parameter.

Our story is almost over. The main point is that a number of very relevant concepts – Shannon entropy, Fisher–Rao metric, mixture families, exponential families and the relative entropy – have appeared in natural succession. But the story can be told in different ways. Every statistical manifold has a totally symmetric skewness tensor built in. Indeed, following Section 2.4, we can use the score vectors l_i and define, using expectation values,

$$g_{ij} = \langle l_i l_j \rangle \quad \text{and} \quad T_{ijk} = \langle l_i l_j l_k \rangle \, . \tag{3.39}$$

In particular, a skewness tensor is naturally available on the space of normal distributions. This space turns out to be (± 1)-flat, although the coordinates that make this property manifest are not those used in Section 2.5. Whenever a totally symmetric tensor is used to define a family of α-connections one can show that

$$X^k \partial_k (g_{ij} Y^i Z^j) = g_{ij} X^k \overset{(\alpha)}{\nabla}_k Y^i Z^j + g_{ij} Y^i X^k \overset{(-\alpha)}{\nabla}_k Z^j \, . \tag{3.40}$$

What this equation says is that the scalar product between two vectors remains constant if the vectors are parallel transported using dual connections (so that their covariant derivatives along the curve with tangent vector X^i vanish by definition). The 0-connection, that is the Levi–Civita connection, is self dual in the sense that the scalar product is preserved if both of them are parallel transported using the

Levi–Civita connection. It is also not difficult to show that if the α-connection is flat for any value of α then the equation

$$\overset{(\alpha)}{R}_{ijkl} = \overset{(-\alpha)}{R}_{ijkl} \tag{3.41}$$

will hold for all values of α. Furthermore it is possible to show that if the space is α-flat then it will be true, in that preferred coordinate system for which the α-connection vanishes, that there exists a potential for the metric in the sense of Eq. (3.96). The point is that Eq. (3.40) then implies that

$$\overset{(-\alpha)}{\Gamma}_{jki} = \partial_k g_{ij} \quad \text{and} \quad \overset{(-\alpha)}{\Gamma}_{[jk]i} = 0 \quad \Rightarrow \quad \partial_j g_{ki} - \partial_k g_{ji} = 0\,. \tag{3.42}$$

The existence of a potential for the metric follows:

$$g_{ij} = \partial_i V_j \quad \text{and} \quad g_{[ij]} = 0 \quad \Rightarrow \quad V_j = \partial_j \Phi\,. \tag{3.43}$$

At the same time it is fair to warn the reader that if a space is compact, it is often impossible to make it globally α-flat (Ay and Tuschmann, 2002).

3.3 Complex, Hermitian and Kähler manifolds

We now return to the study of spheres in the global manner and forget about statistics for the time being. It turns out to matter a lot whether the dimension of the sphere is even or odd. Let us study the even-dimensional case $n = 2m$, and decompose the intrinsic coordinates according to

$$x^i = (x^a, x^{m+a})\,, \tag{3.44}$$

where the range of a goes from 1 to m. Then we can, if we wish, introduce the complex coordinates

$$z^a = x^a + \mathrm{i}x^{m+a}\,, \qquad \bar{z}^{\bar{a}} = x^a - \mathrm{i}x^{m+a}\,. \tag{3.45}$$

We delibarately use two kinds of indices here (barred and unbarred) because we will never contract indices of different kinds. The new coordinates come in pairs connected by complex conjugation,

$$(z^a)^* = \bar{z}^{\bar{a}}\,. \tag{3.46}$$

This equation ensures that the original coordinates take real values. Only the z^a count as coordinates and once they are given the $\bar{z}^{\bar{a}}$ are fully determined. If we choose stereographic coordinates to start with, we find that the round metric becomes

$$\mathrm{d}s^2 \equiv g_{ab}\mathrm{d}z^a\mathrm{d}z^b + 2g_{a\bar{b}}\mathrm{d}z^a\mathrm{d}\bar{z}^{\bar{b}} + g_{\bar{a}\bar{b}}\mathrm{d}\bar{z}^{\bar{a}}\mathrm{d}\bar{z}^{\bar{b}} = \frac{4}{(1+r^2)^2}\delta_{a\bar{b}}\,\mathrm{d}z^a\mathrm{d}\bar{z}^{\bar{b}}\,. \tag{3.47}$$

Note that we do not make the manifold complex. We would obtain the *complexified* sphere by allowing the coordinates x^i to take complex values, in which case the real dimension would be multiplied by two and we would no longer have a real sphere. What we actually did may seem like a cheap trick in comparison, but for the 2-sphere it is anything but cheap, as we will see.

To see if the introduction of complex coordinates is more than a trick we must study what happens when we try to cover the entire space with overlapping coordinate patches. We choose stereographic coordinates and add a patch that is obtained by projection from the north pole; we do it in this way:

$$x'^a = \frac{X^a}{1 - X^0}\,, \qquad x'^{m+a} = \frac{-X^{m+a}}{1 - X^0}\,. \tag{3.48}$$

Now the whole sphere is covered by two coordinate systems. Introducing complex coordinates in both patches, we observe that

$$z'^a = x'^a + \mathrm{i}x'^{m+a} = \frac{X^a - \mathrm{i}X^{m+a}}{1 - X^0} = \frac{2(x^a - \mathrm{i}x^{m+a})}{1 + r^2 - 1 + r^2} = \frac{\bar{z}^a}{r^2}\,. \tag{3.49}$$

These are called the *transition functions* between the two coordinate systems. In the special case of $\mathbf{S}^2$ we can conclude that

$$z'(z) = \frac{1}{z}\,. \tag{3.50}$$

Remarkably, the transition functions between the two patches covering the 2-sphere are holomorphic (that is complex analytic) functions of the complex coordinates. In higher dimensions this simple manoeuvre fails.

There is another peculiar thing that happens for $\mathbf{S}^2$, but not in higher dimensions. Look closely at the metric:

$$g_{z\bar{z}} = \frac{2}{(1 + |z|^2)^2} = \frac{2}{1 + |z|^2}\left(1 - \frac{|z|^2}{1 + |z|^2}\right) = 2\partial_z\partial_{\bar{z}} \ln(1 + |z|^2)\,. \tag{3.51}$$

Again a 'potential' exists for the metric of the 2-sphere. Although superficially similar to Eq. (3.16) there is a difference – Eq. (3.16) is true in very special coordinate systems only, while Eq. (3.51) is true in every complex coordinate system connected to the original one with holomorphic coordinate transformations of the form $z' = z'(z)$. Complex spaces for which all this is true are called *Kähler manifolds*.

The time has come to formalize things. A differentiable manifold is a space which can be covered by coordinate patches in such a way that the transition functions are differentiable. A *complex manifold* is a space which can be covered by coordinate

patches in such a way that the coordinates are complex and the transition functions are holomorphic.[3]

Any even-dimensional manifold can be covered by complex coordinates in such a way that, when the coordinate patches overlap,

$$z' = z'(z, \bar{z}) , \qquad \bar{z}' = \bar{z}'(z, \bar{z}) . \tag{3.52}$$

The manifold is complex if and only if it can be covered by coordinate patches such that

$$z' = z'(z) , \qquad \bar{z}' = \bar{z}'(\bar{z}) . \tag{3.53}$$

Since we are using complex coordinates to describe a real manifold a point in the manifold is specified by the n independent coordinates z^a – we always require that

$$\bar{z}^{\bar{a}} \equiv (z^a)^* . \tag{3.54}$$

A complex manifold is therefore a real manifold that can be described in a particular way. Naturally one could introduce coordinate systems whose transition functions are non-holomorphic, but the idea is to restrict oneself to holomorphic coordinate transformations (just as, on a flat space, it is convenient to restrict oneself to Cartesian coordinate systems).

Complex manifolds have some rather peculiar properties caused ultimately by the 'rigidity properties' of analytic functions. By no means all even-dimensional manifolds are complex, and for those that are there may be several inequivalent ways to turn them into complex manifolds. Examples of complex manifolds are $\mathbb{C}^n = \mathbb{R}^{2n}$ and all orientable two-dimensional spaces, including $\mathbf{S}^2$ as we have seen. An example of a manifold that is not complex is $\mathbf{S}^4$. It may be difficult to decide whether a given manifold is complex or not; an example of a manifold for which this question is open is the 6-sphere.[4]

An example of a manifold that can be turned into a complex manifold in several inequivalent ways is the *torus* $\mathbf{T}^2 = \mathbb{C}/\,\Gamma$. Here Γ is some discrete isometry group generated by two (commuting) translations, and $\mathbf{T}^2$ will inherit the property of being a complex manifold from the complex plane $\mathbb{C}$. A better way to say this is that a flat torus is made from a flat parallelogram by gluing opposite sides together. It means that there is one flat torus for every choice of a pair of vectors. The set of all possible tori can be parametrized by the relative length and the angle between the vectors, and by the total area. Since holomorphic transformations cannot change relative lengths or angles this means that there is a two parameter family of tori

[3] A standard reference on complex manifolds is Chern (1979).

[4] As we go to press, a rumour is afoot that S.-S. Chern proved, just before his untimely death at the age of 93, that $\mathbf{S}^6$ does not admit a complex structure.

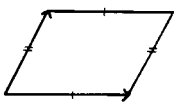


Figure 3.6. A flat torus is a parallelogram with sides identified; it is also defined by a pair of vectors, or by the lattice of points that can be reached by translations with these two vectors.

that are inequivalent as complex manifolds. In other words the 'shape space' of flat tori (technically known as *Teichmüller space*) is two dimensional. Note though that just because two parallelograms look different we cannot conclude right away that they represent inequivalent tori – if one torus is represented by the vectors **u** and **v**, then the torus represented by **u** and **v** + **u** is intrinsically the same.

Tensors on complex manifolds are naturally either real or complex. Consider vectors: since an n complex dimensional complex manifold is a real manifold as well, it has a *real tangent space* $\mathbf{V}$ of dimension $2n$. A real vector (at a point) is an element of $\mathbf{V}$ and can be written as

$$V = V^a \partial_a + \bar{V}^{\bar{a}} \partial_{\bar{a}} \,, \tag{3.55}$$

where $\bar{V}^{\bar{a}}$ is the complex conjugate of V^a. A complex vector is an element of the *complexified tangent space* $\mathbf{V}^{\mathbf{C}}$, and can be written in the same way but with the understanding that $\bar{V}^{\bar{a}}$ is independent of V^a. By definition we say that a real vector space has a *complex structure* if its complexification splits into a direct sum of two complex vector spaces that are related by complex conjugation. This is clearly the case here. We have the direct sum

$$\mathbf{V}^{\mathbf{C}} = \mathbf{V}^{(1,0)} \oplus \mathbf{V}^{(0,1)} \,, \tag{3.56}$$

where

$$V^a \partial_a \in \mathbf{V}^{(1,0)} \qquad \bar{V}^{\bar{a}} \partial_{\bar{a}} \in \mathbf{V}^{(0,1)} \,. \tag{3.57}$$

If $\mathbf{V}$ is the real tangent space of a complex manifold, the space $\mathbf{V}^{(1,0)}$ is known as the *holomorphic tangent space*. This extra structure means that we can talk of vectors of type $(1, 0)$ and $(0, 1)$, respectively; more generally we can define both tensors and differential forms of *type* (p, q). This is well defined because analytic coordinate transformations will not change the type of a tensor, and it is an important part of the theory of complex manifolds.

We still have to understand what happened to the metric of the 2-sphere. We define an *Hermitian manifold* as a complex manifold with a metric tensor of type (1, 1). In complex coordinates it takes the form

$$\mathrm{d}s^2 \;=\; 2g_{a\bar{b}}\,\mathrm{d}z^a\mathrm{d}\bar{z}^{\bar{b}}\;. \tag{3.58}$$

The metric is a symmetric tensor, hence

$$g_{\bar{a}b} \;=\; g_{b\bar{a}}\;. \tag{3.59}$$

The reality of the line element will be ensured if the matrix $g_{a\bar{b}}$ (if we think of it that way) is also Hermitian,

$$(g_{a\bar{b}})^* \;=\; g_{b\bar{a}}\;. \tag{3.60}$$

This is assumed as well.

Just to make sure that you understand what these conditions are, think of the metric as an explicit matrix. Let the real dimension $2n = 4$ in order to be fully explicit: then the metric is

$$\begin{bmatrix} 0 & g_{a\bar{b}} \\ g_{\bar{a}b} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & g_{1\bar{1}} & g_{1\bar{2}} \\ 0 & 0 & g_{2\bar{1}} & g_{2\bar{2}} \\ g_{\bar{1}1} & g_{\bar{1}2} & 0 & 0 \\ g_{\bar{2}1} & g_{\bar{2}2} & 0 & 0 \end{bmatrix}\;. \tag{3.61}$$

It is now easy to see what the conditions on the Hermitian metric really are. By the way $g_{a\bar{b}}$ is not the metric tensor, it is only one block of it.

An Hermitian metric will preserve its form under analytic coordinate transformations, hence the definition is meaningful because the manifold is complex. If the metric is given in advance the property of being Hermitian is non-trivial, but as we have seen $\mathbf{S}^2$ equipped with the round metric provides an example. So does $\mathbb{C}^n$ equipped with its flat metric.

Given an Hermitian metric we can construct a differential form

$$J \;=\; 2\mathrm{i}g_{a\bar{b}}\,\mathrm{d}z^a \wedge \mathrm{d}\bar{z}^{\bar{b}}\;. \tag{3.62}$$

This trick – to use an $n \times n$ matrix to define both a symmetric and an anti-symmetric tensor – works only because the real manifold has even dimension equal to $2n$. The imaginary factor in front of the form J is needed to ensure that the form is a real 2-form. The manifold is *Kähler* – and J is said to be a *Kähler form* – if J is closed, that is to say if

$$\mathrm{d}J = 2\mathrm{i}g_{a\bar{b},c}\,\mathrm{d}z^c \wedge \mathrm{d}z^a \wedge \mathrm{d}\bar{z}^{\bar{b}} + 2\mathrm{i}g_{a\bar{b},\bar{c}}\,\mathrm{d}\bar{z}^{\bar{c}}\mathrm{d}z^a \wedge \mathrm{d}\bar{z}^{\bar{b}} = 0\;, \tag{3.63}$$

where the comma stands for differentiation with respect to the appropriate coordinate. Now this will be true if and only if

$$g_{a\bar{b},\bar{c}} = g_{a\bar{c},\bar{b}} , \qquad g_{a\bar{b},c} = g_{c\bar{b},a} . \tag{3.64}$$

This implies that in the local coordinate system that we are employing there exists a scalar function $K(z, \bar{z})$ such that the metric can be written as

$$g_{a\bar{b}} = \partial_a \partial_{\bar{b}} K . \tag{3.65}$$

This is a highly non-trivial property because it will be true in all allowed coordinate systems, that is in all coordinate systems related to the present one by an equation of the form $z' = z'(z)$. In this sense it is a more striking statement than the superficially similar Eq. (3.16), which holds in a very restricted class of coordinate systems only. The function $K(z, \bar{z})$ is known as the *Kähler potential* and determines both the metric and the Kähler form.

We have seen that $\mathbf{S}^2$ is a Kähler manifold. This happened because any 2-form on a two-dimensional manifold is closed by default (there are no 3-forms), so that every Hermitian two-dimensional manifold has to be Kähler.

The Levi–Civita connection is constructed in such a way that the length of a vector is preserved by parallel transport. On a complex manifold we have more structure worth preserving, namely the complex structure: we would like the type (p, q) of a tensor to be preserved by parallel transport. We must ask if these two requirements can be imposed at the same time. For Kähler manifolds the answer is 'yes'. On a Kähler manifold the only non-vanishing components of the Christoffel symbols are

$$\Gamma_{ab}^{\ c} = g^{c\bar{d}} g_{\bar{d}a,b} , \qquad \Gamma_{\bar{a}\bar{b}}^{\ \bar{c}} = g^{\bar{a}d} g_{d\bar{a},\bar{b}} . \tag{3.66}$$

Now take a holomorphic tangent vector, that is a vector of type $(1, 0)$ (such as $V = V^a \partial_a$). The equation for parallel transport becomes

$$\nabla_X V^a = \dot{V}^a + X^b \Gamma_{bc}^{\ a} V^c = 0 , \tag{3.67}$$

together with its complex conjugate. If we start with a vector whose components $\bar{V}^{\bar{a}}$ vanish and parallel transport it along some curve, then the components $\bar{V}^{\bar{a}}$ will stay zero since certain components of the Christoffel symbols are zero. In other words a vector of type $(1, 0)$ will preserve its type when parallel transported. Hence the complex structures on the tangent spaces at two different points are compatible, and it follows that we can define vector fields of type $(1, 0)$ and $(0, 1)$, respectively, and similarly for tensor fields.

All formulae become simple on a Kähler manifold. Up to index permutations the only non-vanishing components of the Riemann tensor are

$$R_{a\bar{b}c\bar{d}} = g_{\bar{d}d} \Gamma_{ac}^{\ \ d}{}_{,\bar{b}} . \tag{3.68}$$

Finally, a useful concept is that of of *holomorphic sectional curvature*. Choose a 2-plane in the complexified tangent space at the point z such that it is left invariant by complex conjugation. This means that we can choose coordinates such that the plane is spanned by the tangent vectors $\mathrm{d}z^a$ and $\mathrm{d}\bar{z}^{\bar{a}}$ (and here we use the old-fashioned notation according to which $\mathrm{d}z^a$ are the components of a tangent vector rather than a basis element in the cotangent space). Then the holomorphic sectional curvature is defined by

$$R(z, \mathrm{d}z) \;=\; \frac{R_{a\bar{b}c\bar{d}}\,\mathrm{d}z^a \mathrm{d}\bar{z}^{\bar{b}} \mathrm{d}z^c \mathrm{d}\bar{z}^{\bar{d}}}{(\mathrm{d}s^2)^2}\,. \tag{3.69}$$

Holomorphic sectional curvature is clearly analogous to ordinary scalar curvature on real manifolds and (unsurprisingly to those who are familiar with ordinary Riemannian geometry) one can show that, if the holomorphic sectional curvature is everywhere independent of the choice of the 2-plane, then it is independent of the point z as well. Then the space is said to have constant holomorphic sectional curvature. Since there was a restriction on the choice of the 2-planes, constant holomorphic sectional curvature does not imply constant curvature.

3.4 Symplectic manifolds

Kähler manifolds have two kinds of geometry: Riemannian and symplectic. The former concerns itself with a non-degenerate symmetric tensor field, and the latter with a non-degenerate anti-symmetric tensor field that has to be a closed 2-form as well. This is to say that a manifold is *symplectic* only if there exist two tensor fields Ω_{ij} and Ω^{ij} (not related by raising indices with a metric – indeed no metric is assumed) such that

$$\Omega_{ij} = -\Omega_{ji}\,, \qquad \Omega^{ik}\Omega_{kj} = \delta^i_j\,. \tag{3.70}$$

This is a *symplectic 2-form* Ω if it is also closed,

$$\mathrm{d}\,\Omega = 0 \qquad \Leftrightarrow \qquad \Omega_{[ij,k]} = 0\,. \tag{3.71}$$

These requirements are non-trivial: it may well be that a (compact) manifold does not admit a symplectic structure, although it essentially always admits a metric. Indeed $\mathbf{S}^2$ is the only sphere that is also a symplectic manifold.

A manifold may have a symplectic structure even if it is not Kähler. Phase spaces of Hamiltonian systems are symplectic manifolds, so that – at least in the guise of Poisson brackets – symplectic geometry is quite familiar to physicists.[5] The point is that the symplectic form can be used to associate a vector field with any function

[5] For a survey of symplectic geometry by an expert, see Arnold (2000).

$H(x)$ on the manifold through the equation

$$V_H^i \;=\; \Omega^{ij}\partial_j H \;. \tag{3.72}$$

This is known as a *Hamiltonian vector field*, and it generates *canonical transformations*. These transformations preserve the symplectic form, just as isometries generated by Killing vectors preserve the metric. But the space of canonical transformations is always infinite dimensional (since the function H is at our disposal), while the number of linearly independent Killing vectors is always rather small – symplectic geometry and metric geometry are analogous but different. The *Poisson bracket* of two arbitrary functions F and G is defined by

$$\{F, G\} \;=\; \partial_i F \Omega^{ij} \partial_j G \;. \tag{3.73}$$

It is bilinear, anti-symmetric, and obeys the Jacobi identity precisely because the symplectic form is closed. From a geometrical point of view the role of the symplectic form is to associate an area with each pair of tangent vectors. There is also an interesting interpretation of the fact that the symplectic form is closed, namely that the total area assigned by the symplectic form to a closed surface that can be contracted to a point (within the manifold itself) is zero. Every submanifold of a symplectic manifold inherits a 2-form from the manifold in which it sits, but there is no guarantee that the inherited 2-form is non-degenerate. In fact it may vanish. If this happens to a submanifold of dimension equal to one half of the dimension of the symplectic manifold itself, the submanifold is *Lagrangian*. The standard example is the subspace spanned by the coordinates q in a symplectic vector space spanned by the coordinates q and p, in the way familiar from analytical mechanics.

A symplectic form gives rise to a natural notion of volume, invariant under canonical transformations; if the dimension is $2n$ then the volume element is

$$V = \frac{1}{n!}(\frac{1}{2}\Omega) \wedge (\frac{1}{2}\Omega) \wedge \;\ldots\; (\frac{1}{2}\Omega) \,. \tag{3.74}$$

The numerical factor can be chosen at will – unless we are on a Kähler manifold where the choice just made is the only natural one. The point is that a Kähler manifold has both a metric and a symplectic form, so that there will be two notions of volume that we want to be consistent with each other. The symplectic form is precisely the Kähler form from Eq. (3.62), $\Omega = J$. The special feature of Kähler manifolds is that the two kinds of geometry are interwoven with each other and with the complex structure. On the 2-sphere

$$\mathrm{d}V = \frac{1}{2}\Omega = \frac{2i}{(1+|z|^2)^2}\,\mathrm{d}z \wedge \mathrm{d}\bar{z} = \frac{4}{(1+r^2)^2}\,\mathrm{d}x \wedge \mathrm{d}y = \sin\theta\,\mathrm{d}\theta \wedge \mathrm{d}\phi \;. \tag{3.75}$$

This agrees with the volume form as computed using the metric.

3.5 The Hopf fibration of the 3-sphere

The 3-sphere, being odd-dimensional, is neither complex nor symplectic, but like all the odd-dimensional spheres it is a fibre bundle. Unlike all other spheres (except $\mathbf{S}^1$) it is also a Lie group. The theory of fibre bundles was in fact created in response to the treatment that the 3-sphere was given in 1931 by Hopf and by Dirac, and we begin with this part of the story. (By the way, Dirac's concern was with magnetic monopoles.)

The 3-sphere can be defined as the hypersurface

$$X^2 + Y^2 + Z^2 + U^2 = 1 \tag{3.76}$$

embedded in a flat four-dimensional space with (X, Y, Z, U) as its Cartesian coordinates. Using stereographic coordinates (Section 3.1) we can visualize the 3-sphere as $\mathbb{R}^3$ with the south pole ($U = -1$) added as a point 'at infinity'. The equator of the 3-sphere ($U = 0$) will appear in the picture as a unit sphere surrounding the origin. To get used to it we look at geodesics and Killing vectors.

Using Eq. (3.12) it is easy to prove that geodesics appear in our picture either as circles or as straight lines through the origin. Either way – unless they are great circles on the equator – they meet the equator in two antipodal points. Now rotate the sphere in the X–Y plane. The appropriate Killing vector field is

$$J_{XY} = X\partial_Y - Y\partial_X = x\partial_y - y\partial_x \,, \tag{3.77}$$

where x, y (and z) are the stereographic coordinates in our picture. This looks like the flow lines of a rotation in flat space. There is a geodesic line of fixed points along the z-axis. The flow lines are circles around this geodesic, but with one exception they are not themselves geodesics because they do not meet the equator in antipodal points. The Killing vector J_{ZU} behaves intrinsically just like J_{XY}, but it looks quite different in our picture (because we singled out the coordinate U for special treatment). It has fixed points at

$$J_{ZU} = Z\partial_U - U\partial_Z = 0 \quad \Leftrightarrow \quad Z = U = 0 \,. \tag{3.78}$$

This is a geodesic (as it must be), namely a great circle on the equator. By analogy with J_{XY} the flow lines must lie on tori surrounding the line of fixed points. A somewhat boring calculation confirms this; the flow of J_{ZU} leaves the tori of revolution

$$(\rho - a)^2 + z^2 = a^2 - 1 > 0 \,, \qquad \rho^2 \equiv x^2 + y^2 \,, \tag{3.79}$$

invariant for any $a > 1$. So we can draw pictures of these two Killing vector fields, the instructive point of the exercise being that intrinsically these Killing vector fields are really 'the same'.

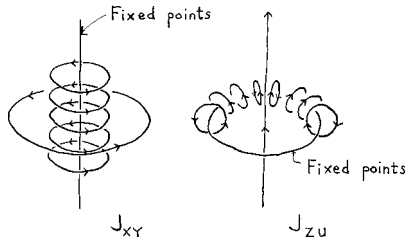


Figure 3.7. Flow lines and fixed points of J_{XY} and J_{ZU}.

A striking thing about the 3-sphere is that there are also Killing vector fields that are everywhere non-vanishing. This is in contrast to the 2-sphere; a well-known theorem in topology states that 'you can't comb a sphere', meaning to say that every vector field on $\mathbf{S}^2$ has to have a fixed point somewhere. An example of a Killing field without fixed points on $\mathbf{S}^3$ is clearly

$$\xi = J_{XY} + J_{ZU} \; ; \qquad ||\xi||^2 = X^2 + Y^2 + Z^2 + U^2 = 1 \; . \tag{3.80}$$

Given our pictures of Killing vector fields it is clear that this combination must have flow lines that lie on the tori that we drew, but which wind once around the z-axis each time they wind around the circle $\rho = 1$. This will be our key to understanding the 3-sphere as a fibre bundle.

Remarkably, all the flow lines of the Killing vector field ξ are geodesics as well. We will prove this in a way that brings complex manifolds back in. The point is that the embedding space $\mathbb{R}^4$ is also the complex vector space $\mathbb{C}^2$. Therefore we can introduce the complex embedding coordinates

$$\begin{bmatrix} Z^1 \\ Z^2 \end{bmatrix} = \begin{bmatrix} X + \mathrm{i}Y \\ Z + \mathrm{i}U \end{bmatrix} \; . \tag{3.81}$$

The generalization to n complex dimensions is immediate. Let us use P, Q, R, ... to denote vectors in complex vector spaces. The scalar product in $\mathbb{R}^{2n}$ becomes an *Hermitian form* on $\mathbb{C}^n$, namely

$$P \cdot \bar{Q} = \delta_{\alpha\bar{\alpha}} P^\alpha \bar{Q}^{\bar{\alpha}} = P^\alpha \bar{Q}_\alpha \; . \tag{3.82}$$

Here we made the obvious move of defining

$$(Z^\alpha)^* = \bar{Z}^{\bar{\alpha}} \equiv \bar{Z}_\alpha \; , \tag{3.83}$$

so that we get rid of the barred indices.

The odd-dimensional sphere $\mathbf{S}^{2n+1}$ is now defined as those points in $\mathbb{C}^{n+1}$ that obey

$$Z \cdot \bar{Z} = 1 \; . \tag{3.84}$$

Translating the formula (3.12) for a general geodesic from the real formulation given in Section 3.1 to the complex one that we are using now we find

$$Z^\alpha(\sigma) = m^\alpha \cos\sigma + n^\alpha \sin\sigma \ , \qquad m\cdot\bar{m} = n\cdot\bar{n} = 1 \ , \ m\cdot\bar{n} + n\cdot\bar{m} = 0 \ , \tag{3.85}$$

where the affine parameter σ measures distance d along the geodesic,

$$d = |\sigma_2 - \sigma_1| \ . \tag{3.86}$$

If we pick two points on the geodesic, say

$$Z_1^\alpha \equiv Z^\alpha(\sigma_1) \qquad Z_2^\alpha \equiv Z^\alpha(\sigma_2) \ , \tag{3.87}$$

then a short calculation reveals that the distance between them is given by

$$\cos d = \frac{1}{2}(Z_1\cdot\bar{Z}_2 + Z_2\cdot\bar{Z}_1) \ . \tag{3.88}$$

This is a useful formula to have.

Now consider the family of geodesics given by

$$n^\alpha = \mathrm{i}m^\alpha \qquad \Rightarrow \qquad Z^\alpha(\sigma) = \mathrm{e}^{\mathrm{i}\sigma}m^\alpha \ . \tag{3.89}$$

Through any point on $\mathbf{S}^{2n+1}$ there will go a geodesic belonging to this family since we are free to let the vector m^α vary. Evidently the equation

$$\dot{Z}^\alpha = \mathrm{i}Z^\alpha \tag{3.90}$$

holds for every geodesic of this kind. Its tangent vector is therefore given by

$$\partial_\sigma = \dot{Z}^\alpha\partial_\alpha + \dot{\bar{Z}}^{\bar{\alpha}}\partial_{\bar{\alpha}} = \mathrm{i}(Z^\alpha\partial_\alpha - \bar{Z}^{\bar{\alpha}}\partial_{\bar{\alpha}}) = J_{XY} + J_{ZU} = \xi \ . \tag{3.91}$$

But this is precisely the everywhere non-vanishing Killing vector field that we found before. So we have found that on $\mathbf{S}^{2n+1}$ there exists a *congruence* – a space-filling family – of curves that are at once geodesics and Killing flow lines. This is a quite remarkable property; flat space has it, but very few curved spaces do.

In flat space the distance between parallel lines remains fixed as we move along the lines, whereas two skew lines eventually diverge from each other. In a positively curved space – like the sphere – parallel geodesics converge, and we ask if it is possible to twist them relative to each other in such a way that this convergence is cancelled by the divergence caused by the fact that they are skew. Then we would have a congruence of geodesics that always stay at the same distance from each other. We will call the geodesics *Clifford parallels* provided that this can be done. A more stringent definition requires a notion of parallel transport that takes tangent vectors of the Clifford parallels into each other; we will touch on this in Section 3.7.

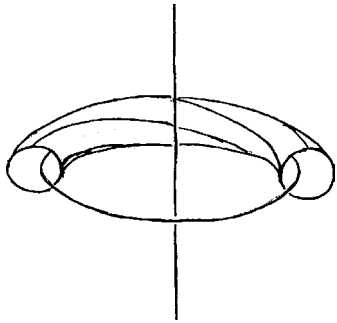


Figure 3.8. The Hopf fibration of the 3-sphere.

Now the congruence of geodesics given by the vector field ξ are Clifford parallels on the 3-sphere. Two points belonging to different geodesics in the congruence must preserve their relative distance as they move along the geodesics, precisely because the geodesics are Killing flow lines as well. It is instructive to prove this directly though. Consider two geodesics defined by

$$P^\alpha = \mathrm{e}^{\mathrm{i}\sigma} P_0^\alpha \,, \qquad Q^\alpha = \mathrm{e}^{\mathrm{i}(\sigma+\sigma_0)} Q_0^\alpha \,. \tag{3.92}$$

We will soon exercise our right to choose the constant σ_0. The scalar product of the constant vectors will be some complex number

$$P_0 \cdot \bar{Q}_0 = r\mathrm{e}^{\mathrm{i}\phi} \,. \tag{3.93}$$

The geodesic distance between two arbitrary points, one on each geodesic, is therefore given by

$$\cos d = \frac{1}{2}(P \cdot \bar{Q} + Q \cdot \bar{P}) = r\cos(\phi - \sigma_0) \,. \tag{3.94}$$

The point is that this is independent of the affine parameter σ, so that the distance does not change as we move along the geodesics (provided of course that we move with the same speed on both). This shows that our congruence of geodesics consists of Clifford parallels.

The perpendicular distance d_0 between a pair of Clifford parallels is obtained by adjusting the zero point σ_0 so that $\cos d_0 = r$, that is so that the distance attains its minimum value. A concise way to express d_0 is by means of the equation

$$\cos^2 d_0 = r^2 = P_0 \cdot \bar{Q}_0 Q_0 \cdot \bar{P}_0 = P \cdot \bar{Q} Q \cdot \bar{P} \,. \tag{3.95}$$

Before we are done we will see that this formula plays an important role in quantum mechanics.

It is time to draw a picture – Figure 3.8 – of the congruence of Clifford parallels. Since we have the pictures of J_{XY} and J_{ZU} already this is straightforward. We draw

a family of tori of revolution surrounding the unit circle in the $z = 0$ plane, and eventually merging into the z-axis. These are precisely the tori defined in Eq. (3.79). Each torus is itself foliated by a one parameter family of circles and they are twisting relative to each other as we expected; indeed any two circles in the congruence are linked (their linking number is one). The whole construction is known as the *Hopf fibration* of the 3-sphere, and the geodesics in the congruence are called *Hopf circles*. It is clear that there exists another Hopf fibration with the opposite twist, that we would arrive at through a simple sign change in Eq. (3.80). By the way the metric induced on the tori by the metric on the 3-sphere is flat (as you can easily check from Eq. (3.98) below, where a torus is given by $\theta =$ constant); you can think of the 3-sphere as a one parameter family of flat tori if you please, or as two solid tori glued together.

Now the interesting question is: how many Hopf circles are there altogether, or more precisely what is the space whose points consists of the Hopf circles? A little thinking gives the answer directly. On each torus there is a one parameter set of Hopf circles, labelled by some periodic cordinate $\phi \in [0, 2\pi[$. There is a one parameter family of tori, labelled by $\theta \in]0, \pi[$. In this way we account for every geodesic in the congruence except the circle in the $z = 0$ plane and the one along the z-axis. These have to be added at the endpoints of the θ-interval, at $\theta = 0$ and $\theta = \pi$, respectively. Evidently what we are describing is a 2-sphere in polar coordinates. So the conclusion is that the space of Hopf circles is a 2-sphere.

It is important to realize that this 2-sphere is not 'sitting inside the 3-sphere' in any natural manner. To find such an embedding of the 2-sphere would entail choosing one point from each Hopf circle in some smooth manner. Equivalently, we want to choose the zero point of the coordinate σ along all the circles in some coherent way. But this is precisely what we cannot do; if we could we would effectively have shown that the topology of $\mathbf{S}^3$ is $\mathbf{S}^2 \otimes \mathbf{S}^1$ and this is not true (because in $\mathbf{S}^2 \otimes \mathbf{S}^1$ there are closed curves that cannot be contracted to a point, while there are no such curves in the $\mathbf{S}^3$). We can almost do it though. For instance, we can select those points where the geodesics are moving down through the $z = 0$ plane. This works fine except for the single geodesic that lies in this plane; we have mapped all of the 2-sphere except one point onto an open unit disc in the picture. It is instructive to make a few more attempts in this vein and see how they always fail to work globally.

The next question is whether the 2-sphere of Hopf circles is a round 2-sphere or not, or indeed whether it has any natural metric at all. The answer turns out to be 'yes'. To see this it is convenient to introduce the *Euler angles*, which are intrinsic coordinates on $\mathbf{S}^3$ adapted to the Hopf fibration. They are defined by

$$\begin{bmatrix} Z^1 \\ Z^2 \end{bmatrix} = \begin{bmatrix} X + \mathrm{i}Y \\ Z + \mathrm{i}U \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{\frac{\mathrm{i}}{2}(\tau+\phi)} \cos\frac{\theta}{2} \\ \mathrm{e}^{\frac{\mathrm{i}}{2}(\tau-\phi)} \sin\frac{\theta}{2} \end{bmatrix}, \tag{3.96}$$

where

$$0 \leq \tau < 4\pi\ , \quad 0 \leq \phi < 2\pi\ , \quad 0 < \theta < \pi\ . \tag{3.97}$$

We have seen that the periodic coordinate τ goes along the Hopf circles, in fact $\tau = 2\sigma$ in the congruence (3.89). (Making τ periodic with period 2π gives the Hopf fibration of real projective 3-space – checking this statement is a good way of making sure that one understands the Hopf fibration.) The coordinate ϕ runs along Hopf circles of the opposite twist. Finally a little calculation verifies that the coordinate θ labels the tori in Eq. (3.79), with $\cos(\theta/2) = 1/a$. The intrinsic metric on the 3-sphere becomes

$$\mathrm{d}s^2 = |\mathrm{d}Z^1|^2 + |\mathrm{d}Z^2|^2 = \frac{1}{4}(\mathrm{d}\tau^2 + \mathrm{d}\theta^2 + \mathrm{d}\phi^2 + 2\cos\theta\ \mathrm{d}\tau\mathrm{d}\phi)\ . \tag{3.98}$$

Here it is easy to see that the tori at $\theta =$ constant are flat.

To continue our argument: since the coordinates θ and ϕ label the 2-sphere's worth of geodesics in the congruence we can try to map this $\mathbf{S}^2$ into $\mathbf{S}^3$ through an equation of the form

$$\tau\ =\ \tau(\theta, \phi). \tag{3.99}$$

But we have already shown that no such map can exist globally, and therefore this is not the way to define a natural metric on our 2-sphere. We still do not know if it is a round 2-sphere in any natural way!

Nevertheless the space of Clifford parallels is naturally a round 2-sphere. We simply define the distance between two arbitrary Clifford parallels as the perpendicular distance d_0 between them. This we have computed already, and it only remains to rewrite Eq. (3.95) in terms of the Euler angles. If the coordinates of the two points on the 2-sphere are (θ_1, ϕ_1) and (θ_2, ϕ_2) we obtain

$$\cos^2 d_0 = \frac{1}{2}\left(1 + \cos\theta_1\cos\theta_2 + \cos(\phi_1 - \phi_2)\sin\theta_1\sin\theta_2\right)\ . \tag{3.100}$$

This formula should be familiar from spherical trigonometry. If the two points are infinitesimally close to each other we can expand the left-hand side as

$$\cos^2 d_0\ \approx\ 1 - d_0^2\ \equiv\ 1 - \mathrm{d}s^2\ . \tag{3.101}$$

A short calculation then verifies that that the metric is

$$\mathrm{d}s^2 = \frac{1}{4}(\mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2)\ . \tag{3.102}$$

Precisely one quarter of the usual round metric. From now on, when we talk about the Hopf fibration it will be understood that the 2- and 3-spheres are equipped with the above metrics.

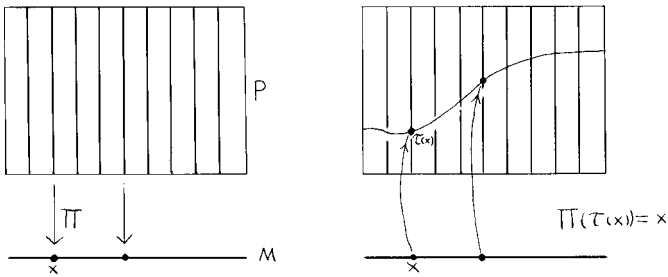


Figure 3.9. Left: the definition of a fibre bundle, including a fibred bundle space P, a base manifold M and a projection Π. Right: a section of the bundle, that is an embedding of the base manifold in the bundle space (which may fail to exist globally).

3.6 Fibre bundles and their connections

Let us now bring in the general theory of fibre bundles.[6] By definition the *fibre* of a map $P \to M$ between two spaces is the set of points in P that are mapped to a given point in M. The situation gets more interesting if all fibres are isomorphic. The definition of a fibre bundle also requires that there is a group acting on the fibres: a *fibre bundle* over a *base manifold* M consists of a *bundle space* P and a *structure group* G acting on the bundle space in such a way that the base manifold is equal to the quotient P/G (a point in M is an orbit of G in P; the orbits are assumed to be isomorphic to each other). In this way we get a canonical projection $\Pi : P \to M$. The set of points that project to a particular point p on M is known as the fibre F over p. It is also required that the bundle space is locally equal to a Cartesian product, that is to say that P can be covered by open sets of the form $U \times F$, where U is an open set in M. A *principal fibre bundle* is a fibre bundle such that the fibres are copies of the group manifold of G.

A *Cartesian product* $M \times F$ (as in $\mathbb{R}^2 \equiv \mathbb{R} \times \mathbb{R}$) is a trivial example of a fibre bundle. The 3-sphere on the other hand is non-trivial because it is not just a Cartesian product of $\mathbf{S}^2$ and $\mathbf{S}^1$, although locally it behaves like that. It should be clear from the previous discussion that the 3-sphere really is a principal fibre bundle with structure group $U(1)$, whose group manifold is indeed a circle. The fact that the structure group is Abelian is a simplifying feature; in general the fibres can be many-dimensional and the structure group can be non-Abelian.

For many purposes it is convenient to introduce a local coordinate system on the bundle that is adapted to the bundle structure. Schematically, let x^a be some

[6] A standard reference for connections and fibre bundles (and for a taste of the precision used in modern differential geometry) is Kobayashi and Nomizu (1963). There are many readable introductions for physicists; examples can be found in Shapere and Wilczek (1989).

coordinates on the base manifold and τ a coordinate along the fibres, assumed one-dimensional for simplicity. A full set of coordinates is then given by $x^i = (\tau, x^a)$. On $\mathbf{S}^3$ the coordinates θ and ϕ play the part of x^a while τ is the fibre coordinate. Now the idea is to restrict oneself to coordinates that can be reached from x^a, τ through coordinate transformations of the general form

$$x'^a = x'^a(x), \qquad \tau' = \tau'(x, \tau). \tag{3.103}$$

Such 'triangular' coordinate transformations appear because there is no natural way of identifying the fibres with each other.

To take a *section* of the bundle means to specify the fibre coordinate τ as a function of x^a,

$$\tau = \tau(x). \tag{3.104}$$

Locally this defines an embedding of the base manifold into the bundle. In coordinate independent terms, a section is defined as an embedding of the base manifold into the bundle such that if we follow it up with the projection down to the base we get back to the point we started out from. For a non-trivial bundle such as the 3-sphere no global section exists, but it is possible to take local sections on the coordinate patches U. In the overlap regions where two different local sections are defined one can go from one to the other provided that one moves along the fibres according to

$$\tau = \tau(x) \quad \rightarrow \quad \tau' = \tau'(x). \tag{3.105}$$

Such a transformation along the fibres is known as a *local gauge transformation*, for reasons that will become clear later. A key theorem is the following:

Theorem 3.1 *A principal fibre bundle admits a global section if and only if it is a trivial Cartesian product.*

Another important fact about fibre bundles is that if one knows all the coordinate patches $U \times F$ as well as the local gauge transformations needed to go from one patch to another, then one can reconstruct the bundle. It is not always the case that one can see what the entire bundle looks like, in the intimate manner that one can see the 3-sphere!

Fibre bundles which are not principal are also of interest. An example is the *vector bundle*, where the fibres consist of vector spaces on which the group G acts, rather than of G itself. Note that the theorem that a fibre bundle has the trivial product topology if and only if it admits a global section holds for principal bundles, and only for principal bundles. A famous example of a non-trivial vector bundle that admits a global section is the Möbius strip; the group G that acts on the fibre is

the finite group $\mathbb{Z}_2$ and indeed the principal bundle does not admit a global section. It is so well known how to construct a Möbius strip made of paper that we simply recommend the reader to make one and check these statements. The *tangent bundle* over a space $\mathbf{M}$ is also a vector bundle. The fibre over a point q is the tangent space $\mathbf{T}_q$ at this point. The group of general linear transformations acts on the fibres of the tangent bundle. A section of the tangent bundle – one element of the fibre at each point in the base space – is just a vector field.

One often wants to *lift* a curve in the base manifold M to the bundle P. Since many curves in the bundle project down to the same curve in M, this requires a further structure, namely a *connection* on the bundle. The idea is to decompose the tangent space of the bundle into *vertical* and *horizontal* directions, that is directions along the fibres and 'perpendicular' to the fibres, respectively. Then the idea is that the *horizontal lift* of a curve threads its way in the horizontal directions only. The problem is that there may be no notion of 'perpendicular' available, so this must be supplied. To a mathematician, a connection is a structure that defines a split of the tangent space into vertical and horizontal. To a physicist, a connection is like the vector potential in electrodynamics, and it takes a little effort to see that the same structure is being referred to. The key fact that must be kept in mind throughout the discussion is that if there is an embedding of one manifold into another, then a differential form on the latter automatically defines a differential form on the former (see Appendix A1.3).

We simplify matters by assuming that the fibres are one-dimensional, and let the coordinate τ run along the fibres. We want to decompose tangent space so that

$$\mathbf{T}_p = \mathbf{V}_p \oplus \mathbf{H}_p \,. \tag{3.106}$$

Any tangent vector can then be written as the sum of one vector belonging to the vertical subspace $\mathbf{V}_p$ and one belonging to the horizontal subspace $\mathbf{H}_p$. Since the vertical vector points along the fibres it must be proportional to

$$\partial_\tau \in \mathbf{V}_p \,. \tag{3.107}$$

But it is not so clear what a horizontal vector should look like. The trick is to choose a special 1-form ω (that is a covariant vector ω_i) and to declare that a horizontal vector h^i is any vector that obeys

$$h^i \in \mathbf{H}_p \qquad \Leftrightarrow \qquad \omega_i h^i = 0 \,. \tag{3.108}$$

Such a statement does not require any metric. But how do we choose ω? The obvious guess $\mathrm{d}\tau$ has a problem since we are permitting coordinate transformations under which

$$\mathrm{d}\tau \;=\; \frac{\partial \tau}{\partial \tau'}\,\mathrm{d}\tau' + \frac{\partial \tau}{\partial x'^a}\,\mathrm{d}x'^a \,. \tag{3.109}$$

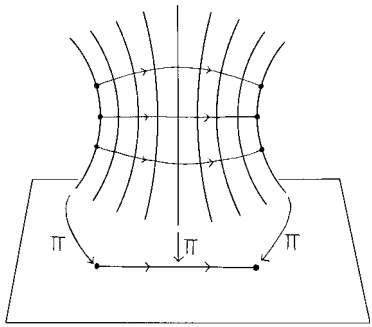


Figure 3.10. Horizontal lifts of a curve in the base manifold, using a notion of 'horizontal subspaces'.

Hence the naive definition of the horizontal subspace is tied to a particular set of coordinates. There will be many ways to perform the split. A priori none of them is better than any other. This is something that we will have to accept, and we therefore simply choose an ω of the general form

$$\omega = \mathrm{d}\tau + U = \mathrm{d}\tau + U_a \mathrm{d}x^a \,. \tag{3.110}$$

In this way a connection, that is a decomposition of the tangent space, is equivalent to the specification of a special 1-form on the bundle.

With a connection in hand, we define *parallel transport* as follows: Fix a curve $x^a(\sigma)$ in the base manifold. Lift this curve to a curve in the bundle by insisting that the 1-form induced on the lifted curve by the connection 1-form ω vanishes. Thus

$$\omega = \mathrm{d}\tau + U = \left(\frac{\mathrm{d}\tau}{\mathrm{d}\sigma} + U_a \frac{\mathrm{d}x^a}{\mathrm{d}\sigma} \right) \mathrm{d}\sigma = 0 \,. \tag{3.111}$$

This leads to an ordinary differential equation that always admits a solution (at least for some range of the parameter σ), and it follows that we can associate a unique family of curves $\big(\tau(\sigma), x^a(\sigma)\big)$ in the bundle to any curve in the base manifold. Through any point in the fibre there passes a unique curve. This is the horizontal lift of the curve in the base manifold.

There are two things to notice. First, that this notion of parallel transport is somewhat different than that discussed in Section 3.2. There we were transporting vectors, here we are transporting the position along the fibres. (Actually, the parallel transport defined in Section 3.2 is a special case, taking place in the tangent bundle.) Second, it may seem that we now have the means to identify the fibres with each other. But this is false in general, because parallel transport may depend on the path – and it may do so even if the bundle is topologically trivial. Indeed the condition that parallel transport be independent of the path, or in other words that the horizontal

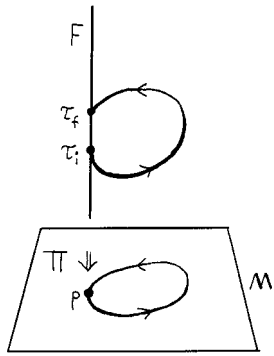


Figure 3.11. Holonomy. For the 3-sphere, the holonomy $\tau_f - \tau_i$ is proportional to the area surrounded by the loop in the base manifold (i.e. the 2-sphere).

lift of a closed curve is itself closed, is (by Stokes' theorem, see Appendix A1.1)

$$\oint_C \omega \;=\; \int_S \mathrm{d}\omega \;=\; 0\,, \tag{3.112}$$

where S is any surface enclosed by a curve C that projects to a given closed curve in the base manifold. The 2-form

$$\Omega \equiv \mathrm{d}\omega = \mathrm{d}U = \frac{1}{2}\,(\partial_a U_b - \partial_b U_a)\,\mathrm{d}x^a \wedge \mathrm{d}x^b \tag{3.113}$$

is known as the *curvature 2-form*. The point is that when the curvature 2–form is non-zero, and when Stokes' theorem applies, Eq. (3.112) will not hold. Then parallel transport along a closed curve will give rise to a shift along the fibre, known to mathematicians as a *holonomy*.

Throughout, the connection is a 1-form that is defined on the bundle space – not on the base manifold. If we take a section of the bundle we have an embedding of the base manifold into the bundle, and the connection on the bundle will induce a definite 1-form on the base manifold. This is what happens in electrodynamics; in this way the physicist's connection is recovered from that of the mathematician. There is a catch: if the section is not globally defined, the connection on the base manifold will be ill-defined in some places. On the other hand there is no such problem with the curvature 2-form – Ω does exist globally on the base manifold.

Formula (3.113) is valid in this simple form for the case of one-dimensional fibres only. But in physics we often need many-dimensional fibres – and a non-Abelian group G. The main novelty is that the connection takes values in the Lie algebra of G; since the dimension of the Lie algebra is equal to the dimension of the fibres this will be enough to single out a unique horizontal subspace. One difficulty that arises is that it is hard to get explicit about holonomies – Stokes' theorem is

no longer helpful because the line integrals must be *path ordered* due to the fact that the connections at different points along the curve do not commute with each other.

Coming back to the 3-sphere, we observe that the fibres are the Hopf circles and the base manifold is a 2-sphere. Now the 3-sphere is not just any fibre bundle. Its bundle space is equipped with a metric, which means that there is a *preferred connection* singled out by the requirement that the vertical and the horizontal subspaces be orthogonal in the sense of that metric. If the horizontal lift of a curve in the base manifold is parametrized by σ its tangent vector is

$$\partial_\sigma \;=\; \dot\tau\,\partial_\tau + \dot\theta\,\partial_\theta + \dot\phi\,\partial_\phi \;\equiv\; h^i\,\partial_i \;. \tag{3.114}$$

The tangent vector of a fibre is $\partial_\tau \equiv v^i\,\partial_i$. We require

$$g_{ij}\,v^i h^j = g_{\tau\tau}\,\dot\tau + g_{\tau\theta}\dot\theta + g_{\tau\phi}\dot\phi = \dot\tau + \cos\theta\,\dot\phi = 0\;, \tag{3.115}$$

where we used Eq. (3.98) for the metric. Hence the metrically preferred connection is

$$\omega \;=\; \mathrm{d}\tau + \cos\theta\,\mathrm{d}\phi\;. \tag{3.116}$$

The curvature of this connection is non-vanishing; indeed if parallel transport occurs along a closed curve ∂S, bounding a surface S in the base manifold, then the holonomy is

$$\tau_{\text{final}} - \tau_{\text{initial}} = \int \mathrm{d}\tau = -\oint_{\partial S} \cos\theta\,\mathrm{d}\phi = \int_S \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\;. \tag{3.117}$$

This is proportional to the area of the enclosed surface on $\mathbf{S}^2$, so the interplay between parallel transport and the metric is intimate.

We will now take a section of $\mathbf{S}^3$, since we wish to view the base manifold $\mathbf{S}^2$ as an embedded submanifold of the 3-sphere so that ω induces a 1-form on the 2-sphere. We will work in some coordinate patch, say using the Euler angles. One local section is defined by the equation

$$\tau \;=\; \phi\;. \tag{3.118}$$

It picks out one point on every fibre except – and here we have to recall how the Euler angles were defined – for that fibre which corresponds to the north pole of the 2-sphere. The form that is induced by the connection on the 2-sphere is

$$\omega^+ = (1+\cos\theta)\,\mathrm{d}\phi\;. \tag{3.119}$$

When looking at this expression we must remember that we are using a coordinate system on $\mathbf{S}^2$ that is ill-defined at the poles. But we are particularly interested in the behaviour at the north pole, where we expect this form to be singular somehow. If,

with a little effort, we transform from θ and ϕ to our standard complex coordinate system on $\mathbf{S}^2$ we find that

$$\omega^+ = \frac{\mathrm{i}}{1+|z|^2}\left(-\frac{\mathrm{d}z}{z}+\frac{\mathrm{d}\bar{z}}{\bar{z}}\right) . \tag{3.120}$$

This is indeed ill-defined at the north pole ($z = 0$). On the other hand there is no problem at the south pole; transforming to complex coordinates that cover this point ($z \to z' = 1/z$) we find that

$$\omega^+ = \frac{\mathrm{i}}{1+|z'|^2}\left(\bar{z}'\mathrm{d}z' - z'\mathrm{d}\bar{z}'\right) , \tag{3.121}$$

and there is no problem at the south pole ($z' = 0$).

By taking a different section (say $\tau = -\phi$) another form is induced on $\mathbf{S}^2$. If we choose $\tau = -\phi$ we obtain

$$\omega^- = (\cos\theta - 1)\,\mathrm{d}\phi . \tag{3.122}$$

This turns out to be well defined everywhere except at the south pole. In the region of the 2-sphere where both sections are well defined – which happens to be the region where the polar coordinates are well defined – the forms are related by

$$\omega^+ - \omega^- = 2\mathrm{d}\phi . \tag{3.123}$$

This is a local gauge transformation in the ordinary sense of the word as used in electrodynamics.

In spite of the little difficulties there is nothing wrong with the connection; we have only confirmed the conclusion that no global section of the bundle exists. The curvature 2-form

$$\Omega_{\theta\phi} = \sin\theta \tag{3.124}$$

is an everywhere regular 2-form on the base manifold, and in fact equal to the symplectic form on $\mathbf{S}^2$. Quite incidentally, if we regard the 2-sphere as a sphere embedded in Euclidean 3-space our curvature tensor corresponds to a radially directed magnetic field. There is a magnetic monopole sitting at the origin – and this is what Dirac was doing in 1931, while Hopf was fibreing the 3-sphere.

3.7 The 3-sphere as a group

The 3-sphere is not only a fibre bundle, it is also a group. More precisely it is a *group manifold*, that is to say a space whose points are the elements of a group. Only two spheres are group manifolds – the other case is $\mathbf{S}^1$, the group manifold of $U(1)$. What we are going to show for the 3-sphere is therefore exceptional as far as spheres are

concerned, but it is typical of all *compact semi-simple* Lie groups. Compact means that the group manifold is compact; semi-simple is a technical condition which is obeyed by the *classical groups* $SO(N)$, $SU(N)$ and $Sp(N)$, with the exception of $SO(2) = U(1)$. What we tell below about $SU(2)$ will be typical of all the classical groups, and this is good enough for us.[7]

The classical groups are *matrix groups* consisting of matrices obeying certain conditions. Thus the group $SU(2)$ is by definition the group of unitary 2×2 matrices of determinant one. An arbitrary group element g in this group can be parametrized by two complex numbers obeying one condition, namely as

$$g = \begin{bmatrix} Z^1 & Z^2 \\ -\bar{Z}^2 & \bar{Z}^1 \end{bmatrix}, \qquad |Z^1|^2 + |Z^2|^2 = 1 \,. \tag{3.125}$$

This means that there is a one-to-one correspondence between $SU(2)$ group elements on the one hand and points on the 3-sphere on the other, that is to say that whenever $X^2 + Y^2 + Z^2 + U^2 = 1$ we have

$$(X, Y, Z, U) \qquad \leftrightarrow \qquad g = \begin{bmatrix} X + \mathrm{i}Y & Z + \mathrm{i}U \\ -Z + \mathrm{i}U & X - \mathrm{i}Y \end{bmatrix} . \tag{3.126}$$

We begin to see the 3-sphere in a new light.

Given two points in a group the group law determines a third point, namely $g_3 = g_1 g_2$. For a matrix group it is evident that the coordinates of g_3 will be given by real analytic functions of the coordinates of g_1 and g_2. This property defines a *Lie group*. Now a unitary matrix g of unit determinant can always be obtained by exponentiating a traceless anti-Hermitian matrix m. We can insert a parameter τ in the exponent and then we get

$$g(\tau) \;=\; \mathrm{e}^{\tau m} \,. \tag{3.127}$$

From one point of view this is a *one parameter subgroup* of $SU(2)$, from another it is a curve in the group manifold $\mathbf{S}^3$. When $\tau = 0$ we are sitting at a point that is rather special from the first viewpoint, namely at the unit element $\mathbb{1}$ of the group. The tangent vector of the curve at this point is

$$\dot{g}(\tau)_{|\tau=0} \;=\; m \,. \tag{3.128}$$

In this way we see that the vector space of traceless anti-Hermitian matrices make up the tangent space at the unit element of the group. (Physicists usually prefer to work with Hermitian matrices, but here anti-Hermitian matrices are more natural – to convert between these preferences just multiply by i. Incidentally, the, at first

[7] There are many good books on group theory available where the reader will find a more complete story; an example that leans towards quantum mechanics is Gilmore (1974).

sight, confusing habit of viewing a matrix as an element of a vector space will grow upon us.) This tangent space is called the *Lie algebra* of the group, and it is part of the magic of Lie groups that many of their properties can be reliably deduced from a study of their Lie algebras. A reminder about Lie algebras can be found in Appendix 2.

A group always acts on itself. A group element g_1 transforms an arbitrary group element g by *left translation* $g \to g_1 g$, by *right translation* $g \to g g_1^{-1}$, and by the *adjoint action* $g \to g_1 g g_1^{-1}$. (The inverse occurs on the right-hand side to ensure that $g \to g g_1^{-1} \to g g_1^{-1} g_2^{-1} = g(g_2 g_1)^{-1}$.) Fix a one parameter subgroup $g_1(\sigma)$ and let it act by left translation; $g \to g_1(\sigma) g$. This defines a curve through every point of $\mathbf{S}^3$, and moreover the group laws ensure that these curves cannot cross each other, so that they form a congruence. The congruence of Clifford parallels that we studied in Section 3.5 can be viewed in this light. To do so consider the one parameter subgroup obtained by exponentiating $m = \mathrm{i}\sigma_3$, where σ_i denotes a Pauli matrix. It gives rise to an infinitesimal transformation $\delta g = mg$ at every point g. If we use the coordinates (X, Y, Z, U) to describe g we find that this equation becomes

$$\begin{bmatrix} \delta(X + \mathrm{i}Y) & \delta(Z + \mathrm{i}U) \\ -\delta(Z + \mathrm{i}U) & \delta(X - \mathrm{i}Y) \end{bmatrix} = \begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{bmatrix} \begin{bmatrix} X + \mathrm{i}Y & Z + \mathrm{i}U \\ -Z + \mathrm{i}U & X - \mathrm{i}Y \end{bmatrix} . \tag{3.129}$$

Working this out we find it to be the same as

$$(\delta X, \delta Y, \delta Z, \delta U) = (J_{XY} + J_{ZU})(X, Y, Z, U) . \tag{3.130}$$

This is precisely the Killing vector field that points along a congruence of Hopf circles. Choosing $m = \mathrm{i}\sigma_1$ and $m = \mathrm{i}\sigma_2$ will give us a total of three nowhere vanishing linearly independent vector fields. They are

$$J_1 = J_{XU} + J_{YZ} , \qquad J_2 = J_{XZ} + J_{UY} , \qquad J_3 = J_{XY} + J_{ZU} \tag{3.131}$$

and they form a representation of the Lie algebra of $SU(2)$. Their number is the same as the dimension of the manifold. We say that the 3-sphere is *parallelizable*. All group manifolds are parallelizable, for the same reason, but among the spheres only $\mathbf{S}^1$, $\mathbf{S}^3$ and $\mathbf{S}^7$ are parallelizable, so this is an exceptional property shared by all group manifolds. Such everywhere non-vanishing vector fields can be used to provide a canonical identification of tangent spaces at different points of the group manifold. One can then modify the rules for parallel transport of vectors so that it respects this identification – in fact we can use Eq. (3.15) but with the totally antisymmetric structure constants of the Lie algebra in the role of the tensor T_{ijk}. For the 3-sphere, this allows us to give a more stringent definition of Clifford parallels than the one we gave in Section 3.5.

So much for left translation. Right translations of the group also give rise, in an analogous way, to a set of three linearly independent vector fields

$$\tilde{J}_1 = J_{XU} - J_{YZ}\,, \qquad \tilde{J}_2 = J_{XZ} - J_{UY}\,, \qquad \tilde{J}_3 = J_{XY} - J_{ZU}\,. \tag{3.132}$$

These three represent another $SU(2)$ Lie algebra and they commute with the previous three, that is those in Eq. (3.131).

A useful fact about group manifolds is that one can always find a Lie algebra valued 1-form that is invariant under (say) right translation. This is the *Maurer–Cartan form* $\mathrm{d}g\,g^{-1}$. That it takes values in the Lie algebra is clear if we write g as an exponential of a Lie algebra element as in Eq. (3.127). That it is invariant under right translation by a fixed group element g_1 is shown by the calculation

$$\mathrm{d}g\;g^{-1} \;\to\; \mathrm{d}(gg_1)\,(gg_1)^{-1} \;=\; \mathrm{d}g\;g_1\,g_1^{-1}\,g^{-1} \;=\; \mathrm{d}g\;g^{-1}\,. \tag{3.133}$$

Given a basis for the Lie algebra the Maurer–Cartan form can be expanded. For $SU(2)$ we use the Pauli matrices multiplied with i for this purpose, and get

$$\mathrm{d}g\;g^{-1} \;=\; \mathrm{i}\sigma_1\Theta_1 + \mathrm{i}\sigma_2\Theta_2 + \mathrm{i}\sigma_3\Theta_3\,. \tag{3.134}$$

Doing the explicit calculation, that is multiplying two 2×2 matrices, we can read off that

$$\begin{aligned}
\Theta_1 &= -U\mathrm{d}X - Z\mathrm{d}Y + Y\mathrm{d}Z + X\mathrm{d}U = \tfrac{1}{2}\,(\sin\tau\;\mathrm{d}\theta - \cos\tau\;\sin\theta\;\mathrm{d}\phi)\\
\Theta_2 &= -Z\mathrm{d}X + U\mathrm{d}Y + X\mathrm{d}Z - Y\mathrm{d}U = \tfrac{1}{2}\,(\cos\tau\;\mathrm{d}\theta + \sin\tau\;\sin\theta\;\mathrm{d}\phi)\\
\Theta_3 &= -Y\mathrm{d}X + X\mathrm{d}Y - U\mathrm{d}Z + Z\mathrm{d}U = \tfrac{1}{2}\,(\mathrm{d}\tau + \cos\theta\;\mathrm{d}\phi)\,.
\end{aligned} \tag{3.135}$$

In the last step we use Euler angles as coordinates; Θ_3 is evidently our familiar friend, the connection from Eq. (3.116). There exists a left invariant Maurer–Cartan form $g^{-1}\,\mathrm{d}g$ as well.

It is now natural to declare that the group manifold of $SU(2)$ is equipped with its standard round metric; the round metric is singled out by the requirement that left and right translations correspond to isometries. Since left and right translations can be performed independently, this means that the isometry group of any Lie group G equipped with its natural metric is $G\times G$, with a discrete factor divided out. For the 3-sphere the isometry group is $SO(4)$, and it obeys the isomorphism

$$SO(4) \;=\; SU(2)\times SU(2)/\mathbb{Z}_2\,. \tag{3.136}$$

The $\mathbb{Z}_2$ in the denominator arises because left and right translations with $-\mathbb{1}$ cancel each other. It is this peculiar isomorphism that explains why the 3-sphere manages to serve as the manifold of a group; no other $SO(N)$ group splits into a product in this way. In general (for the classical groups) the invariant metric is uniquely, up to

a constant, given by

$$\mathrm{d}s^2 = -\frac{1}{2}\,\mathrm{Tr}(\mathrm{d}g\; g^{-1}\mathrm{d}g\; g^{-1}) = -\frac{1}{2}\mathrm{Tr}(g^{-1}\mathrm{d}g g^{-1}\mathrm{d}g) = \Theta_1^2 + \Theta_2^2 + \Theta_3^2\,. \quad (3.137)$$

In the second step we rewrote the metric in terms of the left invariant Maurer–Cartan form, so it is indeed invariant under $G \times G$. The final step is for $SU(2)$ only. The square root of the determinant of this metric is the *Haar measure* on the group manifold, and will figure prominently in Chapter 14.

Entirely by the way, but not uninterestingly, we observe that we can insert a real parameter α into the 3-sphere metric:

$$\mathrm{d}s^2 \;=\; \Theta_1^2 + \Theta_2^2 + \alpha\Theta_3^2\,. \quad (3.138)$$

For $\alpha = 1$ this is the metric on the group and for $\alpha = 0$ it is the metric on the 2-sphere since it does not depend on the fibre coordinate τ. Because it is made from right invariant forms it is an $SO(3)$ invariant metric on $\mathbf{S}^3$ for any intermediate value of α. It is said to be the metric on a *squashed 3-sphere*; 3-spheres that are squashed in this particular way are called *Berger spheres*.

The three everywhere non-vanishing Killing vector fields in Eq. (3.131) do not commute and therefore it is impossible to introduce a coordinate system such that all three coordinate lines coincide with the Killing flow lines. But they are linearly independent, they form a perfectly good basis for the tangent space at every point, and any vector can be written as a linear combination of the J_I, where I runs from one to three. We can introduce a basis Θ_I in the dual cotangent space through

$$\Theta_I(J_J) \;=\; \delta_{IJ}\,. \quad (3.139)$$

The Θ_I are precisely the three forms given in Eq. (3.135). There is no globally defined function whose differential is our connection Θ_3 because we are no longer using a coordinate basis.

We do have a number of coordinate systems especially adapted to the group structure at our disposal. The coordinate lines will arise from exponentiating elements in the Lie algebra. First we try the *canonical parametrization*

$$g = \mathrm{e}^{\mathrm{i}(x\sigma_1 + y\sigma_2 + z\sigma_3)}\,. \quad (3.140)$$

If we switch to spherical polars after performing the exponentiation we obtain

$$g = \begin{bmatrix} X + \mathrm{i}Y & Z + \mathrm{i}U \\ -Z + \mathrm{i}U & X - \mathrm{i}Y \end{bmatrix} = \begin{bmatrix} \cos r + \mathrm{i}\sin r\cos\theta & \mathrm{i}\sin r\sin\theta\;\mathrm{e}^{-\mathrm{i}\phi} \\ \mathrm{i}\sin r\sin\theta\;\mathrm{e}^{\mathrm{i}\phi} & \cos r - \mathrm{i}\sin r\cos\theta \end{bmatrix}. \quad (3.141)$$

The metric when expressed in these coordinates becomes

$$\mathrm{d}s^2 = \mathrm{d}r^2 + \sin^2 r(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\,. \quad (3.142)$$

So these are geodesic polar coordinates – the parameter r is the arc length along geodesics from the origin. The Euler angles from Section 3.5 is another choice; following in Euler's footsteps we write an arbitrary group element as

$$g = \mathrm{e}^{\frac{\mathrm{i}\tau}{2}\sigma_3}\mathrm{e}^{\frac{\mathrm{i}\theta}{2}\sigma_2}\mathrm{e}^{\frac{\mathrm{i}\phi}{2}\sigma_3} = \begin{bmatrix} \mathrm{e}^{\frac{\mathrm{i}}{2}(\tau+\phi)}\cos\frac{\theta}{2} & \mathrm{e}^{\frac{\mathrm{i}}{2}(\tau-\phi)}\sin\frac{\theta}{2} \\ -\mathrm{e}^{-\frac{\mathrm{i}}{2}(\tau-\phi)}\sin\frac{\theta}{2} & \mathrm{e}^{-\frac{\mathrm{i}}{2}(\tau+\phi)}\cos\frac{\theta}{2} \end{bmatrix} . \tag{3.143}$$

Equation (3.96) results. It seems fair to warn the reader that although similar coordinate systems can be erected on all classical group manifolds, the actual calculations involved in expressing, say, the metric tensor in terms of coordinates tend to become very long, except for $SU(2)$.

3.8 Cosets and all that

We have already used the notion of *quotient spaces* such as M/H quite freely. The idea is that we start with a space M and a group H of transformations of M; the quotient space is then a space whose points consists of the *orbits* of G, that is to say that by definition a point in M/H is an equivalence class of points in M that can be transformed into each other by means of transformations belonging to H. We have offered no guarantees that the quotient space is a 'nice' space however. In general no such guarantee can be given. We are in a better position when the space M is a Lie group G and H is some subgroup of G. By definition a *left coset* is a set of elements of a group G that can be written in the form gh, where h is an any element of a fixed subgroup H. This gives a partition of the group into disjoint cosets, and the space of cosets is called the *coset space* and denoted by G/H. The points of the coset space are written schematically as gH; they are in fact the orbits of H in G under right action. For coset spaces there is a comfortable theorem saying that G/H is always a manifold of dimension

$$\dim(G/H) \;=\; \dim(G) - \dim(H)\,. \tag{3.144}$$

Right cosets and right coset spaces can be defined in an analogous way.

A description of a space as a coset space often arises as follows: we start with a space M that is a *homogeneous space*, which means that there is a group G of isometries such that every point in M can be reached from any given point by means of a transformation belonging to G. Hence there is only one G orbit, and, since G is a symmetry group, the space looks the same at each point. The technical term here is that the group G acts *transitively* on M. Now fix a point p and consider the *isotropy* or *little group* at this point; by definition this is the subgroup $H \subset G$ of transformations h that leave p invariant. In this situation it follows that $M = G/H$. At first sight it may seem that this recipe depends on the choice of the point where

the isotropy group was identified but in fact it does not; if we start from another point $p' = gp$ where $g \in G$ then the new little group H' is *conjugated* to H in the sense that $H' = gHg^{-1}$. The coset spaces G/H and G/H' are then identical, it is only the description that has changed a little.

A warning: the notation G/H is ambiguous unless it is specified which particular subgroup H (up to conjugation) that is meant. To see how an ambiguity can arise consider the group $SU(3)$. It has an Abelian subgroup consisting of diagonal matrices

$$h = \begin{bmatrix} e^{i\alpha} & 0 & 0 \\ 0 & e^{i\beta} & 0 \\ 0 & 0 & e^{-i(\alpha+\beta)} \end{bmatrix} . \tag{3.145}$$

In the group manifold this is a two-dimensional surface known as the *Cartan torus*. Now consider the coset space $SU(3)/U(1)$, where the subgroup $U(1)$ forms a circle on the Cartan torus. But there are infinitely many different ways in which a circle can wind around on a 2-torus, and the coset space will get dramatically different properties depending on the choice one makes. On the other hand the space $SU(3)/U(1) \times U(1)$ suffers from no such ambiguities – any two-dimensional Abelian subgroup of $SU(3)$ is related by conjugation to the subgroup of diagonal matrices.

It is interesting to ask again why the 2-sphere of Hopf fibres is naturally a round 2-sphere, this time adopting our new coset space view of things. We can write a point in the coset space as Ωh, where Ω is the *coset representative* and h is an element in H. The group $SU(2)/\mathbb{Z}_2 = SO(3)$ acts naturally on this coset space through left action;

$$g \;\rightarrow\; g_1 g \quad \Rightarrow \quad \Omega h \;\rightarrow\; g_1 \Omega h \, . \tag{3.146}$$

It is therefore natural to select a coset space metric that is invariant under $SO(3)$, and for $\mathbf{S}^2$ the round metric is the answer. In general this construction gives rise also to a uniquely defined measure on the coset space, induced by the Haar measure on the group manifold.

If we take stock of the situation we see that we now have a rich supply of principal fibre bundles to play with, since any coset space G/H is the base manifold of a principal bundle with bundle space G and fibre H. This construction will occur again and again. To see how common it is we show that every sphere is a coset space. We start from an orthogonal matrix belonging to $SO(N)$. Its first column is a normalized vector in $\mathbb{R}^N$ and we can choose coordinates so that it takes the form $(1, 0, \ldots, 0)$. The group $SO(N)$ clearly acts transitively on the space of all normalized vectors (that is, on the sphere $\mathbf{S}^{N-1}$), and the isotropy group of the

chosen point consists of all matrices of the form

$$h = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & SO(N-1) & \\ 0 & & & \end{bmatrix} . \tag{3.147}$$

It follows that

$$\mathbf{S}^{N-1} \;=\; SO(N)/SO(N-1) \;=\; O(N)/O(N-1)\,, \tag{3.148}$$

with the subgroup chosen according to Eq. (3.147). The round metric on $\mathbf{S}^{N-1}$ arises naturally because it is invariant under the group $SO(N)$ that acts on the coset space. A similar argument in $\mathbb{C}^N$, where a column of a unitary matrix is in effect a normalized vector in $\mathbb{R}^{2N}$, shows that

$$\mathbf{S}^{2N-1} \;=\; SU(N)/SU(N-1) \;=\; U(N)/U(N-1)\,. \tag{3.149}$$

And with this observation we finally know enough about spheres.

Problems

Problem 3.1 Derive the intrinsic metric of the sphere using stereographic projection. For the sake of variety, project from the south pole to the tangent plane at the north pole rather than to the equatorial plane.

Problem 3.2 A coordinate patch covering part of the sphere – a map from the sphere to the plane – can be obtained by projecting from an arbitrary point in space. Show geometrically (by drawing a figure!) that the only conformal projection – a projection that preserves angles – is the stereographic projection from a point on the sphere.

Problem 3.3 On the complex plane, identify points connected by $x \to x + 1$ and also $y \to y + 1$. Do the same for $x \to x + 1$ and $y \to y + 2$. Show that the two quotient spaces are inequivalent as complex manifolds.

Problem 3.4 Show that the Poisson brackets obey the Jacobi identity because the symplectic form is closed.

Problem 3.5 Verify that the geodesics on $\mathbf{S}^3$ meet the equator in two antipodal points.

Problem 3.6 The coordinates τ and ϕ run along Hopf circles, and they appear symmetrically in the metric (3.98). But $0 \le \tau < 4\pi$ and $0 \le \phi < 2\pi$. Why this difference?

Problem 3.7 Take sections of the Hopf bundle by choosing $\tau = -\phi$, $\tau = \phi$ and $\tau = \phi + \pi$ respectively, and work out what they look like in the stereographic picture.

Problem 3.8 Everybody knows how to make a Möbius strip by means of a piece of paper and some glue. Convince yourself that the Möbius strip is a fibre bundle and show that it admits a global section. Identify the group acting on its fibres and construct its principal bundle. Show that the principal bundle does not admit a global section, and hence that the bundle is non-trivial.

4

Complex projective spaces

> In the house of mathematics there are many mansions and of these the most elegant is projective geometry.
>
> *Morris Kline*

An attentive reader of Chapter 3 will have noticed that there must exist a Hopf fibration of any odd-dimensional sphere; it is just that we studiously avoided to mention what the resulting base space is. In this chapter it will be revealed that

$$\mathbb{C}\mathbf{P}^n = \mathbf{S}^{2n+1}/\mathbf{S}^1 \,. \tag{4.1}$$

The space on the left-hand side is called *complex projective space* or – when it is used as a space of pure states in quantum mechanics – *projective Hilbert space*. It is not a sphere unless $n = 1$. The study of projective spaces was actually begun by artists during the Renaissance, and this is how we begin our story.

4.1 From art to mathematics

Painters, engaged in projecting the ground (conveniently approximated as a flat plane) onto a flat canvas, discovered that parallel lines on the ground intersect somewhere on a *vanishing line* or *line at infinity* that exists on the canvas but not on the ground, where parallel lines do not meet. Conversely, a line exists on the ground (right below the artist) that does not appear on the canvas at all. When the subject was taken over by mathematicians this led to the theory of the *projective plane* or *real projective 2-space* $\mathbb{R}\mathbf{P}^2$. The idea is to consider a space whose points consists of the one-dimensional subspaces of a three-dimensional vector space $\mathbb{R}^3$. Call them *rays*. The origin is placed at the eye of the artist, and the rays are his lines of sight. The ground and the canvas serve as two *affine coordinate planes* for $\mathbb{R}\mathbf{P}^2$. A *projective line* is a certain one parameter family of points in the projective plane, and is defined as the space of rays belonging to some two-dimensional subspace $\mathbb{R}^2$. In itself, this is the real projective 1-space $\mathbb{R}\mathbf{P}^1$. To see how all this works out

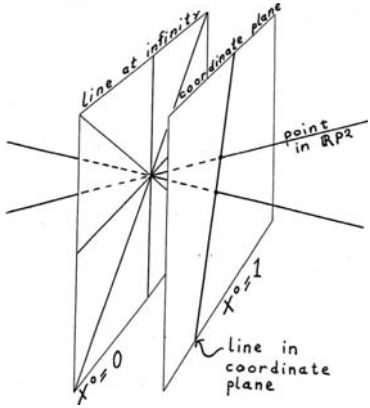


Figure 4.1. Affine coordinates for (almost all of) the projective plane.

in formulae, we observe that any triple of real numbers defines a one-dimensional subspace of $\mathbb{R}^3$. But the latter does not determine the triple of numbers uniquely. Hence a point of $\mathbb{R}\mathbf{P}^2$ is given by an equivalence class

$$(X^0, X^1, X^2) \sim k(X^0, X^1, X^2)\,; \qquad k \in \mathbb{R} \quad k \neq 0\,. \tag{4.2}$$

The numbers X^α are known as *homogeneous coordinates* on $\mathbb{R}\mathbf{P}^2$. If we want to use true coordinates, one possibility is to choose a plane in $\mathbb{R}^3$ such as $X^0 = 1$, representing an infinite canvas, say. We think of this plane as an affine plane (a vector space except that the choice of origin and scalar product is left open). We can then label a ray with the coordinates of the point where it intersects the affine plane. This is an affine coordinate system on the projective plane; it does not cover the entire projective plane since the rays lying in the plane $X^0 = 0$ are missing. But this plane is a vector space $\mathbb{R}^2$, and the space of rays lying in it is the projective line $\mathbb{R}\mathbf{P}^1$. Hence the affine coordinates display the projective plane as an infinite plane to which we must add an extra projective line 'at infinity'.

Why did we define our rays as one-dimensional subspaces, and not as directed lines from the origin? The answer does not matter to the practising artist, but for the mathematician our definition results in simple *incidence properties* for points and lines. The projective plane will now obey the simple axioms that

(1) Any two different points lie on a unique line.
(2) Any two different lines intersect in a unique point.

There are no exceptional cases (such as parallel lines) to worry about. That axiom (2) holds is simply the observation that two planes through the origin in $\mathbb{R}^3$ always intersect, in a unique line.

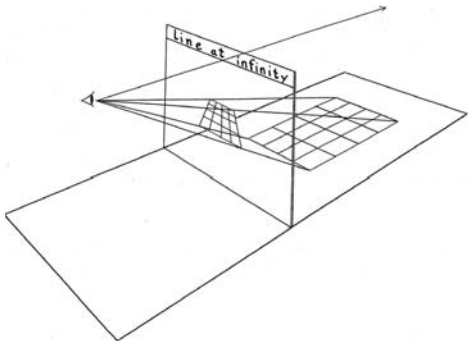


Figure 4.2. Renaissance painter, engaged in changing affine coordinates.

There is nothing special about the line at infinity; the name is just an artefact of a special choice of affine coordinates. Every projective line is the space of rays in some two-dimensional subspace of $\mathbb{R}^3$, and it has the topology of a circle – to its image in the affine plane we must add one point 'at infinity'.

It is interesting to observe that the space of all lines in $\mathbb{R}\mathbf{P}^2$ is another $\mathbb{R}\mathbf{P}^2$ since one can set up a one-to-one correspondence between rays and planes through the origin in $\mathbb{R}^3$. In fact there is a complete *duality* between the space of points and the space of lines in the projective plane. The natural way to select a specific one-to-one correspondence between $\mathbb{R}\mathbf{P}^2$ and the dual $\mathbb{R}\mathbf{P}^2$ is to introduce a metric in the underlying vector space, and to declare that a ray in the vector space is dual to that plane through the origin to which it is orthogonal. But once we think of the vector space as a metric space as well as a linear one we will be led to the round metric as the natural metric to use on $\mathbb{R}\mathbf{P}^2$. Indeed a moment's reflection shows that $\mathbb{R}\mathbf{P}^n$ can be thought of as the sphere $\mathbf{S}^n$ with antipodal points identified,

$$\mathbb{R}\mathbf{P}^n = \mathbf{S}^n/\mathbb{Z}_2 \,. \tag{4.3}$$

This is one of the reasons why the real projective plane is so famous: it is a space of constant curvature where every pair of geodesics intersect once (as opposed to twice, as happens on a sphere). In fact all of Euclid's axioms for points and straight lines are valid, except the fifth, which shows that the parallel axiom is independent of the others.

In general a specific one-to-one map between a projective space and its dual is known as a *polarity*, and choosing a polarity singles out a particular metric on the projective space as being the most natural one.

There are other geometrical figures apart from points and lines. A famous example is that of the *conic section*. Consider a (circular, say) cone of rays with its vertex at the origin of $\mathbb{R}^3$. In one affine coordinate system the points of intersection

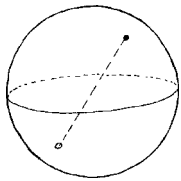


Figure 4.3. Real projective space is the sphere with antipodal points identified.

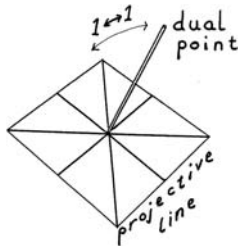


Figure 4.4. Polarity: a metric for the projective plane.

of the cone with the affine plane will appear as a circle, in another as a hyperbola. Hence circles, ellipses and hyperbolae are projectively equivalent. In homogeneous coordinates such a conic section appears as

$$(X^1)^2 + (X^2)^2 = (X^0)^2 \,. \tag{4.4}$$

This equation defines a submanifold of $\mathbb{R}\mathbf{P}^2$ because it is homogeneous in the X^α, and therefore unaffected by scaling. If we use affine coordinates for which $X^0 = 1$ it appears as a circle, but if we use affine coordinates for which $X^1 = 1$ it appears as a hyperbola. We can use this fact to give us some further insight into the somewhat difficult topology of $\mathbb{R}\mathbf{P}^2$: if we draw our space as a disc with antipodal points on the boundary identified – this would be the picture of the topology implied by Figure 4.3 – then we see that the hyperbola becomes a topological circle because we are adding two points at infinity. Its interior is an ordinary disc while its exterior is a *Möbius strip* because of the identifications at the boundary – if you glue a disc to the boundary of a Möbius strip you obtain a projective plane. By the way $\mathbb{R}\mathbf{P}^2$, unlike $\mathbb{R}\mathbf{P}^1$, is a *non-orientable* space; this means that a 'right-handed' basis in the tangent space can be turned into a 'left-handed' one by moving it around so that in fact the distinction between right- and left-handedness cannot be upheld. See Figure 4.5.

If you think a little bit further about the $\mathbb{R}\mathbf{P}^2$ topology you will also notice a – possibly reassuring – fact, namely that there is a topological difference between projective lines and conic sections. Although they are both circles intrinsically, the

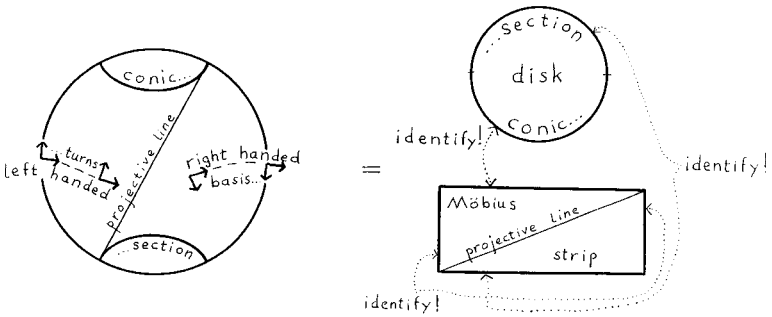


Figure 4.5. The topology of the real projective plane. For a different view, see Figure 4.12.

conic sections have the property that they can be continually deformed to a point within the projective plane, while the lines cannot. Indeed if you cut the projective plane open along a projective line it remains connected, while it splits into two pieces if you cut along a conic section.

We could go on to discuss n-dimensional real projective spaces $\mathbb{R}\mathbf{P}^n$, but we confine ourselves to two topological remarks: that $\mathbb{R}\mathbf{P}^n = \mathbf{S}^n/\mathbb{Z}_2$ (obvious), and that $\mathbb{R}\mathbf{P}^{2n+1}$ is orientable while $\mathbb{R}\mathbf{P}^{2n}$ is not (easy to prove).

We can change our *field* from the real numbers $\mathbb{R}$ to rational, complex or quaternionic numbers. In fact one can also use finite fields, such as the integers modulo p where p is some prime number. If we repeat the construction of the projective plane starting from a three-dimensional vector space over the field of integers modulo p then we will obtain a projective plane with exactly $p + 1$ points on each line. This is an interesting construction, but like real and quaternionic projective spaces it is of only secondary importance in quantum mechanics, where complex projective space occupies the centre of the stage.

4.2 Complex projective geometry

Complex projective space is in some respects easier to study than is its real cousin, mainly because the complex numbers form a *closed field* (that is to say, every polynomial equation has a root). The most effective presentation of $\mathbb{C}\mathbf{P}^n$ starts directly in the vector space $\mathbb{C}^N = \mathbb{C}^{n+1}$, known as *Hilbert space* in quantum mechanics; our convention is always that $N = n + 1$. By definition $\mathbb{C}\mathbf{P}^n$ is the space of rays in $\mathbb{C}^{n+1}$, or equivalently the space of equivalence classes of $n + 1$ complex numbers, not all zero, under

$$(Z^0, Z^1, \ldots, Z^n) \sim \lambda(Z^0, Z^1, \ldots, Z^n)\,; \quad \lambda \in \mathbb{C} \quad \lambda \neq 0\,. \tag{4.5}$$

We will use Greek indices to label the homogeneous coordinates. We can cover $\mathbb{C}\mathbf{P}^n$ with affine coordinate patches like

$$Z^\alpha = (1, z^1, \ldots, z^n)\,, \qquad Z^\alpha = (z^{n'}, 1, z^{1'}, \ldots, z^{(n-1)'}) \tag{4.6}$$

and so on. This leads to two important observations. The first is that one coordinate patch covers all of $\mathbb{C}\mathbf{P}^n$ except that subset which has (say)

$$Z^\alpha = (0, Z^1, \ldots, Z^n)\,. \tag{4.7}$$

But the set of all such rays is a $\mathbb{C}\mathbf{P}^{n-1}$. Hence we conclude that topologically $\mathbb{C}\mathbf{P}^n$ is like $\mathbb{C}^n$ with a $\mathbb{C}\mathbf{P}^{n-1}$ attached 'at infinity'. Iterating this observation we obtain the *cell decomposition*

$$\mathbb{C}\mathbf{P}^n = \mathbb{C}^n \cup \mathbb{C}^{n-1} \cup \cdots \cup \mathbb{C}^0\,. \tag{4.8}$$

Evidently $\mathbb{C}\mathbf{P}^0$ is topologically a point and $\mathbb{C}\mathbf{P}^1$ is a 2-sphere, while for $n > 1$ we get something new that we will have to get used to as we proceed. The second observation is that in a region where the coordinate systems overlap, they are related by (say)

$$z^{a'} = \frac{Z^{a+1}}{Z^1} = \frac{Z^0}{Z^1}\,\frac{Z^{a+1}}{Z^0} = \frac{z^{a+1}}{z^1}\,, \tag{4.9}$$

which is clearly an analytic function on the overlap region. Hence $\mathbb{C}\mathbf{P}^n$ is a complex manifold.

The *linear subspaces* of $\mathbb{C}\mathbf{P}^n$ are of major importance. They are defined as the images of the subspaces of $\mathbb{C}^{n+1}$ under the natural map from the vector space to the projective space. Equivalently they are given by a suitable number of linear equations in the homogeneous coordinates. Thus the *hyperplanes* are $(n-1)$-dimensional submanifolds of $\mathbb{C}\mathbf{P}^n$ defined by the equation

$$P_\alpha Z^\alpha = 0 \tag{4.10}$$

for some fixed set of $n+1$ complex numbers P_α. This definition is unaffected by a change of scale for the homogeneous coordinates, and also by a change of scale in P_α. Hence the $n+1$ complex numbers P_α themselves are homogeneous coordinates for a $\mathbb{C}\mathbf{P}^n$; in other words the hyperplanes in a projective n-space can be regarded as the points of another projective n-space which is dual to the original.

If we impose a set of $m \le n$ independent linear equations on the Z^α we obtain a linear subspace of complex dimension $n-m$ which is itself a $\mathbb{C}\mathbf{P}^{n-m}$. Geometrically this is the intersection of m hyperplanes. The space whose points consist of all such $(n-m)$-dimensional subspaces of $\mathbb{C}\mathbf{P}^n$ is known as a *Grassmannian* – it is only in the case of hyperplanes (and the trivial case of points) that the Grassmannian is itself a $\mathbb{C}\mathbf{P}^n$. A linear subspace of complex dimension one is known as a

complex projective line, a linear subspace of dimension two is a complex projective plane and so on, while $\mathbb{C}\mathbf{P}^0$ is just a point. The complex projective line is a $\mathbb{C}\mathbf{P}^1$ – topologically this is a sphere and this may boggle some minds, but it is a line in the sense that one can draw a unique line between any pair of points (this is essentially the statement that two vectors in $\mathbb{C}^{n+1}$ determine a unique two-dimensional subspace). It also behaves like a line in the sense that two projective lines in a projective space intersect in a unique point if they intersect at all (this is the statement that a pair of two-dimensional subspaces are either disjoint or they share one common ray or they coincide). In general the intersection of two linear subspaces A and B is known as their *meet* $A \cap B$. We can also define their *join* $A \cup B$ by taking the linear span of the two subspaces of the underlying vector space to which A and B correspond, and then go back down to the projective space. These two operations – meet and join – turn the set of linear subspaces into a partially ordered structure known as a *lattice*, in which every pair of elements has a greatest lower bound (the meet) and a least upper bound (the join). This is the starting point of the subject known as quantum logic (Jauch, 1968; Varadarajan, 1985). It is also the second time we encounter a lattice – in Section 1.1 we came across the lattice of faces of a convex body. In Chapter 8 we will find a convex body whose lattice of faces agrees with the lattice of subspaces of a vector space. By then it will be a more interesting lattice, because we will have an inner product on $\mathbb{C}^N$, so that a given subspace can be associated with its orthogonal complement.

But we have not yet introduced any inner product in $\mathbb{C}^N$, or any metric in $\mathbb{C}\mathbf{P}^n$. In fact we will continue to do without it for some time; projective geometry is precisely that part of geometry that can be done without a metric. There is an interesting group theoretical view of this, originated by Felix Klein. All statements about linear subspaces – such as when two linear subspaces intersect – are invariant under general linear transformations of $\mathbb{C}^N$. They form the group $GL(N, \mathbb{C})$, but only a subgroup acts *effectively* on $\mathbb{C}\mathbf{P}^n$. (Recall that $N = n + 1$. A transformation is said to act effectively on some space if it moves at least one point.) Changing the matrix with an overall complex factor does not change the transformation effected on $\mathbb{C}\mathbf{P}^n$, so we can set its determinant equal to one and multiply it with an extra complex Nth root of unity if we wish. The *projective group* is therefore $SL(n + 1, \mathbb{C})/\mathbb{Z}_N$.

According to Klein's conception projective geometry is completely characterized by the projective group; its subgroups include the group of affine transformations that preserves the $\mathbb{C}\mathbf{P}^{n-1}$ at infinity and this subgroup characterizes affine geometry. A helpful fact about the projective group is that any set of $n + 2$ points can be brought to the standard position $(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1), (1, 1, \ldots, 1)$ by means of a projective transformation. For $\mathbb{C}\mathbf{P}^1$ this is the familiar statement that any triple of points on the complex plane can be transformed to 0, 1, and ∞ by a Möbius transformation.

4.3 Complex curves, quadrics and the Segre embedding

The equation that defines a subspace of $\mathbb{C}\mathbf{P}^n$ does not have to be linear; any homogeneous equation in the homogeneous coordinates Z^α gives rise to a well-defined submanifold. Hence in addition to the linear subspaces we have *quadrics, cubics, quartics* and so on, depending on the degree of the defining polynomial. The locus of a number of homogeneous equations

$$w_1(Z) = w_2(Z) = \cdots = w_m(Z) = 0 \tag{4.11}$$

is also a subspace known as an *algebraic* (or projective) *variety*, and we move from projective to *algebraic* geometry.[1] *Chow's theorem* states that every non-singular algebraic variety is a complex submanifold of $\mathbb{C}\mathbf{P}^n$, and conversely every compact complex submanifold is the locus of a set of homogeneous equations. On the other hand it is not true that every complex manifold can be embedded as a complex submanifold in $\mathbb{C}\mathbf{P}^n$.

There are two kinds of submanifolds that are of immediate interest in quantum mechanics. One of them is the *complex curve*; by definition this is a map of $\mathbb{C}\mathbf{P}^1$ into $\mathbb{C}\mathbf{P}^n$. In real terms it is a 2-surface in a $2n$-dimensional space. This does not sound much like a curve, but once it is accepted that $\mathbb{C}\mathbf{P}^1$ deserves the name of line it will be admitted that the name curve is reasonable here. Let us first choose $n = 2$, so that we are looking for a complex curve in the projective plane. Using (u, v) as homogeneous coordinates on $\mathbb{C}\mathbf{P}^1$ we clearly get a map into $\mathbb{C}\mathbf{P}^2$ if we set

$$(u, v) \quad \rightarrow \quad (u^2, uv, v^2)\,. \tag{4.12}$$

(In Section 6.4 we will adjust conventions a little.) This is a well-defined map because the expression is homogeneous in the u and v. Evidently the resulting complex curve in $\mathbb{C}\mathbf{P}^2$ obeys the equation

$$Z^0 Z^2 \;=\; Z^1 Z^1\,. \tag{4.13}$$

Hence it is a quadric hypersurface as well, and indeed any quadric can be brought to this form by projective transformations. In the projective plane a quadric is also known as a conic section.

In higher dimensions we encounter complex curves that are increasingly difficult to grasp since they are not quadrics. The next simplest case is the *twisted cubic* curve in $\mathbb{C}\mathbf{P}^3$, defined by

$$(u, v) \quad \rightarrow \quad (u^3, u^2 v, uv^2, v^3)\,. \tag{4.14}$$

We leave it aside for the moment though.

[1] A standard reference, stressing the setting of complex manifolds, is Griffiths and Harris (1978). For an alternative view see the book by Harris alone (Harris, 1992).

A class of submanifolds that is of particular interest in quantum mechanics arises in the following way. Suppose that the complex vector space is given as a *tensor product*

$$\mathbb{C}^{n+1} \otimes \mathbb{C}^{m+1} \;=\; \mathbb{C}^{(n+1)(m+1)} \,. \tag{4.15}$$

Then there should be an embedded submanifold

$$\mathbb{C}\mathbf{P}^n \times \mathbb{C}\mathbf{P}^m \;\in\; \mathbb{C}\mathbf{P}^{(n+1)(m+1)-1} \,. \tag{4.16}$$

Indeed this is true. In terms of homogeneous coordinates the submanifold can be parametrized as

$$Z^\alpha = Z^{\mu\mu'} = P^\mu Q^{\mu'}, \tag{4.17}$$

in a fairly obvious notation – the P^μ and $Q^{\mu'}$ are homogeneous coordinates on $\mathbb{C}\mathbf{P}^n$ and $\mathbb{C}\mathbf{P}^m$, respectively. The construction is known as the *Segre embedding*. The submanifold is a Cartesian product with (complex) dimension $n + m$, and it follows from Chow's theorem that it can be defined as the locus of nm homogeneous equations in the large space. Indeed it is easy to see from the definition that the submanifold will be the intersection of the quadrics

$$Z^{\mu\mu'} Z^{\nu\nu'} - Z^{\mu\nu'} Z^{\nu\mu'} = 0 \,. \tag{4.18}$$

In quantum mechanics – and in Section 15.2 – the Segre submanifold reappears as the set of separable states of a composite system.

Let us consider the simplest case

$$\mathbb{C}\mathbf{P}^1 \times \mathbb{C}\mathbf{P}^1 \;\in\; \mathbb{C}\mathbf{P}^3 \,, \tag{4.19}$$

when only one equation is needed. Write

$$(Z^0, Z^1, Z^2, Z^3) = (Z^{00'}, Z^{01'}, Z^{10'}, Z^{11'}) \,. \tag{4.20}$$

Then the submanifold $\mathbb{C}\mathbf{P}^1 \times \mathbb{C}\mathbf{P}^1$ is obtained as the quadric surface

$$Z^0 Z^3 - Z^1 Z^2 = 0 \,. \tag{4.21}$$

In general the non-degenerate quadrics in $\mathbb{C}\mathbf{P}^3$ are in one-to-one correspondence to all possible embeddings of $\mathbb{C}\mathbf{P}^1 \times \mathbb{C}\mathbf{P}^1$. It is interesting to consider the projection map from the product manifold to one of its factors. This means that we hold $Q^{\mu'}$ (say) fixed and vary P^μ. Then the fibre of the map – the set of points on the quadric that are projected to the point on $\mathbb{C}\mathbf{P}^m$ that is characterized by that particular $Q^{\mu'}$ – will be given by the equations

$$\frac{Z^{00'}}{Z^{01'}} = \frac{P^0 Q^{0'}}{P^0 Q^{1'}} = \frac{Q^{0'}}{Q^{1'}} = \frac{P^1 Q^{0'}}{P^1 Q^{1'}} = \frac{Z^{10'}}{Z^{11'}} \,. \tag{4.22}$$

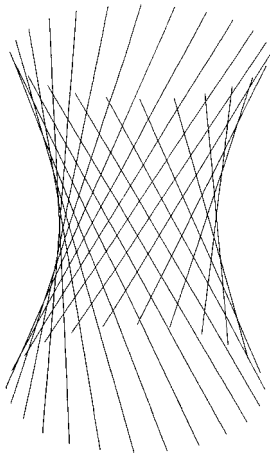


Figure 4.6. The real Segre embedding: a hyperboloid ruled by straight lines.

Since $Q^{\mu'}$ is fixed this implies that there is a complex number λ such that

$$Z^0 = \lambda Z^1 \quad \text{and} \quad Z^2 = \lambda Z^3 \,. \tag{4.23}$$

This pair of linear equations defines a projective line in $\mathbb{C}\mathbf{P}^3$. Projecting down to the other factor leads to a similar conclusion. In this way we see that the quadric is ruled by lines, or in other words that through any point on the quadric there goes a pair of straight lines lying entirely in the quadric.

For visualization, let us consider the real Segre embedding

$$\mathbb{R}\mathbf{P}^1 \times \mathbb{R}\mathbf{P}^1 \;\in\; \mathbb{R}\mathbf{P}^3 \,. \tag{4.24}$$

This time we choose to diagonalize the quadric; define

$$Z^0 = X + U \,, \quad Z^1 = X - U \,, \quad Z^2 = V + Y \,, \quad Z^3 = V - Y \,. \tag{4.25}$$

We then obtain

$$Z^0 Z^3 - Z^1 Z^2 = X^2 + Y^2 - U^2 - V^2 = 0 \,. \tag{4.26}$$

Now let us choose affine coordinates by dividing through with V. The quadric becomes

$$x^2 + y^2 - u^2 \;=\; 1 \,. \tag{4.27}$$

This is a hyperboloid of one sheet sitting in a three-dimensional real space. The fact that such a surface can be *ruled* by straight lines is a surprising fact of elementary geometry (first noted by Sir Christopher Wren).

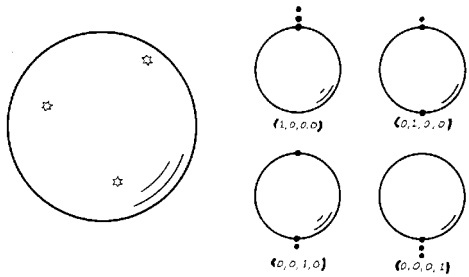


Figure 4.7. Any state from $\mathbb{C}\mathbf{P}^3$ may be represented by three stars on the sphere. Some special cases, where the stars coincide, are shown on the right.

4.4 Stars, spinors and complex curves

The *stellar representation* is a delightful way of visualizing $\mathbb{C}\mathbf{P}^n$ in real terms. It works for any n and simplifies some problems in a remarkable way.[2] Here we will develop it a little from a projective point of view, while Chapter 7 discusses the same construction with the added benefit of a metric. The idea is that vectors in $\mathbb{C}^{n+1}$ are in one-to-one correspondence with the set of nth degree polynomials in one complex variable z, such as

$$w(z) = Z^0 z^n + Z^1 z^{n-1} + \cdots + Z^n \,. \tag{4.28}$$

(The important point is that we have a polynomial of the nth degree; in Chapter 7 we will have occasion to polish the conventions a little.) We can rescale the vector Z^α so that $Z^0 = 1$; therefore points in $\mathbb{C}\mathbf{P}^n$ will be in one-to-one correspondence with unordered sets of n complex numbers, namely with the complex roots of the equation

$$Z^0 z^n + Z^1 z^{n-1} + \cdots + Z^{n+1} = 0 = Z^0 (z - z_1)(z - z_2) \ldots (z - z_n) \,. \tag{4.29}$$

Multiple roots are allowed. If $Z^0 = 0$ then infinity counts as a root (of multiplicity m if $Z^1 = \cdots = Z^{m-1} = 0$). Finally, by means of a stereographic projection the roots can be represented as unordered sets of n points on an ordinary 2-sphere – the points are called 'stars' and thus we have arrived at the stellar representation, in which points in $\mathbb{C}\mathbf{P}^n$ are represented by n unordered stars on a 'celestial' sphere. As a mathematical aside it follows that $\mathbb{C}\mathbf{P}^1 = \mathbf{S}^2$, $\mathbb{C}\mathbf{P}^2 = \mathbf{S}^2 \times \mathbf{S}^2 / S_2$ and in general that $\mathbb{C}\mathbf{P}^n = \mathbf{S}^2 \times \mathbf{S}^2 \times \cdots \times \mathbf{S}^2 / S_n$, where S_n is the *symmetric group* of permutations of n objects.

[2] The first reference appears to be to Majorana (1932). The result has been rediscovered many times within (Bacry, 1974) and without (Penrose, 1960) the context of quantum mechanics.

There is a piece of notation that is conveniently introduced at this juncture. Let us denote the homogeneous coordinates on $\mathbb{C}\mathbf{P}^1$ by

$$\zeta^A = (u, v) \,. \tag{4.30}$$

So we use a capital Latin letter for the index and we will refer to ζ^A as a *spinor*. The overall scale of the spinor is irrelevant to us, so we can introduce an affine coordinate z by

$$z = \frac{v}{u} \;: \qquad \zeta^A \sim (1, z) \,. \tag{4.31}$$

A spinor for which $u = 0$ then corresponds to the south pole on the Riemann sphere. We will use the totally anti-symmetric tensor ϵ_{AB} (the symplectic structure on $\mathbf{S}^2$, in fact) to raise and lower indices according to

$$\zeta_A \equiv \zeta^B \epsilon_{BA} \qquad \Leftrightarrow \qquad \zeta^A = \epsilon^{AB} \zeta_B \,. \tag{4.32}$$

Due to the fact that ϵ_{AB} is anti-symmetric there is a definite risk that sign errors will occur when one uses its inverse ϵ^{AB}. Simply stick to the convention just made and all will be well. Note that $\zeta_A \zeta^A = 0$. So far then a spinor simply denotes a vector in $\mathbb{C}^2$. We do not think of this as a Hilbert space yet because there is no inner product. On the contrary the formalism is manifestly invariant under the full projective group $SL(2, \mathbb{C})/\mathbb{Z}^2$. A special linear transformation on $\mathbb{C}^2$ gives rise to a *Möbius transformation* on the sphere;

$$\zeta^A \to \begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \qquad \Rightarrow \qquad z \to z' = \frac{\alpha z + \beta}{\gamma z + \delta} \,, \tag{4.33}$$

where $\alpha\delta - \beta\gamma = 1$ is the condition (on four complex numbers) that guarantees that the matrix belongs to $SL(2, \mathbb{C})$ and that the transformation preserves ϵ_{AB}. The group of Möbius transformations is exactly the group of projective transformations of $\mathbb{C}\mathbf{P}^1$.

We can go on to consider totally symmetric *multispinors*

$$\Psi^{AB} = \Psi^{(AB)} \,, \qquad \Psi^{ABC} = \Psi^{(ABC)} \tag{4.34}$$

and so on. (The brackets around the indices mean that we are taking the totally symmetric part.) It is easy to see that when the index ranges from zero to one then the number of independent components of a rank n totally symmetric multispinor is $n + 1$, just right so that the multispinor can serve as homogeneous coordinates for $\mathbb{C}\mathbf{P}^n$. To see that this works, consider the equation

$$\Psi_{AB\ldots M} \zeta^A \zeta^B \ldots \zeta^M = 0 \,. \tag{4.35}$$

If we now choose the scale so that $\zeta^A = (1, z)$ then the above equation turns into an nth degree polynomial in z, having n complex roots. We can use this fact together

with Eq. (4.29) to translate between the $\Psi^{AB\ldots M}$ and the Z^α, should this be needed. We can also use the fundamental theorem of algebra to rewrite the polynomial as a product of n factors. In analogy with Eq. (4.29) we find that

$$\Psi_{AB\ldots M}\zeta^A\zeta^B\ldots\zeta^M = 0 = (\alpha_0+\alpha_1 z)(\beta_0+\beta_1 z)\ldots(\mu_0+\mu_1 z)\,. \tag{4.36}$$

The conclusion is that a rank n multispinor can be written – uniquely except for an overall factor of no interest to us – as a symmetrized product of n spinors:

$$\Psi_{AB\ldots M}\zeta^A\zeta^B\ldots\zeta^M = \alpha_A\zeta^A\beta_B\zeta^B\ldots\mu_M\zeta^M \;\Rightarrow\; \Psi_{AB\ldots M} = \alpha_{(A}\beta_B\ldots\mu_{M)}\,. \tag{4.37}$$

The factors are known as *principal spinors*, and they are of course the n unordered points (note the symmetrization) on the sphere in slight disguise.

The stellar representation, or equivalently the spinor notation, deals with linear subspaces in an elegant way. Consider a complex projective line (a $\mathbb{C}\mathbf{P}^1$) in $\mathbb{C}\mathbf{P}^2$ for definiteness. A general point in $\mathbb{C}\mathbf{P}^2$ is described by a pair of unordered points on the 2-sphere, or equivalently as a spinor

$$\Psi^{AB} = \alpha^{(A}\beta^{B)} \equiv \frac{1}{2}\left(\alpha^A\beta^B + \alpha^B\beta^A\right)\,. \tag{4.38}$$

Evidently we get a complex projective line by holding one of the points fixed and letting the other vary, that is by holding one of the principal spinors (say β^A) fixed and letting the other vary.

The spinor notation also deals elegantly with complex curves in general. Thus we get a conic section in $\mathbb{C}\mathbf{P}^2$ as the set of points for which the principal spinors coincide. That is to say that

$$\Psi^{AB} = \Psi^A\Psi^B \tag{4.39}$$

for some spinor Ψ^A. Through any point on the quadric (for which $\Psi^A = \alpha^A$ say) there goes a complex projective line

$$\Psi^{AB} = \zeta^{(A}\alpha^{B)} \tag{4.40}$$

(where ζ^A varies). This line is *tangent* to the quadric since it touches the quadric in a unique point $\alpha^A\alpha^B$. It is moreover rather easy to see that a pair of tangent lines always intersect in a unique point. See Figure 4.8.

4.5 The Fubini–Study metric

Now we want a notion of distance, and indeed a Riemannian metric, in $\mathbb{C}\mathbf{P}^n$. How do we do it? Pick an arbitrary pair of points. The distance between them will be the length of the geodesic curve connecting them. On the other hand we know that there is a unique projective line connecting them; topologically this is $\mathbb{C}\mathbf{P}^1 = \mathbf{S}^2$.

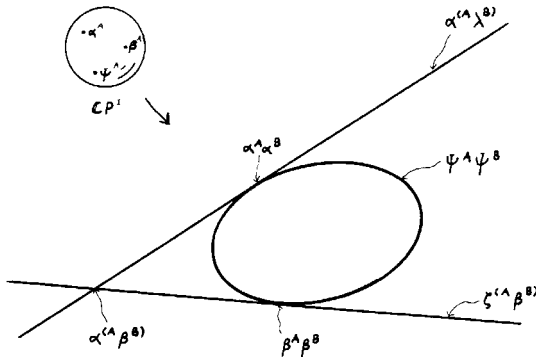


Figure 4.8. A conic section in $\mathbb{C}\mathbf{P}^2$ and a pair of tangent lines.

Suppose that we insist that this 2-sphere is a round 2-sphere, with the ordinary round metric. Let us also insist that the metric on $\mathbb{C}\mathbf{P}^n$ is such that the projective lines are *totally geodesic*. Technically a submanifold is said to be totally geodesic if a geodesic with respect to the induced metric on the submanifold is also a geodesic with respect to the metric on the embedding space, or equivalently if a geodesic that starts out parallel to the submanifold stays in the submanifold. But a geodesic on a complex projective line is simply a great circle on a 2-sphere, so we have found a way to define geodesics between two arbitrary points in $\mathbb{C}\mathbf{P}^n$. The resulting notion of distance is called the *Fubini–Study distance*.[3]

It only remains to make this definition of the metric on $\mathbb{C}\mathbf{P}^n$ explicit. Since the geodesic lives on some complex projective line, we can write down its equation using nothing but the results of Chapter 3. Let us recall Eq. (3.85) for a geodesic on an odd-dimensional sphere:

$$Z^\alpha(\sigma) = Z^\alpha(0)\cos\sigma + \dot{Z}^\alpha(0)\sin\sigma \ , \tag{4.41}$$

where

$$Z(0)\cdot\bar{Z}(0) = \dot{Z}(0)\cdot\dot{\bar{Z}}(0) = 1 \ , \qquad Z(0)\cdot\dot{\bar{Z}}(0) + \dot{Z}(0)\cdot\bar{Z}(0) = 0 \ . \tag{4.42}$$

If $\dot{Z}^\alpha(0) = \mathrm{i}Z^\alpha(0)$ this is a Hopf circle, and projects to a point on $\mathbb{C}\mathbf{P}^1$ $\mathbb{C}\mathbf{P}^1 = \mathbf{S}^2$. In general we can write

$$Z^\alpha(0) \equiv m^\alpha \ , \qquad \dot{Z}^\alpha(0) = m^\alpha\cos a + n^\alpha \sin a \ , \qquad m\cdot\bar{n} = 0 \ . \tag{4.43}$$

Since the constant vectors m^α and n^α have been chosen orthogonal, they lie antipodally on the particular complex projective line that they span. We can now appeal to the fibre bundle perspective (Section 4.8) to see that this curve is horizontal,

[3] It was first studied by two leading geometers from a century ago; Fubini (1903) and Study (1905).

and hence projects to a geodesic on $\mathbb{C}\mathbf{P}^1$, if and only if $\dot{Z} \cdot \bar{Z} = 0$, that is to say if and only if $\cos a = 0$. Alternatively we can use a direct calculation to show that for general a they project to latitude circles, and to great circles if and only if $\cos a = 0$. Either way, we conclude that our definition implies that a geodesic on $\mathbb{C}\mathbf{P}^n$ is given by

$$Z^\alpha(\sigma) = m^\alpha \cos\sigma + n^\alpha \sin\sigma \ , \qquad m \cdot \bar{m} = n \cdot \bar{n} = 1 \ , \ m \cdot \bar{n} = 0 \ . \tag{4.44}$$

More precisely, this is the horizontal lift of a geodesic on $\mathbb{C}\mathbf{P}^n$ to the odd-dimensional sphere $\mathbf{S}^{2n+1}$, in the fibre bundle $\mathbb{C}\mathbf{P}^n = \mathbf{S}^{2n+1}/\mathbf{S}^1$.

We will now use this set-up to define the Fubini–Study distance $D_{\rm FS}$ between any two points on $\mathbb{C}\mathbf{P}^n$. Since everything takes place within a complex projective line, which lifts to a 3-sphere, we can use the expression for the distance arrived at in our discussion of the Hopf fibration of $\mathbf{S}^3$, namely Eq. (3.95). The Fubini–Study distance $D_{\rm FS}$ must be given by

$$\cos^2 D_{\rm FS} \ = \ \kappa \ , \tag{4.45}$$

where the *projective cross-ratio* between two points P^α and Q^α is given by

$$\kappa \ = \ \frac{P \cdot \bar{Q} \ Q \cdot \bar{P}}{P \cdot \bar{P} \ Q \cdot \bar{Q}} \ . \tag{4.46}$$

We give the same formula in standard quantum mechanical notation in Eq. (5.16). There is one new feature: in Section 3.5 we assumed that $P \cdot \bar{P} = Q \cdot \bar{Q} = 1$, but this assumption has now been dropped. As in Section 4.1 there is a polarity – that is to say a map from the space of points to the space of hyperplanes – hidden here: the polarity maps the point P^α to that special hyperplane consisting of all points Q^α with

$$Q \cdot \bar{P} \ \equiv \ Q^\alpha \bar{P}_\alpha \ = \ 0 \ . \tag{4.47}$$

Interestingly complex conjugation appears here for the first time. Another interesting fact is that our definitions imply that $\mathbb{C}\mathbf{P}^n$ is a C_π manifold: all geodesics are closed, and have the same circumference. This is a quite exceptional property, related to the fact that the isotropy group – the subgroup of isometries leaving a given point invariant – is so large that it acts transitively on the space of directions there. Spheres have this property too. The circumference of a geodesic on $\mathbb{C}\mathbf{P}^n$ equals π because $\mathbb{C}\mathbf{P}^1$ is a 2-sphere of radius $1/2$; the geodesic on $\mathbf{S}^{2n+1}$ has circumference 2π and doubly covers the geodesic on $\mathbb{C}\mathbf{P}^n$. We also note that the maximal distance between two points is equal to $\pi/2$ (when the two vectors are orthogonal), and that all geodesics meet again at a point of maximal distance from their point of origin.

To obtain the line element in a local form we assume that

$$Q^\alpha \;=\; P^\alpha + \mathrm{d}P^\alpha \tag{4.48}$$

and expand to second order in the vector $\mathrm{d}P^\alpha$ (and to second order in D_{FS}). The result is the *Fubini–Study metric*

$$\mathrm{d}s^2 = \frac{\mathrm{d}P\cdot\mathrm{d}\bar{P}\;P\cdot\bar{P} - \mathrm{d}P\cdot\bar{P}\;P\cdot\mathrm{d}\bar{P}}{P\cdot\bar{P}\;P\cdot\bar{P}}\,. \tag{4.49}$$

Before we are done we will become very familiar with this expression.

It is already clear that many formulae will be simplified if we normalize our vectors so that $P\cdot\bar{P}=1$. In quantum mechanics, where the vectors are the usual Hilbert space state vectors, this is usually done. In this chapter we will use normalized vectors now and then in calculations, but on festive occasions – say when stating a definition – we do not.

An important definition follows immediately, namely that of the 2-form

$$\Omega \;=\; \mathrm{i}\,\frac{P\cdot\bar{P}\,\mathrm{d}P\cdot\wedge\mathrm{d}\bar{P} - \mathrm{d}P\cdot\bar{P}\wedge P\cdot\mathrm{d}\bar{P}}{P\cdot\bar{P}\;P\cdot\bar{P}}\,. \tag{4.50}$$

This is clearly a relative of the metric and the suggestion is that it is a (closed) symplectic form and that $\mathbb{C}\mathbf{P}^n$ is a Kähler manifold for all n. And this is true. To prove the Kähler property it is helpful to use the affine coordinates from Section 4.2. When expressed in terms of them the Fubini–Study metric becomes

$$\mathrm{d}s^2 = \frac{1}{1+|z|^2}\left(\mathrm{d}z^a\mathrm{d}\bar{z}_a - \frac{\bar{z}_a\mathrm{d}z^a\mathrm{d}\bar{z}_b z^b}{1+|z|^2}\right) \equiv 2g_{a\bar{b}}\,\mathrm{d}z^a\mathrm{d}\bar{z}^{\bar{b}}\,,\qquad |z|^2\equiv z^a\bar{z}_a\,. \tag{4.51}$$

(We are using $\delta_{a\bar{b}}$ to change $\bar{z}^{\bar{a}}$ into $\bar{z}_a$.) Now

$$2g_{a\bar{b}} = \frac{1}{1+|z|^2}\left(\delta_{a\bar{b}} - \frac{\bar{z}_a z_{\bar{b}}}{1+|z|^2}\right) = \partial_a\partial_{\bar{b}}\ln(1+|z|^2)\,. \tag{4.52}$$

Or more elegantly

$$g_{a\bar{b}} \;=\; \frac{1}{2}\partial_a\partial_{\bar{b}}\ln Z\cdot\bar{Z}\,, \tag{4.53}$$

where it is understood that the homogeneous coordinates should be expressed in terms of the affine ones. As we know from Section 3.3 this proves that $\mathbb{C}\mathbf{P}^n$ is a Kähler manifold. Every complex submanifold is Kähler too. For $n=1$ we recognize the metric on $\mathbf{S}^2$ written in stereographic coordinates.

We will use a different choice of coordinate system to explore the space in detail, but the affine coordinates are useful for several purposes, for instance if one wants to perform an explicit check that the curve (4.44) is indeed a geodesic, or when one

wants to compute the Riemann curvature tensor. Thus:

$$g^{a\bar{b}} = 2(1+|z|^2)(\delta^{a\bar{b}} + z^a \bar{z}^{\bar{b}}) \tag{4.54}$$

$$\Gamma_{bc}{}^{a} = -\frac{1}{1+|z|^2}\,(\delta^a_b \bar{z}_c + \delta^a_c \bar{z}_b) \tag{4.55}$$

$$R_{a\bar{b}c\bar{d}} = -2(g_{a\bar{b}} g_{c\bar{d}} + g_{a\bar{d}} g_{c\bar{b}}) \tag{4.56}$$

$$R_{a\bar{b}} = 2(n+1) g_{a\bar{b}}\ . \tag{4.57}$$

All components not related to these through index symmetries or complex conjugation are zero. From the expression for the Ricci tensor $R_{a\bar{b}}$ we see that its traceless part vanishes; the Fubini–Study metric solves the Euclidean Einstein equations with a cosmological constant. The form of the full Riemann tensor $R_{a\bar{b}c\bar{d}}$ shows that the space has constant holomorphic sectional curvature (but it does not show that the curvature is constant, and in fact it is not – our result is weaker basically because $g_{a\bar{b}}$ is just one 'block' of the metric tensor). There is an important theorem that says that a simply connected complex manifold of constant holomorphic sectional curvature is necessarily isometric with $\mathbb{C}\mathbf{P}^n$ or else it has the topology of an open unit ball in $\mathbb{C}^n$, depending on the sign of the curvature. (If it vanishes the space is $\mathbb{C}^n$.) The situation is clearly analogous to that of simply connected two-dimensional manifolds of constant sectional curvature, which are either spheres, flat spaces, or hyperbolic spaces with the topology of a ball.

We observe that both the Riemannian and the symplectic geometry relies on the Hermitian form $Z \cdot \bar{Z} = Z^\alpha \bar{Z}_\alpha$ in $\mathbb{C}^N$. Therefore distances and areas are invariant only under those projective transformations that preserve this form. These are precisely the unitary transformations, with anti-unitary transformations added if we allow the transformations to flip the sign of the symplectic form. In quantum mechanics this theorem goes under the name of Wigner's theorem (Wigner, 1959):

Theorem 4.1 (Wigner's) *All isometries of* $\mathbb{C}\mathbf{P}^n$ *arise from unitary or anti-unitary transformations of* $\mathbb{C}^N$.

Since only a part of the unitary group acts effectively the connected component of the isometry group is in fact $SU(N)/\mathbb{Z}_N$. For $N = 2$ we are dealing with a sphere and $SU(2)/\mathbb{Z}_2 = SO(3)$, that is the group of proper rotations. The full isometry group $O(3)$ includes reflections and arises when anti-unitary transformations are allowed.

We obtain an infinitesimal isometry by choosing an Hermitian matrix – a generator of $U(N)$ – and writing

$$\mathrm{i}\,\dot{Z}^\alpha = H^\alpha{}_\beta Z^\beta\ . \tag{4.58}$$

This equation – to reappear as the *Schrödinger equation* in Section 5.1 – determines a Killing flow on $\mathbb{C}\mathbf{P}^n$. A part of it represents 'gauge', that is changes in the overall phase of the homogeneous coordinates. Therefore we write the projective equation

$$\mathrm{i}Z^{[\alpha}\dot{Z}^{\beta]} \;=\; Z^{[\alpha}H^{\beta]}{}_{\gamma}Z^{\gamma}\;, \tag{4.59}$$

where the brackets denote anti-symmetrization. This equation contains all the information concerning the Killing flow on $\mathbb{C}\mathbf{P}^n$ itself, and is called the *projective Schrödinger equation* (Hughston, 1995, p. 59). The fixed points of the flow occur when the right-hand side vanishes. Because of the anti-symmetrized indices this happens when Z^α is an eigenvector of the Hamiltonian,

$$H^{\alpha}{}_{\beta}Z^{\beta} \;=\; EZ^{\alpha}\;. \tag{4.60}$$

A picture will be given later.

There are other ways of introducing the Fubini–Study metric. We usually understand $\mathbb{C}\mathbf{P}^1$, that is the 2-sphere, as a subset embedded in $\mathbb{R}^3$. Through the Hopf fibration we can also understand it as the base space of a fibre bundle whose total space is either $\mathbf{S}^3$ or $\mathbb{C}^2$ with the origin excluded. Both of these pictures can be generalized to $\mathbb{C}\mathbf{P}^n$ for arbitrary n. The fibre bundle picture will be discussed in more detail later. The embedding picture relies on an embedding into $\mathbb{R}^{(n+1)^2-1}$. The dimension is rather high but we will have to live with this. To see how it works, use homogeneous coordinates to describe $\mathbb{C}\mathbf{P}^n$ and form the matrix

$$\rho^{\alpha}_{\beta} \;\equiv\; \frac{Z^{\alpha}\bar{Z}_{\beta}}{Z\cdot\bar{Z}}\;. \tag{4.61}$$

This is a useful way to represent a point in $\mathbb{C}\mathbf{P}^n$ since the reduncancy of the homogeneous coordinates has disappeared. The resulting matrix is Hermitian and has trace unity. Hence we have an embedding of $\mathbb{C}\mathbf{P}^n$ into the space of Hermitian $N\times N$ matrices with trace unity, which has the dimension advertised, and can be made into a flat space in a natural way. We simply set

$$D^2(A,B) \;=\; \frac{1}{2}\mathrm{Tr}(A-B)^2\;. \tag{4.62}$$

This expression defines a distance D between the matrices A and B (equal to the *Hilbert–Schmidt distance* introduced in Chapter 8). It is analogous to the chordal distance between two points on the sphere when embedded in a flat space in the standard way.

To see that the embedding gives rise to the Fubini–Study metric we take two matrices representing different points of $\mathbb{C}\mathbf{P}^n$, such as

$$\rho^{\alpha}_{P\;\beta} = \frac{P^{\alpha}\bar{P}_{\beta}}{P\cdot\bar{P}} \qquad \rho^{\alpha}_{Q\;\beta} = \frac{Q^{\alpha}\bar{Q}_{\beta}}{Q\cdot\bar{Q}}\;, \tag{4.63}$$

and compute

$$D^2(\rho_P, \rho_Q) = 1 - \kappa = 1 - \cos^2 D_{\mathrm{FS}} \,, \tag{4.64}$$

where κ is the projective cross-ratio (4.46) and D_{FS} is the Fubini–Study distance between P and Q along a curve lying within the embedded $\mathbb{C}\mathbf{P}^n$. If the points are very close we can expand the square of the cosine to second order and obtain

$$D \;=\; D_{\mathrm{FS}} + \text{higher order terms} \,. \tag{4.65}$$

This proves our point. The Riemannian metric

$$\mathrm{d}s^2 = \frac{1}{2} \,\mathrm{Tr}\; \mathrm{d}\rho \,\mathrm{d}\rho, \tag{4.66}$$

is precisely the Fubini–Study metric, provided that it is evaluated at a point where the matrix ρ is an outer product of two vectors – and $\mathrm{d}\rho$ is the 'infinitesimal' difference between two such matrices, or more precisely a tangent vector to $\mathbb{C}\mathbf{P}^n$. This embedding will reappear in Chapter 8 as the embedding of pure states into the space of density matrices.

4.6 $\mathbb{C}\mathbf{P}^n$ illustrated

At this point we still do not know what $\mathbb{C}\mathbf{P}^n$ 'looks like', even in the modest sense that we feel that we know what an n-sphere 'looks like' for arbitrary n. There is a choice of coordinate system that turns out to be surprisingly helpful in this regard (Barros e Sá, 2001a; Bengtsson, Brännlund and Życzkowski, 2002). Define

$$(Z^0, Z^1, \ldots, Z^n) = (n_0, n_1 \mathrm{e}^{\mathrm{i}\nu_1}, \ldots, n_n \mathrm{e}^{\mathrm{i}\nu_n}) \,, \tag{4.67}$$

where $0 \leq \nu_i < 2\pi$ and the real numbers n_0, n_i are non-negative, $n_0 \geq 0$, $n_i \geq 0$, and obey the constraint

$$n_0^2 + n_1^2 + \cdots + n_n^2 = 1 \,. \tag{4.68}$$

We call such coordinates *octant coordinates*, because $n_0, n_1, \ldots$ clearly form the positive hyperoctant of an n-sphere. The phases ν_i form an n-torus, so we already see a picture of the topology of $\mathbb{C}\mathbf{P}^n$ emerging: we have a set of tori parametrized by the points of a hyperoctant, with the proviso that the picture breaks down at the edges of the hyperoctant, where the phases are undefined.[4]

[4] There is an entire branch of mathematics called 'toric geometry' whose subject matter, roughly speaking, consists of spaces that can be described in this way. See Ewald (1996) for more on this.

To go all the way to a local coordinate system we can set

$$\begin{cases} n_0 = \cos\vartheta_1 \sin\vartheta_2 \sin\vartheta_3 \ \dots \ \sin\vartheta_n \\ n_1 = \sin\vartheta_1 \sin\vartheta_2 \sin\vartheta_3 \ \dots \ \sin\vartheta_n \\ n_2 = \quad \cos\vartheta_2 \sin\vartheta_3 \ \dots \ \sin\vartheta_n \\ \vdots \ \vdots \qquad\qquad \vdots \\ n_n = \qquad\quad \cos\vartheta_n \end{cases} \qquad 0 < \vartheta_i < \frac{\pi}{2} \,. \tag{4.69}$$

This is just like Eq. (1.13), except that the range of the coordinates has changed. We can also use n_i as orthographic coordinates. Alternatively we can set

$$y_i \ = \ n_i^2 \tag{4.70}$$

and use the n coordinates y_i. This looks contrived at this stage but will suggest itself later. If we also define $y_0 = n_0^2$ we clearly have

$$y_0 + y_1 + \cdots + y_n = 1 \,. \tag{4.71}$$

A probability interpretation is not far behind ... .

To see how it all works, let us consider $\mathbb{C}\mathbf{P}^2$. The point is that octant coordinates are quite well adapted to the Fubini–Study metric, which becomes

$$\begin{aligned} \mathrm{d}s^2 = {} & \mathrm{d}n_0^2 + \mathrm{d}n_1^2 + \mathrm{d}n_2^2 \\ & + n_1^2(1-n_1^2)\,\mathrm{d}\nu_1^2 + n_2^2(1-n_2^2)\,\mathrm{d}\nu_2^2 - 2n_1^2 n_2^2 \ \mathrm{d}\nu_1\,\mathrm{d}\nu_2 \,. \end{aligned} \tag{4.72}$$

The first piece here, given Eq. (4.68), is recognizable as the ordinary round metric on the sphere. The second part is the metric on a flat torus, whose shape depends on where we are on the octant. Hence we are justified in thinking of $\mathbb{C}\mathbf{P}^2$ as a set of flat 2-tori parametrized by a round octant of a 2-sphere. There is an evident generalization to all n, and in fact we can draw accurate pictures of $\mathbb{C}\mathbf{P}^3$ if we want to.

For $n = 1$ we obtain a one parameter family of circles that degenerate to points at the end of the interval; a moment's thought will convince the reader that this is simply a way to describe a 2-sphere. Just to make sure, use the angular coordinates and find that Eq. (4.72) becomes

$$\mathrm{d}s^2 = \mathrm{d}\vartheta_1^2 + \frac{1}{4}\sin^2(2\vartheta_1)\,\mathrm{d}\nu_1^2 = \frac{1}{4}\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right) , \tag{4.73}$$

where we used $\theta = 2\vartheta_1$ and $\phi = \nu_1$ in the second step.

To make the case $n = 2$ quite clear we make a map of the octant, using a stereographic or a gnomonic projection (Section 3.1). The latter is quite useful here and it does not matter that only half the sphere can be covered since we need to cover only one octant anyway. It is convenient to centre the projection at the centre of the

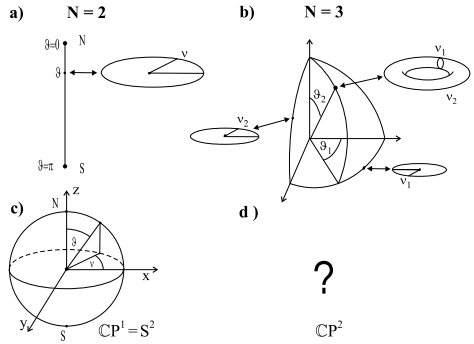


Figure 4.9. $\mathbb{C}\mathbf{P}^2$ may be visualized as the positive octant of a 2-sphere. Each point inside the octant represents a torus T^2 spanned by the phases (ν_1, ν_2) (b). Each point at the edges of the octant denotes a circle, so each of three edges representes a 2-sphere. For comparison we plot $\mathbb{C}\mathbf{P}^1$ in the same manner, in panel (a). A realistic view of $\mathbb{C}\mathbf{P}^1$ is shown in panel (c); an analogous view of $\mathbb{C}\mathbf{P}^2$ would be more difficult (d).

octant and adjust the coordinate plane so that the coordinate distance between a pair of corners of the resulting triangle equals one. The result is shown in Figure 4.10.

This takes care of the octant. We obtain a picture of $\mathbb{C}\mathbf{P}^2$ when we remember that each interior point really represents a flat torus, conveniently regarded as a parallelogram with opposite sides identified. The shape of the parallelogram – discussed in Section 3.3 – is relevant. According to Eq. (4.72) the lengths of the sides are

$$L_1 = \int_0^{2\pi} \mathrm{d}s = 2\pi n_1 \sqrt{1 - n_1^2} \quad \text{and} \quad L_2 = 2\pi n_2 \sqrt{1 - n_2^2}\,. \tag{4.74}$$

The angle between them is given by

$$\cos\theta_{12} = -\frac{n_1 n_2}{\sqrt{1 - n_1^2}\sqrt{1 - n_2^2}}\,. \tag{4.75}$$

The point is that the shape depends on where we are on the octant. So does the total area of the torus,

$$A = L_1 L_2 \sin\theta_{12} = 4\pi^2 n_0 n_1 n_2\,. \tag{4.76}$$

The 'biggest' torus occurs at the centre of the octant. At the boundaries the area of the tori is zero. This is because there the tori degenerate to circles. In effect an edge of the octant is a one parameter family of circles, in other words it is a $\mathbb{C}\mathbf{P}^1$.

It is crucial to realize that there is nothing special going on at the edges and corners of the octant, whatever the impression left by the map may be. Like the

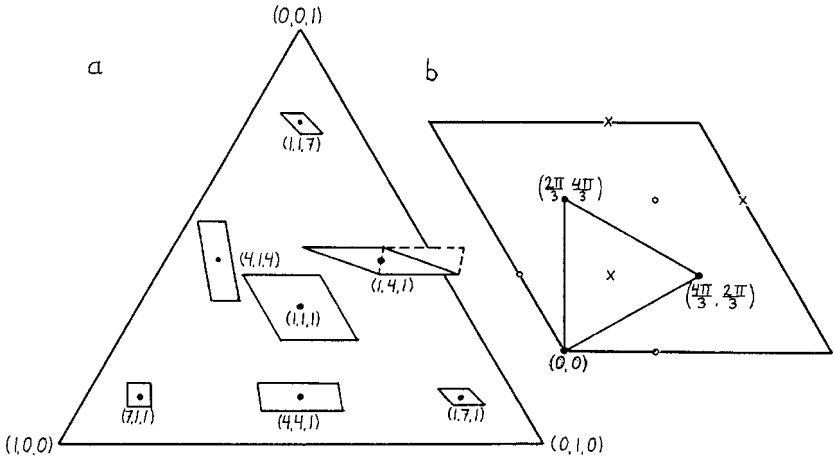


Figure 4.10. In (a) we indicate how the torus lying over each interior point changes with position in the octant. The position in the octant is given by an unnormalized vector. At the edges the tori degenerate to circles so edges are complex projective lines. The corners of the octant represent orthogonal points. It may be convenient to perform some cutting and gluing of the parallelogram before thinking about the shape of the torus it defines, as indicated with dashed lines for the torus lying over the point (1, 4, 1). The size of the octant relative to that of the tori is exaggerated in the picture. To bring this home we show, in (b), the largest torus – the one sitting over (1, 1, 1) – decorated with three times three points (marked with crosses and filled or unfilled dots). Each such triple corresponds to an orthogonal basis. The coordinates (ν_1, ν_2) are given for one of the triples.

sphere, $\mathbb{C}\mathbf{P}^n$ is a homogeneous space and looks the same from every point. To see this, note that any choice of an orthogonal basis in a three-dimensional Hilbert space gives rise to three points separated by the distance $\pi/2$ from each other in $\mathbb{C}\mathbf{P}^2$. By an appropriate choice of coordinates we can make any such triplet of points sit at the corners of an octant in a picture identical to the one shown in Figure 4.11.

To get used to the picture we begin by looking at some one real dimensional curves in it. We choose to look at the flow lines of an isometry – a Killing field. Since the isometries are given by unitary transformations of $\mathbb{C}^N$, we obtain an infinitesimal isometry by choosing an Hermitian matrix. Any such matrix can be diagonalized, so we can choose a basis in which the given Hermitian 3×3 matrix takes the form

$$H^{\alpha}{}_{\beta} = \begin{bmatrix} E_0 & 0 & 0 \\ 0 & E_1 & 0 \\ 0 & 0 & E_2 \end{bmatrix} . \tag{4.77}$$

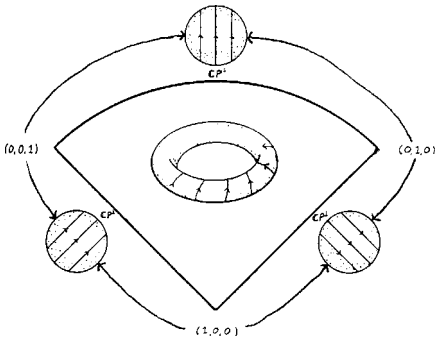


Figure 4.11. Killing vector flows on $\mathbb{C}\mathbf{P}^2$. Stereographic coordinates are used for the octant.

We can therefore arrange our octant picture so that the fixed points of the flow – determined by the eigenvectors – occur at the corners. If we exponentiate Eq. (4.58) we find that the isometry becomes

$$\begin{bmatrix} n^0 \\ n^1 \mathrm{e}^{\mathrm{i}\nu_1} \\ n^2 \mathrm{e}^{\mathrm{i}\nu_2} \end{bmatrix} \quad \rightarrow \quad \begin{bmatrix} \mathrm{e}^{-\mathrm{i}E_0 t} & 0 & 0 \\ 0 & \mathrm{e}^{-\mathrm{i}E_1 t} & 0 \\ 0 & 0 & \mathrm{e}^{-\mathrm{i}E_2 t} \end{bmatrix} \begin{bmatrix} n^0 \\ n^1 \mathrm{e}^{\mathrm{i}\nu_1} \\ n^2 \mathrm{e}^{\mathrm{i}\nu_2} \end{bmatrix} \tag{4.78}$$

where t is a parameter along the flow lines. Taking out an overall phase, we find that this implies that

$$n^I \rightarrow n^I \,, \qquad \nu_1 \rightarrow \nu_1 + (E_0 - E_1)t \,, \qquad \nu_2 \rightarrow \nu_2 + (E_0 - E_2)t \,. \tag{4.79}$$

Hence the position on the octant is preserved by the Killing vector flow; the movement occurs on the tori only. At the edges of the octant (which are spheres, determined by the location of the fixed points) the picture is the expected one. At a generic point the orbits wind around the tori and, unless the frequencies are rational multiples of each other, they will eventually cover the tori densely. Closed orbits are exceptional.

We go on to consider some submanifolds. Every pair of points in the complex projective plane defines a unique complex projective line, that is a $\mathbb{C}\mathbf{P}^1$, containing the pair of points. Conversely a pair of complex projective lines always intersect at a unique point. Through every point there passes a 2-sphere's worth of complex projective lines, conveniently parametrized by the way they intersect the line at infinity, that is the set of points at maximal distance from the given point. This is easily illustrated provided we arrange the picture so that the given point sits in a corner, with the line at infinity represented by the opposite edge. An interesting fact about $\mathbb{C}\mathbf{P}^2$ follows from this, namely that it contains *incontractible* 2-spheres – a complex projective line can be deformed so that its radius grows, but it cannot be

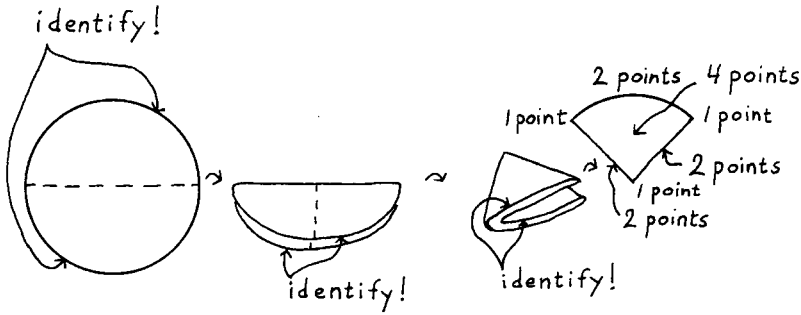


Figure 4.12. Using stereographic coordinates we show how the octant picture of the real submanifold $\mathbb{R}\mathbf{P}^2$ is related to the standard description as a hemisphere with antipodal points on the equator identified.

deformed so that its radius shrinks because it has to intersect the line at infinity in a point. (The topological reasons for this can be seen in the picture.)

Another submanifold is the real projective plane $\mathbb{R}\mathbf{P}^2$. It is defined in a way analogous to the definition of $\mathbb{C}\mathbf{P}^2$ except that real rather than complex numbers are used. The points of $\mathbb{R}\mathbf{P}^2$ are therefore in one-to-one correspondence with the set of lines through the origin in a three-dimensional real vector space and also with the points of $\mathbf{S}^2/\mathbb{Z}_2$, that is to say the sphere with antipodal points identified. In its turn this is a hemisphere with antipodal points on the equator identified. $\mathbb{R}\mathbf{P}^2$ is clearly a subset of $\mathbb{C}\mathbf{P}^2$. It is illuminating to see how the octant picture is obtained, starting from the stereographic projection of a hemisphere (a unit disc) and folding it twice.

Next choose a point. Adjust the picture so that it sits at a corner of the octant. Surround the chosen point with a 3-sphere consisting of points at constant distance from that point. In the picture this will appear as a curve in the octant, with an entire torus sitting over each interior point on the curve. This makes sense: from the Hopf fibration we know that $\mathbf{S}^3$ can be thought of as a one parameter family of tori with circles at the ends. The 3-sphere is round if the tori have a suitable rectangular shape. But as we let our 3-sphere grow, its tori get more and more 'squashed' by the curvature of $\mathbb{C}\mathbf{P}^2$, and the roundness gradually disappears. When the radius reaches its maximum value of $\pi/2$ the 3-sphere has collapsed to a 2-sphere, namely to the projective line at infinity. In equations, set

$$n_1 = \sin r \cos\frac{\theta}{2}\,, \qquad n_2 = \sin r \sin\frac{\theta}{2}\,, \qquad \tau = \nu_1 + \nu_2\,, \qquad \phi = \nu_1 - \nu_2\,, \quad (4.80)$$

where $0 \le r \le \pi/2, 0 \le \theta \le \pi$. When we express the Fubini–Study metric in these coordinates we find that they are geodesic polar coordinates with the coordinate r measuring the distance from the origin – curves with affine parameter equal to r

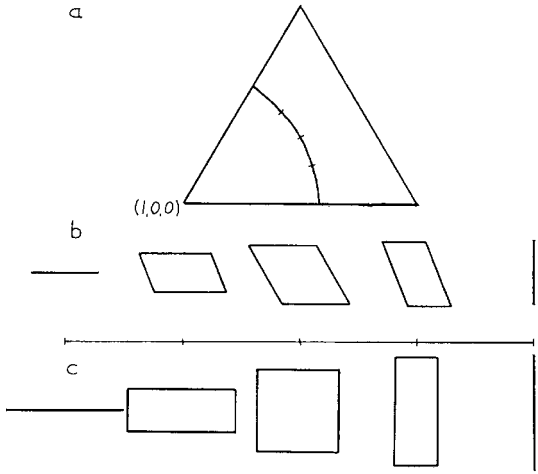


Figure 4.13. The set of points at constant distance from a corner form a squashed 3-sphere. In (a) we show how such a submanifold appears in the octant. All points in the torus lying over a point on the curve are included. In (b) we show how the size and shape of the torus change as we move along the curve in the octant; at the ends of the interval the tori collapse to circles. For comparison, in (c) we show the corresponding picture for a round 3-sphere.

are manifestly geodesics. Indeed

$$\mathrm{d}s^2 \;=\; \mathrm{d}r^2 + \sin^2 r\,(\Theta_1^2 + \Theta_2^2 + \cos^2 r\,\Theta_3^2)\,, \tag{4.81}$$

where the Θ_I are the invariant forms introduced in Section 3.7; our squashed spheres at constant distance are Berger spheres, as defined in the same section.

Finally, a warning: the octant picture distorts distances in many ways. For instance, the distance between two points in a given torus is shorter than it looks, because the shortest path between them is not the same as a straight line within the torus itself. Note also that we have chosen to exaggerate the size of the octant relative to that of the tori in the pictures. To realize how much room there is in the largest torus we note that one can find four sets of orthonormal bases in the Hilbert space $\mathbb{C}^3$ such that the absolute value of the scalar product between members of different bases is $1/\sqrt{3}$ – a kind of sphere packing problem in $\mathbb{C}\mathbf{P}^2$. If one basis is represented by the corners of the octant, the remaining 3×3 basis vectors are situated in the torus over the centre of the octant,[5] as illustrated in Figure 4.10. The message is that the biggest torus is really big.

[5] It is not known to what extent this property generalizes to arbitrary Hilbert space dimension. Ivanović (1981) and Wootters and Fields (1989) have shown that $N + 1$ orthonormal bases with the modulus of the scalar product between members of different bases always equal to $1/\sqrt{N}$ can be found if the dimension $N = p^k$, where p is a prime number.

4.7 Symplectic geometry and the Fubini–Study measure

So far the symplectic form has not been illustrated, but the octant coordinates are very useful for this purpose too. If we use $n_1, n_2, \ldots, n_n$ as orthographic coordinates on the octant we find that the symplectic form (4.50) is

$$\Omega \;=\; 2(n_1 \mathrm{d}n_1 \wedge \mathrm{d}\nu_1 + \cdots + n_n \mathrm{d}n_n \wedge \mathrm{d}\nu_n)\,. \tag{4.82}$$

Even better, we can use $y_i = n_i^2$ as coordinates. Then Ω takes the canonical form

$$\Omega = \mathrm{d}y_1 \wedge \mathrm{d}\nu_1 + \cdots + \mathrm{d}y_n \wedge \mathrm{d}\nu_n\,. \tag{4.83}$$

In effect (y_i, ν_i) are action-angle variables.

Given a symplectic form we can construct a phase space volume by wedging it with itself enough times. This is actually simpler than to compute the determinant of the metric tensor, and the two notions of volume – symplectic and metric – agree because we are on a Kähler manifold. Thus, in form language, the *Fubini–Study volume element* on $\mathbb{C}\mathbf{P}^n$ is

$$\mathrm{d}\tilde{\Omega}_n \;=\; \frac{1}{n!}\,(\frac{1}{2}\Omega) \wedge (\frac{1}{2}\Omega) \wedge \cdots \wedge (\frac{1}{2}\Omega)\,, \tag{4.84}$$

where we take n wedge products. Equivalently we compute the square root of the determinant of the metric. We have decorated the volume element with a tilde because in Section 7.6 we will divide it with the total volume of $\mathbb{C}\mathbf{P}^n$ to obtain the *Fubini–Study measure* $\mathrm{d}\Omega_n$, which is a probability distribution whose integral over $\mathbb{C}\mathbf{P}^n$ equals unity. Any measure related to the Fubini–Study volume element by a constant is distinguished by the fact that it gives the same volume to every ball of given radius, as measured by the Fubini–Study metric, and by the fact that it is unitarily invariant, in the sense that $\mathrm{vol}(A) = \mathrm{vol}\big(U(A)\big)$ for any subset A of $\mathbb{C}\mathbf{P}^n$.

It is a volume element worth studying in several coordinate systems. Using octant coordinates we obtain

$$\mathrm{d}\tilde{\Omega}_n \;=\; n_0 n_1 \ldots n_n \,\mathrm{d}V_{\mathbf{S}^n}\,\mathrm{d}\nu_1 \ldots \mathrm{d}\nu_n\,, \qquad 0 < \nu_i < 2\pi \tag{4.85}$$

where $\mathrm{d}V_{\mathbf{S}^n}$ is the measure on the round n-sphere. If we use orthographic coordinates on the round octant, remember that $\sqrt{g} = 1/n_0$ in these coordinates, and use the coordinates $y_i = n_i^2$, we obtain

$$\mathrm{d}\tilde{\Omega}_n = n_1 \ldots n_n \,\mathrm{d}n_1 \ldots \mathrm{d}n_n \,\mathrm{d}\nu_1 \ldots \mathrm{d}\nu_n = \frac{1}{2^n}\,\mathrm{d}y_1 \ldots \mathrm{d}y_n \,\mathrm{d}\nu_1 \ldots \mathrm{d}\nu_n\,. \tag{4.86}$$

All factors in the measure have cancelled out! As far as calculations of volumes are concerned $\mathbb{C}\mathbf{P}^n$ behaves like a Cartesian product of a flat simplex and a flat torus of a fixed size.

The angular coordinates ϑ can also be used on the octant. We just combine Eqs. (4.85) with Eq. (1.14) for the measure on the round octant. Computing the total volume of $\mathbb{C}\mathbf{P}^n$ with respect to its Fubini–Study volume element then leads to an easy integral:

$$\mathrm{vol}(\mathbb{C}\mathbf{P}^n) = \int_{\mathbb{C}\mathbf{P}^n} \mathrm{d}\tilde{\Omega}_n = \prod_{i=1}^{n} \int_0^{\frac{\pi}{2}} \mathrm{d}\vartheta_i \int_0^{2\pi} \mathrm{d}\nu_i \, \cos\vartheta_i \, \sin^{2i-1}\vartheta_i = \frac{\pi^n}{n!} \,. \quad (4.87)$$

We fixed the linear scale of the space by the requirement that a closed geodesic has circumference π. Then the volume goes to zero when n goes to infinity. Asymptotically the volume of $\mathbb{C}\mathbf{P}^n$ even goes to zero somewhat faster than that of $\mathbf{S}^{2n}$, but the comparison is not fair unless we rescale the sphere until the great circles have circumference π. Having done so we find that the volume of $\mathbb{C}\mathbf{P}^n$ is always larger than that of the sphere (except when $n = 1$, when they coincide).

It is curious to observe that

$$\mathrm{vol}(\mathbb{C}\mathbf{P}^n) = \mathrm{vol}(\Delta_n) \times \mathrm{vol}(\mathbf{T}^n) = \frac{\mathrm{vol}(\mathbf{S}^{2n+1})}{\mathrm{vol}(\mathbf{S}^1)} \,. \quad (4.88)$$

The first equality expresses the volume as the product of the volumes of a flat simplex and a flat torus and is quite surprising, while the second is the volume of $\mathbf{S}^{2n+1}$, given in Eq. (1.17), divided by the volume 2π of a Hopf circle as one would perhaps expect from the Hopf fibration. The lesson is that the volume does not feel any of the more subtle topological properties involved in the fibre bundle construction of $\mathbb{C}\mathbf{P}^n$.

4.8 Fibre bundle aspects

Although we did not stress it so far, it is clear that $\mathbb{C}\mathbf{P}^n$ is the base manifold of a bundle whose total space is $\mathbf{S}^{2n+1}$ or even $\mathbb{C}^N = \mathbb{C}^{n+1}$ with its origin deleted. The latter bundle is known as the *tautological bundle* for fairly obvious reasons. A lightning review of our discussion of fibre bundles in Chapter 3 consists, in effect, of the observation that our expression for the Fubini–Study metric is invariant under

$$\mathrm{d}P^\alpha \;\to\; \mathrm{d}P^\alpha + z P^\alpha \,, \quad (4.89)$$

where $z = x + \mathrm{i}y$ is some complex number. But the vector $x P^\alpha$ is orthogonal to the sphere, while $\mathrm{i}y P^\alpha$ points along its Hopf fibres. Because it is unaffected by changes in these directions the Fubini–Study metric really is a metric on the space of Hopf fibres. In some ways it is convenient to normalize our vectors to $Z \cdot \bar{Z} = 1$ here because then we are dealing with a principal bundle. The tautological bundle is not a principal bundle; technically it is called a *Hermitian line bundle*.

A fibre bundle can always be equipped with a connection that allows us to lift curves in the base manifold to the bundle space in a unique manner. Here the preferred choice is

$$\omega \;=\; \frac{\mathrm{i}}{2}\,\frac{Z\cdot \mathrm{d}\bar{Z} - \mathrm{d}Z\cdot \bar{Z}}{Z\cdot \bar{Z}} \;=\; -\mathrm{i}\,\mathrm{d}Z\cdot \bar{Z}\;. \tag{4.90}$$

In the last step we used normalized vectors. The equation $\omega = 0$ now expresses the requirement that $\mathrm{d}Z^\alpha$ be orthogonal to Z^α, so that the lifted curve will be perpendicular to the fibres of the bundle. A minor calculation confirms that

$$\mathrm{d}\omega \;=\; \Omega\;, \tag{4.91}$$

where Ω is the symplectic 2-form defined in Eq. (4.50).

Let us lift some curves on $\mathbb{C}\mathbf{P}^n$ to Hilbert space. We have already done so for a geodesic on the base manifold – the result is given in Eq. (4.44). Since $\dot{Z}\cdot\bar{Z} = 0$ along this curve, it follows that $\omega = 0$ along it, so indeed it is a horizontal lift. Another example that we have encountered already: in Eq. (4.58) we wrote the Schrödinger equation, or equivalently the flow line of some isometry, as

$$\mathrm{i}\,\mathrm{d}Z^\alpha = H^\alpha{}_\beta Z^\beta \mathrm{d}t\;. \tag{4.92}$$

Evidently

$$\omega = \frac{1}{2}(Z^\alpha \bar{H}_\alpha{}^\beta \bar{Z}_\beta - \bar{Z}_\alpha H^\alpha{}_\beta Z^\beta)\,\mathrm{d}t = 0 \tag{4.93}$$

along the resulting curve in the bundle, so that this is the horizontal lift of a curve defined in the base manifold by the projective Schrödinger equation (4.59).

Now consider an arbitrary curve $C(\sigma)$ in $\mathbb{C}\mathbf{P}^n$, and an arbitrary lift of it to $\mathbb{C}^N$. Let the curve run from σ_1 to σ_2. Define its *geometric phase* as

$$\phi_{\mathrm{g}} = \arg\big(Z(\sigma_2)\cdot\bar{Z}(\sigma_1)\big) - \int_C \omega\;. \tag{4.94}$$

From now on we assume that the length of the curve is less than $\pi/2$ so that the argument that defines the *total phase* – the first term on the right-hand side – of the curve is well defined. Suppose that we change the lift of the curve by means of the transformation

$$Z^\alpha \to \mathrm{e}^{\mathrm{i}\lambda} Z^\alpha \qquad \Rightarrow \qquad \omega \to \omega + \mathrm{d}\lambda\;, \tag{4.95}$$

where λ is some function of σ. Although both the connection and the total phase change under this transformation, the geometric phase does not. Indeed it is easy to see that

$$\phi_{\mathrm{g}} \to \phi_{\mathrm{g}} + \lambda(\sigma_2) - \lambda(\sigma_1) - \int_{\sigma_1}^{\sigma_2} \mathrm{d}\lambda = \phi_{\mathrm{g}}\;. \tag{4.96}$$

The conclusion is that the geometric phase does not depend on the particular lift that we use. Hence it is genuinely a function of the original curve in $\mathbb{C}\mathbf{P}^n$ itself. The total phase on the other hand is not – although it equals the geometric phase for a horizontal lift of the curve.

An interesting observation, with consequences, is that if we compute the total phase along the horizontal lift (4.44) of a geodesic in $\mathbb{C}\mathbf{P}^n$ we find that it vanishes. This means that the geometrical phase is zero for any lift of a geodesic. A geodesic is therefore a *null phase curve*. The converse is not true – we will find further examples of null phase curves in Chapter 6. To see what the consequences are, consider a *geodesic triangle* in $\mathbb{C}\mathbf{P}^n$, that is three points P, Q and R connected with geodesic arcs not longer than $\pi/2$. In flat space the lengths of the sides determine the triangle uniquely up to isometries, but in curved spaces this need not be so. As it happens a geodesic triangle in $\mathbb{C}\mathbf{P}^n$ is determined up to isometries by its side lengths and its *symplectic area*, that is the integral of the symplectic form Ω over any area bounded by the three geodesic arcs (Brehm, 1990). To such a triangle we can associate the *Bargmann invariant* [6]

$$\Delta_3(P, Q, R) = \frac{P \cdot \bar{Q}Q \cdot \bar{R}R \cdot \bar{P}}{P \cdot \bar{P}Q \cdot \bar{Q}R \cdot \bar{R}} = \cos D_{PQ} \cos D_{QR} \cos D_{RP} \mathrm{e}^{-\mathrm{i}\Phi} \,. \quad (4.97)$$

Here D_{PQ} is the Fubini–Study distance between the points whose homogeneous coordinates are P and Q, and so on. But what is the phase Φ? Since each side of the triangle is a geodesic arc the geometric phase of each side is zero, and by Eq. (4.94) the total phase of the side is equal to the integral along the arc of ω. Adding the contribution of all the sides (and normalizing the vectors) we get

$$\arg P \cdot \bar{Q}Q \cdot \bar{R}R \cdot \bar{P} = -\arg Q \cdot \bar{P} - \arg R \cdot \bar{Q} - \arg P \cdot \bar{R} = -\oint_{\partial\Delta} \omega \,. \quad (4.98)$$

It follows from Stokes' theorem that the phase Φ is the symplectic area of the triangle,

$$\Phi = \int_{\Delta} \Omega \,. \quad (4.99)$$

We can go on to define Bargmann invariants for four, five and more points in the same cyclic manner and – by triangulation – we find that their phases are given by the symplectic area of the geodesic polygon spanned by the points.

[6] It is invariant under phase changes of the homogenous coordinates and attracted Bargmann's attention (Bargmann, 1964) because, unlike the Fubini–Study distance, it is a complex valued function defined on $\mathbb{C}\mathbf{P}^n$. Blaschke and Terheggen (1939) considered this early on. That its argument is the area of a triangle was proved by both physicists (Mukunda and Simon, 1993) and mathematicians (Hangan and Masala, 1994).

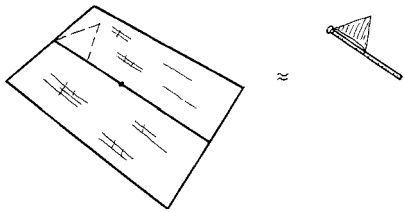


Figure 4.14. A flag may consist of a one-dimensional subspace of a two-dimensional subspace of a three-dimensional space – it is clear why it is called that!

The geometric phase also goes under the name of the *Berry phase*; the case that Berry studied was that of an eigenstate of a Hamiltonian H that is carried around a loop by adiabatic cycling of H.[7]

4.9 Grassmannians and flag manifolds

Projective space is only the first of its kind. Starting from an N-dimensional vector space V, real or complex, we can consider nested sequences of subspaces V_i of dimension d_i, such that $V_1 \subset V_2 \subset \cdots \subset V_r$. This is called a *flag* of subspaces (as explained in Figure 4.14). The space of all flags of a given kind is known as a *flag manifold* and denoted by $\mathbf{F}^{(N)}_{d_1 \ldots d_r}$; projective space is the easy case where a flag consists of a single one-dimensional subspace only. The next simplest case is that where the flags consist of a single M-dimensional subspace; such a flag manifold is known as the *Grassmannian* $\mathbf{F}^{(N)}_M = \mathbf{Gr}^{\mathbb{C}}_{M,N}$. The notation also tells us which field ($\mathbb{R}$ or $\mathbb{C}$) we are using; if there is no label then the complex numbers are implied.

To see how this works let us consider the case of $\mathbf{Gr}_{2,4}$, which is the space of 2-planes in a four-dimensional vector space. We fix a 2-plane in $\mathbb{C}^4$ by fixing two linearly independent vectors spanning that plane, and collect them into a rank two $N \times 2$ matrix

$$\begin{bmatrix} Z_{0,0} & Z_{0,1} \\ Z_{1,0} & Z_{1,1} \\ Z_{2,0} & Z_{2,1} \\ Z_{3,0} & Z_{3,1} \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ z_{2,0} & z_{2,1} \\ z_{3,0} & z_{3,1} \end{bmatrix} . \tag{4.100}$$

The entries in the matrix on the left are homogeneous coordinates on the Grassmannian. Then we exercised our right to perform linear combinations and rescalings of the columns to get the matrix in a standard form. The remaining four complex

[7] The original paper by Berry (1984) is reprinted in a book edited by Shapere and Wilczek (1989). See Anandan and Aharonov (1990) and Mukunda and Simon (1993) for the point of view that we take here, and Chruściński and Jamiołkowski (2004) for much more.

numbers serve as affine coordinates on the four complex dimensional Grassmannian. Note that if the upper two rows were linearly dependent this form could not have been reached, but since the matrix as a whole has rank two we can introduce a similar coordinate system by singling out another pair of rows. (Indeed it is not hard to see that the Grassmannian is a complex manifold that can be completely covered by six coordinate patches of this kind.) From a geometrical point of view it is sometimes advantageous to think of the Grassmannian of 2-planes as the space of projective lines in a projective space of dimension $N-1$; it is an interesting exercise to convince oneself directly that the space of lines in 3-space is four-dimensional. On the other hand the use of an $N \times M$ matrix as homogeneous coordinates for $\mathbf{Gr}_{M,N}$ has advantages too and leads to an immediate proof that it is a complex manifold of $M(N-M)$ complex dimensions.[8]

When we get to more involved examples of flag manifolds any kind of explicit description of the space will get involved, too. For this reason it is a good idea to fall back on a description as coset spaces. Let us begin with $\mathbf{F}_1^{(4)} = \mathbb{C}\mathbf{P}^3$. The group $GL(4,\mathbb{C})$ acts transitively on the underlying vector space. Pick any point, say $p = (1,0,0,0)$ and find its isotropy group – that subgroup of the transitive group that leaves the given point invariant. Here the isotropy group consists of all matrices h of the form

$$h = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet \end{bmatrix}, \tag{4.101}$$

where $\bullet$ is any complex number. This defines a subgroup of $GL(4,\mathbb{C})$, and we can now define $\mathbb{C}\mathbf{P}^3$ as a coset space following the recipe in Section 3.8. We can use $SL(4,\mathbb{C})$ as our starting point instead; the resulting coset will be the same once the form of h is restricted so that it has unit determinant. Call the resulting isotropy group $P_1^{(4)}$. Indeed we can restrict ourselves to $U(N)$ since the unitary subgroup of $GL(4,\mathbb{C})$ also acts transitively; when h is restricted to belong to $U(N)$ it becomes a block diagonal matrix representing $U(1) \times U(3)$. The argument clearly generalizes to any dimension, so that we have proved that

$$\mathbf{F}_1^{(N)} \equiv \mathbb{C}\mathbf{P}^{N-1} = \frac{SL(N,\mathbb{C})}{P_1^{(N)}} = \frac{U(N)}{U(1)\times U(N-1)}\,. \tag{4.102}$$

For real projective space we just have to replace $SL(\mathbb{C})$ with $SL(\mathbb{R})$ and the unitary groups by orthogonal groups.

[8] The affine coordinates are particularly useful if one wants to discuss the natural (Fubini–Study like) metric properties of the Grassmannians. See Wong (1967) for this.

The same argument repeated for the Grassmannian $\mathbf{Gr}_{2,4}$ reveals that the isotropy group can be written as the set $P_2^{(4)}$ of special linear matrices of the form

$$h = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet \end{bmatrix} , \tag{4.103}$$

or as block diagonal unitary matrices belonging to $U(2) \times U(2)$ if only unitary transformations are considered. For the flag manifold $\mathbf{F}_{1,2,3}^{(N)}$ the isotropy group is a *Borel subgroup* B of $SL(4, \mathbb{C})$; by definition the *standard Borel subgroup* is the group of upper triangular matrices and a general Borel subgroup is given by conjugation of the standard one. If the group is $U(N)$ the Borel subgroup is the subgroup $U(1) \times U(1) \times U(1) \times U(1)$, given by diagonal matrices in the standard case. The isotropy groups considered above are examples of *parabolic subgroups* P of the group G, represented by 'block upper triangular' matrices such that $B \subset P \subset G$. Thus we can write, in complete generality,

$$\mathbf{F}_{d_1 d_2 \ldots d_r}^{(N)} = \frac{SL(N, \mathbb{C})}{P_{d_1 d_2 \ldots d_r}^{(N)}} = \frac{U(N)}{[U(k_1) \times U(k_2) \ldots U(k_{r+1})]} , \tag{4.104}$$

where $k_1 = d_1$ and $k_{i+1} = d_{i+1} - d_i$ (and $d_{r+1} = N$). Given that the real dimension of $U(N)$ is N^2 the real dimension of an arbitrary flag manifold is

$$\dim(\mathbf{F}_{d_1 d_2 \ldots d_r}^{(N)}) = N^2 - \sum_{i=1}^{r+1} k_i^2 . \tag{4.105}$$

Equivalently it is simply twice the number of zeroes in the complex matrices representing the parabolic subgroup $P_{d_1 d_2 \ldots d_r}^{(N)}$. For real flag manifolds, replace the unitary groups with orthogonal groups.

The two descriptions of flag manifolds given in Eq. (4.104) are useful in different ways. To see why we quote two facts: the quotient of two complex groups is a complex manifold, and the quotient of two compact groups is a compact manifold. Then the conclusion is that all flag manifolds are compact complex manifolds. Indeed they are also Kähler manifolds. This has to do with another way of arriving at them, based on *adjoint orbits* in a Lie algebra. This aspect of flag manifolds will be discussed at length in Chapter 8 but let us reveal the point already here. The Lie algebra of the unitary group is the set of Hermitian matrices, and such matrices can always be diagonalized by unitary transformations. The set of Hermitian matrices with a given spectrum $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ can therefore be written as

$$H_{\lambda_1 \lambda_2 \ldots \lambda_N} = U H_{\text{diag}} U^{-1} . \tag{4.106}$$

This is an adjoint orbit and a little bit of thinking reveals that it is also a flag manifold $\mathbf{F}^{(N)}$; which particular one depends on whether there are degeneracies in the spectrum. For the special case when $\lambda_1 = 1$ and all the others are zero we came across this fact when we embedded $\mathbb{C}\mathbf{P}^n$ in the flat vector space of Hermitian matrices; see Eq. (4.61). The theme will be further pursued in section 8.5.[9]

A variation on the theme deserves mention too. A *Stiefel manifold* is by definition the space of sets of M orthonormal vectors in an N-dimensional vector space. It is not hard to see that these are the complex (real) homogeneous spaces

$$\mathcal{S}t_{M,N}^{(\mathbb{C})} = \frac{U(N)}{U(N-M)} \quad \text{and} \quad \mathcal{S}t_{M,N}^{(\mathbb{R})} = \frac{O(N)}{O(N-M)}. \tag{4.107}$$

This is not quite the same as a Grassmannian since the definition of the latter is insensitive to the choice of basis in the M-dimensional subspaces that are the points of the Grassmannian. As special cases we get $\mathcal{S}t_{1,N}^{(\mathbb{C})} = \mathbf{S}^{2N-1}$ and $\mathcal{S}t_{1,N}^{(\mathbb{R})} = \mathbf{S}^{N-1}$; see Section 3.8.

Problems

Problem 4.1 Consider $n+2$ ordered points on the plane, not all of which coincide. Consider two such sets equivalent if they can be transformed into each other by translations, rotations and scalings. Show that the topology of the resulting set is that of $\mathbb{C}\mathbf{P}^n$. What does $\mathbb{C}\mathbf{P}^n$ have to do with archaeology?

Problem 4.2 If you manage to glue together a Möbius strip and a hemisphere you get $\mathbb{R}\mathbf{P}^2$. What will you obtain if you glue two Möbius strips together?

Problem 4.3 Carry through the argument needed to prove that a complex projective line cannot be shrunk to a point within $\mathbb{C}\mathbf{P}^n$, using formulae, and using the octant picture.

[9] For further information on flag manifolds, in a form accessible to physicists, consult Picken (1990). The adjoint orbits should really be coadjoint, but the distinction is irrelevant here.

5

Outline of quantum mechanics

> Quantum mechanics is like a pot: it is almost indestructible and extremely rigid, but also very flexible because you can use any ingredients for your soup.
>
> *Göran Lindblad*

5.1 Quantum mechanics

Although our first four chapters have been very mathematical, quantum mechanics has never been very far away. Let us recall how the mathematical structure of quantum mechanics is usually summarized at the end of a first course in the subject. First of all the pure states are given by vectors in a Hilbert space. If that Hilbert space is finite dimensional it is simply the vector space $\mathbb{C}^N$ equipped with a scalar product of the particular kind that we called a Hermitian form in Eq. (3.82). Actually a pure state corresponds to an entire equivalence class of vectors; this is usually treated in such a way that the vectors are normalized to have length one and afterwards vectors differing by an overall phase $e^{i\phi}$ are regarded as physically equivalent. In effect then the space of pure states is the complex projective space $\mathbb{C}\mathbf{P}^n$; as always in this book $n = N - 1$. The notation used in quantum mechanics differs from what we have used so far. We have denoted vectors in $\mathbb{C}^N$ by Z^α, while in quantum mechanics they are usually denoted by a *ket* vector $|\psi\rangle$. We can think of the index α as just a label telling us that Z^α is a vector, and then these two notations are in fact exactly equivalent. It is more common to regard the index as taking N different values, and then Z^α stands for the N components of the vector with respect to some chosen basis, that is

$$|\psi\rangle \;=\; \sum_{\alpha=0}^{n} Z^\alpha \, |e_\alpha\rangle \; . \tag{5.1}$$

To each vector Z^α there corresponds in a canonical way a dual vector $\bar{Z}_\alpha$; in Dirac's notation this is a *bra* vector $\langle\psi|$. Given two vectors there are then two kinds of objects that can be formed: a complex number $\langle\phi|\psi\rangle$, and an operator $|\psi\rangle\langle\phi|$ that can be represented by a square matrix of size N.

The full set of states includes both pure and mixed states. To form mixtures we first write all pure states as operators, that is we define

$$\rho = |\psi\rangle\langle\psi| \quad \Leftrightarrow \quad \rho^\alpha{}_\beta = Z^\alpha \bar{Z}_\beta \,, \tag{5.2}$$

where the state vector is assumed to be normalized to length one and our two equivalent notations have been used. This is a Hermitian matrix of trace one and rank one, in other words it is a projection operator onto the linear subspace spanned by the original state vector; note that the unphysical phase of the state vector has dropped out of this formula so that we have a one-to-one correspondence between pure states and projection operators onto rays. Next we take convex combinations of K pure states, and obtain expressions of the form

$$\rho \;=\; \sum_{i=1}^{K} p_i \, |\psi_i\rangle\langle\psi_i| \,. \tag{5.3}$$

This is a *density matrix*, written as a convex mixture of pure states.[1] The set $\mathcal{M}$ of all states, pure or mixed, coincides with the set of Hermitian matrices with non-negative eigenvalues and unit trace, and the pure states as defined above form the extreme elements of this set (in the sense of Section 1.1).

The time evolution of pure states is given by the *Schrödinger equation*. It is unitary in the sense that the scalar product of two vectors is preserved by it. In the two equivalent notations it is given by

$$\mathrm{i}\hbar\,\partial_t|\psi\rangle = H|\psi\rangle \quad \Leftrightarrow \quad \mathrm{i}\hbar\,\dot{Z}^\alpha = H^\alpha{}_\beta\, Z^\beta \,, \tag{5.4}$$

where H is an Hermitian matrix that has been chosen as the Hamiltonian of the system and Planck's constant $\hbar$ has been explicitly included. We will follow the usual custom of defining our units so that $\hbar = 1$. In the geometrical language of Section 4.6 the Hamiltonian is a generator of a Killing field on $\mathbb{C}\mathbf{P}^n$. By linearity, the unitary time evolution of a density matrix is given by

$$\mathrm{i}\dot{\rho} = [H, \rho] = H\rho - \rho H \,. \tag{5.5}$$

A system that interacts with the external world can change its state in other and non-unitary ways as well; how this happens is discussed in Chapter 10.

[1] The density matrix was introduced by von Neumann (1927). He refers to it as the statistical operator (von Neumann, 1955). The word density matrix was first used by Dirac, in a slightly different sense. According to Coleman (1963) quantum chemists and statistical physicists came to an agreement on nomenclature in the early 1960s.

It remains to extract physical statements out of this formalism. One way to do this is to associate the projection operators (5.2) to elementary 'yes/no' questions. In this view, the *transition probability* $|\langle\psi|\phi\rangle|^2$ is the probability that a system in the state $|\phi\rangle$ answers 'yes' to the question represented by $|\psi\rangle$ – or the other way around, the expression is symmetric in $|\phi\rangle$ and $|\psi\rangle$. Hermitian operators can be thought of as weighted sums of projection operators – namely those projection operators that project onto the eigenvectors of the operator. They can also be thought of as random variables. What one ends up with is an association between physical observables, or measurements, on the one hand, and Hermitian operators on the other hand, such that the possible outcomes of a measurement are in one-to-one correspondence with the eigenvalues of the operator, and such that the expectation value of the measurement is

$$\langle A\rangle \;=\; \mathrm{Tr}\rho A\;. \tag{5.6}$$

When the state is pure, as in Eq. (5.2), this reduces to

$$\langle A\rangle \;=\; \langle\psi|A|\psi\rangle\;. \tag{5.7}$$

There is more to be said about measurements – and we will say a little bit more about this tangled question, in Section 10.1.

The role of the state is to assign probabilities to measurements. Note that, even if the state is pure, probabilities not equal to one or zero are present as soon as $[\rho, A] \neq 0$. There is a sample space for every non-degenerate Hermitian matrix, and the quantum state ρ has to assign a probability distribution for each and every one of these. A pure quantum state ρ is classically pure only with respect to the very special observables that commute with ρ. This is the reason (or at least the beginning of a reason) for why quantum mechanics is so much more subtle than classical statistics, where the set of pure states itself has the trivial structure of a discrete set of points.

Two disclaimers must be made. First, if infinite-dimensional Hilbert spaces are allowed then one must be more careful with the mathematical formulation. Second, the interpretation of quantum mechanics is a difficult subject. For the moment we ignore both points, but we do observe that using optical devices it is possible to design an experimental realization of arbitrary $N \times N$ unitary matrices (Reck, Zeilinger, Bernstein and Bertani, 1994). Therefore the formalism must be taken seriously in its entirety.

5.2 Qubits and Bloch spheres

It is not enough to say that $\mathbb{C}\mathbf{P}^n$ is the space of pure states. We have to know what physical states are being referred to. Like the sphere, $\mathbb{C}\mathbf{P}^n$ is a homogeneous space

in which a priori all points are equivalent. Therefore it cannot serve as a space of states without further embellishments, just as a sphere is a poor model of the surface of the Earth until we have decided which particular points are to represent Kraków, Stockholm, and so on. The claim is that every physical system can be modelled by $\mathbb{C}\mathbf{P}^n$ for some (possibly infinite) value of n, provided that a definite correspondence between the system and the points of $\mathbb{C}\mathbf{P}^n$ is set up.

But how do we set up such a correspondence, or in other words how do we tie the physics of some particular system to the mathematical framework? The answer depends very much on the system. One case which we understand very well is that of a particle of spin $1/2$ that we can perform Stern–Gerlach experiments on. The idea is that to every pure state of this particle there corresponds a unique direction in space such that the spin is 'up' in that direction. Therefore its space of pure states is isomorphic to the set of all directions in space – ordinary physical Space – and hence to $\mathbf{S}^2 = \mathbb{C}\mathbf{P}^1$. In this connection the sphere is known as the *Bloch sphere*; from the way it is defined it is evident that antipodal points on the Bloch sphere correspond to states that have spin up in opposite directions, that is to orthogonal states as usually defined. The *Bloch ball*, whose boundary is the Bloch sphere, is also of interest and in fact it corresponds to the space of density matrices in this case. Indeed an arbitrary Hermitian matrix of unit trace can be parametrized as

$$\rho = \begin{bmatrix} \frac{1}{2} + z & x - \mathrm{i}y \\ x + \mathrm{i}y & \frac{1}{2} - z \end{bmatrix} . \tag{5.8}$$

It is customary to regard this as an expansion in terms of the Pauli matrices $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, so that

$$\rho \;=\; \frac{1}{2}\,\mathbb{1} + \vec{\tau} \cdot \vec{\sigma} \; . \tag{5.9}$$

The vector $\vec{\tau}$ is known as the *Bloch vector*, and its components are coordinates in the space of matrices with unit trace. The matrix ρ is a density matrix if and only if its eigenvalues are non-negative, and this is so if and only if

$$x^2 + y^2 + z^2 \;\leq\; \frac{1}{4} \, . \tag{5.10}$$

This is indeed a ball, with the Bloch sphere as its surface. We will study density matrices in depth in Chapter 8; meanwhile let us observe that if the physics singles out some particular spatial direction as being of special importance – say because the particle is in a magnetic field – then the homogeneous Bloch sphere begins to acquire an interesting geography.

Note that we are clearly oversimplifying things here. A real spin $1/2$ particle has degrees of freedom such as position and momentum that we ignore. Our description

of the spin 1/2 particle is therefore in a way a caricature, but like all good caricatures it does tell us something essential about its subject.

Another well-understood case is that of a photon of fixed momentum. Here we can measure the polarization of the photon, and it is found that the states of the photon are in one-to-one correspondence to the set of all oriented ellipses (including the degenerate cases corresponding to circular and linear polarization). This set at first sight does not seem to be a sphere. What we can do is regard every ellipse as the projection down to a plane of an oriented great circle on the sphere, and associate a point on the sphere to a vector through the origin that is orthogonal to the given great circle. Unfortunately to every oriented ellipse – except the two special ones representing circular polarization – there now correspond two points on the sphere, so this cannot be quite right. The solution is simple. Start with the Riemann sphere and represent its points with a complex number

$$z = x + \mathrm{i}y = \frac{\sin\theta \mathrm{e}^{\mathrm{i}\phi}}{1+\cos\theta} = \tan\frac{\theta}{2}\mathrm{e}^{\mathrm{i}\phi} \,. \tag{5.11}$$

Here x and y are the stereographic coordinates defined in Section 3.1, while ϕ and θ are the latitude and longitude (counted from the north pole). Let 0 and ∞ – the north and south poles of the sphere – correspond to circular polarization. Now take the square root $w = \sqrt{z}$ and introduce another Riemann sphere whose points are labelled with w. For every point except 0 and ∞ there are two distinct points w for every z. Finally associate an oriented ellipse with the second sphere as originally indicated. The pair of points w that correspond to the same z now give rise to the same ellipse. In this way we obtain a one-to-one correspondence between oriented ellipses, that is to the states of the photon, and the points z on the first of the Riemann spheres. Hence the state space is again $\mathbb{C}\mathbf{P}^1$, in this connection known as the *Poincaré sphere*. Note that antipodal points on the equator of the Poincaré sphere correspond to states that are linearly polarized in perpendicular directions. For an arbitrary state the ratio of the minor to the major axis of the polarization ellipse is (with the angle α defined in Figure 5.1)

$$\cos\alpha = \frac{1-\tan^2\alpha}{1+\tan^2\alpha} = \frac{1-\tan\frac{\theta}{2}}{1+\tan\frac{\theta}{2}} = \frac{1-|z|}{1+|z|} \,. \tag{5.12}$$

The reader may now recognize the *Stokes' parameters* from textbooks on optics (Born and Wolf, 1987).

The spin 1/2 particle and the photon are examples of *two-level systems*, whose defining property is that a (special kind of) measurement that we can perform on the system yields one of two possible results. In this respect a two-level system behaves like the 'bit' of computer science. However, quantum mechanics dictates that the full space of states of such a system is a $\mathbb{C}\mathbf{P}^1$, a much richer state space

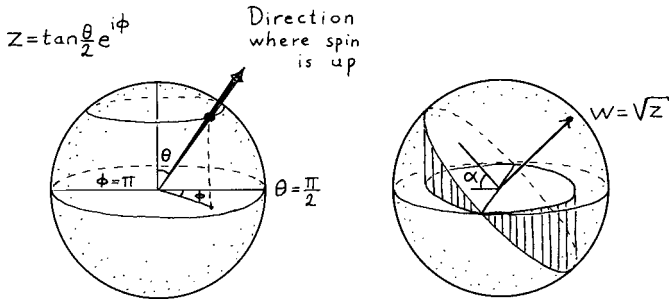


Figure 5.1. The Bloch ball, with a reminder about the coordinates that we are using. Also the auxiliary sphere used to establish the correspondence to photon polarization.

than that of a classical bit. It is a *qubit*. Two-level systems often appear when a sufficiently drastic approximation of a physical system is made; typically we have a potential with two degenerate minima and a high barrier between them. There are then four points on the sphere that have a clear interpretation: the north and south poles correspond to states where the system sits in one of the minima of the potential, and the east and west poles correspond to eigenstates of the Hamiltonian. We may not have sufficient control over the system to interpret the remaining states of the qubit, but quantum mechanics dictates them to be there.

When the dimension of the Hilbert space goes up we encounter first *three-level systems* or *qutrits* and then *N-level systems* or *quNits*.[2] The quantum mechanical formalism treats all quNits in the same way, but its meaning depends on how the formalism is tied to the physics of the problem; it may be a spin system, it may be an atom with N relevant energy levels, or the number N may appear simply because we have binned our experimental data into N separate bins.

5.3 The statistical and the Fubini–Study distances

We now turn to the physical interpretation of the Fubini–Study metric. There is a very good reason to consider this particular geometry, namely that it gives the statistical distance between states, in the sense of Section 2.5. But the definition of statistical distance in quantum mechanics requires some thought. Suppose that we wish to distinguish between two quantum states by means of a finite set of experiments. It will then be necessary to choose some specific measurement to

[2] The name 'qubit' was born in conversations between Wootters and Schumacher, and first used in print by the latter (Schumacher, 1995). The name 'bit' is due to John Tukey, consulting for Bell Labs in the 1940s. Whether anyone deserves credit for the name 'quNit' is unclear. Authors who work with Hilbert spaces of dimension d talk of 'qudits', which is perhaps marginally better as a name.

perform, or in mathematical terms to choose some Hermitian operator to describe it. We can use the result of this measurement to define the statistical distance between the given states, but it is clear that this distance will depend on the operator as well as on the states. By varying the operator we should be able to define the least possible statistical distance between the states in a unique manner, and by definition this will be the distance between the states. What we see here reflects the fact that each non-degenerate observable defines its own sample space. The situation is simple only if the states are orthogonal. Then the optimal measurement is one having the two states as eigenvectors. In general no such operator will exist, and the pure states therefore have a much more interesting geometry than the pure states in classical statistics, which are just an equidistant set of corners of a simplex.

Suppose that the two states are $|\psi\rangle$ and $|\phi\rangle$. Choose an operator A, and observe that it has $n+1$ orthogonal eigenstates $|i\rangle$ in terms of which we can expand

$$|\psi\rangle = \sum_{i=0}^{n} c_i |i\rangle \quad \text{and} \quad |\phi\rangle = \sum_{i=0}^{n} d_i |i\rangle \ . \tag{5.13}$$

For convenience we normalize all state vectors to unity. The probability to obtain a given outcome of the measurement is given by the standard interpretation of quantum mechanics – for the ith outcome to occur when the state is $|\psi\rangle$ it is $|c_i|^2$. According to Section 2.5 the statistical Bhattacharyya distance between the given states, given the operator, can be computed from the square roots of the probabilities:

$$\cos d_A = \sum_{i=0}^{n} |c_i|\,|d_i| = \sum_{i=0}^{n} |\langle\psi|i\rangle|\,|\langle\phi|i\rangle| \ . \tag{5.14}$$

According to the definition of distance between quantum mechanical states we should now choose the operator A in such a way that d_A becomes as small as possible, that is to say that the right-hand side should be as close to one as it can get. As noted in Problem 5.1 there are several operators that give the same answer; one solution is to let A have either $|\psi\rangle$ or $|\phi\rangle$ as one of its eigenstates, in which case the expression collapses to

$$\cos D_{\mathrm{FS}} = |\langle\psi|\phi\rangle| \ . \tag{5.15}$$

But this is precisely the geodesic distance between the two states as computed by means of the Fubini–Study metric. It was given in projectively invariant form in Eqs. (4.45)–(4.46); in the present notation

$$\cos^2 D_{\mathrm{FS}} = \kappa = \frac{|\langle\psi|\phi\rangle|^2}{\langle\psi|\psi\rangle\ \langle\phi|\phi\rangle} \ . \tag{5.16}$$

We have therefore established that the Fubini–Study metric measures the distinguishability of pure quantum states in the sense of statistical distance.[3] The projective cross-ratio κ is more often referred to as the transition probability or, in quantum communication theory, as the *fidelity function* (Jozsa, 1994).

More precisely the Fubini–Study distance measures the experimental distinguishability of two quantum states under the assumption that there are no limitations on the kind of experiments we can do. In practice a laboratory may be equipped with measurement apparatus corresponding to a small subset of Hermitian operators only, and this apparatus may have various imperfections. An atomic physicist confronted with three orthogonal states of the hydrogen atom, say the ground state $n = 1$, and two states with $n = 100$ and $n = 101$, respectively, may justifiably feel that the latter two are in some sense closer even though all three are equidistant according to the Fubini–Study metric. (They are closer indeed with respect to the Monge metric discussed in Section 7.7.) The Fubini–Study geometry remains interesting because it concerns what we can know in general, without any knowledge of the specific physical system.

An instructive sidelight on the role of the Fubini–Study metric as a distinguishability measure is thrown by its appearance in the Aharonov–Anandan time–energy uncertainty relation (Anandan and Aharonov, 1990). In geometrical language this is a statement about the velocity of the Killing flow. Using homogeneous coordinates to express the metric as (4.49) together with the projective Schrödinger equation (4.59) we can write

$$\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 = 2\frac{Z^{[\alpha}\mathrm{d}\dot{Z}^{\beta]}\,\bar{Z}_{[\alpha}\mathrm{d}\dot{\bar{Z}}_{\beta]}}{\bar{Z}\cdot Z\,\bar{Z}\cdot Z} = \langle H^2\rangle - \langle H\rangle^2\,, \tag{5.17}$$

where brackets denote anti-symmetrization, a minor calculation precedes the last step and

$$\langle H\rangle \;\equiv\; \frac{\bar{Z}_\alpha H^\alpha{}_\beta\, Z^\beta}{\bar{Z}\cdot Z} \tag{5.18}$$

and so on. The final result is pleasing:

$$\frac{\mathrm{d}s}{\mathrm{d}t} \;=\; \sqrt{\langle H^2\rangle - \langle H\rangle^2}\,. \tag{5.19}$$

This is indeed a precise version of the time–energy uncertainty relation: the system is moving quickly through regions where the uncertainty in energy is large.

[3] This very precise interpretation of the Fubini–Study metric was first given by Wootters (1981), although similar but less precise interpretations were given earlier.

5.4 A real look at quantum dynamics

Why is the Hilbert space complex? This is a grand question that we will not answer. But an interesting point of view can be found if we 'take away' the imaginary number i and think of the complex vector space $\mathbb{C}^N$ as the real vector space $\mathbb{R}^{2N}$ with some extra structure on it.[4] The notation that is used for complex vector spaces, whether bras or kets or our index notation, actually hides some features that become transparent when we use real notation and keep careful track of whether the indices are upstairs or downstairs. The thing to watch is how the observables manage to play two different roles. They form an algebra, and in this respect they can cause transformations of the states, but they also provide a map from the states to the real numbers. Let us call the vectors X^I. A linear observable capable of transforming this vector to another vector must then be written as $A^I{}_J$ so that $X^I \to A^I{}_J X^J$. On the other hand the matrix should be able to transform the vector to a real number, too. This can be done if we regard the observable as a quadratic form, in which case it must have both its indices downstairs: $X^I \to X^I A_{IJ} X^J$. We therefore need a way of raising and lowering indices on the observable. A metric tensor g_{IJ} seems to be called for, but there is a problem with this. The metric tensor must be a very special object, and therefore it seems natural to insist that it is not changed by transformations caused by the observables. In effect then the observables should belong to the Lie algebra of the orthogonal group $SO(2N)$. But using the metric to lower indices does not work because then we have

$$0 = \delta g_{IJ} = g_{IK} A^K{}_J + g_{KJ} A^K{}_I \quad \Leftrightarrow \quad A_{IJ} = -A_{JI} \,. \tag{5.20}$$

The problem is that in its other role as a quadratic form A_{IJ} needs to be symmetric rather than anti-symmetric in its indices. The metric will play a key role, but it cannot play this particular one!

But let us begin at the beginning. We split the complex numbers Z^α into real and imaginary parts and collect them into a $2N$-dimensional real vector X^I:

$$Z^\alpha = x^\alpha + \mathrm{i} y^\alpha \quad \to \quad X^I = \begin{bmatrix} x^\alpha \\ y^\alpha \end{bmatrix} . \tag{5.21}$$

The imaginary number i then turns into a matrix $J^I{}_J$:

$$J^I{}_J = \begin{bmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{bmatrix} \quad \Rightarrow \quad J^2 = -1 \,. \tag{5.22}$$

The existence of a matrix J that squares to minus one provides the $2N$-dimensional vector space with a *complex structure*. (This definition is equivalent to, and more straightforward than, the definition given in Section 3.3.)

[4] This has been urged by, among others, Dyson (1962), Gibbons and Pohle (1993) and Hughston (1995).

If we write everything out in real terms further structure emerges. The complex valued scalar product of two vectors becomes

$$\langle X|Y\rangle = X^I g_{IJ} Y^J + \mathrm{i} X^I \Omega_{IJ} Y^J \,, \tag{5.23}$$

where we had to introduce two new tensors

$$g_{IJ} = \begin{bmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{bmatrix} \qquad \Omega_{IJ} = \begin{bmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{bmatrix} . \tag{5.24}$$

The first is a metric tensor, the second is a symplectic form (in the sense of Section 3.4). Using the inverse of the symplectic form,

$$\Omega^{IK}\Omega_{KJ} = \delta^I_J \,, \tag{5.25}$$

we observe the key equation

$$J^I{}_J = \Omega^{IK} g_{KJ} \qquad \Leftrightarrow \qquad g_{IJ} = \Omega_{IK} J^K_J \,. \tag{5.26}$$

Thus, the complex, metric, and symplectic structures depend on each other. In fact we are redoing (for the case of a flat vector space) the definition of Kähler manifolds that we gave in Section 3.3.

We can use the symplectic structure to define Poisson brackets according to Eq. (3.73). But we will be interested in Poisson brackets of a rather special set of functions on our phase space. Let $O_{IJ} = O_{JI}$ be a symmetric tensor and define

$$\langle O\rangle \;\equiv\; X^I O_{IJ} X^J \,. \tag{5.27}$$

Such tensors can then play one of the two key roles for observables, namely to provide a map from the state vectors to the real numbers. But observables play another role too. They transform states into new states. According to the rules for how tensor indices may be contracted this means that they must have one index upstairs. We therefore need a way of raising and lowering indices on the observables-to-be. We have seen that the metric is not useful for this purpose, so we try the symplectic form:

$$\tilde{O}^I{}_J \;\equiv\; \Omega^{IK} O_{KJ} \,. \tag{5.28}$$

The observables now form an algebra because they can be multiplied together, $\tilde{O}_1\tilde{O}_2 = \tilde{O}^I_{1K}\tilde{O}^K_{2J}$. Note that, according to Eq. (5.26), the complex structure is the transformation matrix corresponding to the metric tensor: $J = \tilde{g}$.

We can work out some Poisson brackets. Using Eq. (3.73) we find that

$$\{X^I, \langle O\rangle\} = 2\tilde{O}^I{}_J X^J = \Omega^{IJ}\partial_J\langle O\rangle \tag{5.29}$$

$$\{\langle O_1\rangle, \langle O_2\rangle\} = 2\langle[\tilde{O}_1, \tilde{O}_2]\rangle \quad \text{where} \quad [\tilde{O}_1, \tilde{O}_2] \equiv \tilde{O}_1\tilde{O}_2 - \tilde{O}_2\tilde{O}_1 \,. \tag{5.30}$$

The Poisson bracket algebra is isomorphic to the commutator algebra of the matrices $\tilde{O}$ – even though there is no classical limit involved.

The observables as defined so far are too general. We know that we should confine ourselves to Hermitian operators, but why? What is missing is a requirement that is built into the complex formalism, namely that operators commute with the number i. In real terms this means that our observables should obey

$$[J, \tilde{O}] = 0 \quad \Leftrightarrow \quad O_{IJ} = \begin{bmatrix} m & -a \\ a & m \end{bmatrix}, \tag{5.31}$$

where m is a symmetric and a an anti-symmetric matrix. From now on only such observables are admitted. We call them *Hermitian*. There is a 'classical' way of looking at this condition. We impose the constraint that the states be normalized,

$$\langle g \rangle - 1 \equiv X^I g_{IJ} X^J - 1 = 0 . \tag{5.32}$$

According to the rules of constrained Hamiltonian dynamics we must then restrict our observables to be those that Poisson commute with this constraint, that is to say that Eq. (5.30) forces us to impose Eq. (5.31). Technically, this is a *moment map*.

Finally we select a specific observable, call it H, and write down Hamilton's equations:

$$\dot{X}^I = \{X^i, \langle H \rangle\} = 2\tilde{H}^I{}_J X^J = \Omega^{IJ} \partial_J \langle H \rangle . \tag{5.33}$$

Using Eqs. (5.21) and (5.31) reveals that, in complex notation, this is precisely

$$\dot{Z}^\alpha = \dot{x}^\alpha + \mathrm{i}\dot{y}^\alpha = 2\mathrm{i}(m + \mathrm{i}a)^\alpha{}_\beta (x^\beta + \mathrm{i}y^\beta) \equiv -\mathrm{i}H^\alpha{}_\beta Z^\beta . \tag{5.34}$$

This is Schrödinger's equation, with a complex matrix H that is Hermitian in the ordinary sense. In this way we learn that Schrödinger's equation is simply Hamilton's equation in disguise.[5] Note also that the constraint (5.32) generates a Hamiltonian flow

$$\dot{X}^I = \{X^I, \langle g \rangle - 1\} = 2J^I{}_J X^J \quad \Leftrightarrow \quad \dot{Z}^\alpha = 2\mathrm{i}Z^\alpha . \tag{5.35}$$

According to the theory of constrained Hamiltonian systems this is an unobservable *gauge motion*, as indeed it is. It just changes the unphysical phase of the complex state vector.

In a sense then quantum mechanics is classical mechanics on $\mathbb{C}\mathbf{P}^n$, formulated as a constrained Hamiltonian system on $\mathbb{C}^N$. But there is a key difference that we have so far glossed over. In classical mechanics arbitrary functions on phase space are used as observables, and each such function defines a Hamiltonian flow

[5] It appears that this interesting point was first made by Kibble (1979). Further studies were made by, among others, Gibbons (1992), Ashtekar and Schilling (1998) and Brody and Hughston (2001).

via Hamilton's equations. In our discussion we restricted ourselves at the outset to quadratic functions of the state vectors, leading to linear equations of motion. The reason for this restriction is that we require the Hamiltonian flow to leave not only the symplectic form but also the metric invariant. From Section 4.5 we know that the Killing vector fields of $\mathbb{C}\mathbf{P}^n$ are generated by Hermitian matrices, and this is precisely the set of Hamiltonian flows that we ended up with. The unitary group $U(N)$, or the isometry group that acts on Hilbert space, is a subgroup of the orthogonal group $SO(2N)$; instead of the unfortunate Eq. (5.20) we now have

$$g_{IK}\,\tilde{O}^K{}_J + g_{KJ}\,\tilde{O}^K{}_I = 0 \quad \Leftrightarrow \quad [J, \tilde{O}] = 0\,. \tag{5.36}$$

Hence the transformations effected by our observables preserve the metric. From this point of view then quantum mechanics appears as a special case of classical mechanics, where the phase space is special ($\mathbb{C}\mathbf{P}^n$) and the set of observables is restricted to be those that give rise to Hamiltonian vector fields that are also Killing vector fields. Unlike classical mechanics, quantum mechanics has a metric worth preserving.

The restriction that leads to quantum mechanics – that the Hamiltonian flow preserves the metric – can be imposed on any Kähler manifold, not just on $\mathbb{C}\mathbf{P}^n$. It has often been asked whether this could lead to a viable generalization of quantum mechanics. In fact several problems arise almost immediately. If the resulting formalism is to be of any interest one needs a reasonable large number of Hamiltonian Killing vector fields to work with. Some Kähler manifolds do not have any. If the dimension is fixed the maximal number of linearly independent Killing vector fields occurs if the Kähler space has constant holomorphic sectional curvature. But as noted in Section 4.5 if the space is also compact and simply connected this condition singles out $\mathbb{C}\mathbf{P}^n$ uniquely.

There are other reasons why $\mathbb{C}\mathbf{P}^n$ appears to be preferred over general Kähler spaces, notably the existence of the Segre embedding (Section 4.3). The fact that this embedding is always available makes it possible to treat composite systems in just the way that is peculiar to quantum mechanics, where the dimension of the state space of the composite system is surprisingly large. These dimensions are being used to keep track of *entanglement* – correlations between the subsystems that have no classical counterpart – and it is entanglement that gives quantum mechanics much of its special flavour (see Chapter 15). In some sense it is also the large dimension of the quantum mechanical state space that enables general symplectic manifolds to arise in the classical limit; although we have argued that quantum mechanics is (in a way) a special case of classical mechanics we must remember that quantum mechanics uses a much larger phase space than the classical theory – when the Hilbert space is finite dimensional the set of pure states of the corresponding classical theory is a finite set of points, while theories with

finite-dimensional classical phase spaces require an infinite-dimensional quantum 'phase space'.

5.5 Time reversals

The operation of time reversal is represented by an anti-unitary rather than a unitary operator, that is to say by an isometry that is a reflection rather than a rotation.[6] A complex conjugation is then involved, and it is again helpful to think of the state space as a $2N$-dimensional real manifold. In some situations we can see that the choice of a complex structure on the real vector space is connected to the direction of time. The point is that any matrix J that squares to -1 can be choosen as a complex structure on a real $2N$-dimensional vector space. Regarded as an $SO(2N)$ transformation such a matrix can be thought of as producing quarter turns in N suitably chosen orthogonal two-dimensional planes. But the Hamiltonian is an $SO(2N)$ matrix as well, and it is chosen to commute with J, so that it generates rotations in the same N planes. If the Hamiltonian is also positive definite it defines a sense of direction for all these rotations, and it is natural to require that the quarter turns effected by J takes place in the same directions. This then singles out a unique J. It follows that if we reverse the direction of time then the complex structure changes sign – which is why time reversal will be represented in quantum mechanics by an anti-unitary transformation.[7]

Wigner's theorem states that every isometry of $\mathbb{C}\mathbf{P}^n$ can be effected by a transformation of $\mathbb{C}^N$ that is either unitary or anti-unitary. The latter possibility arises only when discrete isometries are concerned (since the square of an anti-unitary transformation is unitary), but as we have seen it includes the interesting case of time reversal. Let us see what this looks like in real terms. Since all we require is that the Fubini–Study metric is preserved it is enough to ensure that the projective cross-ratio as derived from the Hermitian form is preserved. This will be so if the transformation is effected by a matrix which obeys either

$$UgU^T = g\,, \qquad U\Omega U^T = \Omega \tag{5.37}$$

(the unitary case) or

$$\Theta g\Theta^T = g\,, \qquad \Theta\Omega\Theta^T = -\Omega \tag{5.38}$$

(the anti-unitary case). Hence anti-unitary transformations are anti-canonical. Note by the way that the equation that defines an anti-unitary transformation, namely

$$\langle \Theta X | \Theta Y \rangle = \langle Y | X \rangle\,, \tag{5.39}$$

indeed follows from the definition (5.23) of the Hermitian form.

[6] The standard reference on time reversal is Wigner (1959). The usefulness of the real point of view was emphasized by Dyson (1962).

[7] See Ashtekar and Magnon (1975) and Gibbons and Pohle (1993) for elaborations of this point.

Two options are still open for Θ:

$$\Theta^2 = \pm 1 \,. \tag{5.40}$$

The choice of sign depends on the system. For spin systems the choice of sign is made in an interesting way. The physical interpretation requires that the angular momentum operators $\mathbf{J}$ are odd under time reversals, so that

$$\Theta \mathbf{J} + \mathbf{J}\Theta = 0 \,. \tag{5.41}$$

An anti-unitary operator can always be written in the form

$$\Theta = UK \,, \tag{5.42}$$

where K denotes the operation of complex conjugation and U is a unitary operator. In the standard representation of the angular momentum operators – Eqs. (B.4)–(B.5), where J_y is imaginary and the others real – the unitary operator U must obey

$$UJ_x + J_x U = UJ_z + J_z U = UJ_y - J_y U = 0 \,. \tag{5.43}$$

These equations determine U uniquely up to an irrelevant phase factor; the answer is

$$U = \begin{bmatrix} 0 & 0 & 0 & 0 & (-1)^N \\ \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} . \tag{5.44}$$

It then follows, with no ambiguity, that

$$N \text{ odd, } n \text{ even}; \quad \Rightarrow \quad \Theta^2 = 1 \qquad N \text{ even, } n \text{ odd} \quad \Rightarrow \quad \Theta^2 = -1 \,. \tag{5.45}$$

Time reversals therefore work quite differently depending on whether the spin is integer (even n) or half-integer (odd n). For even n there will be a subspace of states that are left invariant by time reversal. In fact they form the real projective space $\mathbb{R}\mathbf{P}^n$.

Odd n is a different matter. There can be no Θ invariant states. This means that $|P\rangle$ and $\Theta|P\rangle$ must be distinct points in $\mathbb{C}\mathbf{P}^{2n+1}$ and therefore they define a unique projective line L_P; in fact they are placed on opposite poles of this projective line, regarded as a 2-sphere. Moreover it is not hard to show that the resulting projective line transforms onto itself under Θ. Now consider a point $|Q\rangle$ not lying on L_P. Together with $\Theta|Q\rangle$ it determines another projective line L_Q. But L_P and L_Q cannot have a point in common (because if they had then their intersection would also contain the time-reversed point, and therefore they would coincide, contrary to

assumption). It follows that $\mathbb{C}\mathbf{P}^{2n+1}$ can be foliated by complex projective lines that never touch each other. It is a little hard to see this directly, but a similar statement is true for real projective spaces of odd dimension. Then the real projective lines that never touch each other are precisely the Clifford parallels in the Hopf fibration of $\mathbf{S}^{2n+1}$, restricted to $\mathbb{R}\mathbf{P}^{2n+1}$; see Section 3.5.

Keeping n odd, suppose that we restrict ourselves throughout to time-reversal invariant operators. This means that we decide to probe the system by means of time-reversal invariant observables only. Such observables obey

$$[\tilde{O}, \Theta] = 0 \,. \tag{5.46}$$

The theory when subject to a restriction of this kind is said to be obey a *superselection rule*. The first observation is that all such observables are degenerate, since all points on a time-reversal invariant 2-sphere have the same eigenvalues of $\tilde{O}$. (This is known as *Kramer's degeneracy*.) The second observation is that this restriction of the observables effectively restricts the state space to be the base manifold of the fibre bundle whose bundle space is $\mathbb{C}\mathbf{P}^{2n+1}$ and whose fibres are $\mathbb{C}\mathbf{P}^1$. This is analogous to how $\mathbb{C}\mathbf{P}^n$ arises as the state space when we restrict ourselves to observables that commute with the complex structure.

What is this new state space? To see this, let us go back to the real vector space of states that we analysed in Section 5.4. We know that

$$\Theta J + J\Theta = 0 \,, \tag{5.47}$$

where J is the complex structure. We define

$$\mathbf{i} = J \quad \mathbf{j} = \Theta \quad \mathbf{k} = J\Theta \,. \tag{5.48}$$

It follows that

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1 \,, \quad \mathbf{ij} = \mathbf{k} \quad \mathbf{jk} = \mathbf{i} \quad \mathbf{ki} = \mathbf{j} \,. \tag{5.49}$$

This is the algebra of *quaternions*.[8] Quaternions are best regarded as a natural generalization of complex numbers. A general quaternion can be written in terms of four real numbers as

$$\mathbf{q} = a_0 + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \,. \tag{5.50}$$

For every quaternion there is a conjugate quaternion

$$\bar{\mathbf{q}} = a_0 - a_1\mathbf{i} - a_2\mathbf{j} - a_3\mathbf{k} \tag{5.51}$$

[8] Denoted $\mathbb{H}$ after Hamilton who invented them. It was Dyson (1962) who first noted that quaternions arise naturally in standard quantum mechanics when time reversals are considered. For further elaborations see Avron, Sadun, Segert and Simon (1989) and Uhlmann (1996). For a review of just about every other aspect of quaternions, see Gürsey and Tze (1980).

and the quantity $|\mathbf{q}|^2 \equiv \bar{\mathbf{q}}\mathbf{q} = \mathbf{q}\bar{\mathbf{q}}$ is a real number called the absolute value squared of the quaternion. The absolute value of a product is the product of the absolute values of the factors. In general two quaternions do not commute. Like the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$ the quaternions form a normed associative *division algebra*, that is to say that they share the algebraic structure of the real number field except that multiplication is not commutative. Moreover these three are the only division algebras over the real numbers that exist – a statement known as *Frobenius' theorem*. (Multiplication of *octonions* is not associative.) For the moment the main point is that we can form quaternionic vector spaces and quaternionic projective spaces too, provided we agree that multiplication with scalars – that are now quaternions – always takes place from the left (say); we must be careful to observe this rule since the scalars no longer commute with each other.

Since we are dealing with $\mathbb{C}\mathbf{P}^{2n+2}$ we are working in a real vector space of dimension $4n+4$. We can regard this as a quaternionic vector space of quaternionic dimension $n+1$. For $n=1$ this means that an arbitrary vector is written as a two component object

$$\mathbf{Z}^a = (Z^0 + Z^3\mathbf{j},\, Z^2 + Z^1\mathbf{j})\,, \tag{5.52}$$

where the Z^α are complex numbers and we must remember that the imaginary unit $\mathbf{i}$ anticommutes with $\mathbf{j}$. It is straightforward to check that time reversal as defined above is effected by a left multiplication of the vector by $\mathbf{j}$. Since we have imposed the superselection rule that multiplication of the vector with an overall quaternion leaves the state unchanged we can now form the quaternionic projective space $\mathbb{H}\mathbf{P}^1$. Topologically this is $\mathbb{R}^4 + \infty = \mathbf{S}^4$.

Everything works out in such close analogy to the real and complex projective spaces that we need not give the details. Let us just quote the salient points: starting from a real vector space of dimension $4n+4$ we first normalize the vectors to unity, and then we impose the superselection rules

$$[\tilde{O}, \mathbf{i}] = 0 \;\Rightarrow\; \mathbb{C}\mathbf{P}^n \qquad [\tilde{O}, \mathbf{j}] = 0 \;\Rightarrow\; \mathbb{H}\mathbf{P}^n\,. \tag{5.53}$$

This is known as the Hopf fibration of the $4n+3$ sphere. In two steps $\mathbf{S}^{4n+3} \to \mathbb{C}\mathbf{P}^{2n+1}$ with fibre $\mathbf{S}^1$ and $\mathbb{C}\mathbf{P}^{2n+1} \to \mathbb{H}\mathbf{P}^n$ with fibre $\mathbf{S}^2$, or in one step $\mathbf{S}^{4n+3} \to \mathbb{H}\mathbf{P}^n$ with fibre $\mathbf{S}^3$. The base spaces inherit natural metrics from this construction. In the case of $\mathbb{H}\mathbf{P}^1$ this happens to be the round metric of the 4-sphere. When equipped with their natural metrics the projective spaces $\mathbb{R}\mathbf{P}^n$, $\mathbb{C}\mathbf{P}^n$ and $\mathbb{H}\mathbf{P}^n$ share some crucial features, notably that all their geodesics are closed.

Occasionally it is suggested that quaternionic quantum mechanics can offer an alternative to the standard formalism, but we see that the quaternionic projective space has a role to play as a state space also within the latter.

Table 5.1. *Classical and quantum states: a comparison*

Framework	States	Positivity	Normalization
i) probability vectors	$\vec{p} \in \mathbb{R}^N$	$p_i \geq 0$	$\sum_i p_i = 1$
ii) probability measures	$\mu \in \Omega(X)$	$\mu(Y) \geq 0$	$\int_X \mathrm{d}\mu(x) = 1$
iii) density operators	$\rho \in \mathcal{M}^{(N)}$	$\rho \geq 0$	$\mathrm{Tr}\rho = 1$
iv) states on C^* algebra	ω on $\mathcal{A}$	$\omega(x^*x) \geq 0$	$\omega(\mathbb{1}) = 1$

5.6 Classical and quantum states: a unified approach

Some remarks on how quantum mechanics can be axiomatized may not be out of place – even if they will be very incomplete. There is a choice whether observables or states should be regarded as primary. The same choice occurs when classical mechanics is defined, for the same reason: observables and states cooperate in producing the real numbers that constitute the predictions of the theory. These real numbers are the result of real-valued linear maps applied to a vector space, and the space of such maps is itself a vector space. The end result is a duality between observables and states, so that either can be taken as primary. Just as one can take either the hen or the egg as primary. Let us try to summarize similarities and differences between states in classical and quantum mechanics, or perhaps more accurately in classical and quantum probability theory. A classical state is described by a probability vector or by an element from the set of all probability measures on the classical phase space. The space of quantum states consists of density matrices. What features are common for these different frameworks?

Let us start by considering a convex cone V^+ in a vector space V. Elements x belonging to V^+ will be called positive. Let $e : V \to \mathbb{R}$ be a linear functional on V. The space of all states is then defined as a cross section of the positive cone, consisting of elements that obey the normalization condition $e(x) = 1$. As shown in Table 5.1 both classical and quantum states fit well into this scheme. Let us be a little bit more explicit: in the classical case the vector space has arbitrary dimension and V^+ consists of positive vectors. The functional e is given by the l_1-norm of the vector $\vec{p}$, or by the integral over phase space in the continuous case; in the table Y denotes an arbitrary subset of the classical phase space X. This is the case discussed at length in Chapter 2. In the quantum case the vector space is the space of Hermitian matrices, V^+ is the space of positive operators, and the functional e is the trace. Classical probability theory can be obtained from quantum mechanics through a restriction to diagonal matrices.

The similarities between the classical and quantum cases are made very transparent in Segal's axiomatic formulation.[9] Then the axioms are about a set $\mathcal{A}$ of objects called *observables*. The axioms determine what an observable is – reflection on what one observes in experiments comes in at a later stage. The set $\mathcal{A}$ has a distinguished subset $\mathcal{M}$ whose elements ρ are called *states*. We assume

I) The observables form a real linear space.
II) The observables form a commutative algebra.
III) There exists a bilinear map from $\mathcal{M} \times \mathcal{A}$ to the real numbers.
IV) The observables form a Lie algebra.

We see that a state can be thought of as a special kind of linear map from the observables to the real numbers – and this is characteristic of Segal's approach, even if at first sight it appears to be taking things in the wrong order!

These axioms do not characterize quantum mechanics completely. Indeed they are obeyed by both classical and quantum mechanics. (As far as we are aware no one knows if there is a third kind of mechanics also obeying Segal's axioms.) In classical mechanics states are probability distributions on phase space, observables are general functions on phase space, and their algebras are given by pointwise multiplication (axiom II) and by Poisson brackets (axiom IV). This means that the extreme points of the convex set form a symplectic manifold; unless rather peculiar measures are taken (Wootters, 1987) this is a continuous manifold and the set of all states becomes infinite dimensional.

In quantum mechanics states are density matrices acting on some Hilbert space. The observables A are Hermitian matrices – since we assume that the Hilbert space is finite dimensional we need no further precision. The algebra in axiom II is given by a commutative but non-associative *Jordan algebra*, to be discussed in Section 8.6. The Lie bracket of axiom IV is given by the imaginary unit i times the commutator of the observables. The bilinear map required by axiom III is the trace of the product of ρ and A; this is the quantum mechanical expectation value $\langle A \rangle = \mathrm{Tr}\rho A$. When we think of states as real-valued maps from the observables, we write $\rho(A) = \mathrm{Tr}\rho A$. Note that $\mathrm{Tr}\rho = 1$ then becomes the statement that $\rho(\mathbb{1}) = 1$. In both classical and quantum mechanics the states are defined so that they form a convex set, and it becomes important to identify the pure states. In quantum mechanics they are projection operators onto rays in a complex Hilbert space, and can therefore also be thought of as equivalence classes of vectors.

[9] For further study we recommend the original paper by Segal (1947). For a simplified account, see Currie, Jordan and Sudarshan (1963).

Segal's axioms are rather even handed in their choice between observables and states. In the C^*-algebra approach the algebra of observables occupy the centre of the stage. The states are defined as positive and normalized linear functionals ω on a suitable algebra $\mathcal{A}$. Both classical and quantum mechanics fit into this mould, since the states may also be viewed as functionals. In case (i), the state $\vec{p}$ maps the vector $\vec{y} \in \mathbb{R}^N$ to $\vec{p} \cdot \vec{y} \in \mathbb{R}$. In case (ii), the state μ acts on an arbitrary integrable function, $\mu : f(X) \to \int_X f(x)\,\mathrm{d}\mu(x)$, while in case (iii), the state ρ acts on an Hermitian operator, $\rho : A \to \mathrm{Tr}A\rho$. The most essential difference between the classical cases (i) or (ii) and the quantum case (iii) concerns commutativity: classical observables commute, but this is not so in quantum mechanics.[10]

Although we do not intend to pursue the algebraic approach, we observe that once states are defined as linear functionals on a suitable algebra, the set of all states will be a convex set. We are then inevitably drawn to the converse question: what properties must a convex set have, if it is to arise as the state space of an operator algebra? The answer to this question is known, but it is not a short answer and this is not the place to give it.[11]

To demonstrate further similarities between the classical and the quantum cases let us replace the sum in Eq. (5.3) by an integral over the space of pure states,

$$\hat{\rho} \;=\; \int_{\mathbb{C}P^{N-1}} \rho^s(\psi)\,|\psi\rangle\langle\psi|\;\mathrm{d}\Omega(\psi)\;. \tag{5.54}$$

Here $\mathrm{d}\Omega(\psi)$ stands for the Fubini–Study measure (see Section 4.7), while $\rho^s(\psi)$ is a normalized probability distribution, and we decorated the density matrix itself with a hat – a notation that we use occasionally to distinguish operators. Thus a state may be considered as a probability measure on the space of pure states (Mielnik, 1969). If $\rho^s(\psi)$ is a δ-function then Eq. (5.54) represents a pure state. The unitary time evolution of the density matrix,

$$\frac{\mathrm{d}\hat{\rho}}{\mathrm{d}t} \;=\; \mathrm{i}\,[\hat{\rho},\, H]\;, \tag{5.55}$$

is equivalent to the Liouville equation for the distribution function ρ^s,

$$\frac{\mathrm{d}\rho^s}{\mathrm{d}t} \;=\; \{\langle H\rangle,\, \rho^s\}\;. \tag{5.56}$$

So far classical and quantum mechanics look similar.

But there are differences too. One key difference concerns the pure states. The classical set of pure states can be discrete. The set of pure quantum states is always continuous, and indeed it is always a symplectic manifold. In particular there always

[10] For an account of the algebraic approach to quantum mechanics consult Emch (1972). See Section 8.6 for a glimpse.

[11] We refer to Alfsen and Shultz (2001) and Alfsen and Shultz (2003). Their answer takes 843 pages!

exists a continuous reversible transformation joining any two pure quantum states.[12] A related fact is that the representation (5.54) is not unique. There are many ρ^s for each $\hat{\rho}$; quantum theory identifies two probability distributions if their barycentres are equal. The result is a kind of projection of the infinite-dimensional space of probability distributions ρ^s into the compact (N^2-1)-dimensional space $\mathcal{M}^{(N)}$ of density matrices. This is completely foreign to the classical theory, and it happens because of the severe restriction on allowed observables that is imposed in quantum mechanics. It is similar to the situation encountered in the theory of colours: the detector system of the eye causes a projection of the infinite-dimensional space of all possible spectral distributions down to the three-dimensional cone of colours.

Note that we just studied a *hidden variable theory*. By definition, this is a classical model from which quantum mechanics arises through some projection reminiscent of the projection in colour theory. Indeed we studied this model in detail when we developed quantum mechanics as a form of classical mechanics in Section 5.4. Quantum mechanics arose once we declared that most classical variables were hidden. But for various reasons this model is not the model that people dream about.[13]

To end our slightly unsystematic tour of fundamentals, we point out that the amount of choice that one has in choosing how probability distributions emerge from the formalism is severely limited. There is a theorem that comes into play once it has been decided that we are in the Hilbert space framework, and try to assign a probability distribution over its subspaces. More precisely the probabilities are functions on the (real or complex) projective space, that is of the rays in Hilbert space. The assumptions that we have to make are the following:

- **Normalization**. The elements $|e_i\rangle$ of every orthonormal basis are assigned probabilities p_i such that

$$\sum_{i=1}^{N} p_i |e_i\rangle\langle e_i| \;=\; \mathbb{1}\,. \tag{5.57}$$

- **Non-contextuality**. Every vector is an element of many orthonormal bases. The probability of its ray is independent of how the remaining vectors of the basis are chosen.

That is all (although it may be remarked that the non-contextuality assumption is not really a self-evident one). Nevertheless we now have the following:

[12] An axiomatic approach to quantum mechanics developed by Hardy (n.d.) takes this observation as its starting point.

[13] For an interesting discussion of hidden variables from this point of view, see Holevo (2001).

Theorem 5.1 (Gleason's) *Under the conditions stated, and provided the dimension N of the Hilbert space obeys $N > 2$, there exists a density matrix ρ such that $p_i = Tr\rho\,|e_i\rangle\langle e_i|$.*

So the density matrix and the trace rule (5.6) have been forced upon us – probability can enter the picture only in precisely the way that it does enter in conventional quantum mechanics. The remarkable thing is that no further assumptions are needed to prove the theorem – it is proved, not assumed, that the probability distribution is a continuous function on the projective space.[14]

Problem

Problem 5.1 Two pure states sit on the Bloch sphere separated by an angle θ. Choose an operator A whose eigenstates sit on the same great circle as the two pure states; the diameter defined by A makes an angle θ_A with the nearest of the pure states. Compute the Bhattacharyya distance between the two states for the measurement associated with A.

[14] The theorem is due to Gleason (1957). Its proof is famous for being difficult; a version that is comparatively easy to follow was given by Pitowsky (1998).

6

Coherent states and group actions

Coherent states are the natural language of quantum theory.
John R. Klauder

In this chapter we study how groups act on the space of pure states, and especially the coherent states that can be associated to the group.

6.1 Canonical coherent states

The term 'coherent state' means different things to different people, but all agree that the *canonical coherent states*[1] form the canonical example, and with this we begin – even though it is somewhat against our rules in this book because now the Hilbert space is infinite dimensional. A key feature is that there is a group known as the *Heisenberg–Weyl group* that acts irreducibly on our Hilbert space.[2] The coherent states form a subset of states that can be reached from a special *reference* state by means of transformations belonging to the group. The group theoretical way of looking at things has the advantage that generalized coherent states can be defined in an analogous way as soon as one has a group acting irreducibly on a Hilbert space.

But let us begin at the beginning. The *Heisenberg algebra* of the operators $\hat{q}$ and $\hat{p}$ together with the unit operator $\mathbb{1}$ is defined by

$$[\hat{q}, \hat{p}] \;=\; \mathrm{i}\hbar\,\mathbb{1}\;. \tag{6.1}$$

[1] Also known as Glauber states. They were first described by Schrödinger (1926a), and then, after an interval, by Glauber (1963). For their use in quantum optics see Klauder and Sudarshan (1968) and Mandel and Wolf (1995).

[2] This is the point of view of Perelomov and Gilmore, the inventors of generalized coherent states. Useful reviews include Perelomov (1977), Zhang, Feng and Gilmore (1990) and Ali, Antoine and Gazeau (2000); a valuable reprint collection is Klauder and Skagerstam (1985).

It acts on the infinite-dimensional Hilbert space $\mathcal{H}_\infty$ of square integrable functions on the real line. In this chapter we will have hats on all the operators (because we will see much of the c-numbers q and p as well). Planck's constant $\hbar$ is written explicitly because in this chapter we will be interested in the limit in which $\hbar$ cannot be distinguished from zero. It is a dimensionless number because it is assumed that we have fixed units of length and momentum and use these to rescale the operators $\hat{q}$ and $\hat{p}$ so that they are dimensionless as well. If the units are chosen so that the measurement precision relative to this scale is of order unity, and if $\hbar$ is very small, then $\hbar$ can be safely set to zero. In SI units $\hbar = 1.054 \cdot 10^{-34}$ joule seconds. Here our policy is to set $\hbar = 1$, which means that classical behaviour may set in when measurements can distinguish only points with a very large separation in phase space.

Equation (6.1) is the Lie algebra of a group. First we recall the Baker–Hausdorff formula

$$\mathrm{e}^{\hat{A}}\mathrm{e}^{\hat{B}} = \mathrm{e}^{\frac{1}{2}[\hat{A},\hat{B}]}\mathrm{e}^{\hat{A}+\hat{B}} = \mathrm{e}^{[\hat{A},\hat{B}]}\mathrm{e}^{\hat{B}}\mathrm{e}^{\hat{A}}\,, \tag{6.2}$$

which is valid whenever $[\hat{A}, \hat{B}]$ commutes with $\hat{A}$ and $\hat{B}$. Thus equipped we form the unitary group elements

$$\hat{U}(q, p) \;\equiv\; \mathrm{e}^{\mathrm{i}(p\hat{q}-q\hat{p})}\,. \tag{6.3}$$

To find out what group they belong to we use the Baker–Hausdorff formula to find that

$$\hat{U}(q_1, p_1)\hat{U}(q_2, p_2) \;=\; \mathrm{e}^{-\mathrm{i}(q_1p_2-p_1q_2)}\,\hat{U}(q_2, p_2)\,\hat{U}(q_1, p_1)\,. \tag{6.4}$$

This equation effectively defines a faithful representation of the *Heisenberg group*.[3] This group acts irreducibly on the Hilbert space $\mathcal{H}_\infty$ (and this happens to be the only unitary and irreducible representation that exists, although this fact is incidental to our purposes). Since the phase factor is irrelevant in the underlying projective space of states it is also a *projective representation* of the Abelian group of translations in two dimensions.

Now we can form creation and annihilation operators in the standard way,

$$\hat{a} = \frac{1}{\sqrt{2}}(\hat{q} + \mathrm{i}\hat{p})\,, \qquad \hat{a}^\dagger = \frac{1}{\sqrt{2}}(\hat{q} - \mathrm{i}\hat{p})\,, \qquad [a, a^\dagger] = 1\,, \tag{6.5}$$

we can define the vacuum state $|0\rangle$ as the state that is annihilated by $\hat{a}$, and finally we can consider the two-dimensional manifold of states of the form

$$|q, p\rangle \;=\; \hat{U}(q, p)|0\rangle\,. \tag{6.6}$$

[3] This is really a three-dimensional group, including a phase factor. The name *Weyl group* is more appropriate, but too many things are named after Weyl already. The Heisenberg algebra was discovered by Born and is engraved on his tombstone.

These are the canonical coherent states with the vacuum state serving as the reference state, and q and p serve as coordinates on the space of coherent states. Our question is: why are coherent states interesting? To answer it we must get to know them better.

Two important facts follow immediately from the irreducibility of the representation. First, the coherent states are complete in the sense that any state can be obtained by superposing coherent states. Indeed they form an *overcomplete* set because they are much more numerous than the elements of an orthonormal set would be – hence they are not orthogonal and do overlap. Second, we have the resolution of the identity

$$\frac{1}{2\pi}\int \mathrm{d}q\,\mathrm{d}p\,|q,p\rangle\langle q,p| \;=\; \mathbb{1}\;. \tag{6.7}$$

The overall numerical factor must be calculated, but otherwise this equation follows immediately from the easily ascertained fact that the operator on the left-hand side commutes with $\hat{U}(q,p)$; because the representation of the group is irreducible Schur's lemma implies that the operator must be proportional to the identity. Resolutions of identity will take on added importance in Section 10.1, where they will be referred to as 'POVMs'.

The coherent states form a Kähler manifold (see Section 3.3). To see this we first bring in a connection to complex analyticity that is very helpful in calculations. We trade $\hat{q}$ and $\hat{p}$ for the creation and annihilation operators and define the complex coordinate

$$z = \frac{1}{\sqrt{2}}(q+\mathrm{i}p)\;. \tag{6.8}$$

With some help from the Baker–Hausdorff formula, the submanifold of coherent states becomes

$$|q,p\rangle = |z\rangle = \mathrm{e}^{z\hat{a}^\dagger - z^*\hat{a}}|0\rangle = \mathrm{e}^{-|z|^2/2}\sum_{n=0}^{\infty}\frac{z^n}{\sqrt{n!}}|n\rangle\;. \tag{6.9}$$

We assume that the reader is familiar with the orthonormal basis spanned by the *number* or *Fock states* $|n\rangle$ (see, for example, Leonhardt, 1997).

We have reached a convenient platform from which to prove a number of elementary facts about coherent states. We can check that the states $|z\rangle$ are eigenstates of the annihilation operator:

$$\hat{a}|z\rangle \;=\; z|z\rangle\;. \tag{6.10}$$

Their usefulness in quantum optics has to do with this fact since light is usually measured by absorption of photons. In fact a high quality laser produces

coherent states. A low quality laser produces a statistical mixture of coherent states – producing anything else is rather more difficult.

In x-space a coherent state wave function labelled by q and p is

$$\psi(x; q, p) = \langle x|q, p\rangle = \pi^{-1/4} \mathrm{e}^{-\mathrm{i}pq/2 + \mathrm{i}px - (x-q)^2/2} . \tag{6.11}$$

The shape is a Gaussian centred at $x = q$. The overlap between two coherent states is

$$\langle q_2, p_2|q_1, p_1\rangle = \mathrm{e}^{-\mathrm{i}(q_1 p_2 - p_1 q_2)/2} \mathrm{e}^{-[(q_2-q_1)^2 + (p_2-p_1)^2]/4} . \tag{6.12}$$

It shrinks rapidly as the coordinate distance between the two points increases.

Let us now think about the space of coherent states itself. The choice of labels (q and p) is not accidental because we intend to regard the space of coherent states as being in some sense an embedding of the phase space of the classical system – whose quantum version we are studying – into the space of quantum states. Certainly the coherent states form a two-dimensional space embedded in the infinite-dimensional space of quantum states and it will therefore inherit both a metric and a symplectic form from the latter. We know that the absolute value of the overlap is the cosine of the Fubini–Study distance D_{FS} between the two states (see Section 5.3), and for infinitesimally nearby coherent states we can read off the intrinsic metric $\mathrm{d}s^2$ on the embedded submanifold. From Eq. (6.12) we see that the metric on the space of coherent states is

$$\mathrm{d}s^2 = \mathrm{d}z\mathrm{d}\bar{z} = \frac{1}{2}(\mathrm{d}q^2 + \mathrm{d}p^2) . \tag{6.13}$$

It is a flat space – indeed a flat vector space since the vacuum state forms a natural point of origin. From the phase of the overlap we can read off the symplectic form induced by the embedding on the submanifold of coherent states. It is non-degenerate:

$$\Omega = \mathrm{i}\mathrm{d}z \wedge \mathrm{d}\bar{z} = \mathrm{d}q \wedge \mathrm{d}p . \tag{6.14}$$

It is the non-degenerate symplectic form that enables us to write down Poisson brackets and think of the space of coherent states as a phase space, isomorphic to the ordinary classical phase space spanned by q and p. The metric and the symplectic form are related to each other in precisely the way that is required for a Kähler manifold – although in a classical phase space the metric plays no particular role in the formalism. It is clearly tempting to try to argue that in some sense the space of coherent states is the classical phase space, embedded in the state space of its quantum version. A point in the classical phase space corresponds to a coherent state. The metric on phase space has a role to play here because Eq. (6.12) allows us to say that if the distance between the two points is large as measured by the metric,

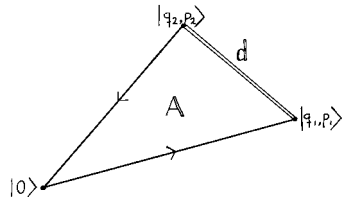


Figure 6.1. The overlap of two coherent states is determined by geometry: its modulus by the Euclidean distance d between the states and its phase by the (oriented) Euclidean area A of the triangle formed by the two states together with the reference state.

then the overlap between the two coherent states is small so that they interfere very little. Classical behaviour is clearly setting in; we will elaborate this point later on. Meanwhile we observe that the overlap can be written

$$\langle q_2, p_2|q_1, p_1\rangle = \mathrm{e}^{-\mathrm{i}2(\text{area of triangle})}\,\mathrm{e}^{-\frac{1}{2}(\text{distance})^2}\ , \tag{6.15}$$

where the triangle is defined by the two states together with the reference state. This is clearly reminiscent of the properties of geodesic triangles in $\mathbb{C}\mathbf{P}^n$ that we studied in Section 4.8, but the present triangle lies within the space of coherent states itself. The reason why the phase equals an area is the same in both cases, namely that geodesics in $\mathbb{C}\mathbf{P}^n$ as well as geodesics within the embedded subspace of canonical coherent states share the property of being null phase curves (in the sense of Section 4.8) (Rabei, Arvind, Simon and Mukunda, 1990).

There is a large class of observables – self adjoint operators on Hilbert space – that can be associated with functions on phase space in a natural way. In general we define the *covariant symbol*[4] of the operator $\hat{A}$ as

$$A(q, p) \ = \ \langle q, p|\hat{A}|q, p\rangle\ . \tag{6.16}$$

This is a function on the space of coherent states, that is on the would-be classical phase space. It is easy to compute the symbol of any operator that can be expressed as a polynomial in the creation and annihilation operators. In particular

$$\langle q, p|\hat{q}|q, p\rangle = q \qquad \langle q, p|\hat{q}^2|q, p\rangle = q^2 + \frac{1}{2} \tag{6.17}$$

(and similarly for $\hat{p}$). This implies that the variance, when the state is coherent, is

$$(\Delta q)^2 = \langle\hat{q}^2\rangle - \langle\hat{q}\rangle^2 = \frac{1}{2} \tag{6.18}$$

and similarly for $(\Delta p)^2$, so it follows that $\Delta q\,\Delta p = 1/2$; in words, the coherent states are states of minimal uncertainty in the sense that they saturate Heisenberg's

[4] The contravariant symbol will appear presently.

inequality. This confirms our suspicion that there is 'something classical' about the coherent states. Actually the coherent states are not the only states that saturate the uncertainty relation; the coherent states are singled out by the extra requirement that $\Delta q = \Delta p$.

We have not yet given a complete answer to the question why coherent states are interesting – to get such an answer it is necessary to see how they can be used in some interesting application – but we do have enough hints. Let us try to gather together some key features:

- The coherent states form a complete set and there is a resolution of unity.
- There is a one-to-one mapping of a classical phase space onto the space of coherent states.
- There is an interesting set of observables whose expectation values in a coherent state match the values of the corresponding classical observables.
- The coherent states saturate an uncertainty relation and are in this sense as classical as they can be.

These are properties that we want any set of states to have if they are to be called coherent states. The generalized coherent states defined by Perelomov (1977) and Zhang et al. (1990) do share these properties. The basic idea is to identify a group G that acts irreducibly on the Hilbert space and define the coherent states as a particular *orbit* of the group. If the orbit is chosen suitably the resulting space of coherent states is a Kähler manifold. We will see how later; meanwhile let us observe that there are many alternative ways to define generalized coherent states. Sometimes any Kähler submanifold of $\mathbb{C}\mathbf{P}^n$ is referred to as coherent, regardless of whether there is a group in the game or not. Other times the group is there but the coherent states are not required to form a Kähler space. Here we require both simply because it is interesting to do so. Certainly a *coherence group* formed by all the observables of interest for some particular problem arises in many applications, and the irreducible representation is just the minimal Hilbert space that allows us to describe the action of that group. Note that the coherence group is basically of kinematical origin; it is not supposed to be a symmetry group.

6.2 Quasi-probability distributions on the plane

The covariant symbol of an operator, as defined in Eq. (6.16), gives us the means to associate a function on phase space to any 'observable' in quantum mechanics. In classical physics an observable is precisely a function on phase space, and moreover the classical state of the system is represented by a particular such function – namely by a probability distribution on phase space. Curiously similar schemes can work in quantum mechanics too. It is interesting to think of music in this connection. Music, as produced by orchestras and sold by record companies, is a certain function of

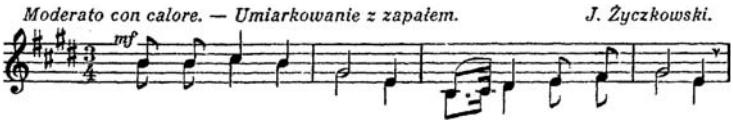


Figure 6.2. Music scores resemble Wigner functions.

time. But this is not how it is described by composers, who think of music[5] as a function of both time and frequency. Like the classical physicist, the composer can afford to ignore the limitations imposed by the uncertainty relation that holds in Fourier theory. The various quasi-probability distributions that we will discuss in this section are musical scores for quantum mechanics, and, remarkably, nothing is lost in this transcription. For the classical phase space we use the coordinates q and p. They may denote the position and momentum of a particle, but they may also define an electromagnetic field through its 'quadratures'.[6]

Quantum mechanics does provide a function on phase space that gives the probability distribution for obtaining the value q in a measurement associated to the special operator $\hat{q}$. This is just the familiar probability density for a pure state, or $\langle q|\hat{\rho}|q\rangle$ for a general mixed state. We ask for a function $W(q, p)$ such that this probability distribution can be recovered as a marginal distribution, in the sense that

$$\langle q|\hat{\rho}|q\rangle \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}p\; W(q, p)\,. \tag{6.19}$$

This can be rewritten as an equation for the probability to find that the value q lies in an infinite strip bounded by the parallel lines $q = q_1$ and $q = q_2$, namely

$$P(q_1 \le q \le q_2) \;=\; \frac{1}{2\pi}\int_{strip} \mathrm{d}q\,\mathrm{d}p\; W(q, p)\,. \tag{6.20}$$

In classical physics (as sketched in Section 5.6) we would go on to demand that the probability to find that the values of q and p are confined to an arbitrary phase space region Ω is given by the integral of $W(q, p)$ over Ω, and we would end up with a function $W(q, p)$ that serves as a joint probability distribution for both variables. This cannot be done in quantum mechanics. But it turns out that a requirement somewhat in-between Eq. (6.20) and the classical requirement can be met, and indeed uniquely determines the function $W(q, p)$, although the function will not qualify as a joint probability distribution because it may fail to be positive.

[5] The sample shown is by Józef Życzkowski, 1895–1967.

[6] A good general reference for this section is the survey of phase space methods given in the beautifully illustrated book by Leonhardt (1997). Note that factors of 2π are distributed differently throughout the literature.

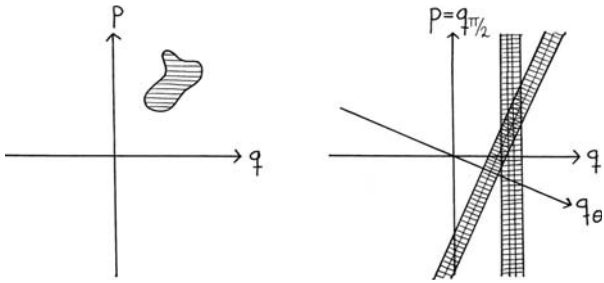


Figure 6.3. Left: in classical mechanics there is a phase space density such that we obtain the probability that p and q is confined to any region in phase space by integrating the density over that region. Right: in quantum mechanics we obtain the probability that p and q is confined to any infinite strip by integrating the Wigner function over that strip.

For this purpose consider the operators

$$\hat{q}_\theta = \hat{q}\cos\theta + \hat{p}\sin\theta \qquad \hat{p}_\theta = -\hat{q}\sin\theta + \hat{p}\cos\theta\ . \tag{6.21}$$

Note that $\hat{q}_\theta$ may be set equal to either $\hat{q}$ or $\hat{p}$ through a choice of the phase θ, and also that the commutator is independent of θ. The eigenvalues of $\hat{q}_\theta$ are denoted by q_θ. These operators gain in interest when one learns that the phase can actually be controlled in quantum optics experiments. We now have the following theorem (Bertrand and Bertrand, 1987):

Theorem 6.1 (Bertrand and Bertrand's) *The function $W(q, p)$ is uniquely determined by the requirement that*

$$\langle q_\theta|\hat{\rho}|q_\theta\rangle \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}p_\theta\; W_\theta(q_\theta, p_\theta) \tag{6.22}$$

for all values of θ. Here $W_\theta(q_\theta, p_\theta) = W\big(q(q_\theta, p_\theta), p(q_\theta, p_\theta)\big)$.

That is to say, as explained in Figure 6.3, we now require that all infinite strips are treated on the same footing. We will not prove uniqueness here, but we will see that the *Wigner function* $W(q, p)$ has the stated property.

A convenient definition of the Wigner function is

$$W(q, p) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}u\,\mathrm{d}v\; \widetilde{W}(u, v)\,\mathrm{e}^{\mathrm{i}uq+\mathrm{i}vp}\ , \tag{6.23}$$

where the *characteristic function* is a 'quantum Fourier transformation' of the density matrix,

$$\widetilde{W}(u, v) \;=\; \mathrm{Tr}\hat{\rho}\,\mathrm{e}^{-\mathrm{i}u\hat{q}-\mathrm{i}v\hat{p}}\ . \tag{6.24}$$

To express this in the q-representation we use the Baker–Hausdorff formula and insert a resolution of unity to deduce that

$$\mathrm{e}^{-\mathrm{i}u\hat{q}-\mathrm{i}v\hat{p}} = \mathrm{e}^{\frac{\mathrm{i}}{2}uv}\mathrm{e}^{-\mathrm{i}u\hat{q}}\mathrm{e}^{-\mathrm{i}v\hat{p}} = \int_{-\infty}^{\infty} \mathrm{d}q\ \mathrm{e}^{-\mathrm{i}uq}|q+\frac{v}{2}\rangle\langle q-\frac{v}{2}| \,. \tag{6.25}$$

We can just as well work with the operators in Eq. (6.21) and express everything in the q_θ-representation. We assume that this has been done – effectively it just means that we add a subscript θ to the eigenvalues. We then arrive, in a few steps, at Wigner's formula[7]

$$W_\theta(q_\theta, p_\theta) = \int_{-\infty}^{\infty} \mathrm{d}x\ \langle q_\theta - \frac{x}{2}|\hat{\rho}|q_\theta + \frac{x}{2}\rangle\ \mathrm{e}^{\mathrm{i}xp_\theta} \,. \tag{6.26}$$

Integration over p_θ immediately yields Eq. (6.22).

It is interesting to play with the definition a little. Let us look for a *phase point operator* $\hat{A}_{qp}$ such that

$$W(q,p) = \mathrm{Tr}\hat{\rho}\hat{A}_{qp} = \int_{-\infty}^{\infty}\mathrm{d}q' \int_{-\infty}^{\infty}\mathrm{d}q'\ \langle q'|\hat{\rho}|q'\rangle\ \langle q'|\hat{A}_{qp}|q'\rangle \,. \tag{6.27}$$

That is to say that we will define the phase point operator through its matrix elements in the q-representation. The solution is

$$\langle q'|\hat{A}_{qp}|q'\rangle \ = \ \delta\Big(q - \frac{q'+q'}{2}\Big)\mathrm{e}^{\mathrm{i}(q'-q')p} \,. \tag{6.28}$$

This permits us to write the density matrix in terms of the Wigner function as

$$\hat{\rho} \ = \ \frac{1}{2\pi}\int \mathrm{d}q\,\mathrm{d}p\ W(q,p)\,\hat{A}_{qp} \tag{6.29}$$

(as one can check by looking at the matrix elements). Hence the fact that the density matrix and the Wigner function determine each other has been made manifest. This is interesting because, given an ensemble of identically prepared systems, the various marginal probability distributions tied to the rotated position (or quadrature) operators in Eq. (6.21) can be measured – or at least a sufficient number of them can, for selected values of the phase θ – and then the Wigner function can be reconstructed using an appropriate (inverse) *Radon transformation*. This is known as *quantum state tomography* and is actually being performed in laboratories.[8]

The Wigner function has some remarkable properties. First of all it is clear that we can associate a function W_A to an arbitrary operator $\hat{A}$ if we replace the operator

[7] Wigner (1932) originally introduced this formula, with $\theta = 0$, as 'the simplest expression' that he (and Szilard) could think of.

[8] The Wigner function was first measured experimentally by Smithey, Beck, Raymer and Faridani (1993) in 1993, and negative values of W were reported soon after. See Nogues, Rauschenbeutel, Osnaghi, Bertet, Brune, Raimond, Haroche, Lutterbach and Davidovich (2000) and references therein.

$\hat{\rho}$ by the operator $\hat{A}$ in Eq. (6.24). This is known as the *Weyl symbol* of the operator and is very important in mathematics. If no subscript is used it is understood that the operator $\hat{\rho}$ is meant, that is $W \equiv W_\rho$. Now it is straightforward to show that the expectation value for measurements associated to the operator $\hat{A}$ is

$$\langle \hat{A} \rangle_{\hat{\rho}} \equiv \mathrm{Tr}\hat{\rho}\hat{A} = \frac{1}{2\pi} \int \mathrm{d}q\, \mathrm{d}p\; W_A(q, p) W(q, p) \,. \tag{6.30}$$

This is the *overlap formula*. Thus the formula for computing expectation values is the same as in classical physics: integrate the function corresponding to the observable against the state distribution function over the classical phase space. Classical and quantum mechanics are nevertheless very different. To see this, choose two pure states $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$ and form the corresponding Wigner functions. It follows (as a special case of the preceding formula, in fact) that

$$|\langle\psi_1|\psi_2\rangle|^2 = \frac{1}{2\pi} \int \mathrm{d}q\, \mathrm{d}p\; W_1(q, p) W_2(q, p) \,. \tag{6.31}$$

If the two states are orthogonal the integral has to vanish. From this we conclude that the Wigner function cannot be a positive function in general. Therefore, even though it is normalized, it is not a probability distribution. But somehow it is 'used' by the theory as if it were.

On the other hand the Wigner function is subject to restrictions in ways that classical probability distributions are not – this must be so since we are using a function of two variables to express the content of the wave function, which depends on only one variable. For instance, using the Cauchy–Schwarz inequality one can show that

$$|W(q, p)| \leq 2 \,. \tag{6.32}$$

It appears that the only economical way to state all the restrictions is to say that the Wigner function arises from a Hermitian operator with trace unity and positive eigenvalues via Eq. (6.24). We can formulate quantum mechanics in terms of the Wigner function, but it is difficult to make this formulation stand on its own legs.

To clarify how Wigner's formulation associates operators with functions we look at the moments of the characteristic function. Specifically, we observe that

$$\mathrm{Tr}\hat{\rho}(u\hat{q} + v\hat{p})^k = \mathrm{i}^k \left(\frac{\mathrm{d}}{\mathrm{d}\sigma}\right)^k \mathrm{Tr}\hat{\rho}\, \mathrm{e}^{-\mathrm{i}\sigma(u\hat{q}+v\hat{p})}|_{\sigma=0} = \mathrm{i}^k \left(\frac{\mathrm{d}}{\mathrm{d}\sigma}\right)^k \widetilde{W}(\sigma u, \sigma v)|_{\sigma=0} \,. \tag{6.33}$$

But if we undo the Fourier transformation we can conclude that

$$\mathrm{Tr}\hat{\rho}(u\hat{q} + v\hat{p})^k = \frac{1}{2\pi} \int \mathrm{d}q\, \mathrm{d}p\; (uq + vp)^k\, W(q, p) \,. \tag{6.34}$$

By comparing the coefficients we see that the moments of the Wigner function give the expectation values of symmetrized products of operators, that is to say that

$$\mathrm{Tr}\hat{\rho}(\hat{q}^m\hat{p}^n)_{\mathrm{sym}} = \frac{1}{2\pi}\int \mathrm{d}q\,\mathrm{d}p\; W(q,p)\,q^m p^n\;, \tag{6.35}$$

where $(\hat{q}\hat{p})_{\mathrm{sym}} = (\hat{q}\hat{p}+\hat{p}\hat{q})/2$ and so on. Symmetric ordering is also known as *Weyl ordering*, and the precise statement is that Weyl ordered polynomials in $\hat{q}$ and $\hat{p}$ are associated to polynomials in q and p.

Finally, let us take a look at some examples. For the special case of a coherent state $|q_0, p_0\rangle$, with the wavefunction given in Eq. (6.11), the Wigner function is a Gaussian,

$$W_{|q_0,p_0\rangle}(q,p) \;=\; 2\,\mathrm{e}^{-(q-q_0)^2-(p-p_0)^2}\;. \tag{6.36}$$

That the Wigner function of a coherent state is positive again confirms that there is 'something classical' about coherent states. Actually the coherent states are not the only states for which the Wigner function is positive – this property is characteristic of a class of states known as *squeezed states*.[9] If we superpose two coherent states the Wigner function will show two roughly Gaussian peaks with a 'wavy' structure in between, where both positive and negative values occur; in the quantum optics literature such states are known (perhaps somewhat optimistically) as *Schrödinger cat states*. For the number (or *Fock*) states $|n\rangle$, the Wigner function is

$$W_{|n\rangle}(q,p) = 2\,(-1)^n\,\mathrm{e}^{-q^2-p^2}L_n(2q^2+2p^2)\,, \tag{6.37}$$

where the L_n are Laguerre polynomials. They have n zeroes, so we obtain $n+1$ circular bands of alternating signs surrounding the origin, concentrated within a radius of about $q^2+p^2 = 2n+1$. Note that both examples saturate the bound (6.32) somewhere.

To see how the Wigner function relates to other quasi-probability distributions that are in use we again look at its characteristic function, and introduce a one-parameter family of characteristic functions by changing the high frequency behaviour (Cahill and Glauber, 1969):

$$\widetilde{W}^{(s)}(u,v) \;=\; \widetilde{W}(u,v)\,\mathrm{e}^{s(u^2+v^2)/4}\;. \tag{6.38}$$

This leads to a family of phase space distributions

$$W^{(s)}(q,p) \;=\; \frac{1}{2\pi}\int \mathrm{d}u\,\mathrm{d}v\;\widetilde{W}^{(s)}(u,v)\,\mathrm{e}^{\mathrm{i}uq+\mathrm{i}vp}\;. \tag{6.39}$$

[9] This was shown in 1974 by Hudson (1974). As the name suggests squeezed states are Gaussian, but 'squeezed'. A precise definition is $|\eta,z\rangle \equiv \exp[\eta(\hat{a}^\dagger)^2 - \eta^*\hat{a}^2]|z\rangle$, with $\eta\in\mathbb{C}$ (see Leonhardt, 1997).

For $s = 0$ we recover the Wigner function, $W = W^{(0)}$. We are also interested in the two 'dual' cases $s = -1$, leading to the *Husimi* or Q-function (Husimi, 1940), and $s = 1$, leading to the *Glauber–Sudarshan* or P-function (Glauber, 1963; Sudarshan, 1963). Note that, when $s > 0$, the Fourier transformation of the characteristic function may not converge to a function, so the P-function will have a distributional meaning. Using the Baker–Hausdorff formula, and the definition (6.24) of $\widetilde{W}(u, v)$, it is easily seen that the characteristic functions of the Q- and P-functions are, respectively,

$$\widetilde{Q}(u, v) \equiv \widetilde{W}^{(-1)}(u, v) = \mathrm{Tr}\hat{\rho}\, \mathrm{e}^{-\mathrm{i}\eta^* \hat{a}} \mathrm{e}^{-\mathrm{i}\eta \hat{a}^\dagger} \tag{6.40}$$

$$\widetilde{P}(u, v) \equiv \widetilde{W}^{(1)}(u, v) = \mathrm{Tr}\hat{\rho}\, \mathrm{e}^{-\mathrm{i}\eta \hat{a}^\dagger} \mathrm{e}^{-\mathrm{i}\eta^* \hat{a}}, \tag{6.41}$$

where $\eta \equiv (u + \mathrm{i}v)/\sqrt{2}$. Equation (6.35) is now to be replaced by

$$\mathrm{Tr}\hat{\rho}\, \hat{a}^n \hat{a}^{\dagger m} = \frac{1}{2\pi} \int \mathrm{d}q\, \mathrm{d}p\; Q(z, \bar{z})\, z^n \bar{z}^m \tag{6.42}$$

$$\mathrm{Tr}\hat{\rho}\, \hat{a}^{\dagger m} \hat{a}^n = \frac{1}{2\pi} \int \mathrm{d}q\, \mathrm{d}p\; P(z, \bar{z})\, z^n \bar{z}^m \tag{6.43}$$

where $z \equiv (q + \mathrm{i}p)/\sqrt{2}$ as usual. Thus the Wigner, Q- and P-functions correspond to different ordering prescriptions (symmetric, anti-normal and normal, respectively).

The Q-function is a smoothed Wigner function,

$$Q(q, p) \;=\; \frac{1}{2\pi} \int \mathrm{d}q'\, \mathrm{d}p'\; W(q', p')\, 2\, \mathrm{e}^{-(q-q')^2 - (p-p')^2}\,, \tag{6.44}$$

as was to be expected because its high frequency behaviour was suppressed. It is also a familiar object. Using Eq. (6.36) for the Wigner function of a coherent state we see that

$$Q(q, p) \;=\; \frac{1}{2\pi} \int \mathrm{d}q'\, \mathrm{d}p'\; W_\rho(q', p')\, W_{|q,p\rangle}(q', p') \tag{6.45}$$

Using the overlap formula (6.30) this is

$$Q(q, p) \;=\; \mathrm{Tr}\hat{\rho}\, |q, p\rangle\langle q, p| \;=\; \langle q, p|\hat{\rho}|q, p\rangle\,. \tag{6.46}$$

This is the symbol of the density operator as defined in Eq. (6.16). The Q-function has some desirable properties that the Wigner function does not have, in particular it is everywhere positive. Actually, as should be clear from the overlap formula together with the fact that density matrices are positive operators, we can use the Wigner function of an arbitrary state to smooth a given Wigner function and we will always obtain a positive distribution. We concentrate on the Q-function because

in that case the smoothing has been done with a particularly interesting reference state.

Since the integral of Q over phase space equals one the Husimi function is a genuine probability distribution. But it is a probability distribution of a somewhat peculiar kind, since it is not a probability density for mutually exclusive events. Instead $Q(q, p)$ is the probability that the system, if measured, would be found in a coherent state whose probability density has its mean at (q, p). Such 'events' are not mutually exclusive because the coherent states overlap. This has in its train that the overlap formula is not as simple as (6.30). If Q_A is the Q-function corresponding to an operator $\hat{A}$, and P_B the P-function corresponding to an operator $\hat{B}$, then

$$\mathrm{Tr}\hat{A}\hat{B} = \frac{1}{2\pi}\int \mathrm{d}q\,\mathrm{d}p\; Q_A(q, p)\, P_B(q, p)\,. \tag{6.47}$$

This explains why the Q-function is known as a *covariant symbol* – it is dual to the P-function which is then the *contravariant symbol* of the operator. The relation of the P-function to the density matrix is now not hard to see (although unfortunately not in a constructive way). It must be true that

$$\hat{\rho} = \frac{1}{2\pi}\int \mathrm{d}q\,\mathrm{d}p\, |q, p\rangle P(q, p)\langle q, p|\,. \tag{6.48}$$

This does not mean that the density matrix is a convex mixture of coherent states since the P-function may fail to be positive. Indeed in general it is not a function, and may fail to exist even as a tempered distribution. Apart from this difficulty we can think of the P-function as analogous to the barycentric coordinates introduced in Section 1.1.

Compared to the Wigner function the Q-function has the disadvantage that one does not recover the right marginals, say $|\psi(q)|^2$ by integrating over p. Moreover the definition of the Q-function (and the P-function) depends on the definition of the coherent states, and hence on some special reference state in the Hilbert space. This is clearly seen in Eq. (6.45), where the Wigner function of the reference state appears as a filter that is smoothing the Wigner function. But this peculiarity can be turned to an advantage. The Q-function may be the relevant probability distribution to use in a situation where the measurement device introduces a 'noise' that can be modelled by the reference state used to define the Q-function.[10] And the Q-function does have the advantage that it is a probability distribution. Unlike classical probability distributions, which obey no further constraints, it is also bounded from above by $1/2\pi$. This is an interesting property that can be exploited

[10] There is a discussion of this, with many references, in Leonhardt (1997) (and a very brief glimpse in our Section 10.1).

to define an entropy associated to any given density matrix, namely

$$S_W = -\frac{1}{2\pi}\int \mathrm{d}q\,\mathrm{d}p\; Q(q,p)\ln Q(q,p)\,. \tag{6.49}$$

This is the *Wehrl entropy* (Wehrl, 1978). It is a concave function of the density matrix ρ as it should be, and it has a number of other desirable properties as well. Unlike the classical Boltzmann entropy, which may assume the value $-\infty$, the Wehrl entropy obeys $S_W \geq 1$, and attains its lower bound if and only if the density matrix is a coherent pure state.[11] If we take the view that coherent states are classical states then this means that the Wehrl entropy somehow quantifies the departure from classicality of a given state. It will be compared to the quantum mechanical von Neumann entropy in Section 12.4.

6.3 Bloch coherent states

We will now study the *Bloch coherent states*.[12] In fact we have already done so – they are the complex curves, with topology $\mathbf{S}^2$, that were mentioned in Section 4.3. But this time we will develop them along the same lines as the canonical coherent states were developed. Our coherence group will be $SU(2)$, and our Hilbert space will be any finite-dimensional Hilbert space in which $SU(2)$ acts irreducibly. The physical system may be a spin system of total spin j, but it can also be a collection of n two-level atoms. The mathematics is the same, provided that $n = 2j$; a good example of the flexibility of quantum mechanics. In the latter application the angular momentum eigenstates $|j, m\rangle$ are referred to as *Dicke states*, and the quantum number m is interpreted as half the difference between the number of excited and unexcited atoms. The dimension of Hilbert space is N, and throughout $N = n + 1 = 2j + 1$.

We need a little group theory to get started. We will choose our reference state to be $|j, j\rangle$, that is it has spin up along the z-axis. Then the coherent states are all states of the form $\mathcal{D}|j, j\rangle$, where $\mathcal{D}$ is a Wigner rotation matrix. Using our standard representation of the angular momentum operators (in Appendix 2) the reference state is described by the vector $(1, 0, \ldots, 0)$, so the coherent states are described by the first column of $\mathcal{D}$. The rotation matrix can still be coordinatized in various ways. The Euler angle parametrization is a common choice, but we will use a different

[11] This was conjectured by Wehrl (1978) and proved by Lieb (1978). The original proof is quite difficult and depends on some hard theorems in Fourier analysis. The simplest proof so far is due to Luo (2000), who relied on properties of the Heisenberg group. For some explicit expressions for selected states, see Orłowski (1993).

[12] Also known as spin, atomic, or $SU(2)$ coherent states. They were first studied by Klauder (1960) and Radcliffe (1971). Bloch had, as far as we know, nothing to do with them but we call them 'Bloch' so as to not prejudge which physical application we have in mind.

one that brings out the complex analyticity that is waiting for us. We set

$$\mathcal{D} \;=\; \mathrm{e}^{zJ_-}\mathrm{e}^{-\ln(1+|z|^2)J_3}\mathrm{e}^{-\bar{z}J_+}\mathrm{e}^{\mathrm{i}\tau J_3}\;. \tag{6.50}$$

Because our group is defined using 2×2 matrices, we can prove this statement using 2×2 matrices; it will be true for all representations. Using the Pauli matrices from Appendix 2 we can see that

$$\mathrm{e}^{zJ_-}\mathrm{e}^{-\ln(1+|z|^2)J_3}\mathrm{e}^{-\bar{z}J_+} = \frac{1}{\sqrt{1+|z|^2}}\begin{bmatrix}1 & -\bar{z}\\ z & 1\end{bmatrix} \tag{6.51}$$

and we just need to multiply this from the right with $\mathrm{e}^{\mathrm{i}\tau J_3}$ to see that we have a general $SU(2)$ matrix. Of course the complex number z is going to be a stereographic coordinate on the 2-sphere.

The final factor in Eq. (6.50) is actually irrelevant: we observe that when $\mathrm{e}^{\mathrm{i}\tau J_3}$ is acting on the reference state $|j,\,j\rangle$ it just contributes an overall constant phase to the coherent states. In $\mathbb{C}\mathbf{P}^n$ the reference state is a *fixed point* of the $U(1)$ subgroup represented by $\mathrm{e}^{\mathrm{i}\tau J_3}$. In the terminology of Section 3.8 the isotropy group of the reference state is $U(1)$, and the $SU(2)$ orbit that we obtain by acting on it is the coset space $SU(2)/U(1) = \mathbf{S}^2$. This coset space will be coordinatized by the complex stereographic coordinate z. We choose the overall phase of the reference state to be zero. Since the reference state is annihilated by J_+ the complex conjugate $\bar{z}$ does not enter, and the coherent states are

$$|z\rangle = \mathrm{e}^{zJ_-}\mathrm{e}^{-\ln(1+|z|^2)J_3}\mathrm{e}^{-\bar{z}J_+}|j,\,j\rangle = \frac{1}{(1+|z|^2)^j}\,\mathrm{e}^{zJ_-}|j,\,j\rangle\;. \tag{6.52}$$

Using $z = \tan\frac{\theta}{2}\mathrm{e}^{\mathrm{i}\phi}$ we are always ready to express the coherent states as functions of the polar angles; $|z\rangle = |\theta,\,\phi\rangle$.

Since J_- is a lower triangular matrix, that obeys $(J_-)^{2j+1} = (J_-)^{n+1} = 0$, it is straightforward to express the unnormalized state in components. Using Eqs. (A2.4)–(A2.6) from Appendix 2 we get

$$\mathrm{e}^{zJ_-}|j,\,j\rangle = \sum_{m=-j}^{j} z^k\sqrt{\binom{2j}{j+m}}\,|j,\,m\rangle\;. \tag{6.53}$$

That is, the homogeneous coordinates – that do not involve the normalization factor – for coherent states are

$$Z^\alpha = (1,\;\sqrt{2j}z,\ldots,\;\sqrt{\binom{2j}{j+m}}z^{j+m},\ldots,z^{2j})\;. \tag{6.54}$$

We can use this expression to prove that the coherent states form a sphere of radius $\sqrt{j/2}$, embedded in $\mathbb{C}\mathbf{P}^{2j}$. There is an intelligent way to do so, using the Kähler

property of the metrics (see Section 3.3). First we compare with Eq. (4.6), and read off the affine coordinates $z^a(z)$ of the coherent states regarded as a complex curve embedded in $\mathbb{C}\mathbf{P}^{2j}$. For the coherent states we obtain

$$Z(z) \cdot \bar{Z}(\bar{z}) = (1 + |z|^2)^{2j} \tag{6.55}$$

(so that, with the normalization factor included, $\langle z|z\rangle = 1$). The Kähler potential for the metric is the logarithm of this expression, and the Kähler potential determines the metric as explained in Section 4.5. With no effort therefore, we see that on the coherent states the Fubini–Study metric induces the metric

$$\mathrm{d}s^2 = \partial_a \bar{\partial}_b \ln\big(Z(z) \cdot \bar{Z}(\bar{z})\big)\, \mathrm{d}z^a \mathrm{d}\bar{z}^b = \partial\bar{\partial} \ln(1 + |z|^2)^{2j} \mathrm{d}z\mathrm{d}\bar{z} = \frac{j}{2}\mathrm{d}\Omega^2 \,, \tag{6.56}$$

where we used the chain rule to see that $\mathrm{d}z^a\partial_a = \mathrm{d}z\partial_z$, and $\mathrm{d}\Omega^2$ is the metric on the unit 2-sphere written in stereographic coordinates. This proves that the embedding into $\mathbb{C}\mathbf{P}^{2j}$ turns the space of coherent states into a sphere of radius $\sqrt{j/2}$, as was to be shown. It is indeed a complex curve as defined in Section 4.3. The symplectic form on the space of coherent states is obtained from the same Kähler potential.

This is an important observation. At the end of Section 6.1 we listed four requirements that we want a set of coherent states to meet. One of them is that we should be able to think of the coherent states as forming a classical phase space embedded in the space of quantum mechanical pure states. In the case of canonical coherent states that phase space is the phase space spanned by q and p. The 2-sphere can also serve as a classical phase space because, as a Kähler manifold, it has a symplectic form that can be used to define Poisson brackets between any pair of functions on $\mathbf{S}^2$ (see Section 3.4). So all is well on this score. We also note that as j increases the sphere grows, and will in some sense approximate the flat plane with better and better accuracy.

Another requirement, listed in Section 6.1, is that there should exist a resolution of unity, so that an arbitrary state can be expressed as a linear combination of coherent states. This also works here. Using Eq. (6.53), this time with the normalization factor from Eq. (6.52) included, we can prove that

$$\frac{2j+1}{4\pi}\int \mathrm{d}\Omega |z\rangle\langle z| = \frac{2j+1}{4\pi}\int_0^\infty \mathrm{d}r \int_0^{2\pi} \mathrm{d}\phi\, \frac{4r}{(1+r^2)^2}|z\rangle\langle z| = \mathbb{1} \,, \tag{6.57}$$

where $\mathrm{d}\Omega$ is the round measure on the unit 2-sphere, that we wrote out using the fact that z is a stereographic coordinate and the definition $z = r^{\mathrm{i}\phi}$. It follows that the coherent states form a complete, and indeed an overcomplete, set. Next we require a correspondence between selected quantum mechanical observables on the one hand and classical observables on the other. Here we can use the symbol

of an operator, defined in analogy with the definition for canonical coherent states. In particular the symbols of the generators of the coherence group are the classical phase space functions

$$J_i(\theta,\phi) \;=\; \langle z|\hat{J}_i|z\rangle \;=\; jn_i(\theta,\phi)\,, \tag{6.58}$$

where $n_i(\theta,\phi)$ is a unit vector pointing in the direction labelled by the angles θ and ϕ. This is the symbol of the operator $\hat{J}_i$.

Our final requirement is that coherent states saturate an uncertainty relation. But there are several uncertainty relations in use for spin systems, a popular choice being

$$(\Delta J_x)^2(\Delta J_y)^2 \;\geq\; -\frac{1}{4}\,\langle[\hat{J}_x,\,\hat{J}_y]\rangle^2 \;=\; \frac{\langle\hat{J}_z\rangle^2}{4}\,. \tag{6.59}$$

(Here $(\Delta J_x)^2 \equiv \langle\hat{J}_x^2\rangle - \langle\hat{J}_x\rangle^2$ as usual.) States that saturate this relation are known (Aragone, Gueri, Salamó and Tani, 1974) as *intelligent* states – but since the right-hand side involves an operator this does not mean that the left-hand side is minimized. The relation itself may be interesting if, say, a magnetic field singles out the z-direction for attention. We observe that a coherent state that has spin up in the z-direction satisfies this relation, but for a general coherent state the uncertainty relation itself has to be rotated before it is satisfied. Another measure of uncertainty is

$$\Delta^2 \equiv (\Delta J_x)^2 + (\Delta J_y)^2 + (\Delta J_z)^2 = \langle\hat{J}^2\rangle - \langle\hat{J}_i\rangle\langle\hat{J}_i\rangle\,. \tag{6.60}$$

This has the advantage that Δ^2 is invariant under $SU(2)$, and takes the same value on all states in a given $SU(2)$ orbit in Hilbert space. This follows because $\langle\hat{J}_i\rangle$ transforms like an $SO(3)$ vector when the state is subject to an $SU(2)$ transformation.

One can now prove[13] that

$$j \;\leq\; \Delta^2 \;\leq\; j(j+1)\,. \tag{6.61}$$

It is quite simple. We know that $\langle\hat{J}^2\rangle = j(j+1)$. Moreover, in any given orbit we can use $SU(2)$ rotations to bring the vector $\langle\hat{J}_i\rangle$ to the form

$$\langle\hat{J}_i\rangle \;=\; \langle\hat{J}_z\rangle\,\delta_{i3}\,. \tag{6.62}$$

Expanding the state we see that

$$|\psi\rangle = \sum_{m=-j}^{j} c_m|m\rangle \quad\Rightarrow\quad \langle\hat{J}_z\rangle = \sum_{m=-j}^{j} m|c_m|^2 \quad\Rightarrow\quad 0 \leq \langle\hat{J}_z\rangle \leq j \tag{6.63}$$

[13] As was done by Delbourgo (1977) and further considered by Barros e Sá (2001b).

and the result follows in easy steps. It also follows that the lower bound in Eq. (6.61) is saturated if and only if the state is a Bloch coherent state, for which $\langle \hat{L}_z \rangle = j$. The upper bound will be saturated by states in the $\mathbb{R}\mathbf{P}^2$ orbit when it exists, and also by some other states. It can be shown that Δ^2 when averaged over $\mathbb{C}\mathbf{P}^n$, using the Fubini–Study measure from Section 4.7, is $j(j+\frac{1}{2})$. Hence for large values of j the average state is close to the upper bound in uncertainty.

In conclusion, Bloch coherent states obey all four requirements that we want coherent states to obey. There are further analogies to canonical coherent states to be drawn. Remembering the normalization we obtain the overlap of two coherent states as

$$\langle z'|z\rangle = \frac{\sum_{k=0}^{n}\binom{n}{k}(z\bar{z}')^k}{(\sqrt{1+|z|^2}\sqrt{1+|z'|^2})^n} = \left(\frac{1+z\bar{z}'}{\sqrt{1+|z|^2}\sqrt{1+|z'|^2}}\right)^{2j} . \tag{6.64}$$

The factorization is interesting. On the one hand we can write

$$\langle z'|z\rangle \;=\; \mathrm{e}^{-2\mathrm{i}A}\,\cos D_{\mathrm{FS}}\;, \tag{6.65}$$

where D_{FS} is the Fubini–Study distance between the two states and A is a phase factor that, for the moment, is not determined. On the other hand the quantity within brackets has a natural interpretation for $j = 1/2$, that is, on $\mathbb{C}\mathbf{P}^1$. Indeed

$$\langle z'|z\rangle \;=\; \left(\langle z'|z\rangle_{|j=\frac{1}{2}}\right)^{2j} . \tag{6.66}$$

But for the phase factor inside the brackets it is true that

$$\arg\langle z'|z\rangle_{|j=\frac{1}{2}} = \arg\langle z'|z\rangle_{|j=\frac{1}{2}}\langle z|+\rangle_{|j=\frac{1}{2}}\langle +|z'\rangle_{|j=\frac{1}{2}} = -2A_1\;, \tag{6.67}$$

where $|+\rangle$ is the reference state for spin $1/2$, Eq. (4.97) was used, and A_1 is the area of a triangle on $\mathbb{C}\mathbf{P}^1$ with vertices at the three points indicated. Comparing the two expressions we see that $A = 2jA_1$ and it follows that A is the area of a triangle on a sphere of radius $\sqrt{2j}/2 = \sqrt{j/2}$, that is the area of a triangle on the space of coherent states itself. The analogy with Eq. (6.15) for canonical coherent states is evident. This is a non-trivial result and has to do with our quite special choice of reference state; technically it happens because geodesics within the embedded 2-sphere of coherent states are null phase curves in the sense of Section 4.8, as a pleasant calculation confirms.[14]

Quasi-probability distributions on the sphere can be defined in analogy to those on the plane. In particular, the Wigner function can be defined and it is found to have similar properties to that on the plane. For instance, a Bloch coherent state $|\theta,\phi\rangle$ has a positive Wigner function centred around the point (θ,ϕ). We refer to

[14] This statement remains true for the $SU(3)$ coherent states discussed in Section 6.4; Berceanu (2001) has investigated things in more generality.

Table 6.1. *Comparison of the canonical coherent states on the plane and the Bloch coherent states on the sphere, defined by the Wigner rotation matrix $\mathcal{D}^{(j)}_{\theta,\phi}$. The overlap between two canonical (Bloch) coherent states is a function of the distance between two points on the plane (sphere), while the phase is determined by the area A of a flat (spherical) triangle.*

Hilbert space $\mathcal{H}$	Infinite	Finite, $N = 2j + 1$
phase space	plane $\mathbb{R}^2$	sphere S^2
commutation relations	$[\hat{q}, \hat{p}] = \mathrm{i}$	$[J_i, J_j] = \mathrm{i}\,\epsilon_{ijk}\,J_k$
basis	Fock $\{\|0\rangle, \|1\rangle, \dots\}$	J_z eigenstates $\|j, m\rangle$, $m = (-j, \dots, j)$
reference state	vacuum $\|0\rangle$	north pole $\|\kappa\rangle = \|j, j\rangle$
coherent states	$\|q, p\rangle = \exp[\mathrm{i}(p\hat{q} - q\hat{p})]\,\|0\rangle$	$\|\theta, \phi\rangle = \mathcal{D}^{(j)}_{\theta,\phi}\,\|j, j\rangle$
POVM	$\frac{1}{2\pi}\int_{\mathbb{R}^2} \|q, p\rangle\langle q, p|\,\mathrm{d}q\,\mathrm{d}p = \mathbb{1}$	$\frac{2j+1}{4\pi}\int_\Omega \|\theta, \phi\rangle\langle\theta, \phi|\,\mathrm{d}\Omega = \mathbb{1}$
overlap	$\mathrm{e}^{-2\mathrm{i}A}\exp\left[-\frac{1}{2}\left(D_E\right)^2\right]$	$\mathrm{e}^{-2\mathrm{i}A}\left[\cos\left(\frac{1}{2}D_R\right)\right]^{2j}$
Husimi representation	$Q_\rho(q, p) = \langle q, p\|\rho\|q, p\rangle$	$Q_\rho(\theta, \phi) = \langle\theta, \phi\|\rho\|\theta, \phi\rangle$
Wehrl entropy S_W	$-\frac{1}{2\pi}\int_{\mathbb{R}^2}\mathrm{d}q\,\mathrm{d}p\;Q_\rho \ln[Q_\rho]$	$-\frac{2j+1}{4\pi}\int_\Omega \mathrm{d}\Omega\;Q_\rho \ln[Q_\rho]$
Wehrl–Lieb conjecture	$S_W(\|\psi\rangle\langle\psi|) \geq 1$	$S_W(\|\psi\rangle\langle\psi|) \geq \frac{2j}{2j+1}$

the literature for details (Agarwal, 1981; Dowling, Agarwal and Schleich, 1994). The Husimi Q-function on the sphere will be given a detailed treatment in the next chapter. The Glauber–Sudarshan P-function exists for the Bloch coherent states whatever the dimension of the Hilbert space (Mukunda, Arvind, Chaturvedi and Simon, 2003); again a non-trivial statement because it requires the complex curve to 'wiggle around' enough inside $\mathbb{C}\mathbf{P}^{2j}$ so that, once the latter is embedded in the flat vector space of Hermitian matrices, it fills out all the dimensions of the latter. It is like the circle that one can see embedded in the surface of a tennis ball. The P-function will be positive for all mixed states in the convex cover of the Bloch coherent states.

To wind up the story so far we compare the canonical and the $SU(2)$ Bloch coherent states in Table 6.1.

6.4 From complex curves to $SU(K)$ coherent states

In the previous section we played down the viewpoint that regards the Bloch coherent states as a complex curve, but now we come back to it. Physically, what we have to do (for a spin system) is to assign a state to each direction in space. These states then serve as 'spin up' states for given directions. Mathematically this is a

map from $\mathbf{S}^2 = \mathbb{C}\mathbf{P}^1$ into $\mathbb{C}\mathbf{P}^n$, with $n = 2j = N - 1$, that is a complex curve in the sense of Section 4.3. To describe the sphere of directions in space we use the homogeneous coordinates

$$(u, v) \sim (\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\mathrm{e}^{\mathrm{i}\phi}) \sim (1, \tan\frac{\theta}{2}\mathrm{e}^{\mathrm{i}\phi}) \,. \tag{6.68}$$

As we know already, the solution to our problem is

$$(u, v) \quad \rightarrow \quad (u^n, \sqrt{n}u^{n-1}v, \ldots, \sqrt{\binom{n}{k}}u^{n-k}v^k, \ldots, v^n) \,. \tag{6.69}$$

As we also know, the Fubini–Study metric on $\mathbb{C}\mathbf{P}^n$ induces the metric

$$\mathrm{d}s^2 = \frac{n}{4}(\mathrm{d}\theta^2 + \sin^2\theta\ \mathrm{d}\phi^2) \tag{6.70}$$

on this curve, so it is a round sphere with a radius of curvature equal to $\sqrt{n}/2$. The fact that already for modest values of n the radius of curvature becomes larger than the longest geodesic distance between two points in $\mathbb{C}\mathbf{P}^n$ is not a problem since this sphere cannot be deformed to a point, and therefore it has no centre.

We have now specified the $m = j$ eigenstate for each possible spatial direction by means of a specific complex curve. It is remarkable that the location of all the other eigenstates is determined by the projective geometry of the curve. The location of the $m = -j$ state is evidently the antipodal point on the curve, as defined by the metric just defined. The other states lie off the curve and their location requires more work to describe. In the simple case of $n = 2$ (that is $j = 1$) the complex curve is a conic section, and the $m = 0$ state lies at the intersection of the unique pair of lines that are tangent to the curve at $m = \pm 1$, as described in Section 4.3. Note that the space of singlet states is an $\mathbb{R}\mathbf{P}^2$, since any two antipodal points on the $m = 1$ complex curve defines the same $m = 0$ state. The physical interpretation of the points of $\mathbb{C}\mathbf{P}^2$ is now fixed. Unfortunately it becomes increasingly difficult to work out the details for higher n. [15]

So far we have stuck to the simplest compact Lie algebra $SU(2)$. But, since the full isometry group of $\mathbb{C}\mathbf{P}^n$ is $SU(n+1)/\mathbb{Z}^{n+1}$, it is clear that all the special unitary groups are of interest to us. For any K, a physical application of $SU(K)$ coherent states may be a collection of K-level atoms.[16] For a single K-level atom we would use a K-dimensional Hilbert space, and for the collection the dimension can be much larger. But how do we find the orbits under, say, $SU(3)$ in some $\mathbb{C}\mathbf{P}^n$, and more especially what is the $SU(3)$ analogue of the Bloch coherent states? The simplest answer (Gitman and Shelepin, 1993) is obtained by a straightforward

[15] See Brody and Hughston (2001) for the $n = 3$ case.
[16] Here we refer to $SU(K)$ rather than $SU(N)$ because the letter $N = n + 1$ is otherwise engaged.

generalization of the approach just taken for $SU(2)$: since $SU(3)$ acts naturally on $\mathbb{C}\mathbf{P}^2$ this means that we should ask for an embedding of $\mathbb{C}\mathbf{P}^2$ into $\mathbb{C}\mathbf{P}^n$. Let the homogeneous coordinates of a point in $\mathbb{C}\mathbf{P}^2$ be $P^\alpha = (u, v, w)$. We embed this into $\mathbb{C}\mathbf{P}^n$ through the equation

$$(u, v, w) \to (u^m, \ldots, \sqrt{\frac{m!}{k_1!k_2!k_3!}} u^{k_1} v^{k_2} w^{k_3}, \ldots); \quad k_1 + k_2 + k_3 = m\,. \tag{6.71}$$

Actually this puts a restriction on the allowed values of n, namely

$$N = n + 1 = \frac{1}{2}(m + 1)(m + 2)\,. \tag{6.72}$$

For these values of n we can choose the components of a symmetric tensor of rank m as homogeneous coordinates for $\mathbb{C}\mathbf{P}^n$. The map from $P^\alpha \in \mathbb{C}\mathbf{P}^2$ to a point in $\mathbb{C}\mathbf{P}^5$ is then defined by

$$P^\alpha \to T^{\alpha\beta} = P^{(\alpha} P^{\beta)}\,. \tag{6.73}$$

In effect we are dealing with the symmetric tensor representation of $SU(3)$. (The brackets mean that we are taking the totally symmetric part; compare this to the symmetric multispinors used in Section 4.4.)

Anyway we now have an orbit of $SU(3)$ in $\mathbb{C}\mathbf{P}^n$ for special values of n. To compute the intrinsic metric on this orbit (as defined by the Fubini–Study metric in the embedding space) we again take the short cut via the Kähler potential. We first observe that

$$Z \cdot \bar{Z} = \sum_{k_1+k_2+k_3=m} \frac{m!}{k_1!k_2!k_3!} |u|^{2k_1} |v|^{2k_2} |w|^{2k_3} = (|u|^2 + |v|^2 + |w|^2)^m = (P \cdot \bar{P})^m\,. \tag{6.74}$$

Since the logarithm of this expression is the Kähler potential expressed in affine coordinates, we find that the induced metric on the orbits becomes

$$\mathrm{d}s^2 = \partial_a \partial_{\bar{b}} \ln \left(P(z) \cdot \bar{P}(\bar{z})\right)^m \mathrm{d}z^a \mathrm{d}\bar{z}^{\bar{b}} = m\, \mathrm{d}\hat{s}^2\,, \tag{6.75}$$

where $\mathrm{d}\hat{s}^2$ is the Fubini–Study metric on $\mathbb{C}\mathbf{P}^2$ written in affine coordinates. Hence, just as for the Bloch coherent states, we find that the intrinsic metric on the orbit is just a rescaled Fubini–Study metric. Since the space of coherent states is Kähler the symplectic form can be obtained from the same Kähler potential, using the recipe in Section 3.3.

The generalization to $SU(K)$ with an arbitrary K should be obvious. The Hilbert spaces in which we can represent $SU(K)$ using symmetric tensors of rank m have dimension

$$N_{K,m} \equiv \dim(\mathcal{H}_{K,m}) = \binom{K + m - 1}{m} = \frac{(K + m - 1)!}{m!\,(K - 1)!}\,. \tag{6.76}$$

which is the number of ways of distributing m identical objects in K boxes. For $K = 3$ it reduces to Eq. (6.72). The coherent states manifold itself is now $\mathbb{C}\mathbf{P}^{K-1}$, and the construction embeds it into $\mathbb{C}\mathbf{P}^{N_{K,m}-1}$.

But the story of coherent states for $SU(K)$ is much richer than this for every $K > 2$.

6.5 $SU(3)$ coherent states

Let us recall some group theory. In this book we deal mostly with the *classical groups* $SU(K)$, $SO(K)$ and $Sp(K)$ and in fact mostly with the special unitary groups $SU(K)$. There are several reasons for this. For one thing the isometry group of $\mathbb{C}\mathbf{P}^{K-1}$ is $SU(K)/\mathbb{Z}^K$, for $\mathbb{R}\mathbf{P}^{K-1}$ it is the special orthogonal group $SO(K)$ and for the quaternionic projective space $\mathbb{H}\mathbf{P}^{K-1}$ it is the symplectic group $Sp(K)/\mathbb{Z}^2$, so these groups are always there. Also they are all, in the technical sense, simple and compact groups and have in many respects analogous properties. In particular, most of their properties can be read off from their Lie algebras, and their complexified Lie algebras can be brought to the standard form

$$[H_i, H_j] = 0\,, \qquad [H_i, E_\alpha] = \alpha_i E_\alpha\,, \tag{6.77}$$

$$[E_\alpha, E_\beta] = N_{\alpha\beta} E_{\alpha+\beta}\,, \qquad [E_\alpha, E_{-\alpha}] = \alpha^i H_i\,. \tag{6.78}$$

where α_i is a member of the set of *root vectors* and $N_{\alpha\beta} = 0$ if $\alpha_i + \beta_i$ is not a root vector. The H_i form a maximal commuting set of operators and span what is known as the *Cartan subalgebra*. Of course α_i and $N_{\alpha\beta}$ depend on the group; readers not familiar with group theory will at least be able to see that $SU(2)$ fits into this scheme (if necessary, consult Appendix 2, or a book on group theory (Gilmore, 1974)). A catalogue of the irreducible unitary representations can now be made by specifying a *highest weight vector* $|\mu\rangle$ in the Hilbert space, with the property that it is annihilated by the 'raising operators' E_α (for all positive roots), and it is labelled by the eigenvalues of the commuting operators H_i. Thus

$$E_\alpha|\mu\rangle = 0\,, \qquad \alpha > 0\,; \qquad H_i|\mu\rangle = \frac{1}{2} m_i |\mu\rangle\,, \tag{6.79}$$

where m_i are the components of a *weight vector*. For $SU(2)$ we expect every reader to be familiar with this result. For $SU(3)$ we obtain representations labelled by two integers m_1 and m_2, with the dimension of the representation being

$$\dim\mathcal{H}_{[m_1,m_2]} = \frac{1}{2}(m_1+1)(m_2+1)(m_1+m_2+2)\,. \tag{6.80}$$

We will concentrate on $SU(3)$ because the main conceptual novelty – compared to $SU(2)$ – can be seen already in this case.

In accordance with our general scheme we obtain $SU(3)$ coherent states by acting with $SU(3)$ on some reference state. It turns out that the resulting orbit is a Kähler manifold if and only if the reference state is a highest weight vector of the representation (Perelomov, 1977; Zhang et al., 1990) – indeed the reason why the $\mathbf{S}^2$ orbit of $SU(2)$ is distinguished can now be explained as a consequence of its reference state being a highest weight vector. What is new compared to $SU(2)$ is that there are several qualitatively different choices of highest weight vectors. There is more news on a practical level: whereas the calculations in the $SU(2)$ case are straightforward, they become quite lengthy already for $SU(3)$. For this reason we confine ourselves to a sketch.[17] We begin in the defining representation of $SU(3)$ by introducing the 3×3 matrices

$$S_{ij} \;=\; |i\rangle\langle j| \;. \tag{6.81}$$

If $i < j$ we have a 'raising operator' E_α with positive root, if $i > j$ we have a 'lowering operator' $E_{-\alpha}$ with negative root. We exponentiate the latter and define

$$b_-(z) = \mathrm{e}^{z_3 S_{31}}\, \mathrm{e}^{z_1 S_{21}}\, \mathrm{e}^{z_2 S_{32}} \;, \tag{6.82}$$

where the γ_i are complex numbers and no particular representation is assumed. In the defining representation this is the lower triangular 3×3 matrix,

$$b_-(z) = \begin{bmatrix} 1 & 0 & 0 \\ z_1 & 1 & 0 \\ z_3 & z_2 & 1 \end{bmatrix} \;. \tag{6.83}$$

Upper triangular matrices b_+ are defined analogously (or by Hermitian conjugation) and will annihilate the reference state that we are about to choose. Then we use that the fact that almost all (in the sense of the Haar measure) $SU(3)$ matrices can be written in the Gauss form

$$A = b_- D b_+ \;, \tag{6.84}$$

where D is a diagonal matrix (obtained by exponentiating the elements of the Cartan subalgebra). Finally we define the coherent states by

$$|z\rangle = \mathcal{N}(z)\, b_-(z)\, |\mu_{[m_1,m_2]}\rangle \;, \tag{6.85}$$

[17] For full details consult Gnutzmann and Kuś (1998), from whom everything that we say here has been taken.

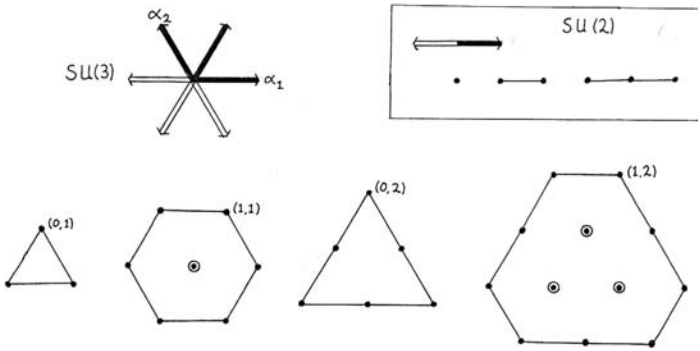


Figure 6.4. A reminder about representation theory: we show the root vectors of $SU(3)$ and four representations – the idea is that one can get to every point (state) by subtracting one of the simple root vectors α_1 or α_2 from the highest weight vector. For the degenerate representations $(0, m)$ there is no way to go when subtracting α_1; this is the reason why Eq. (6.90) holds. The corresponding picture for $SU(2)$ is shown inserted.

where the reference state is a highest weight vector for the irreducible representation that we have chosen. This formula is clearly analogous to Eq. (6.52). The calculation of the normalizing factor $\mathcal{N}$ is again straightforward but is somewhat cumbersome compared to the calculation in the $K = 2$ case. Let us define

$$\gamma_1 \equiv 1 + |z_1|^2 + |z_3|^2 \tag{6.86}$$

$$\gamma_2 \equiv 1 + |z_2|^2 + |z_3 - z_1 z_2|^2 \,. \tag{6.87}$$

Then the result is

$$\mathcal{N}(z) = \sqrt{\gamma_1^{-m_1} \gamma_2^{-m_2}} \,. \tag{6.88}$$

With a view to employing affine coordinates in $\mathbb{C}\mathbf{P}^n$ we may prefer to write the state vector in the form $Z^\alpha = (1, \dots)$ instead. Then we find

$$Z(z) \cdot \bar{Z}(\bar{z}) = \gamma_1^{m_1} \gamma_2^{m_2} \,. \tag{6.89}$$

Equipped with this Kähler potential we can easily write down the metric and the symplectic form that is induced on the submanifold of coherent states.

There is, however, a subtlety. For degenerate representations, namely when either m_1 or m_2 equals zero, the reference state is annihilated not only by S_{ij} for $i < j$ but also by an additional operator. Thus

$$j_2 = 0 \quad \Rightarrow \quad S_{31}|\mu\rangle = 0 \quad \text{and} \quad m_1 = 0 \quad \Rightarrow \quad S_{32}|\mu\rangle = 0 \,. \tag{6.90}$$

Readers who are familiar with the representation theory of $SU(3)$ see this immediately from Figure 6.4. This means that the isotropy subgroup is larger for the degenerate representations than in the generic case. Generically the state vector is left invariant up to a phase by group elements belonging to the Cartan subgroup $U(1) \times U(1)$, but in the degenerate case the isotropy subgroup grows to $SU(2) \times U(1)$.

The conclusion is that for a general representation the space of coherent states is the six-dimensional space $SU(3)/U(1) \times U(1)$, with metric and symplectic form derived from the Kähler potential given in Eq. (6.89). However, if $m_2 = 0$ the space of coherent states is the four-dimensional space $SU(3)/SU(2) \times U(1) = \mathbb{C}\mathbf{P}^2$, with metric and symplectic form again given by Eq. (6.89). This agrees with what we found in the previous section, where the Kähler potential was given by Eq. (6.74). For $m_1 = 0$ the space of coherent states is again $\mathbb{C}\mathbf{P}^2$; the Kähler potential is obtained from Eq. (6.89) by setting $z_1 = 0$. Interestingly, the 'classicality' of coherent states now gives rise to classical dynamics on manifolds of either four or six dimensions (Gnutzmann, Haake and Kuś, 2000).

The partition of unity – or, the POVM – becomes

$$\mathbb{1} = \frac{(m_1+1)(m_2+1)(m_1+m_2+2)}{\pi^3} \int \mathrm{d}^2 z_1 \, \mathrm{d}^2 z_2 \, \mathrm{d}^2 z_3 \, \frac{1}{\gamma_1^2 \gamma_2^2} \, |z\rangle\langle z| \qquad (6.91)$$

in the generic case and

$$\mathbb{1} = \frac{(m_1+1)(m_1+2)}{\pi^2} \int \mathrm{d}^2 z_1 \, \mathrm{d}^2 z_3 \, \frac{1}{\gamma_1^3} \, |z\rangle\langle z| \quad \text{for } m_2 = 0 \,, \qquad (6.92)$$

$$\mathbb{1} = \frac{(m_2+1)(m_2+2)}{\pi^2} \int \mathrm{d}^2 z_2 \, \mathrm{d}^2 z_3 \, \frac{1}{\gamma_2^3} \, |z\rangle\langle z| \quad \text{for } m_1 = 0 \,, \qquad (6.93)$$

where the last integral is evaluated at $z_1 = 0$.

In conclusion the $SU(3)$ coherent states differ from the $SU(2)$ coherent states primarily in that there is more variety in the choice of representation, and hence more variety in the possible coherent state spaces that occur. As it is easy to guess, the same situation occurs in the general case of $SU(K)$ coherent states. Let us also emphasize, that it may be useful to define generalized coherent states for some more complicated groups. For instance, in Chapter 15 we analyse pure product states of a composite $N \times M$ system, which may be regarded as coherent with respect to the group $SU(N) \times SU(M)$. The key point to observe is that if we use a maximal weight vector as the reference state from which we build coherent states of a compact Lie group then the space of coherent states is Kähler, so it can serve as a classical phase space, and many properties of the Bloch coherent states recur, for

example there is an invariant measure of uncertainty using the quadratic Casimir operator, and coherent states saturate the lower bound of that uncertainty relation (Delbourgo and Fox, 1977).

Problems

Problem 6.1 Compute the Q- and P-distributions for the one-photon Fock state.

Problem 6.2 Compute the Wehrl entropy for the Fock states $|n\rangle$.

7

The stellar representation

We are all in the gutter, but some of us are looking at the stars.
Oscar Wilde

We have already, in Section 4.4, touched on the possibility of regarding points in complex projective space $\mathbb{C}\mathbf{P}^{N-1}$ as unordered sets of $n = N - 1$ stars on a 'celestial sphere'. There is an equivalent description in terms of the n zeros of the Husimi function. Formulated either way, the stellar representation illuminates the orbits of $SU(2)$, the properties of the Husimi function, and the nature of 'typical' and 'random' quantum states.

7.1 The stellar representation in quantum mechanics

Our previous discussion of the stellar represenation was based on projective geometry only. When such points are thought of as quantum states we will wish to take the Fubini–Study metric into account as well. This means that we will want to restrict the transformations acting on the celestial sphere, from the Möbius group $SL(2,\mathbb{C})/\mathbb{Z}_2$ to the distance preserving subgroup $SO(3) = SU(2)/\mathbb{Z}_2$. So the transformations that we consider are

$$z \;\rightarrow\; z' \;=\; \frac{\alpha z - \beta}{\beta^* z + \alpha^*}\,, \tag{7.1}$$

where it is understood that $z = \tan\frac{\theta}{2}\mathrm{e}^{\mathrm{i}\phi}$ is a stereographic coordinate on the 2-sphere. Recall that the idea in Section 4.4 was to associate a polynomial to each vector in $\mathbb{C}^N$, and the roots of that polynomial to the corresponding point in $\mathbb{C}\mathbf{P}^{N-1}$. The roots are the stars. We continue to use this idea, but this time we want to make sure that an $SU(2)$ transformation really corresponds to an ordinary rotation of the sphere on which we have placed our stars. For this purpose we need to polish our conventions a little: we want a state of spin 'up' in the direction given by the unit vector $\mathbf{n}$ to be represented by $n = 2j$ points sitting at the point where $\mathbf{n}$ meets the

sphere, and more generally a state that is an eigenstate of $\mathbf{n}\cdot\mathbf{L}$ with eigenvalue m to be represented by $j+m$ points at this point and $j-m$ points at the antipode. Now consider a spin $j=1$ particle and place two points at the east pole of the sphere. In stereographic coordinates the east pole is at $z=\tan\frac{\pi}{4}=1$, so the east pole polynomial is

$$w(z)=(z-1)^2=z^2-2z+1\ . \tag{7.2}$$

The eigenvector of L_x with eigenvalue $+1$ is $Z^\alpha=(1,\sqrt{2},1)$, so if we are to read off this vector from Eq. (7.2) we must set

$$w(z)\ \equiv\ Z^0z^2-\sqrt{2}Z^1z+Z^2\ . \tag{7.3}$$

After a little experimentation like this it becomes clear that to any point in $\mathbb{C}\mathbf{P}^n$, given by the homogeneous coordinates Z^α, we want to associate the n unordered roots of the polynomial

$$w(z)\ \equiv\ \sum_{\alpha=0}^{n}(-1)^\alpha Z^\alpha\sqrt{\binom{n}{\alpha}}z^{n-\alpha}\ . \tag{7.4}$$

The convention for when ∞ counts as a root is as described in Section 4.4. The factors and signs have been chosen precisely so that the eigenstate of the operator $\mathbf{n}\cdot\mathbf{L}$ with eigenvalue m, where $\mathbf{n}=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$, is represented by $j+m$ points at $z=\tan\frac{\theta}{2}e^{i\phi}$ and $j-m$ points at the antipodal point (see Figure 4.7). It is interesting to notice that the location of the stars has an operational significance. A spin system, say, cannot be observed to have spin up along a direction that points away from a star on our celestial sphere.

With these conventions the stellar representation behaves nicely under rotations, in the sense that if we apply a rotation operator to $\mathbb{C}\mathbf{P}^n$ the effect in the picture is simply to rotate the sphere containing the n unordered points.[1] The action of a general unitary transformation, not belonging to the $SU(2)$ subgroup that we have singled out for attention, is of course not transparent in the stellar representation. On the other hand the anti-unitary operation of time reversal, as defined in Section 5.5, is nicely described. For $n=1$ we find that

$$\Theta\begin{bmatrix}Z^0\\Z^1\end{bmatrix}=\begin{bmatrix}-\bar{Z}^1\\\bar{Z}^0\end{bmatrix}\quad\Rightarrow\quad\Theta z=-\frac{1}{\bar{z}} \tag{7.5}$$

(since $z\equiv Z^1/Z^0$). This is just an inversion of the sphere through the origin. But this works for all n. From the transformation properties of Z^α together with Eq. (7.4) it follows that a state that is pictured as n points located at the n positions z_i will go

[1] The resemblance to Schwinger's harmonic oscillator representation of $SU(2)$ (Schwinger, 1965) is not accidental. He was led to his representation by Majorana's description of the stellar representation (Majorana, 1932).

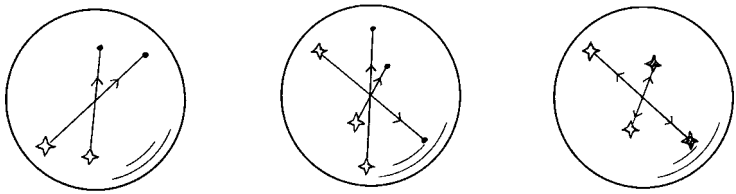


Figure 7.1. In the stellar representation a time reversal moves the stars to their antipodal positions; time-reversal invariant states can therefore occur only when the number of stars is even (as it is in the rightmost case, representing a point in $\mathbb{R}\mathbf{P}^4$).

over to the state that is pictured by n points located at the inverted positions $-1/\bar{z}_i$. Since no configuration of an odd number of points can be invariant under such a transformation it immediately follows that there are no states invariant under time reversal when n is odd.

For even n there will be a subspace of states left invariant under time reversal. For $n = 2$ it is evident that this subspace is the real projective space $\mathbb{R}\mathbf{P}^2$, because the stellar representation of a time reversal invariant state is a pair of points in antipodal position on $\mathbf{S}^2$. This is not the $\mathbb{R}\mathbf{P}^2$ that we would obtain by choosing all coordinates real, rather it is the $\mathbb{R}\mathbf{P}^2$ of all possible $m = 0$ states. For higher n the story is less transparent, but it is still true that the invariant subspace is $\mathbb{R}\mathbf{P}^n$, and we obtain a stellar representation of real projective space into the bargain – a point in the latter corresponds to a configuration of stars on the sphere that is symmetric under inversion through the origin.

7.2 Orbits and coherent states

The stellar representation shows its strength when we decide to classify all possible orbits of $SU(2)/\mathbb{Z}^2 = SO(3)$ in $\mathbb{C}\mathbf{P}^n$.[2] The general problem is that of a group G acting on a manifold M; the set of points that can be reached from a given point is called a G-orbit. In itself the orbit is the coset space G/H, where H is the subgroup of transformations leaving the given point invariant. H is called the *isotropy group*. We want to know what kinds of orbits there are and how M is partitioned into G-orbits. The properties of the orbits will depend on the isotropy group H. A part of the manifold M where all orbits are similar is called a *stratum*, and in general M becomes a *stratified manifold* foliated by orbits of different kinds. We can also define the *orbit space* as the space whose points are the orbits of G in M; a function of M is called *G-invariant* if it takes the same value at all points of a

[2] This was done by Bacry (1974). By the way his paper contains no references whatsoever to prior work.

given orbit, which means that it is automatically a well-defined function of the orbit space.

In Section 6.3 we had $M = \mathbb{C}\mathbf{P}^n$, and, by choosing a particular reference state, we selected a particular orbit as the space of Bloch coherent states. This orbit had the intrinsic structure of $SO(3)/SO(2) = \mathbf{S}^2$. It was a successful choice, but it is interesting to investigate if other choices would have worked as well. For $\mathbb{C}\mathbf{P}^1$ the problem is trivial: we have one star and can rotate it to any position, so there is only one orbit, namely $\mathbb{C}\mathbf{P}^1$ itself. For $\mathbb{C}\mathbf{P}^2$ it gets more interesting. We now have two stars on the sphere. Suppose first that they coincide. The little group – the subgroup of rotations that leaves the configuration of stars invariant – consists of rotations around the axis where the pair is situated. Therefore the orbit becomes $SO(3)/SO(2) = O(3)/O(2) = \mathbf{S}^2$. Every state represented by a pair of coinciding points lies on this orbit. Referring back to Section 6.4, we note that the states on this orbit can be regarded as states that have spin up in some direction, and we already know that these form a sphere inside $\mathbb{C}\mathbf{P}^2$. But now we know it in a new way. The next case to consider is when the pair of stars are placed antipodally on the sphere. This configuration is invariant under rotations around the axis defined by the stars, but also under an extra turn that interchanges the two points. Hence the little group is $SO(2) \times \mathbb{Z}^2 = O(2)$ and the orbit is $\mathbf{S}^2/\mathbb{Z}^2 = O(3)/[O(2) \times O(1)] = \mathbb{R}\mathbf{P}^2$. For any other pair of stars the little group has a single element, namely a discrete rotation that interchanges the two. Hence the generic orbit is $SO(3)/\mathbb{Z}^2 = O(3)/[O(1) \times O(1)]$. Since $SO(3) = \mathbb{R}\mathbf{P}^3 = \mathbf{S}^3/\mathbb{Z}^2$ we can also think of this as a space of the form $\mathbf{S}^3/\Gamma$, where Γ is a discrete subgroup of the isometry group of the 3-sphere. Spaces of this particular kind are called *lens spaces* by mathematicians.

To solve the classification problem for arbitrary n we first recall that the complete list of subgroups of $SO(3)$ consists of e (the trivial subgroup consisting of just the identity); the discrete groups C_p (the cyclic groups, with p some integer), D_p (the dihedral groups), T (the symmetry group of the tetrahedron), O (the symmetry group of the octahedron and the cube) and Y (the symmetry group of the icosahedron and the dodecahedron); also the continuous groups $SO(2)$ and $O(2)$. This is well known to crystallographers and to mathematicians who have studied the regular polyhedra. Recall, moreover, that the tetrahedron has four vertices, six edges and four faces so that we may denote it by $\{4, 6, 4\}$. Similarly the octahedron is $\{6, 12, 8\}$, the cube $\{8, 12, 6\}$, the dodecahedron $\{12, 30, 20\}$ and the icosahedron is $\{20, 30, 12\}$. The question is: given the number n, does there exist a configuration of n stars on the sphere invariant under the subgroup Γ? The most interesting case is for $\Gamma = SO(2)$ which occurs for all n, for instance when all the stars coincide. For $O(2)$ the stars must divide themselves equally between two antipodal positions, which can happen for all even n. The cyclic group C_p occurs for all $n \geq p$, the

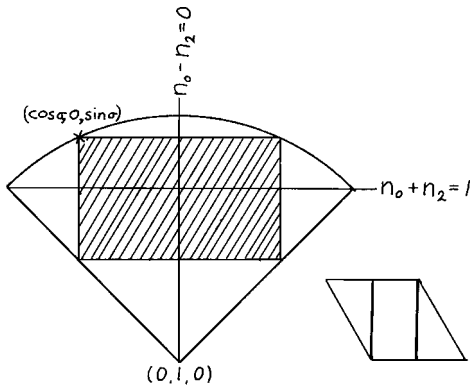


Figure 7.2. An orbit of $SU(2)$ acting on $\mathbb{C}\mathbf{P}^2$. We use orthographic coordinates to show the octant, in which the orbit fills out a two-dimensional rectangle. Our reference state is at its upper left-hand corner. In the tori there is a circle and we show how it winds around its torus; its precise position varies. When $\sigma \to 0$ the orbit collapses to $\mathbf{S}^2$ and when $\sigma \to \pi/4$ it collapses to $\mathbb{R}\mathbf{P}^2$.

groups D_2 and D_4 for all even $n \geq 4$, and the remaining dihedral groups D_p when $n = p + pa + 2b$ with a and b non-negative integers. For the tetrahedral group T, we must have $n = 4a + 6b$ with a non-negative (this may be a configuration of a stars at each corner of the tetrahedron and b stars sitting 'above' each edge midpoint – if the latter stars only are present the symmetry group becomes O). Similarly the octahedral group O occurs when $n = 6a + 8b$ and the icosahedral group Y when $n = 12a + 20b + 30c$, a, b and c being integers. Finally configurations with no symmetry at all appear for all $n \geq 3$. The possible orbits are of the form $SO(3)/\Gamma$; if Γ is one of the discrete groups this is a three-dimensional manifold. Indeed among the orbits only the exceptional $SO(3)/SO(2) = \mathbf{S}^2$ orbit is a Kähler manifold, and it is the only orbit that can serve as a classical phase space. Hence this is the orbit that we will use to form coherent states.

Since the orbits have rather small dimensions the story of how $\mathbb{C}\mathbf{P}^n$ can be partitioned into $SU(2)$ orbits is rather involved when n is large, but for $n = 2$ it can be told quite elegantly. There will be a one parameter family of three-dimensional orbits, and correspondingly a one parameter choice of reference vectors. A possible choice is

$$Z_0^\alpha(\sigma) = \begin{bmatrix} \cos\sigma \\ 0 \\ \sin\sigma \end{bmatrix}, \qquad 0 \leq \sigma \leq \frac{\pi}{4}. \tag{7.6}$$

The corresponding polynomial is $w(z) = z^2 + \tan\sigma$ and its roots will go from coinciding to antipodally placed as σ grows. If we act on this vector with a general

3×3 three-rotation matrix $\mathcal{D}$ – parametrized say by Euler angles as in Eq. (3.143), but this time for the three-dimensional representation – we will obtain the state vector

$$Z^{\alpha}(\sigma, \tau, \theta, \phi) = \mathcal{D}^{\alpha}_{\ \beta}(\tau, \theta, \phi) Z^{\beta}_{0}(\sigma) \ , \tag{7.7}$$

where σ labels the orbit and the Euler angles serve as coordinates within the orbit. The range of τ turns out to be $[0, \pi[$. Together these four parameters serve as a coordinate system for $\mathbb{C}\mathbf{P}^2$.

By means of lengthy calculations we can express the Fubini–Study metric in these coordinates; in the notation of Section 3.7

$$\mathrm{d}s^2 = \mathrm{d}\sigma^2 + 2(1 + \sin 2\sigma)\Theta_1^2 + 2(1 - \sin 2\sigma)\Theta_2^2 + 4\sin^2 2\sigma\, \Theta_3^2 \ . \tag{7.8}$$

On a given orbit σ is constant and the metric becomes the metric of a 3-sphere that has been squashed in a particular way. It is in fact a lens space rather than a 3-sphere because of the restricted range of the periodic coordinate τ. When $\sigma = 0$ the orbit degenerates to a (round) 2-sphere, and when $\sigma = \pi/4$ to real projective 2-space. The parameter σ measures the distance to the $\mathbf{S}^2$ orbit.

Another way to look at this is to see what the orbits look like in the octant picture (Section 4.6). The answer turns out to be quite striking (Barros e Sá, 2001a) and is given in Figure 7.2.

7.3 The Husimi function

Scanning our list of orbits we see that the only orbit that is a symplectic space, and can serve as a classical phase space, is the exceptional $SO(3)/SO(2) = \mathbf{S}^2$ orbit. These are the Bloch coherent states that we will use, and we will now proceed to the Husimi or Q-function for such coherent states. In Section 6.2 it was explained that the Husimi function is a genuine probability distribution on the classical phase space and that, at least in theory, it allows us to reconstruct the quantum state. Our notation will differ somewhat from that of Section 6.3, so before we introduce it we recapitulate what we know so far. The dimension of the Hilbert space $\mathcal{H}_N$ is $N = n + 1 = 2j + 1$. Using the basis states $|e_\alpha\rangle = |j, m\rangle$, a general pure state can be written in the form

$$|\psi\rangle \ = \ \sum_{\alpha=0}^{n} Z^{\alpha} |e_\alpha\rangle \tag{7.9}$$

and a (normalized) Bloch coherent state in the form

$$|z\rangle \ = \ \frac{1}{(1 + |z|^2)^{\frac{n}{2}}} \sum_{\alpha=0}^{n} \binom{n}{\alpha} z^{\alpha} \, |e_\alpha\rangle \ . \tag{7.10}$$

(The notation here is inconsistent – Z^α is a component of a vector, while z^α is the complex number z raised to a power – but quite convenient.) At this point we introduce the *Bargmann function*. Up to a factor it is again an nth order polynomial uniquely associated to any given state. By definition

$$\psi(z) = \langle\psi|z\rangle = \frac{1}{(1+|z|^2)^{\frac{n}{2}}} \sum_{\alpha=0}^{n} \bar{Z}_\alpha \sqrt{\binom{n}{\alpha}}\, z^\alpha \,. \tag{7.11}$$

It is convenient to regard our Hilbert space as the space of functions of this form, with the scalar product

$$\langle\psi|\phi\rangle = \frac{n+1}{4\pi}\int \mathrm{d}\Omega\; \psi\bar{\phi} \equiv \frac{n+1}{4\pi}\int \frac{4\,\mathrm{d}^2 z}{(1+|z|^2)^2}\,\psi\,\bar{\phi}\,. \tag{7.12}$$

So $\mathrm{d}\Omega$ is the usual measure on the unit 2-sphere. That this is equivalent to the usual scalar product follows from our formula (6.57) for the resolution of unity.[3]

Being a polynomial the Bargmann function can be factorized. Then we obtain

$$\psi(z) = \frac{\bar{Z}_n}{(1+|z|^2)^{\frac{n}{2}}}\,(z-\omega_1)(z-\omega_2)\ldots(z-\omega_n)\,. \tag{7.13}$$

The state vector is uniquely characterized by the zeros of the Bargmann function, so again we have stars on the celestial sphere to describe our states. But the association is not quite the same that we have used so far. For instance, we used to describe a coherent state by the polynomial

$$w(z) \;=\; (z-z_0)^n\,. \tag{7.14}$$

(And we know how to read off the components Z^α from this expression.) But the Bargmann function of the same coherent state $|z_0\rangle$ is

$$\psi_{z_0}(z) = \langle z_0|z\rangle = \frac{\bar{z}_0^{\,n}}{(1+|z|^2)^{\frac{n}{2}}(1+|z_0|^2)^{\frac{n}{2}}}\,(z+\frac{1}{\bar{z}_0})^n\,. \tag{7.15}$$

Hence $\omega_0 = -1/\bar{z}_0$. In general the zeros of the Bargmann function are antipodally placed with respect to our stars. As long as there is no confusion, no harm is done.

With the Husimi function for the canonical coherent states in mind, we rely on the Bloch coherent states to define the Husimi function as[4]

$$Q_\psi(z) = |\langle\psi|z\rangle|^2 = \frac{|Z^n|^2}{(1+|z|^2)^n}\,|z-\omega_1|^2|z-\omega_2|^2\ldots|z-\omega_n|^2\,. \tag{7.16}$$

It is by now obvious that the state $|\psi\rangle$ is uniquely defined by the zeros of its Husimi function. It is also obvious that Q is positive and, from Eqs. (7.12) and (6.57), that

[3] This Hilbert space was presented by V. Bargmann (1961) in an influential paper.
[4] Some authors prefer the definition $Q_\psi(z) = (n+1)|\langle\psi|z\rangle|^2$.

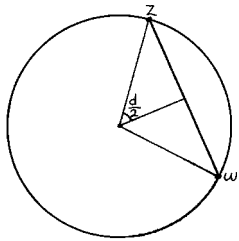


Figure 7.3. We know that $\sigma(z,\omega) = \sin^2\frac{d}{2}$; here we see that $\sin\frac{d}{2}$ equals one half of the chordal distance $d_{\rm ch}$ between the two points.

it is normalized to one:

$$\frac{n+1}{4\pi}\int \mathrm{d}\Omega\; Q_\psi(z) \;=\; 1\;. \tag{7.17}$$

Hence it provides a genuine probability distribution on the 2-sphere. It is bounded from above. Its maximum value has an interesting interpretation: the Fubini–Study distance between $|\psi\rangle$ and $|z\rangle$ is given by $D_{\rm FS} = \arccos\sqrt{\kappa}$, where $\kappa = |\langle\psi|z\rangle|^2 = Q_\psi(z)$, so the maximum of $Q_\psi(z)$ determines the minimum distance between $|\psi\rangle$ and the orbit of coherent states.

A convenient way to rewrite the Husimi function is

$$Q(z) = k_n\,\sigma(z,\omega_1)\sigma(z,\omega_2)\ldots\sigma(z,\omega_n)\;, \tag{7.18}$$

where

$$\sigma(z,\omega) \equiv \frac{|z-\omega|^2}{(1+|z|^2)(1+|\omega|^2)} = \frac{1-\cos d}{2} = \sin^2\frac{d}{2} = \frac{d_{\rm ch}^2}{4}\;, \tag{7.19}$$

d is the geodesic and $d_{\rm ch}$ is the chordal distance between the two points z and ω. (To show this, set $z = \tan\frac{\theta}{2}\mathrm{e}^{\mathrm{i}\phi}$ and $\omega = 0$. Simple geometry now tells us that $\sigma(z,\omega)$ is one quarter of the square of the chordal distance $d_{\rm ch}$ between the two points, assuming the sphere to be of unit radius. See Figure 7.3.) The factor k_n in Eq. (7.18) is a z-independent normalizing factor. Unfortunately it is a somewhat involved business to actually calculate k_n when n is large. For low values of n one finds, using the notation $\sigma_{kl} \equiv \sigma(\omega_k,\omega_l)$, that

$$k_2^{-1} = 1 - \frac{1}{2}\sigma_{12} \tag{7.20}$$

$$k_3^{-1} = 1 - \frac{1}{3}(\sigma_{12}+\sigma_{23}+\sigma_{31}) \tag{7.21}$$

$$\begin{aligned} k_4^{-1} = 1 &- \frac{1}{4}(\sigma_{12}+\sigma_{23}+\sigma_{31}+\sigma_{14}+\sigma_{24}+\sigma_{34}) \\ &+ \frac{1}{12}(\sigma_{12}\sigma_{34}+\sigma_{13}\sigma_{24}+\sigma_{14}\sigma_{23})\;. \end{aligned} \tag{7.22}$$

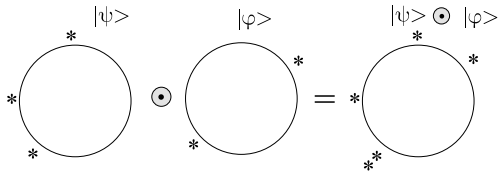


Figure 7.4. Composing states by adding stars.

For general n the answer can be given as a sum of a set of symmetric functions of the squared chordal distances σ_{kl} (Lee, 1988).

To get some preliminary feeling for Q we compute it for the Dicke states $|\psi_k\rangle$, that is for states that have the single component $Z^k = 1$ and all others zero. We find

$$Q_{|\psi_k\rangle}(z) = \binom{n}{k} \frac{|z|^{2k}}{(1+|z|^2)^n} = \binom{n}{k} \Big(\cos\big(\frac{\theta}{2}\big)\Big)^{2(n-k)} \Big(\sin\big(\frac{\theta}{2}\big)\Big)^{2k} , \tag{7.23}$$

where we switched to polar coordinates on the sphere in the last step. The zeros sit at $z = 0$ and at $z = \infty$, that is at the north and south poles of the sphere – as we knew they would. When $k = 0$ we have an $m = j$ state, all the zeros coincide, and the function is concentrated around the north pole. This is a coherent state. If n is even and $k = n/2$ we have an $m = 0$ state, and the function is concentrated in a band along the equator. The indication, then, is that the Husimi function tends to be more spread out the more the state differs from being a coherent one. As a first step towards confirming this conjecture we will compute the moments of the Husimi function. But before doing so, let us discuss how it can be used to *compose* two states.

The tensor product of an $N = (n+1)$-dimensional Hilbert space with itself is

$$\mathcal{H}_N \otimes \mathcal{H}_N = \mathcal{H}_{2N-1} \oplus \mathcal{H}_{2N-3} \oplus \cdots \oplus \mathcal{H}_1 . \tag{7.24}$$

Given two states $|\psi_1\rangle$ and $|\psi_2\rangle$ in $\mathcal{H}_N$ we can define a state $|\psi_1\rangle \odot |\psi_2\rangle$ in the tensor product space by the equation

$$Q_{|\psi_1\rangle\odot|\psi_2\rangle} \propto Q_{|\psi_1\rangle} Q_{|\psi_2\rangle} . \tag{7.25}$$

We simply add the stars of the original states to obtain a state described by $2(N-1)$ stars. This state clearly sits in the subspace $\mathcal{H}_{2N-1}$ of the tensor product space. This operation becomes particularly interesting when we compose a coherent state with itself. The result is again a state with all stars coinciding, and moreover all states with $2N-1$ coinciding stars arise in this way. Reasoning along this line one quickly sees that

$$\frac{2n+1}{4\pi} \int \mathrm{d}\Omega \, |z\rangle|z\rangle\langle z|\langle z| = \mathbb{1}_{2N-1} , \tag{7.26}$$

where $\mathbb{1}_{2N-1}$ is the projector, in $\mathcal{H}_N \otimes \mathcal{H}_N$, onto $\mathcal{H}_{2N-1}$. Similarly

$$\frac{3n+1}{4\pi}\int \mathrm{d}\Omega\ |z\rangle|z\rangle|z\rangle\langle z|\langle z|\langle z| = \mathbb{1}_{3N-2}\ . \tag{7.27}$$

And so on. These are useful facts.

Thus equipped we turn to the second moment of the Husimi function:

$$\frac{n+1}{4\pi}\int \mathrm{d}\Omega\ Q^2 = \frac{n+1}{2n+1}\langle\psi|\langle\psi|\frac{2n+1}{4\pi}\int \mathrm{d}\Omega\ |z\rangle|z\rangle\langle z|\langle z|\psi\rangle|\psi\rangle \le \frac{n+1}{2n+1} \tag{7.28}$$

with equality if and only if $|\psi\rangle|\psi\rangle \in \mathcal{H}_{2N-1}$, that is by the preceding argument if and only if $|\psi\rangle$ is a coherent state. For the higher moments one shows in the same way that

$$\frac{n+1}{4\pi}\int \mathrm{d}\Omega\ Q^p \le \frac{n+1}{pn+1} \quad \text{i.e.} \quad \frac{pn+1}{4\pi}\int \mathrm{d}\Omega\ Q^p \le 1\ . \tag{7.29}$$

We will use these results later.[5] For the moment we simply observe that if we define the *Wehrl participation number* as

$$R = \left(\frac{n+1}{4\pi}\int \mathrm{d}\Omega\ Q^2\right)^{-1}\ , \tag{7.30}$$

and if we take this as a first measure of delocalization, then the coherent states have the least delocalized Husimi functions (Schupp, 1999; Gnutzmann and Życzkowski, 2001).

The Husimi function can be defined for mixed states as well, by

$$Q_\rho(z) = \langle z|\rho|z\rangle\ . \tag{7.31}$$

The density matrix ρ can be written as a convex sum of projectors $|\psi_i\rangle\langle\psi_i|$, so the Husimi function of a mixed state is a sum of polynomials up to a common factor. It has no zeros, unless there is a zero that is shared by all the polynomials in the sum. Let us order the eigenvalues of ρ in descending order, $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_N$. The largest (smallest) eigenvalue gives a bound for the largest (smallest) projection onto a pure state. Therefore it will be true that

$$\max_{z\in S^2} Q_\rho(z) \le \lambda_1 \quad \text{and} \quad \min_{z\in S^2} Q_\rho(z) \ge \lambda_N\ . \tag{7.32}$$

These inequalites are saturated if the eigenstate of ρ corresponding to the largest (smallest) eigenvalue happens to be a coherent state. The main conclusion is that the Husimi function of a mixed state tends to be flatter than that of a pure state, and generically it is nowhere zero.

[5] Note that a somewhat stronger result is available; see Bodmann (2004).

7.4 Wehrl entropy and the Lieb conjecture

We can go on to define the Wehrl entropy (Wehrl, 1978; Wehrl, 1979) of the state $|\psi\rangle$ by

$$S_W(|\psi\rangle\langle\psi|) \equiv -\frac{n+1}{4\pi}\int_\Omega \mathrm{d}\Omega\; Q_\psi(z)\ln Q_\psi(z)\,. \tag{7.33}$$

One of the key properties of the Q-function on the plane was that (as proved by Lieb) the Wehrl entropy attains its minimum for the coherent states. Clearly we would like to know whether the same is true for the Q-function on the sphere. Consider the coherent state $|z_0\rangle = |j, j\rangle$ with all stars at the south pole. Its Husimi function is given in Eq. (7.23), with $k = 0$. The integration (7.33) is easily performed (try the substitution $x = \cos^2\!\left(\frac{\theta}{2}\right)$!) and one finds that the Wehrl entropy of a coherent state is $n/(n+1)$. The **Lieb conjecture** (Lieb, 1978) states that

$$S_W(|\psi\rangle\langle\psi|) \geq \frac{n}{n+1} = \frac{2j}{2j+1} \tag{7.34}$$

with equality if and only if $|\psi\rangle$ is a coherent state. It is also easy to see that the Wehrl entropy of the maximally mixed state $\rho_* = \frac{1}{n+1}\mathbb{1}$ is $S_W(\rho_*) = \ln(n+1)$; given that that the Wehrl entropy is concave in ρ this provides us with a rough upper bound.

The integral that defines S_W can be calculated because the logarithm factorizes the integral.[6] In effect

$$S_W = -\frac{n+1}{4\pi}\int_\Omega \mathrm{d}\Omega\; Q(z)\left(\ln k_n + \sum_{i=1}^{n}\ln\bigl(\sigma(z,\omega_i)\bigr)\right)\,. \tag{7.35}$$

The answer is again given in terms of various symmetric functions of the squares of the chordal distances. We make the definitions:

$$\textcircled{1} = \sum_{i<j}\sigma_{ij} \tag{7.36}$$

$$\textcircled{2} = \sum_{k=1}^{n}\sum_{i<j}\sigma_{ik}\sigma_{jk} \tag{7.37}$$

$$\textcircled{3} = \sum_{l=1}^{n}\sum_{i<j<k}\sigma_{il}\sigma_{jl}\sigma_{kl}\,. \tag{7.38}$$

The notation is intended to make it easier to remember the structure of the various functions (Schupp, 1999). As n grows we will need more of them. For $n = 4$:

$$\textcircled{11} = \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23} \tag{7.39}$$

[6] This feat was performed by Lee (1988). Unfortunately the answer looks so complicated that we do not quote it in full here. We will however sketch his proof that S_W assumes a local minimum at the coherent states.

For arbitrary n, sum over all quadratic terms such that all indices are different; but it is becoming evident that it will be labourious even to write down all the symmetric functions that occur for high values of n (Lee, 1988). Anyway, with this notation

$$n = 2: \quad S_W = k_2 \left(\frac{2}{3} + \frac{1}{6} \sigma_{12} \right) - \ln k_2 = k_2 \left(\frac{2}{3} + \frac{1}{6} \textcircled{1} \right) - \ln k_2 \,, \tag{7.40}$$

$$n = 3: \quad S_W = k_3 \left(\frac{3}{4} + \frac{1}{12} \textcircled{1} \;\; - \frac{1}{6} \textcircled{2} \right) - \ln k_3 \,, \tag{7.41}$$

$$n = 4: \quad S_W = k_4 \left(\frac{4}{5} + \frac{1}{20} \textcircled{1} \;\; - \frac{13}{180} \textcircled{3} \;\; - \frac{1}{12} \textcircled{2} \;\; - \frac{1}{24} \textcircled{4} \right) - \ln k_4 \,. \tag{7.42}$$

For $n = 2$ it is easy to see that S_W assumes its minimum when $\sigma_{12} = 0$, that is when the zeros coincide and the state is coherent (Scutaru, n.d.). In fact one can also show that Lieb's conjecture is true for $n = 3$ (Schupp, 1999). The first non-trivial case is $n = 4$: we are facing a very difficult optimization problem because the σ_{kl} are constrained by the requirement that they can be given in terms of the chordal distances between n points on a sphere.

From a different direction Bodmann (2004) has shown that

$$S_W \;\geq\; n \ln \left(1 + \frac{1}{n+1} \right) . \tag{7.43}$$

The conjecture therefore holds in the limit of large n, not surprisingly since in some sense it then goes over to the known result for canonical coherent states. But in general the Lieb conjecture remains open.

A complementary problem is to ask for states that maximize the Wehrl entropy. For $n = 1$ all pure states are coherent, so the question does not arise; $S_{\min} = S_{\max} = 1/2$. For $n = 2$ the maximal Wehrl entropy $S_{\max} = 5/3 - \ln 2$ is achieved for states whose stars are placed antipodally on the sphere. For $n = 3$ it is states whose stars are located on an equilateral triangle inscribed in a great circle and $S_{\max} = 21/8 - 2 \ln 2$. Problem 7.2 provides some further information, but the general problem of finding such maximally delocalized states for arbitrary n is still open. In this direction let us also observe that three stars placed on an equilateral triangle on a great circle of the sphere correspond to states that saturate the upper limit of the uncertainty relation (6.61). This is also true for four stars placed at the corner of a regular tetrahedron, but the story becomes more complicated when $n \geq 5$ (Davis, Delbourgo and Jarvis, 2000).

It may be remarked that considerable effort has been spent on a problem with a somewhat similar flavour, namely that of optimizing the potential energy

$$E_s \;=\; \sum_{k<l} \sigma_{kl}^s \tag{7.44}$$

for n points on the sphere. The case $s = -\frac{1}{2}$ corresponds to electrostatic interaction and is of interest both to physicists concerned with Thomson's 'plum pudding' model of the atom (assuming anyone is left) and to chemists concerned with buckminster fullerenes (molecules like C_{60}).[7] Although this problem appears to be much simpler than the Lieb conjecture it has many open ends. The minima do tend to be regular configurations, but as a matter of fact neither the cube (for $n = 8$) nor the dodecahedron (for $n = 12$) are minima. It is also known that when the number of point charges is large there tends to be many local minima of nearly degenerate energy. If the experience gained from this problem, and others like it, is to be trusted then we expect states that maximize the Wehrl entropy to form interesting and rather regular patterns when looked at in the stellar representation.

With this background information in mind we can sketch Lee's proof that the coherent states provide local minima of the Wehrl entropy. For this it is enough to expand S_W to second order in the σ_{ij} and this can be done for all n. The answer is

$$S_W = \frac{n}{n+1} + \frac{1}{2n^2}\textcircled{1}\ \textcircled{1} \ -\frac{n-2}{n(n-1)^2}\textcircled{11} \ -\frac{1}{n(n-1)}\textcircled{2} \ + o\big(\sigma_{il}^3\big)\,. \quad (7.45)$$

Next we expand the position (θ_i, ϕ_i) of the n zeros around their average position (θ_0, ϕ_0), using polar angles as coordinates. Thus

$$\Delta\theta_i = \theta_i - \theta_0 \quad \Rightarrow \quad \sum_i \Delta\theta_i = \sum_i \theta_i - n\theta_0 = 0\,, \quad (7.46)$$

and similarly for $\Delta\phi_i$. To lowest non-trivial order a short calculation gives

$$\sigma_{il} = \sin^2\frac{d_{il}}{2} \approx \frac{1}{4}(\Delta\theta_i - \Delta\theta_l)^2 + \frac{1}{4}\sin^2\theta_0(\Delta\phi_i - \Delta\phi_l)^2\,. \quad (7.47)$$

Finally a long calculation gives (Lee, 1988)

$$S_W \approx \frac{n}{n+1} + \frac{1}{32(n-1)^2}(F_1 - F_3)^2 + \frac{1}{8(n-1)^2}F_2^2\,, \quad (7.48)$$

where

$$F_1 = \sum_i (\Delta\theta_i)^2, \quad F_2 = \sin\theta_0 \sum_i (\Delta\theta_i)(\Delta\phi_i), \quad F_3 = \sin^2\theta_0 \sum_i (\Delta\phi_i)^2\,. \quad (7.49)$$

Evidently Eq. (7.48) implies that the coherent states form a (quite shallow) local minimum of S_W.

[7] For a review of this problem, with entries to the literature, see Saff and Kuijlaars (1997).

7.5 Generalized Wehrl entropies

One can also formulate the Lieb conjecture for the generalized entropies discussed in Section 2.7, which provide alternative measures of localization of a quantum state in the phase space. All generalized entropies depend on the shape of the Husimi function only, and not on where it may be localized.

For instance, one may consider the Rényi–Wehrl entropies, defined according to (2.79),

$$S_q^{RW}(\psi) \;=\; \frac{1}{1-q}\ln\left[\frac{n+1}{4\pi}\int_\Omega \mathrm{d}\Omega\,\big(Q_\psi(z)\big)^q\right], \tag{7.50}$$

and conjecture that their minima are attained for coherent states. A proof for all q would imply the proof of the original Lieb conjecture (in the limit $q \to 1$) and would be difficult, but it is encouraging that this modified Lieb conjecture has been proved for $q = 2,\ 3, \ldots$, and moreover in two different ways (Schupp, 1999; Gnutzmann and Życzkowski, 2001).

The easy way is to rely on Eq. (7.29). For q positive the Rényi–Wehrl entropy is smallest when the qth moment of the Q-function is maximal, and for $q = 2,\ 3, \ldots$ we already know that this happens if and only if the state is coherent. In this way we get

$$S_q^{RW}(|z\rangle) \;\geq\; \frac{1}{1-q}\ln\Big(\frac{n+1}{qn+1}\Big), \qquad q = 2, 3, \ldots , \tag{7.51}$$

with equality if and only if the state is coherent.

Before we go on, let us define the *digamma function*

$$\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}\,. \tag{7.52}$$

For any integer $m > n$ it enjoys the property that

$$\Psi(m) - \Psi(n) = \sum_{k=n}^{m-1}\frac{1}{k}\,. \tag{7.53}$$

When x is large it is true that $\Psi(x+1) \sim \ln x + 1/2x$ (Spanier and Oldham, 1987).

To get some feeling for how the Wehrl entropies behave, let us look at the eigenstates $|j, m\rangle$ of the angular momentum operator J_z. We computed their Husimi functions in eq. (7.23). They do not depend on the azimuthal angle ϕ, which simplifies things. Direct integration (Gnutzmann and Życzkowski, 2001) gives

$$\begin{aligned} S_W(|j,m\rangle) = \frac{2j}{2j+1} - \ln\binom{2j}{j-m} + 2j\,\Psi(2j+1) \\ - (j+m)\Psi(j+m+1) - (j-m)\Psi(j-m+1)\,, \end{aligned} \tag{7.54}$$

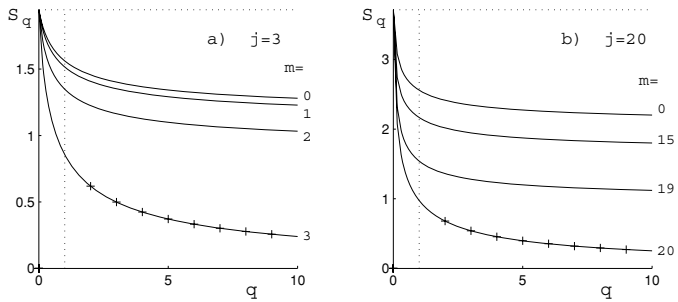


Figure 7.5. Rényi–Wehrl entropy S_q of the J_z eigenstates $|j,m\rangle$ for a) $j=3$ and b) $j=20$. For $m=j$ the states are coherent, and the crosses show integer values of q, for which the generalized Lieb conjecture is proven.

and

$$S_q^{RW}(|j,m\rangle) = \frac{1}{1-q}\ln\Big[\frac{2j+1}{2qj+1}\binom{2j}{j-m}^q \frac{\Gamma[q(j+m)+1]\Gamma[q(j-m)+1]}{\Gamma(2qj+1)}\Big], \tag{7.55}$$

For $m=0$ the state is localized on the equator of the Bloch sphere. A Stirling-like expansion (Spanier and Oldham, 1987), $\ln(k!)\approx(k+1/2)\ln k - k + \ln\sqrt{2\pi}$, allows us to find the asymptotes, $S_{|m=0\rangle}\sim \frac{1}{2}\ln N + \frac{1}{2}(1+\ln\pi/2)$. Interestingly, the mean Wehrl entropy of the eigenstates of J_z behaves similarly;

$$\frac{1}{2j+1}\sum_{m=-j}^{m=j} S_W(|j,m\rangle) = j - \frac{1}{2j+1}\sum_{m=-j}^{m=j}\ln\binom{2j}{j-m} \approx \frac{1}{2}\ln N + \ln(2\pi) - \frac{1}{2}. \tag{7.56}$$

A plot of some Rényi–Wehrl entropies obtained for the angular momentum eigenstates is shown in Figure 7.5. Any Rényi entropy is a continuous, non-increasing function of the Rényi parameter q. The fact that it attains minimal values for coherent states, may suggest that for these very states the absolute value of the derivative $\mathrm{d}S/\mathrm{d}q|_{q=0}$ is maximal.

So far we had $SU(2)$ coherent states in mind, but the treatment is easily generalized to the (rather special) $SU(K)$ coherent states $|z^{(K)}\rangle$ from Section 6.4. Then our 'skies' are $\mathbb{C}\mathbf{P}^{K-1}$, and they contain m stars if the dimension of the representation is $N_{K,m}$, as defined in Eq. (6.76). The Husimi function is normalized by

$$N_{K,m}\int_{\Omega_{K-1}} \mathrm{d}\Omega_{K-1}(z)\,|\langle\psi|z^{(K)}\rangle|^2 = 1\,, \tag{7.57}$$

where $\mathrm{d}\Omega_{K-1}$ is the FS measure on $\mathbb{C}\mathbf{P}^{K-1}$, normalized so that $\int_{\Omega_n}\mathrm{d}\Omega_n = 1$.

The moments of the Husimi function can be bounded from above, using the same method that was used for $SU(2)$. In particular

$$M_2(\psi) \equiv N_{K,n} \int_{\Omega_{K-1}} \mathrm{d}\Omega_{K-1}(z) |\langle z|\psi\rangle|^4 \tag{7.58}$$
$$= \frac{N_{K,n}}{N_{K,2n}} N_{K,2n} \int_{\Omega_{K-1}} \mathrm{d}\Omega_{K-1}(z) |\langle z \otimes z|\psi \otimes \psi\rangle|^2 = \frac{N_{K,n}}{N_{K,2n}} ||P_{2n}|\psi \otimes \psi\rangle||^2,$$

where P_{2n} projects the space $N_{K,n} \otimes N_{K,n}$ into $N_{K,2n}$. The norm of the projection $||P_{2n}|\psi \otimes \psi\rangle||^2$ is smaller than $||\psi \otimes \psi\rangle||^2 = 1$ unless $|\psi\rangle$ is coherent, in which case $|\psi \otimes \psi\rangle = |\psi \odot \psi\rangle \in \mathcal{H}_{N_{K,2n}}$. Therefore $M_2(|\psi\rangle) \leq N_{K,n}/N_{K,2n}$. The same trick (Schupp, 1999; Sugita, 2003; Sugita, 2002) works for any integer $q \geq 2$, and we obtain

$$M_q(\psi) = N_{K,n} \int_{\Omega_{K-1}} \mathrm{d}\Omega_{K-1}(z) |\langle z|\psi\rangle|^{2q} \leq \frac{N_{K,n}}{N_{K,qn}} = \frac{\binom{K+n-1}{n}}{\binom{K+qn-1}{qn}}, \tag{7.59}$$

with equality if and only if the state $|\psi\rangle$ is coherent. By analytical continuation one finds that the Rényi–Wehrl entropy of an $SU(K)$ coherent state is

$$S_q^{RW}(|z^{(K)}\rangle) = \frac{1}{1-q} \ln \left[\frac{\Gamma(K+n)\Gamma(qn+1)}{\Gamma(K+qn)\Gamma(n+1)} \right] . \tag{7.60}$$

In the limit $q \to 1$ one obtains the Wehrl entropy of a coherent state (Słomczyński and Życzkowski, 1998; Jones, 1990),

$$S_W\left(|z^{(K)}\rangle\right) = n\left[\Psi(n+K) - \Psi(n+1)\right] . \tag{7.61}$$

Using Eq. (7.53) for the digamma function, the right-hand side equals $n/(n+1)$ for $K = 2$, as we knew already. It is natural to conjecture that this is an upper bound for the Wehrl entropy also for $K > 2$.

7.6 Random pure states

Up until now we were concerned with properties of individual pure states. But one may also define an ensemble of pure states and ask about the properties of a typical (random) state. In effect we are asking for the analogue of Jeffrey's prior (Section 2.6) for pure quantum states. There is such an analogue, determined uniquely by unitary invariance, namely the Fubini–Study measure that we studied in Section 4.7, this time normalized so that its integral over $\mathbb{C}\mathbf{P}^n$ equals one. Interestingly the measure splits into two factors, one describing a flat simplex spanned by the moduli squared $y_i = |\langle e_i|\psi\rangle|^2$ of the state vector, and another describing a flat torus of a fixed size, parametrized by the phases. In fact

$$\mathrm{d}\Omega_n = \frac{n!}{(2\pi)^n} \mathrm{d}y_1 \ldots \mathrm{d}y_n \, \mathrm{d}\nu_1 \ldots \mathrm{d}\nu_n \, , \tag{7.62}$$

or using our octant coordinates (4.69),

$$\mathrm{d}\Omega_n = \frac{1}{\pi^n}\prod_{k=1}^{n}\cos\vartheta_k\,(\sin\vartheta_k)^{2k-1}\,\mathrm{d}\vartheta_k\,\mathrm{d}\nu_k\ . \tag{7.63}$$

Integrating out all but one variable we obtain the probability distribution $P(y) = n(1-y)^{n-1}$, which is a special, $\beta = 2$ case of the general formula

$$P_\beta(y) = \frac{\Gamma\big((n+1)\beta/2\big)}{\Gamma\big(\beta/2\big)\Gamma\big(n\beta/2\big)}y^{\beta/2-1}(1-y)^{n\beta/2-1}\ . \tag{7.64}$$

For $\beta = 1$ this gives the distribution of components of real random vectors distributed according to the round measure on $\mathbb{R}\mathbf{P}^n$, while for $\beta = 4$ and $n+1$ even it describes the quaternionic case (Kuś, Mostowski and Haake, 1988). For large $N = n+1$ these distributions approach the χ^2 distributions with β degrees of freedom,

$$P_\beta(y) \approx N\chi^2_\beta(Ny) = \Big(\frac{\beta N}{2}\Big)^{\beta/2}\frac{1}{\Gamma(\beta/2)}y^{\beta/2-1}\exp\Big(-\frac{\beta N}{2}y\Big)\ . \tag{7.65}$$

When we use octant coordinates the distribution of independent random variables reads

$$P(\vartheta_k) = k\sin(2\vartheta_k)(\sin\vartheta_k)^{2k-2},\quad P(\nu_k) = \frac{1}{2\pi}\ . \tag{7.66}$$

One may convince oneself that $y = \cos^2\vartheta_n$, the last component of Eq. (4.69), is distributed according to $P\big(y\big) = n(\sin^2\vartheta_n)^{n-1} = (N-1)(1-y)^{N-2}$, in agreement with Eq. (7.64) for $\beta = 2$. To generate a random quantum state it is convenient to use auxiliary independent random variables ξ_k distributed uniformly in $[0, 1]$ and to set $\vartheta_k = \arcsin\big(\xi_k^{1/2k}\big)$.

Random pure states may be generated by:

- selecting $2n$ random variables according to Eq. (7.66) and using them as octant coordinates;
- taking the first row (column) of a random unitary matrix U distributed according to the Haar measure on $U(N)$;
- taking an eigenvector of a random Hermitian (unitary) matrix pertaining to GUE[8] or CUE[9] and multiplying it with a random phase;
- picking a set of N independent complex random numbers z_i, drawn according to the normal distribution, and rescaling them as $c_i = z_i/\big(\sum_{i=1}^{N}|z_i|^2\big)^{1/2}$ (Życzkowski and Sommers, 2001);

[8] Gaussian unitary ensemble of Hermitian matrices consisting of independent Gaussian entries; such a probability measure is invariant with respect to unitary rotations. As an introduction to random matrices we recommend the book by Mehta (1991) or a more recent opus by Forrester (2005).

[9] Circular unitary ensemble of unitary matrices distributed according to Haar measure on $U(N)$.

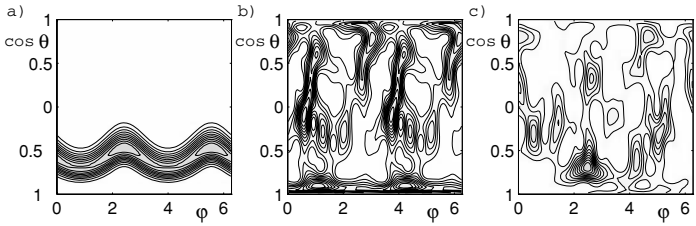


Figure 7.6. Husimi functions on the sphere (in Mercator's projection) for a typical eigenstate of a unitary operator describing (a) regular dynamics, (b) chaotic dynamics, and (c) a random pure state for $N = 60$. The Wehrl entropies read 2.05, 3.67, and 3.68, respectively. The maximal entropy is $\ln N \approx 4.094$.

- taking a row (column) of a random Hermitian matrix pertaining to GUE, normalizing it as above, and multiplying by a random phase.

To obtain real random vectors, distributed according to the Fubini–Study measure on $\mathbb{R}\mathbf{P}^n$, we use orthogonal matrices, or eigenvectors of random symmetric matrices, or we take a vector of N independent real Gaussian variables and normalize it.

Random pure states may be analysed in the Husimi representation. If the Fubini–Study measure on $\mathbb{C}\mathbf{P}^n$ is used, the states should not distinguish any part of the sphere; therefore the averaged density of zeros of the Husimi function for a random state will be uniform on the sphere (Lebœuf, 1991; Bogomolny, Bohigas and Lebœuf, 1992; Bogomolny, Bohigas and Lebœuf, 1996). However, this does not imply that each zero may be obtained as an independent random variable distributed uniformly on the sphere. On the contrary, the zeros of a random state are correlated as shown by Hannay (1996). In particular, the probability to find two zeros at small distance s behaves as $P(s) \sim s^2$ (Lebœuf and Shukla, 1996), while for independent random variables the probability grows linearly, $P(s) \sim s$.

The statistical properties of the Husimi zeros for random states are shared by the zeros representing eigenstates of unitary operators that give the one-step time evolution of quantized versions of classically chaotic maps (Lebœuf, 1991). Such states are delocalized, and their statistical properties (Haake, 2001) coincide with the properties of random pure states. Eigenstates of unitary operators giving the one-step time evolution of a regular dynamical system behave very differently, and their Husimi functions tend to be concentrated along curves in phase space (Lebœuf and Voros, 1990; Życzkowski, 2001). Figure 7.6 shows a selected eigenstate of the unitary operator $U = \exp(\mathrm{i}pJ_z)\exp(\mathrm{i}kJ_x^2/2j)$, representing a periodically kicked top (Kuś et al., 1988; Haake, 2001; Życzkowski, 2001) for $j = 29, 5$, $p = 1.7$ and the kicking strength $k = 0.7$ (a) and $k = 10.0$ (b). When $k = 0$ we have an angular momentum eigenstate, in which case the Husimi function assumes its maximum on a latitude circle; as the kicking strength grows the curve where the maximum

occurs begins to wiggle and the zeros start to diffuse away from the poles – and eventually the picture becomes similar to that for a random state.[10]

If the system in question enjoys a time-reversal symmetry, the evolution operator pertains to the circular orthogonal ensemble (COE), and the distribution of Husimi zeros of the eigenstates is no more uniform. The time-reversal symmetry induces a symmetry in the coefficients of the Bargmann polynomial (7.13), causing some zeros to cluster along a symmetry line on the sphere (Bogomolny et al., 1992; Prosen, 1996a; Braun, Kuś and Życzkowski, 1997).[11]

The time evolution of a pure state may be translated into a set of equations of motion for each star (Lebœuf, 1991). Parametric statistics of stars was initiated by Prosen (1996b), who found the distribution of velocities of stars of random pure states, while Hannay (1998) has shown how the Berry phase (discussed in Section 4.8) can be related to the loops formed by the stars during a cyclic variation of the state. Let us emphasize that the number of the stars of any quantum state in a finite-dimensional Hilbert space is constant during the time evolution (although they may coalesce), while for the canonical harmonic oscillator coherent states some zeros of the Husimi function may 'jump to infinity', so that the number of zeros may vary in time (see e.g. Korsch, Müller and Wiescher, 1997).

Let us now compute the generalized Wehrl entropy of a typical random pure state $|\psi\rangle \in \mathcal{H}_N$. In other words we are going to average the entropy of a pure state with respect to the Fubini–Study measure on $\mathbb{C}\mathbf{P}^n$. The normalization constant $N_{n+1,1}$ is defined in (6.76) and the Husimi function is computed with respect to the standard $SU(2)$ coherent states, so the average Wehrl moments are

$$\bar{M}_q = N_{n+1,1} \int_{\mathbb{C}\mathbf{P}^n} \mathrm{d}\Omega_n(\psi) M_q(\psi) = N_{n+1,1} \int_{\mathbb{C}\mathbf{P}^n} \mathrm{d}\Omega_n(\psi) \Big[N_{2,n} \int_{\mathbb{C}\mathbf{P}^1} \mathrm{d}\Omega_1(z) |\langle z|\psi\rangle|^{2q} \Big]. \tag{7.67}$$

It is now sufficient to change the order of integrations

$$\bar{M}_q = N_{2,n} \int_{\mathbb{C}\mathbf{P}^1} \mathrm{d}\Omega_1(z) \Big[N_{n+1,1} \int_{\mathbb{C}\mathbf{P}^n} \mathrm{d}\Omega_n(\psi) |\langle z|\psi\rangle|^{2q} \Big] = M_q(|z^{(K)}\rangle) \,, \tag{7.68}$$

and to notice that the integrals factorize: the latter gives the value of the Rényi moment (7.59) of a $SU(K)$ coherent state with $K = n + 1$, while the former is equal to unity due to the normalization condition. This result has a simple explanation: any state in an $N = (n + 1)$-dimensional Hilbert space can be obtained by the action of $SU(N)$ on the maximal weight state. Indeed all states of size N are $SU(N)$

[10] Similar observations were reported for quantization of dynamical systems on the torus (Lebœuf and Voros, 1990; Nonnenmacher, 1989).

[11] Mark Kac (1943) has considered a closely related problem: what is the expected number of real roots of a random polynomial with real coefficients? For a more recent discussion of this issue consult the very readable paper by Edelman and Kostlan (1995).

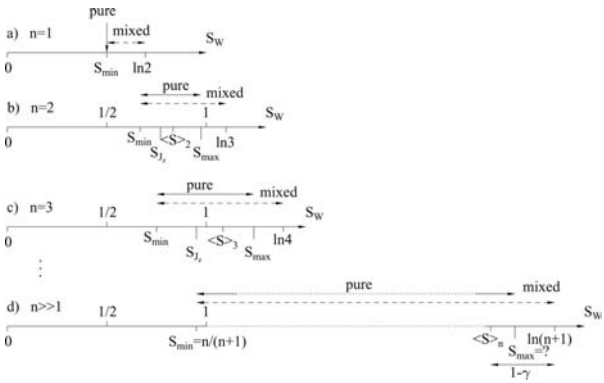


Figure 7.7. Range of the Wehrl entropy S_W for different dimensions $N = n + 1$ of the Hilbert space; γ is Euler's constant ≈ 0.58.

coherent. Yet another clarification: the state $|\psi\rangle$ is specified by $n = N - 1$ points on the sphere, or by a single point on $\mathbb{C}\mathbf{P}^n$. Equation (7.68) holds for any integer Rényi parameter q and due to concavity of the logarithmical function we obtain an upper bound for the average entropies of random pure states in $\mathbb{C}\mathbf{P}^n$

$$\bar{S}_q \equiv \left\langle S_q\left(|\psi\rangle\right)\right\rangle_{\mathbb{C}\mathbf{P}^{N-1}} \leq S_q(|z^{(N)}\rangle) = \frac{1}{1-q}\ln\left[\frac{\Gamma(N+1)\Gamma(q+1)}{\Gamma(N+q)}\right], \quad (7.69)$$

where the explicit formula (7.60) was used. Thus the Wehrl participation number of a typical random state is $1/\bar{M}_2 = (N+1)/2$. In the limit $q \to 1$ we can actually perform the averaging analytically by differentiating the averaged moment with respect to the Rényi parameter q,

$$\bar{S}_W = \langle -\lim_{q\to 1}\frac{\partial}{\partial q}M_q(\psi)\rangle = \frac{\partial}{\partial q}|_{q=1}\langle M_q(\psi)\rangle = \Psi(N+1) - \Psi(2) = \sum_{k=2}^{N}\frac{1}{k}. \quad (7.70)$$

This result gives the mean Wehrl entropy of a random pure state (Słomczyński and Życzkowski, 1998). In the asymptotic limit $N \to \infty$ the mean entropy $\bar{S}_W$ behaves as $\ln N + \gamma - 1$, where $\gamma \approx 0.5772\ldots$ is the Euler constant. Hence the average Wehrl entropy comes close to its maximal value $\ln N$, attained for the maximally mixed state.

The Wehrl entropy S_W allows us to quantify the *global* properties of a state in the classical phase space, say of an eigenstate $|\phi_k\rangle$ of an evolution operator U. To get information concerning local properties one may expand a given coherent state $|z\rangle$ in the eigenbasis of U. The Shannon entropy $S(\vec{y}_z)$ of the vector $y_k = |\langle z|\phi_k\rangle|^2$ provides some information on the character of the dynamics in the vicinity of the

classical point z (Życzkowski, 1990; Wootters, 1990; Haake, 2001). As shown by Jones (1990) the mean entropy, $\langle S\rangle_{\mathbb{C}\mathbf{P}^{N-1}}$, averaged over the FS measure on the set of complex random vectors given by Eq. (7.63) is just equal to Eq. (7.70), which happens to be the continuous, Boltzmann entropy (2.40) of the distribution (7.64) with $\beta = 2$. Since the variance $\langle(\Delta S)^2\rangle_{\mathbb{C}\mathbf{P}^{N-1}}$ behaves like $(\pi^2/3 - 3)/N$ (Wootters, 1990; Mirbach and Korsch, 1998), the relative fluctuations of the entropy decrease with the size N of the Hilbert space. Based on these results one may conjecture that if the corresponding classical dynamics in vicinity of the point z is chaotic, then the Shannon entropy of the expansion coefficients of the coherent state $|z\rangle$ is of order of $\langle S\rangle_{\mathbb{C}\mathbf{P}^{N-1}}$. In the opposite case, a value of $S(\vec{y}_z)$ much smaller than the mean value minus $\Delta S \sim 1/\sqrt{N}$ may be considered as a signature of a classically regular dynamics in the fragment of the phase space distinguished by the point $z \in \Omega$.

A generic evolution operator U has a non-degenerate spectrum, and the entropy $S(\vec{y}_z)$ acquires another useful interpretation (Thiele and Stone, 1984; Mirbach and Korsch, 1998): it is equal to the von Neumann entropy of the mixed state $\bar{\rho}$ obtained by the time average over the trajectory initiated from the coherent state,

$$\bar{\rho} \;=\; \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} U^t |z\rangle\langle z|(U^\dagger)^t \;. \tag{7.71}$$

To show this it is sufficient to expand the coherent state $|z\rangle$ in the eigenbasis of the evolution operator U and observe that the diagonal terms only do not dephase.

Let us return to the analysis of global properties of pure states in the classical phase space Ω. We have shown that for $N \gg 1$ the random pure states are delocalized in Ω. One can ask, if the Husimi distribution of a random pure state tends to the uniform distribution on the sphere in the semiclassical limit $N \to \infty$. Since strong convergence is excluded by the presence of exactly $N - 1$ zeros of the Husimi distribution, we will consider weak convergence only. To characterize this convergence quantitatively we introduce the L_2 distance between the Husimi distribution of the analysed state $|\psi\rangle$ and the uniform distribution $Q_* = 1/N$, normalized as $N \int Q_* \mathrm{d}\Omega_1 = 1$, representing the maximally mixed state $\rho_* = \mathbb{1}/N$,

$$L_2(\psi) \equiv L_2(Q_\psi, Q_*) = \Big(N \int_\Omega \Big[Q_\psi(z) - \frac{1}{N}\Big]^2 \mathrm{d}\Omega_1(z)\Big)^{1/2} = M_2(\psi) - \frac{1}{2j+1} \;. \tag{7.72}$$

Applying the previous result (7.68) we see that the mean L_2 distance to the uniform distribution tends to zero in the semiclassical limit,

$$\langle L_2(\psi)\rangle_{\mathbb{C}\mathbf{P}^{2j}} = \frac{2j}{(2j+1)(2j+2)} \;\to\; 0 \quad \text{as} \quad j \to \infty. \tag{7.73}$$

Thus the Husimi distribution of a typical random state tends, in the weak sense, to the uniform distribution on the sphere.

The L_2 distance to Q_* may be readily computed for any coherent state $|z\rangle$,

$$L_2(|z\rangle) \equiv L_2(Q_{|z\rangle}, Q_*) = \frac{j}{2j+1} \;\to\; \frac{1}{2}, \quad \text{as} \quad j \to \infty. \tag{7.74}$$

This outcome differs qualitatively from (7.73), which emphasizes the fact that coherent states are exceedingly non-typical states in a large dimensional Hilbert space. For random states the n stars cover the sphere uniformly, but the stars must coalesce in a single point to give rise to a coherent state. From the generalized Lieb conjecture proved for $q = 2$ it follows that the L_2 distance achieves its maximum for coherent states.

7.7 From the transport problem to the Monge distance

The stellar representation allows us to introduce a simple metric in the manifold of pure states with an interesting 'classical' property: the distance between two spin coherent states is equal to the Riemannian distance between the two points on the sphere where the stars of the states are situated. We begin the story be formulating a version of the famous **transport problem**:

Let n stars be situated at the n not necessarily distinct points x_i. The stars are moved so that they occupy n not necessarily distinct points y_i. The cost of moving a star from x_i to y_j is c_{ij}. How should the stars be moved if we want to minimize the total cost

$$T = \sum_{i=1}^{n} c_{i\pi(i)} \tag{7.75}$$

where $\pi(i)$ is some permutation telling us which star goes where?[12]

To solve this problem it is convenient to relax it, and consider the linear programming problem of minimizing

$$\tilde{T} = \sum_{i,j=1}^{N} c_{ij} B_{ij} \ , \tag{7.76}$$

where B_{ij} is any bistochastic matrix (Section 2.1). The minimum always occurs at a corner of the convex polytope of bistochastic matrices, that is for some permutation matrix, so a solution of the relaxed problem automatically solves the original problem. Since there are $n!$ permutation matrices altogether one may worry about the

[12] The problem is often referred to as the assignment problem. Clearly 'cost' suggests some application to economics; during the Second World War the setting was transport of munitions to fighting brigades.

growth of computer time with n. However, efficient algorithms exist where the amount of computer time grows as n^3 only (Wu and Coppins, 1981).

We can now try to define the distance between two pure states in a finite-dimensional Hilbert space as the minimal total cost $T_{\min}$ of transporting their stars into each other, where the cost c_{ij} is given by the geodesic distance between the stars on a sphere of an appropriate radius (chosen to be $1/2j$ if the dimension of the Hilbert space is $2j+1$). If we wish we can formulate this as a distance between two discrete probability distributions,

$$D_{\mathrm{dM}}(|\psi\rangle, |\phi\rangle) \equiv T_{\min}\big(P_1(|\psi_1\rangle), P_2(|\psi_2\rangle)\big) , \tag{7.77}$$

where any state is associated with the distribution $P = \sum_{i=1}^{n} \frac{1}{n}\delta(z - z_i)$, where z_i are the zeros of the Husimi function Q_ψ. This is a legitimate notion of distance because, by construction, it obeys the triangle inequality (Rachev and Rüschendorf, 1998). We refer to this as the *discrete Monge distance* D_{dM} between the states (Życzkowski and Słomczyński, 2001), since it is a discrete analogue of the continuous Monge distance that we will soon introduce. We observe that the discrete Monge distance between two coherent states (each represented by $2j$ coinciding stars) then becomes equal to the usual distance between two points on the unit sphere. In fact, locally the set of pure states will now have the same geometry as a product of 2-spheres almost everywhere, although the global properties are quite different.

Some further properties are worth noting. The discrete Monge distance between the states $|j, m\rangle$ and $|j, m'\rangle$ becomes

$$D_{\mathrm{dM}}(|j, m\rangle, |j, m'\rangle) = \frac{\pi}{2j}|m - m'| . \tag{7.78}$$

The set of eigenstates of J_z therefore form a metric line with respect to D_{dM}, while they form the corners of a simplex with respect to the Fubini–Study distance. It is also easy to see that all the eigenstates of J_x are situated at the distance $\pi/2$ from any eigenstate of J_z. Finally, consider two uncorrelated random states $|\psi_{\mathrm{rand}}\rangle$ and $|\phi_{\mathrm{rand}}\rangle$. Since the states are random their respective stars, carrying the weight $1/2j$ each, will be evenly distributed on the sphere. Its surface can be divided into $2j$ cells of diameter $\sim (2j)^{-1/2}$ and given a star from one of the states the distance to the nearest star of the other will be of the same order. Hence the discrete Monge distance between two random states will behave as

$$D_{\mathrm{dM}}(|\psi_{\mathrm{rand}}\rangle, |\phi_{\mathrm{rand}}\rangle) \sim \frac{2j}{(2j)^{3/2}} = \frac{1}{(2j)^{1/2}} \to 0 \qquad \text{as} \quad j \to \infty . \tag{7.79}$$

The discrete Monge distance between two random states goes to zero in the semi-classical limit.

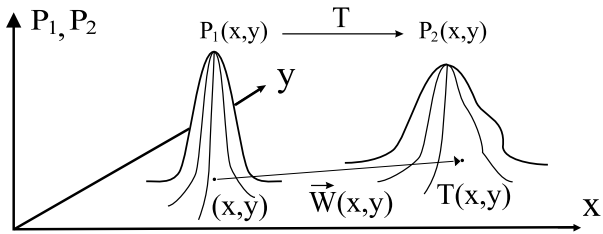


Figure 7.8. Monge transport problem: How to shovel a pile of soil $P_1(x, y)$ into a new location $P_2(x, y)$, minimizing the work done?

Using similar ideas, a more sophisticated notion of distance can be introduced between continuous probability distributions. The original **Monge problem**, formulated in 1781, emerged from studying the most effective way of transporting soil (Rachev, 1991):

Split two equally large volumes of soil into infinitely small particles and then associate them with each other so that the sum of the paths of the particles over the volume is least. Along which paths must the particles be transported and what is the smallest transportation cost?

Let $P_1(x, y)$ and $P_2(x, y)$ denote two probability densities, defined on the Euclidean plane, that describe the initial and the final location of the soil. Let the sets Ω_1 and Ω_2 denote the supports of both probability distributions. Consider smooth one-to-one maps $T : \Omega \to \Omega$ which generate volume preserving transformations Ω_1 into Ω_2, that is

$$P_1(x, y) \;=\; P_2\big(T(x, y)\big)\,\big|T'(x, y)\big| \tag{7.80}$$

for all points in Ω, where $T'(x, y)$ denotes the Jacobian of the map T at the point (x, y). We shall look for a transformation giving the minimal displacement integral (see Figure 7.8), and then we define the *Monge distance* between two probability distributions as (Rachev, 1991)

$$D_{\mathrm{M}}(P_1, P_2) \;\equiv\; \inf \int_{\Omega} |\vec{W}(x, y)|\; P_1(x, y)\, \mathrm{d}x\, \mathrm{d}y\ , \tag{7.81}$$

where $|\vec{W}(x, y)| = |(x, y) - T(x, y)|$ denotes the length of the path travelled – the Euclidean distance between a point in Ω_1 and its image. The infimum is taken over all volume preserving transformations T. The optimal transformation need not be unique; its existence can be guaranteed under reasonably general conditions. The generalization to curved spaces is straightforward; we just replace the Euclidean distance with the appropriate Riemannian one.

Note that in this formulation of the problem the vertical component of the soil movement is neglected, so perhaps the definition does not capture the essence of moving soil around. But we can use it to define the Monge distance between two probability distributions, as well as the Monge distance between pairs of quantum states as the Monge distance between their Husimi functions. A decided advantage is that, since the Husimi function exists for any density matrix, the definition applies to pure and mixed states alike:

$$D_{\text{Mon}}(\rho, \sigma) \equiv D_M\big(Q_\rho(z), Q_\sigma(z)\big) \,. \tag{7.82}$$

The definition depends on the choice of reference state for a set of coherent states, and it is not unitarily invariant. But in some circumstances this may be a desirable feature.

A price we have to pay is that the Monge distance is, in general, difficult to compute (Rachev, 1991; Rachev and Rüschendorf, 1998). Some things can be said right away however. If two Husimi functions have the same shape then the optimal transformation T is just a rigid translation, and the Monge distance becomes equal to the Riemannian distance with which our classical phase space is equipped. In particular this will be the case for pairs of coherent states. In some cases one may use symmetry to reduce the two-dimensional Monge problem to the one-dimensional case, for which an analytical solution due to Salvemini is known (see Problem 7.6). In this way it is possible to compute the Monge distance between two arbitrary Fock states (Życzkowski and Słomczyński, 1998). An asymptotic result, $D_{\text{Mon}}\big(|n\rangle, |m\rangle\big) \approx \frac{1}{2}\sum_{k=n+1}^{m} 1/\sqrt{k}$, has an intuitive interpretation: although all Fock states are orthogonal, and hence equidistant according to the Fubini–Study metric, the Monge distance shows that the state $|100\rangle$ is much closer to $|101\rangle$ than to the vacuum state $|0\rangle$. Using Bloch coherent states to define the Monge distance in finite-dimensional state spaces, one finds that the eigenstates of J_z form a metric line in the sense that

$$D_{\text{Mon}}\big(|j,-j\rangle, |j,j\rangle\big) = \sum_{m=-j+1}^{j} D_{\text{Mon}}\big(|j,m-1\rangle, |j,m\rangle\big) = \pi\Big[1 - \binom{2N}{N} 2^{1-2N}\Big] \,. \tag{7.83}$$

This is similar to the discrete Monge distance defined above (and very different from the Fubini–Study distance).

As in the discrete case the task of computing the Monge distance is facilitated by a relaxation of the problem (in this case due to Kantorovich (1942)), which is explicitly symmetric with respect to the two distributions. Using discretization one can make use of the efficient numerical methods that have been developed for the discrete Monge problem.

So what is the point? One reason why the Monge distance may be of interest is precisely the fact that it is not invariant under unitary transformations. This resembles the classical situation, in which two points in the phase space may drift away under the action of a given Hamiltonian system. Hence the discrete Monge distance (for pure states) and the Monge distance (for pure and mixed states) may be useful to construct a quantum analogue of the classical Lyapunov exponent and to elucidate various aspects of the quantum–classical correspondence.[13]

Problems

Problem 7.1 For the angular momentum eigenstates $|j, m\rangle$, find the minimal Fubini–Study distance to the set of coherent states.

Problem 7.2 For $N = 2, 3, 4, 5$, compute the Wehrl entropy S_W and the Wehrl participation number R for the J_z eigenstates, and for the states $|\psi_{\triangle}\rangle$, three stars in a equilateral triangle, and $|\psi_{\text{tetr.}}\rangle$, four stars forming a regular tetrahedron.

Problem 7.3 Show that the squared moduli of components $y_i = |\langle i|\psi\rangle|^2$ of a complex random vector (7.66) in $\mathbb{C}\mathbf{P}^n$ are distributed uniformly in Δ_n.

Problem 7.4 Show that the mean Fubini–Study distance of a random state $|\psi_{\text{rand}}\rangle$ to any selected pure state $|1\rangle$ reads

$$\big\langle D_{\text{FS}}(|1\rangle, |\psi_{\text{rand}}\rangle)\big\rangle_{\mathbb{C}\mathbf{P}^n} = \frac{\pi}{2} - \frac{\sqrt{\pi}\,\Gamma(n+1/2)}{2\,\Gamma(n+1)}\,, \tag{7.84}$$

which equals $\pi/4$ for $n = 1$ and tends to $\pi/2$ in for $n \to \infty$.

Problem 7.5 Show that the average discrete Monge distance of a random state $|\psi_{\text{rand}}\rangle$ to the eigenstates $|j, m\rangle$ of J_z reads (Życzkowski and Słomczyński, 2001)

$$\big\langle D_{\text{dM}}(|j, m\rangle, |\psi_{\text{rand}}\rangle)\big\rangle_{\mathbb{C}\mathbf{P}^{2j}} = \chi \sin\chi + \cos\chi, \quad \text{where} \quad \chi = \frac{m\pi}{2j}\,. \tag{7.85}$$

Problem 7.6 Prove that the Monge distance, in the one-dimensional case, $\Omega = \mathbb{R}$, is given by the *Salvemini solution* (Salvemini, 1943; Rachev, 1991; Życzkowski and Słomczyński, 1998).

$$D_{\text{M}}(P_1, P_2) = \int_{-\infty}^{+\infty} |F_1(x) - F_2(x)|\,\mathrm{d}x\,, \tag{7.86}$$

where the distribution functions are $F_i(x) \equiv \int_{-\infty}^{x} P_i(t)\,\mathrm{d}t$ for $i = 1, 2$.

[13] For further details consult Życzkowski and Słomczyński (2001).

Problem 7.7 Making use of the Salvemini formula (7.86) analyse the space of $N = 2$ mixed states form the point of view of the Monge distance: (a) prove that $D_{\mathrm{Mon}}(\rho_+, \rho_-) = \pi/4$, $D_{\mathrm{Mon}}(\rho_+, \rho_*) = \pi/8$, and in general (b) prove that the Monge distance between any two mixed states of a qubit is proportional to their Hilbert–Schmidt distance and generates the geometry of the Bloch ball (Życzkowski and Słomczyński, 2001).

8

The space of density matrices

> Over the years, the mathematics of quantum mechanics has become more abstract and, consequently, simpler.
>
> *V. S. Varadarajan*

We have already introduced density matrices and made use of some of their properties. In general a complex $N \times N$ matrix is a density matrix if it is

$$\begin{array}{lcll} \text{i)} & \text{Hermitian,} & \rho = \rho^\dagger , & \\ \text{ii)} & \text{positive,} & \rho \geq 0 , & (8.1) \\ \text{iii)} & \text{normalized,} & \mathrm{Tr}\rho = 1 . & \end{array}$$

The middle equation is shorthand for the statement that all the eigenvalues of ρ are non-negative. The set of density matrices will be denoted $\mathcal{M}^{(N)}$. It is a convex set sitting in the vector space of Hermitian matrices, and its pure states are density matrices obeying $\rho^2 = \rho$. As explained in Section 4.5 the pure states form a complex projective space.

Why should we consider this particular convex set for our state space? One possible answer is that there is a point in choosing vectors that are also matrices, or a vector space that is also an algebra (a different way of saying the same thing). In effect our vectors now have an intrinsic structure, namely their spectra when regarded as matrices, and this enables us to define a positive cone, and a set of pure states, in a more interesting way than that used in classical probability theory. We will cast a glance at the algebraic framework towards the end of this chapter, but first we explore the structure as it is given to us.

8.1 Hilbert–Schmidt space and positive operators

This section will be devoted to some basic facts about complex matrices, leading up to the definition of the space $\mathcal{M}^{(N)}$. We begin with an N complex-dimensional

Hilbert space $\mathcal{H}$. There is then always a dual Hilbert space $\mathcal{H}^*$ defined as the space of linear maps from $\mathcal{H}$ to the complex numbers; in the finite-dimensional case these two spaces are isomorphic. Another space that is always available is the space of operators on $\mathcal{H}$. This is actually a Hilbert space in its own right when it is equipped with the Hermitian form

$$\langle A, B\rangle = c\mathrm{Tr}A^\dagger B \,, \tag{8.2}$$

where c is a real number that sets the scale. This is *Hilbert–Schmidt space* $\mathcal{HS}$; an alternative notation is $\mathcal{B}(\mathcal{H})$ where $\mathcal{B}$ stands for 'bounded operators'. All our operators are bounded, and all traces exist, but this is so only because all our Hilbert spaces are finite dimensional. The scalar product gives rise to an Euclidean distance, the *Hilbert–Schmidt distance*

$$D_2^2(A, B) \equiv \frac{1}{2}\mathrm{Tr}[(A-B)(A^\dagger - B^\dagger)] \equiv \frac{1}{2}D_{\mathrm{HS}}^2(A, B) \,. \tag{8.3}$$

In this chapter we set the scale with $c = 1/2$, and work with the distance D_2, while in Chapter 9 we use D_{HS}. As explained in Section 4.5 this ensures that we agree with standard conventions in quantum mechanics.

Chapter 9 will hinge on the fact that

$$\mathcal{HS} = \mathcal{H} \otimes \mathcal{H}^* \,. \tag{8.4}$$

This is a way of saying that any operator can be written in the form

$$A = a|P\rangle\langle Q| + b|R\rangle\langle S| + \cdots \,, \tag{8.5}$$

provided that enough terms are included in the sum. Given $\mathcal{H}$, the N^2 complex-dimensional space $\mathcal{HS}$ – also known as the algebra of complex matrices – is always with us.

Some brief reminders about (linear) operators may prove useful. First we recall that an operator A can be diagonalized by a unitary change of bases, if and only if it is *normal*, that is if and only if $[A, A^\dagger] = 0$. Examples include *Hermitian* operators for which $A^\dagger = A$ and *unitary* operators for which $A^\dagger = A^{-1}$. A normal operator is Hermitian if and only if it has real eigenvalues. Any normal operator can be written in the form

$$A = \sum_{i=1}^{r} z_i |e_i\rangle\langle e_i| \,, \tag{8.6}$$

where the sum includes orthogonal projectors corresponding to the r non-vanishing eigenvalues z_i. The eigenvectors $|e_i\rangle$ span a linear subspace $\mathrm{supp}(A)$ known as the *support* (or *range*) of the operator. There is an orthogonal subspace $\mathrm{kern}(A)$ called the *kernel*, consisting of all vectors $|\psi\rangle$ such that $A|\psi\rangle = 0$. For normal operators

the kernel is the subspace spanned by all eigenvectors with zero eigenvalues, and it has dimension $N - r$. The full Hilbert space may be expressed as the direct sum $\mathcal{H} = \text{kern}(A) \oplus \text{supp}(A)$.

The vector space $\mathcal{HM}$ of Hermitian operators is an N^2 real-dimensional subspace of Hilbert–Schmidt space. It can also – and this will turn out to be important – be thought of as the Lie algebra of $U(N)$. The $(N^2 - 1)$-dimensional subspace of Hermitian operators with zero trace is the Lie algebra of the group $SU(N)$. In Appendix 2 we give an explicit basis for the latter vector space, orthonormal with respect to the scalar product (8.2). If we add the unit matrix we see that a general Hermitian matrix can be written in the form

$$A = \tau_0 \sqrt{\frac{2}{N}} \mathbb{1} + \sum_{i=1}^{N^2-1} \tau_i \sigma_i \quad \Leftrightarrow \quad \tau_0 = \frac{\text{Tr}A}{\sqrt{2N}} , \quad \tau_i = \frac{1}{2}\text{Tr}\sigma_i A , \tag{8.7}$$

where σ_i are generators of $SU(N)$ obeying

$$\sigma_i \sigma_j = \frac{2}{N}\delta_{ij} + d_{ijk}\sigma_k + \mathrm{i} f_{ijk}\sigma_k . \tag{8.8}$$

Here f_{ijk} is totally anti-symmetric in its indices and d_{ijk} is totally symmetric (and vanishing if and only if $N = 2$). The use of the notation (τ_0, τ_i) for the Cartesian coordinates in this basis is standard. Of course there is nothing sacred about the basis that we have chosen; an orthonormal basis consisting of N^2 elements of equal trace may be a better choice.

We will now use the 'internal structure' of our vectors to define a positive cone. By definition, a *positive* operator P is an operator such that $\langle\psi|P|\psi\rangle$ is real and non-negative for all vectors $|\psi\rangle$ in the Hilbert space. This condition actually implies that the operator is Hermitian, so we can equivalently define a positive operator as a Hermitian matrix with non-negative eigenvalues. (To see this, observe that any matrix P can be written as $P = X + \mathrm{i}Y$, where X and Y are Hermitian – the argument fails if the vector space is real, in which case a positive operator is usually defined as a symmetric matric with non-negative eigenvalues.) The condition that an operator be positive can be rewritten in a useful way:

$$P \geq 0 \quad \Leftrightarrow \quad \langle\psi|P|\psi\rangle \geq 0 \quad \Leftrightarrow \quad P = AA^\dagger , \tag{8.9}$$

for all vectors $|\psi\rangle$ and for some matrix A. The set $\mathcal{P}$ of positive operators obeys

$$\mathcal{P} \subset \mathcal{HM} \qquad \dim[\mathcal{P}] = \dim[\mathcal{HM}] = N^2 . \tag{8.10}$$

Since it is easy to see that a convex combination of positive operators is again a positive operator, the set $\mathcal{P}$ is a convex cone (in the sense of Section 1.1).

A positive operator admits a unique positive square root $\sqrt{P}$,

$$(\sqrt{P})^2 = P . \tag{8.11}$$

For an arbitrary (not necessarily Hermitian) operator A we can define the positive operator $AA^\dagger$, and then the *absolute value* $|A|$:

$$|A| \equiv \sqrt{AA^\dagger} \tag{8.12}$$

This is in itself a positive operator. Furthermore, any linear operator A can be decomposed into polar form, which means that

$$A = |A|U = \sqrt{AA^\dagger}\,U , \tag{8.13}$$

where U is a unitary operator which is unique if A is invertible. This *polar decomposition* is analogous to the representation $z = re^{i\phi}$ of a complex number. The modulus r is non-negative; so are the eigenvalues of the positive operator $|A|$, which are known as the *singular values* of A. (There exists an alternative left polar decomposition, $A = U\sqrt{A^\dagger A}$, which we will not use. It leads to the same singular values.) The polar decomposition can be used to prove that any operator, normal or not, admits a *singular values decomposition* (SVD)

$$A = UDV , \tag{8.14}$$

where U and V are unitary, and D is diagonal with non-negative entries which are precisely the singular values of A.[1]

Evidently the question whether a given Hermitian matrix is positive is raising its head – an ugly head since generally speaking it is not easy to diagonalize a matrix and investigate its spectrum. A helpful fact is that the spectrum can be determined by the traces of the first N powers of the matrix. To see how, look at the *characteristic equation*

$$\det(\lambda\mathbb{1} - A) = \lambda^N - s_1\lambda^{N-1} + s_2\lambda^{N-2} - \cdots + (-1)^N s_N = 0 . \tag{8.15}$$

If the matrix is in diagonal form the coefficients s_k – that clearly determine the spectrum – are the elementary symmetric functions of the eigenvalues,

$$s_1 = \sum_i \lambda_i , \qquad s_2 = \sum_{i<j} \lambda_i\lambda_j , \qquad s_3 = \sum_{i<j<k} \lambda_i\lambda_j\lambda_k , \quad \ldots \tag{8.16}$$

and so on. If the matrix is not in diagonal form we can still write

$$s_1 = \mathrm{Tr}A \qquad s_2 = \frac{1}{2}(s_1\mathrm{Tr}A - \mathrm{Tr}A^2) \tag{8.17}$$

and in general, iteratively,

$$s_k = \frac{1}{k}(s_{k-1}\mathrm{Tr}A - s_{k-2}\mathrm{Tr}A^2 + \cdots + (-1)^{k-1}\mathrm{Tr}A^k) . \tag{8.18}$$

[1] In popular numerical routines for computing the SVD of an arbitrary matrix one obtains the vector of singular values as well as the matrices U and V. An interesting review of properties of SVD may be found in Horn and Johnson (1985, 1991).

(Proof: Diagonalize the matrix. Since the traces are not affected by diagonalization it is only a matter of comparing our two expressions for the coefficients.) Returning to the question whether A is positive, it can be proved that A is positive if and only if all the coefficients s_k in the characteristic equation are positive. This is encouraging since the criterion only requires the calculation of traces, but it remains a lengthy business to apply it to a given A. There is no easy way.

8.2 The set of mixed states

Finally we come to the density matrices. The set of density matrices consists of all positive operators ρ with unit trace, $\mathrm{Tr}\rho = 1$. We denote it by $\mathcal{M}$, or by $\mathcal{M}^{(N)}$ if we want to emphasize that it consists of $N \times N$ matrices. $\mathcal{M}$ is the intersection, in the space of Hermitian matrices, of the positive cone $\mathcal{P}$ with a hyperplane parallel to the linear subspace of traceless operators. It is a convex set in its own right, whose pure states are projectors onto one-dimensional subspaces in $\mathcal{H}$, that is density matrices of the form

$$\rho = |\psi\rangle\langle\psi| \quad \Leftrightarrow \quad \rho^{\alpha}_{\ \beta} = Z^{\alpha}\,\bar{Z}_{\beta} \quad \Leftrightarrow \quad \rho^2 = \rho\ . \tag{8.19}$$

In this chapter we assume that the vectors are normalized. As we observed in Section 4.5 this is an isometric embedding of $\mathbb{C}\mathbf{P}^{N-1}$ in $\mathcal{HM}$.

Two different ways of coordinatizing $\mathcal{M}^{(N)}$ spring to mind. We can set

$$\rho = \frac{1}{N}\,\mathbb{1} + \sum_{i=1}^{N^2-1} \tau_i \sigma_i\ . \tag{8.20}$$

What we have done is to identify $\mathcal{M}^{(N)}$ with a subset of the Lie algebra of $SU(N)$, by shifting the origin of $\mathcal{HM}$ from the zero matrix to the matrix

$$\rho_\star \ \equiv\ \frac{1}{N}\,\mathbb{1}\ . \tag{8.21}$$

This is a special density matrix known as the *maximally mixed state*; it is also known as the *tracial state* or, familiarly, as the 'matrix of ignorance'. The components τ_i of the *Bloch vector* serve as Cartesian coordinates in $\mathcal{M}^{(N)}$ and are also known as *mixture coordinates*. A convex combination of two density matrices lies on a manifestly straight line, when these coordinates are used. Moreover the definition (8.3), with (A2.9), implies that

$$D_2(\rho, \rho') = D_2\left(\sum_i \tau_i\sigma_i, \sum_j \tau'_j\sigma_j\right) \ =\ \sqrt{\sum_i (\tau_i - \tau'_i)^2}\ , \tag{8.22}$$

which is the familiar formula in an Euclidean space.

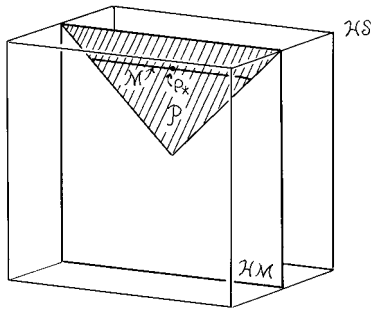


Figure 8.1. The spaces discussed in this chapter. Hilbert–Schmidt space $\mathcal{HS}$ only serves as background for the linear subspace $\mathcal{HM}$ of Hermitian matrices, that contains the positive cone $\mathcal{P}$ and the set of density matrices $\mathcal{M}$. But compare with Figure 9.1 in the next chapter!

There is another way of identifying $\mathcal{M}^{(N)}$ with a subset of the Lie algebra of $SU(N)$, namely to set

$$\rho = \frac{\mathrm{e}^{-\beta H}}{\mathrm{Tr}\mathrm{e}^{-\beta H}} ; \qquad H = \sum_{i=1}^{N^2-1} x_i \sigma_i \ . \tag{8.23}$$

The coordinates x_i are known as *exponential coordinates*; the real number β is known as the inverse temperature and serves as a reminder of the role these coordinates play in statistical mechanics. They are consistent with their own notion of straight line – the analogue of the exponential families of classical statistics that we studied in Section 3.2.

We would like to draw a more detailed picture of the set of density matrices than that offered in Figure 8.1. If we choose to work over the real rather than the complex numbers, then we can investigate the space of real symmetric 2×2 matrices. This has only three dimensions. The condition that the trace be unity defines a two-dimensional plane in this space, and we can literally see how it intersects the convex cone of positive matrices in a circular disc, which is the two-dimensional space of *rebits*. Physics requires complex qubits but rebits are simpler to look at. But already for $N = 3$ the set of real density matrices has dimension 5, which is too high for easy visualization.

Moving on to qubits, we find that the space $\mathcal{M}^{(2)}$ has three real dimensions. This is the Bloch ball discussed in Section 5.2. There is no particular difficulty in understanding a ball as a convex set. Physically however there is much to think about, because we now have two different ways of adding two pure states together. We can form a complex superposition (another pure state) and we can form a statistical mixture (a mixed state). In the other direction, any given point in the

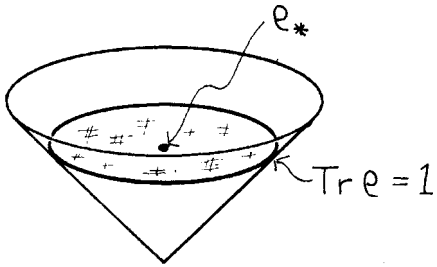


Figure 8.2. The cone of positive real symmetric 2×2 matrices and its intersection with the plane of Hermitian matrices of unit trace.

interior can be obtained as a mixture of pure states in many different ways. We are confronted with an issue that does not arise in classical statistics at all: any mixed state can be expressed as a mixture of pure states in many different ways, indeed in as many ways as a point in a ball can be thought of as the 'centre of mass' of a mass distribution on the surface of the ball. Physically it is a basic tenet of quantum mechanics that there does not exist an operational procedure to distinguish different ensembles of pure states if they yield the same density matrix – otherwise quantum correlations between separated systems could be used for instantaneous signalling (Herbert, 1982).

Quantum mechanics is a significant generalization of classical probability theory. When $N = 2$ there are two possible outcomes of any measurement described by a Hermitian operator, or put in another way the sample space belonging to a given observable consists of two points. They correspond to two orthogonal pure states, placed antipodally on the surface of the Bloch ball. Each pair of antipodal points on the surface corresponds to a new sample space coexisting with the original.

The Bloch ball is a convenient example to keep in mind since it is easy to visualize, but in some respects it is quite misleadingly simple. We begin to see this if we ask what conditions one has to put on the Bloch vector in order for a density matrix to describe a pure state. Using Eq. (8.8) it is straightforward to deduce that

$$\rho^2 = \rho \quad \Leftrightarrow \quad \begin{cases} \tau^2 = \frac{N-1}{2N} \\ (\vec{\tau} \star \vec{\tau})_i \equiv d_{ijk}\tau_j\tau_k = \frac{N-2}{N}\tau_i \end{cases} . \tag{8.24}$$

The first condition implies that the Bloch vector of a pure state is confined to an $(N^2 - 2)$-dimensional outsphere. The second condition arises only for $N > 2$, and it says that the pure states form a certain well-defined subset of the surface of

the outsphere.[2] We know that this subset is a complex projective space, of real dimension $2(N-1)$. When $N > 2$ this is a rather small subset of the outsphere.

8.3 Unitary transformations

It is convenient to think of the set of density matrices as a rigid body in $\mathbb{R}^{N^2-1}$. We think of $\mathbb{R}^{N^2-1}$ as an Euclidean space. When $N > 2$ our rigid body is not spherical, and we must try to understand its shape. The first question one asks about a rigid body is: what is its symmetry? What subgroup of rotations leaves it invariant? Unless the body is spherical, the symmetry group is a proper subgroup of the group of rotations. For the Platonic solids, the symmetry groups are discrete subgroups of $SO(3)$. Our case is subtler; although pure quantum states form only a small subset of the outsphere, they do form a continuous manifold, and the symmetry group is not discrete. As it happens there is an analogue of Wigner's theorem (Section 4.5) that applies to density matrices, and answers our question. Its assumptions concern only the convex structure, not the Euclidean distance:

Theorem 8.1 (Kadison's) *Let there be a map Φ from $\mathcal{M}$ to $\mathcal{M}$ which is one-to-one and onto, and which preserves the convex structure in the sense that*

$$\Phi\big(p\rho_1 + (1-p)\rho_2\big) = p\Phi(\rho_1) + (1-p)\Phi(\rho_2)\,. \tag{8.25}$$

Then the map must take the form

$$\Phi(\rho) \;=\; U\rho\,U^{-1}\,, \tag{8.26}$$

where the operator U is either unitary or anti-unitary.

In infinitesimal form a unitary transformation takes the form

$$\dot\rho \;=\; \mathrm{i}\,[\rho, H]\,, \tag{8.27}$$

where H is an Hermitian operator (say, a Hamiltonian) that determines the one parameter family of unitary operators $U(t) = \mathrm{e}^{-\mathrm{i}Ht}$.

It is easy to see why something like this must be true.[3] Because the map preserves the convex structure it must be an affine map, and it must map pure states to pure states. For $N = 2$ the pure states form a sphere, and the only affine maps that preserve the sphere are rotations; the map must belong to $SO(3) = SU(2)/\mathbb{Z}_2$, that is, it must take the form (8.26). For $N > 3$ the pure states form just a subset of a

[2] The elegant 'star product' used here occurs now and then in the literature, for instance in Arvind, Mallesh and Mukunda (1997). As a general reference for this section we suggest Mahler and Weberruß (1995, (2nd. ed.) 1998).

[3] We will not give a strict proof of Kadison's theorem; a complete and elementary proof was given by Hunziker (1972).

sphere, and the rotation must be a rather special one – in particular, it must give rise to an isometry of the space of its pure states. The last condition points right at the unitary group.

In a sense Kadison's theorem answers all our questions, but we must make sure that we understand the answer. Because the body of density matrices sits in a vector space that we have identified with the Lie algebra of $SU(N)$, and because the only unitary transformations that have any effect on density matrices are those that belong to $SU(N)$, we have here an example of an *adjoint action* of a group on its own Lie algebra. What sets the scene is that $SU(N)/\mathbb{Z}_N$ is a subgroup of $SO(N^2-1)$. An explicit way to see this goes as follows: let ρ be defined as in Eq. (8.20) and let

$$\rho' \;=\; U\rho\, U^\dagger = \frac{1}{N}\,\mathbb{1} + \sum_i \tau_i U\sigma_i U^\dagger \;\equiv\; \frac{1}{N}\,\mathbb{1} + \sum_i \tau_i'\sigma_i \;. \tag{8.28}$$

Here τ_i' is the rotated Bloch vector. We compute that

$$\tau_i' = \frac{1}{2}\mathrm{Tr}\rho'\sigma_i = \frac{1}{2}\sum_j \mathrm{Tr}(\sigma_i U\sigma_j U^\dagger)\tau_j \equiv \sum_j O_{ij}\tau_j \;, \tag{8.29}$$

where the matrix O by definition has the matrix elements

$$O_{ij} \equiv \mathrm{Tr}\big(\sigma_i U\sigma_j U^\dagger\big) \;. \tag{8.30}$$

This must be an orthogonal matrix, and indeed it is. It is easy to check that the elements are real. Using the completeness relation (A2.10) for the generators it is fairly easy to check also that

$$(O\,O^T)_{ij} \;=\; \sum_k O_{ik}\,O_{jk} \;=\; \delta_{ij} \;. \tag{8.31}$$

In this way we have exhibited $SU(N)/\mathbb{Z}_N$ as a subgroup of the rotation group $SO(N^2-1)$.

Another way to see what goes on is to observe that rotations preserve the distance to the origin – in this case, to ρ_*. This means that they preserve

$$D_2^2(\rho,\rho_\star) = \frac{1}{2}\,\mathrm{Tr}\,(\rho-\rho_\star)^2 = \frac{1}{2}\,\mathrm{Tr}\rho^2 - \frac{1}{2N} \;. \tag{8.32}$$

Therefore all rotations preserve $\mathrm{Tr}\rho^2$. But unitary transformations preserve the additional traces $\mathrm{Tr}\rho$, $\mathrm{Tr}\rho^3$, $\mathrm{Tr}\rho^4$ and so on, up to the last independent trace $\mathrm{Tr}\rho^N$. We are dealing with very special rotations that preserve the spectrum of every density matrix.

Since rotations in arbitrary dimensions may be unfamiliar, let us first discuss the action of a generic rotation matrix. With a suitable choice of basis it can always

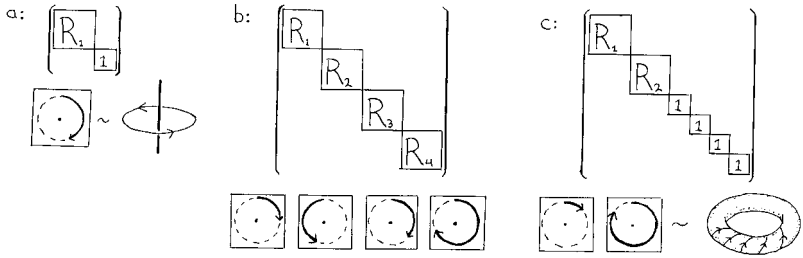


Figure 8.3. Rotations can be represented with 2×2 rotation matrices along the diagonal, or pictorially as rotations in totally orthogonal 2-planes. We show (a) an $SO(3)$ rotation, (b) a generic $SO(8)$ rotation, and (c) a generic $SU(3) \in SO(8)$ rotation.

be block diagonalized, that is to say it can be written with 2×2 rotation matrices occurring along the diagonal and zeros elsewhere. If the dimension is odd, we add an extra 1 on the diagonal. What this means is that for any rotation of $\mathbb{R}^n$ we can choose a set of totally orthogonal 2-planes such that the rotation can be described as independent rotations in these 2-planes; if n is odd there will always be, as Euler pointed out, an axis that is not affected by the rotation at all. A typical flow line is not a circle. If the dimension is either $2n$ or $2n + 1$ it will wind around a flat torus of dimension n, since there are n totally orthogonal 2-planes involved. Unless the rotations in these 2-planes are carefully adjusted, the resulting curve will not even be closed. (We came across this kind of thing in Section 4.6, and tried to draw it in Figure 4.11. Now we offer Figure 8.3.) A generic rotation has only one fixed point if the dimension of the space is even, and only one fixed axis if it is odd.

The $SU(N)/\mathbb{Z}_N$ subgroup of $SO(N^2 - 1)$ that we are dealing with does not describe generic rotations, and the picture changes as follows: after choosing a basis in which the $SU(N)$ matrix is diagonal, the latter belongs to a Cartan subgroup of dimension $N - 1$. Generically therefore its flow lines will lie on a torus of dimension $N - 1$; quite a bit smaller than the torus that occurs for a generic $SO(N^2 - 1)$ rotation. The set of fixed points consists of all density matrices that commute with the given $SU(N)$ matrix. When all the eigenvalues of the latter are different, the set of fixed points consists of all density matrices that are diagonal in the chosen basis. This set is the $(N - 1)$-dimensional *eigenvalue simplex*, to be discussed in the next section. The eigenvalue simplex contains only N pure states: the eigenstates of the $SU(N)$ matrix that describes the rotation. The set of fixed points is of larger dimension if degeneracies occur in the spectrum of the $SU(N)$ matrix.

The action of $SU(N)$ on $\mathcal{M}^{(N)}$ contains a number of intricacies that will occupy us later in this chapter, but at least now we have made a start.

8.4 The space of density matrices as a convex set

Let us state some general facts and definitions.[4] The dimension of $\mathcal{M}^{(N)}$ is $N^2 - 1$. The pure states are the projectors

$$\rho^2 = \rho \quad \Leftrightarrow \quad \rho = |\psi\rangle\langle\psi| \,. \tag{8.33}$$

The space of pure states is $\mathbb{C}\mathbf{P}^{N-1}$. As we observed in Section 4.5 this space is isometrically embedded in the set of Hermitian matrices provided that we define distance as the Hilbert–Schmidt distance (8.3). The pure states form a $2(N-1)$-dimensional subset of the (N^2-2)-dimensional boundary of $\mathcal{M}^{(N)}$. To see whether a given density matrix belongs to the boundary or not, we diagonalize it and check its eigenvalues. If one of them equals zero we are on the boundary.

Any density matrix can be diagonalized. The set of density matrices that are diagonal in a given basis $\{|e_i\rangle\}$ can be written as

$$\rho = \sum_{i=1}^{N} \lambda_i |e_i\rangle\langle e_i| \,, \qquad \rho|e_i\rangle = \lambda_i |e_i\rangle \,, \qquad \sum_{i=1}^{N} \lambda_i = 1 \,. \tag{8.34}$$

This set is known as the *eigenensemble* or as the *eigenvalue simplex*. It forms a special $(N-1)$-dimensional cut through the set of density matrices, and every density matrix sits in some eigenvalue simplex. It is a simplex since the eigenvalues are positive and add to one – indeed it is a copy of the $(N-1)$-dimensional simplex with N corners that we studied in classical probability theory. It is centred at the maximally mixed state $\rho_\star$.

In Section 1.1 we defined the rank of a point in a convex set as the minimum number r of pure points that are needed to express it as a convex combination of pure states. It is comforting to observe that this definition coincides with the usual definition of the *rank* of a Hermitian matrix: a density matrix of matrix rank r can be written as a convex sum of no less than r projectors (as is obvious when the matrix is diagonalized). Hence the maximal rank of a mixed state is equal to N, much less than the upper bound N^2 provided by Carathéodory's theorem (Section 1.1).

The distance between an arbitrary density matrix ρ and the centre $\rho_\star$ was given in Eq. (8.32), where we saw that $\mathrm{Tr}\rho^2$ determines the distance from $\rho_\star$. It happens that $\mathrm{Tr}\rho^2 \leq 1$ with equality if and only if the state is pure, so as expected the pure states lie at maximal distance from $\rho_\star$. This observation determines the radius R_{out} of the outsphere or circumsphere (the smallest ball containing $\mathcal{M}$). To compute the radius r_{in} of the insphere (the largest ball inscribed in $\mathcal{M}$), we observe that every

[4] Some elementary properties of the convex set of density matrices were discussed by Bloore (1976) and by Harriman (1978).

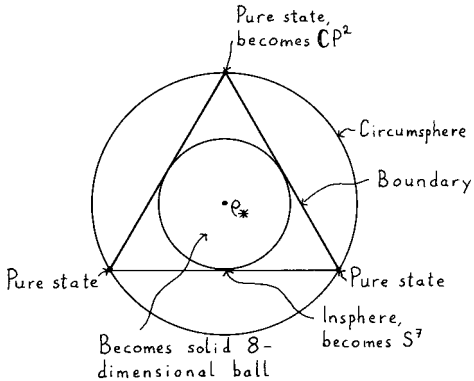


Figure 8.4. This picture shows the eigenvalue simplex for $N = 3$ with its insphere and its circumsphere, and indicates what happens if we apply $SU(3)/\mathbb{Z}_3$ transformations to obtain the eight-dimensional space of density matrices.

density matrix belongs to some eigenvalue simplex. It follows that the radius of the insphere of $\mathcal{M}^{(N)}$ will equal that of a simplex Δ_{N-1}, and this we computed in Section 1.2. So we find

$$R_{\text{out}} = \sqrt{\frac{N-1}{2N}} \quad \text{and} \quad r_{\text{in}} = \sqrt{\frac{1}{2N(N-1)}}\,. \tag{8.35}$$

The maximally mixed state $\rho_\star$ is surrounded by a ball of radius r_{in} consisting entirely of density matrices. These deliberations are illustrated in Figure 8.4, which shows the eigenvalue simplex for a qutrit.

In Eq. (1.26) we computed the angle subtended by two corners of a simplex and found that it approaches 90° when N becomes large. The corners of the eigenvalue simplex represent pure states at maximal distance from each other, so if χ denotes the angle subtended by two pure states at $\rho_\star$ then

$$\cos\chi \;\leq\; -\frac{1}{N-1}\,, \tag{8.36}$$

with equality if and only if the states are at maximal distance from each other, that is to say if and only if they are orthogonal. When N is large this means that all pure states in the hemisphere opposite to a given pure state (with respect to $\rho_\star$) lie very close to the equator. This is not surprising since almost all the area of the circumsphere is concentrated around the equator for large N (see Section 1.2). But it is very different from the $N = 2$ case, where to every pure state there corresponds an antipodal pure state.

For the $N = 3$ case we have to think in eight dimensions. This is not easy, but we can try. By looking at Figure 8.4 we can see the in- and outspheres, but

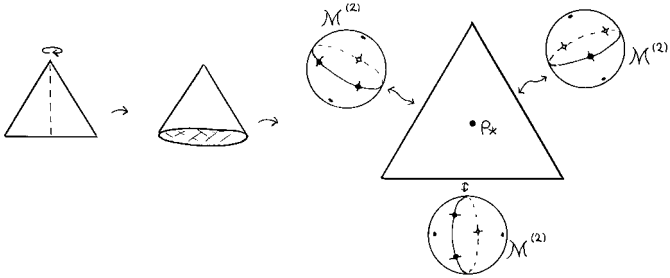


Figure 8.5. An attempt to visualize $\mathcal{M}^{(3)}$. We rotate the eigenvalue simplex to obtain a cone, then we rotate it in another dimension to turn the base of the cone into a Bloch ball rather than a disc. That is a maximal face of $\mathcal{M}^{(3)}$. On the right, we imagine that we have done this to all the three edges of the simplex. In each maximal face we have placed three equidistant points – it happens that when these points are placed correctly on all the three spheres, they form a regular simplex inscribed in $\mathcal{M}^{(3)}$.

because $\mathcal{M}^{(3)}$ is left invariant only under quite special rotations – those that move a given corner of the simplex through a two complex-dimensional projective space embedded in $\mathbb{R}^8$ – it is not so easy to imagine what the full structure looks like. An example of an allowed rotation is a rotation of the eigenvalue simplex around an axis joining a corner to ρ_*. This turns the simplex into a cone (see Figure 8.5). In fact, if we could imagine just one more dimension, we could see a four-dimensional slice of $\mathcal{M}^{(3)}$, which would be a cone whose base is a three-dimensional ball, having one of the edges of the simplex as its diameter. This ball is one of the *faces* of $\mathcal{M}^{(3)}$.

Perhaps we should recall, at this point, that a face of a convex body is defined as a convex subset stable under purification and mixing. Contemplation of Figure 8.5 shows that there is a face opposite to each pure state, consisting of all density matrices that are mixtures of pure states that are orthogonal to the given pure state. This is to say that $\mathcal{M}^{(3)}$ has faces that in themselves are Bloch balls. There is an interesting way to look at this, which goes as follows: Pick a pure state $|Z\rangle$ and study the equation

$$\mathrm{Tr}\rho\,|Z\rangle\langle Z| \;=\; \bar{Z}\rho Z = c\,, \qquad c \in [0,1]\,. \tag{8.37}$$

For fixed $|Z\rangle$ this is an affine functional representing a family of parallel hyperplanes in the space of Hermitian matrices of unit trace; when $c = 1$ the hyperplane intersects $\mathcal{M}^{(3)}$ in a single pure state and when $c = 0$ in the face opposite to that pure state. It is interesting to observe that for intermediate values of c the hyperplane intersects the pure states in one of the squashed Berger spheres, depicted in Figure 4.13.

It is clear that we are far from understanding the shape of the set of qutrit states – one can try various devices, such as looking at two-dimensional (Jakóbczyk and Siennicki, 2001) or three-dimensional (Bloore, 1976) sections through the body, but the dimension is too high for easy comprehension. Since $N = 3$ does not bring any great simplification, we will from now on discuss the case of arbitrary N.

The maximally mixed state $\rho_\star$ can be obtained as a mixture of pure states by putting equal weight (in the sense of the Fubini–Study measure) on all pure states, or by putting equal weight on each corner in an eigenvalue simplex, and also in many other ways. A similar non-uniqueness afflicts all mixed states. Interestingly this non-uniqueness can be expressed in a precise way as follows:[5]

Theorem 8.2 (Schrödinger's mixture) *A density matrix ρ having the diagonal form*

$$\rho = \sum_{i=1}^{N} \lambda_i \, |e_i\rangle\langle e_i| \tag{8.38}$$

can be written in the form

$$\rho = \sum_{i=1}^{M} p_i |\psi_i\rangle\langle\psi_i| \, , \qquad \sum_{i=1}^{M} p_i = 1 \, , \qquad p_i \geq 0 \tag{8.39}$$

if and only if there exists a unitary $M \times M$ matrix U such that

$$|\psi_i\rangle = \frac{1}{\sqrt{p_i}} \sum_{j=1}^{N} U_{ij} \sqrt{\lambda_j} \, |e_j\rangle \, . \tag{8.40}$$

Here all states are normalized to unit length but they need not be orthogonal to each other.

Given ρ, this theorem supplies all the ways in which ρ can be expressed as an ensemble of pure states. Observe that the matrix U does not act on the Hilbert space but on vectors whose components are state vectors, and also that we may well have $M > N$. But only the first N columns of U appear in the equation – the remaining $M - N$ columns are just added in order to allow us to refer to the matrix U as a unitary matrix. What the theorem basically tells us is that the pure states that make up an ensemble are linearly dependent on the N vectors $|e_i\rangle$ that make up the eigenensemble. Moreover an arbitrary state in that linear span can be included. For definiteness we assume that all density matrices have rank N so that we consider ensembles of M pure states in an N-dimensional Hilbert space.

[5] This theorem has an interesting history. Schrödinger (1936) published it with no claim to priority. When the time was ripe it was rediscovered by (among others) Gisin (1989) and Hughston, Jozsa and Wootters (1993); it is now often known as the GHJW lemma.

It is straightforward to prove that Eq. (8.39) and Eq. (8.40) imply Eq. (8.38). To prove the converse, define the first N columns of the unitary matrix U by

$$U_{ij} \equiv \frac{\sqrt{p_i}}{\sqrt{\lambda_j}} \langle e_j|\psi_i\rangle \quad \Rightarrow \quad \sum_{j=1}^{N} U_{ij}\sqrt{\lambda_j}\,|e_j\rangle = \sqrt{p_i}\,|\psi_i\rangle\,. \tag{8.41}$$

The remaining $N - M$ columns can be chosen at will, consistent with unitarity of the matrix. The matrix will be unitary because

$$\sum_{i=1}^{M} U^{\dagger}_{ki}U_{ij} = \sum_{i=1}^{M} \frac{p_i}{\sqrt{\lambda_j\lambda_k}} \langle e_j|\psi_i\rangle\langle\psi_i|e_k\rangle = \frac{1}{\sqrt{\lambda_j\lambda_k}} \langle e_j|\rho|e_k\rangle = \delta_{jk}\,. \tag{8.42}$$

This concludes the proof of the mixture theorem. Since only a rectangular submatrix of U is actually used in the theorem we could leave it like that and refer to U as a 'right unitary' or *isometry* matrix if we wanted to, but the extra columns do no harm. In fact they are helpful. We can deduce that

$$p_i = \sum_{j=1}^{N} |U_{ij}|^2\,\lambda_j\,. \tag{8.43}$$

Thus $\vec{p} = B\vec{\lambda}$, where B is a unistochastic, and hence bistochastic, matrix. In the language of Section 2.1, the vector $\vec{p}$ is majorized by the eigenvalue vector $\vec{\lambda}$.

To see how useful the mixture theorem is, consider the face structure of $\mathcal{M}^{(N)}$. Recall (from Section 1.1) that a face is a convex subset of a convex set that is stable under mixing and purification. But the mixture theorem tells us that a density matrix is always a mixture of pure states belonging to that subspace of Hilbert space that is spanned by its eigenvectors. What this means is that every face is a copy of $\mathcal{M}^{(K)}$, the body of density matrices for a system whose Hilbert space has dimension K. If $K < N$ the face is a proper face, and belongs to the boundary of $\mathcal{M}^{(N)}$.

On closer inspection, we see that the face structure of $\mathcal{M}^{(N)}$ reveals 'crystalline' regularities. A given face corresponds to some subspace $\mathcal{H}_K$ of Hilbert space. Introduce a projector E that projects down to that subspace. In full analogy to Eq. (8.37) we can consider the affine functional

$$\mathrm{Tr}\rho E = c\,, \qquad c \in [0,1]\,. \tag{8.44}$$

Again, this defines a family of parallel hyperplanes in the space of Hermitian matrices of unit trace. For $c = 1$ it defines a face of density matrices with support in $\mathcal{H}_K$, and for $c = 0$ it defines an opposing face with support in the orthogonal complement of that subspace. There is a straight line between the 'centres of mass' of these two faces, passing through the 'center of mass' of $\mathcal{M}^{(N)}$ (i.e. through $\rho_\star$).

Next, we recall from Section 1.1 that the faces of a convex body always form a partially ordered structure known as a lattice, and from Section 4.2 that the set of subspaces of a Hilbert space also forms a lattice. The following is true:

Theorem 8.3 *The lattice of faces of $\mathcal{M}^{(N)}$ is identical to the orthocomplemented lattice of subspaces in $\mathcal{H}_N$.*

The identity of the two lattices is by now obvious, but interesting nevertheless. A lattice is said to be *orthocomplemented* if and only if there is a map $a \to a'$ of the lattice L onto itself, such that

$$(a')' = a \quad a \leq b \;\Rightarrow\; b' \leq a' \quad a \cap a' = 0 \quad a \cup a' = L \tag{8.45}$$

for all $a, b \in L$, where $\cap$ and $\cup$ denote, respectively, the greatest lower and the smallest upper bound of a pair of elements (for the lattice of subspaces of $\mathcal{H}_N$, they are, respectively, the intersection and the linear span of the union of a pair of subspaces). In our lattice, two opposing faces of $\mathcal{M}^{(N)}$ are indeed related by a map $a \to a'$. It is possible to single out further properties of this lattice, and to use them for an axiomatic formulation of quantum mechanics – this is the viewpoint of quantum logic.[6]

Let us mention in passing that there is another angle from which we can try to view the structure: we choose a convex polytope that we feel comfortable with, let it have the same outradius as the body of density matrices itself, and ask if it can be inscribed in $\mathcal{M}^{(N)}$. The obvious choice is a regular simplex. The simplex Δ_8 can be inscribed in $\mathcal{M}^{(3)}$, and Figure 8.5 indicates how. (To avoid misunderstanding: this Δ_8 does not have edges of unit length. The largest simplex with edge lengths equal to one that can be placed inside $\mathcal{M}^{(N)}$ is Δ_{N-1}.) Oddly enough it is not as easy to do this for $N > 3$, but at least for moderately small N it can be done, and so our intuition has a little more material to work with.[7]

8.5 Stratification

Yet another way to organize our impressions of $\mathcal{M}^{(N)}$ is to study how it is partitioned into orbits of the unitary group (recall Section 7.2). We will see that each individual orbit is a flag manifold (Section 4.9) and that the space of orbits has a transparent structure.[8] We begin from the fact that any Hermitian matrix can be diagonalized

[6] Further details can be found in the books by Jauch (1968) and Varadarajan (1985); for a version of the story that emphasizes the geometry of the convex set one can profitably consult Mielnik (1981).

[7] The N^2 corners of such a simplex define what is known as a symmetric informationally complete POVM. See Renes, Blume-Kohout, Scott and Caves (2004); it is likely, but not proven, that such POVMs exist for any N. See Wootters and Fields (1989) for another choice of the convex polytope.

[8] A general reference for this section is Adelman, Corbett and Hurst (1993).

by a unitary rotation,

$$\rho \;=\; V\Lambda V^{\dagger}\,. \tag{8.46}$$

where Λ is a diagonal density matrix that fixes a point in the eigenvalue simplex. We obtain a $U(N)$ orbit from Eq. (8.46) by letting the matrix V range over $U(N)$. Before we can tell what the result is, we must see to what extent different choices of V can lead to the same ρ. Let B be a diagonal unitary matrix. It commutes with Λ, so

$$\rho = V\Lambda V^{\dagger} = VB\Lambda B^{\dagger}V^{\dagger}. \tag{8.47}$$

In the case of non-degenerate spectrum this is all there is; the matrix V is determined up to the N arbitrary phases entering B, and the orbit will be the coset space $U(N)/U(1)\times U(1)\times\cdots\times U(1)$. From Section 4.9 we recognize this as the flag manifold $\mathbf{F}^{(N)}_{1,2,\ldots,N-1}$. If degeneracies occur in the spectrum of ρ, the matrix B need not be diagonal in order to commute with Λ, and various special cases ensue. In the language of Section 7.2, the isotropy group changes, and so does the nature of the orbit.

Let us discuss the case of $N = 3$ in detail to see what happens; our deliberations are illustrated in Figure 8.6(b). The space of diagonal density matrices is the simplex Δ_2. However, using unitary permutation matrices we can change the order in which the eigenvalues occur, so that without loss of generality we may assume that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$. This corresponds to dividing the simplex Δ_2 into 3! parts, and to picking one of them. Denote it by $\widetilde{\Delta}_2$. We call it a *Weyl chamber*, with a terminology borrowed from group theory. It is the Weyl chamber that forms the space of orbits of $U(3)$ in $\mathcal{M}^{(3)}$.

The nature of the orbit will depend on its location in the Weyl chamber. Depending on the degeneracy of the spectrum, we decompose $\widetilde{\Delta}_2$ into four parts (Adelman et al., 1993; Życzkowski and Słomczyński, 2001) (see also Figure 8.6(b)):

- **(a)** a point K_3 representing the state ρ_* with triple degeneracy $\{1/3, 1/3, 1/3\}$; the isotropy group is $U(3)$;
- **(b)** a one-dimensional line K_{12} representing the states with double degeneracy, $\lambda_2 = \lambda_3$; the isotropy group is $U(1)\times U(2)$;
- **(c)** a one-dimensional line K_{21} representing the states with double degeneracy, $\lambda_1 = \lambda_2$; the isotropy group is $U(2)\times U(1)$;
- **(d)** the two-dimensional part K_{111} of the generic points of the simplex, for which no degeneracy occurs; the isotropy group is $U(1)\times U(1)\times U(1)$.

Since the degeneracy of the spectrum determines the isotropy group it also determines the topology of the $U(3)$ orbit. In case (a) the orbit is $U(3)/U(3)$, that is, a single point, namely the maximally mixed state ρ_*. In the cases (b) and (c) the

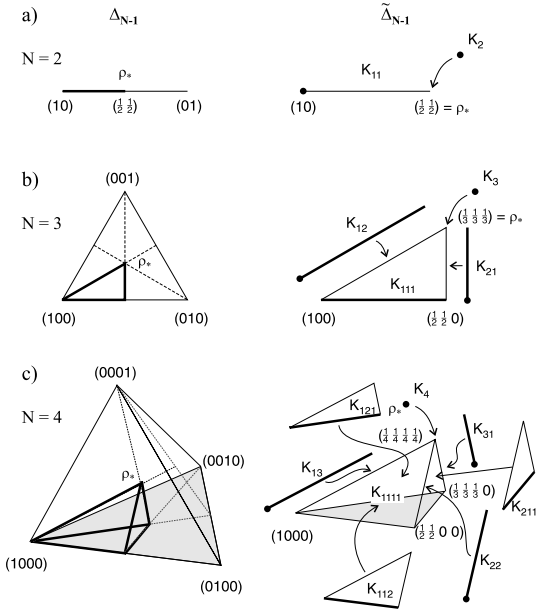


Figure 8.6. The eigenvalue simplex and the Weyl chamber for $N = 2, 3$ and 4. The Weyl chamber $\widetilde{\Delta}_{N-1}$, enlarged on the right-hand side, can be decomposed according to the degeneracy into 2^{N-1} parts.

orbit is $U(3)/[U(1) \times U(2)] = \mathbf{F}_1^{(3)} = \mathbb{C}\mathbf{P}^2$. In the generic case (d) we obtain the generic flag manifold $\mathbf{F}_{1,2}^{(3)}$.

Now we are ready to tackle the general problem of $N \times N$ density matrices. There are two things to watch: the boundary of $\mathcal{M}^{(N)}$, and the stratification of $\mathcal{M}^{(N)}$ by qualitatively different orbits under $U(N)$. It will be easier to follow the discussion if Figure 8.6 is kept in mind.

The diagonal density matrices form a simplex Δ_{N-1}. It can be divided into $N!$ parts depending on the ordering of the eigenvalues, and we can select one of these parts to be the Weyl chamber $\widetilde{\Delta}_{N-1}$. The Weyl chamber is the $(N-1)$-dimensional space of orbits under $U(N)$. The nature of the orbits is determined by the degeneracy of the spectrum, so we decompose the Weyl chamber into parts $K_{k_1,\ldots,k_m}$ where the largest eigenvalue has degeneracy k_1, the second largest degeneracy k_2, and so on. Clearly $k_1 + \cdots + k_m = N$. Each of these parts parametrize a stratum (see Section 7.2) of $\mathcal{M}^{(N)}$, where each orbit is a flag manifold

$$\mathbf{F}_{k_1,k_2,\ldots,k_{m-1}}^{(N)} = \frac{U(N)}{U(k_1) \times \cdots \times U(k_m)}\,. \tag{8.48}$$

See Section 4.9 for the notation. The generic case is $K_{1,1,\ldots,1}$ consisting of the interior of $\widetilde{\Delta}_{N-1}$ together with one of its (open) facets corresponding to the case of one vanishing eigenvalue. This means that, except for a set of measure zero, the space $\mathcal{M}^{(N)}$ is equal to

$$\mathcal{M}_{1,\ldots,1} \sim \left[\frac{U(N)}{T^N}\right] \times K_{1,1,\ldots,1} = \mathbf{F}^{(N)}_{1,2,\ldots,N-1} \times G_N. \tag{8.49}$$

Here we used T^N to denote the product of N factors $U(1)$, topologically a torus, and we also used $G_N \equiv K_{1,1,\ldots,1}$. The equality holds in a topological sense only, but, as we will see in Chapter 14, it is also accurate when computing volumes in $\mathcal{M}$.

There are exceptional places in the Weyl chamber where the spectrum is degenerate. In fact there are 2^{N-1} different possibilities for $K_{k_1,\ldots,k_m}$, because there are $N-1$ choices between 'larger than' or 'equal' when the eigenvalues are ordered. Each $K_{k_1,\ldots,k_m}$ forms an $(m-1)$-dimensional (irregular) simplex that we denote by G_m. Each G_m can be realized in $\binom{N-1}{m-1}$ different ways (e.g. for $N=4$ the set G_2 can be realized as $K_{3,1}$, $K_{2,2}$, $K_{1,3}$). In this way we get a decomposition of the Weyl chamber as

$$\widetilde{\Delta}_{N-1} = \bigcup_{k_1+\cdots+k_m=N} K_{k_1,\ldots,k_m}\,, \tag{8.50}$$

and a topological decomposition of the space of density matrices as

$$\mathcal{M}^{(N)} \sim \bigcup_{k_1+\cdots+k_m=N} \left[\frac{U(N)}{U(k_1)\times\cdots\times U(k_m)}\right] \times K_{k_1,\ldots,k_m}\,, \tag{8.51}$$

where the sum ranges over all partitions of N into sums of positive integers. The total number of such partitions is 2^{N-1}. However, the orbits sitting over, say, $K_{1,2}$ and $K_{2,1}$ will be flag manifolds of the same topology. To count the number of qualitatively different flag manifolds that appear, we must count the number of partitions of N with no regard to ordering, that is we must compute the number $p(N)$ of different representations of the number N as the sum of positive integers. For $N = 1, 2, \ldots, 10$ the number $p(N)$ is equal to 1, 2, 3, 5, 7, 11, 15, 22, 30, and 42, while for large N the asymptotic Hardy–Ramanujan formula (Hardy and Ramanujan, 1918) gives $p(N) \simeq \exp\left(\pi\sqrt{2N/3}\right)/4\sqrt{3}N$. Figure 8.6 and Table 8.1 summarize these deliberations for $N \leq 4$.

Let us now take a look at the boundary of $\mathcal{M}^{(N)}$. It consists of all density matrices of less than maximal rank. The boundary as a whole consists of a continuous family of maximal faces, and each maximal face is a copy of $\mathcal{M}^{(N-1)}$. To every pure state there corresponds an opposing maximal face, so the family of maximal faces can be parametrized by $\mathbb{C}\mathbf{P}^{N-1}$. It is simpler to describe the orbit space of the boundary,

Table 8.1. *Stratification of $\mathcal{M}^{(N)}$. $U(N)$ is the unitary group, T^k is a k-dimensional torus, and G_m stands for a part of the eigenvalue simplex defined in the text. The dimension d of the strata equals $d_F + d_S$, where d_F is the even dimension of the complex flag manifold $\mathbf{F}$, while $d_S = m - 1$ is the dimension of G_m.*

N	Label	Subspace	Part of the Weyl chamber	Topological structure	Flag manifold	Dimension $d = d_F + d_S$
1	$\mathcal{M}_1$	λ_1	point	$[U(1)/U(1)] \times G_1 = \{\rho_*\}$	$\mathbf{F}_0^{(1)}$	$0 = 0 + 0$
2	$\mathcal{M}_{11}$	$\lambda_1 > \lambda_2$	line with left edge	$[U(2)/T^2] \times G_2$	$\mathbf{F}_1^{(2)}$	$3 = 2 + 1$
	$\mathcal{M}_2$	$\lambda_1 = \lambda_2$	right edge	$[U(2)/U(2)] \times G_1 = \{\rho_*\}$	$\mathbf{F}_0^{(2)}$	$0 = 0 + 0$
3	$\mathcal{M}_{111}$	$\lambda_1 > \lambda_2 > \lambda_3$	triangle with base without corners	$[U(3)/T^3] \times G_3$	$\mathbf{F}_{12}^{(3)}$	$8 = 6 + 2$
	$\mathcal{M}_{12}$	$\lambda_1 > \lambda_2 = \lambda_3$	edges with lower corners	$[U(3)/(U(2) \times T)] \times G_2$	$\mathbf{F}_1^{(3)}$	$5 = 4 + 1$
	$\mathcal{M}_{21}$	$\lambda_1 = \lambda_2 > \lambda_3$			$\mathbf{F}_2^{(3)}$	
	$\mathcal{M}_3$	$\lambda_1 = \lambda_2 = \lambda_3$	upper corner	$[U(3)/U(3)] \times G_1 = \{\rho_*\}$	$\mathbf{F}_0^{(3)}$	$0 = 0 + 0$
4	$\mathcal{M}_{1111}$	$\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4$	interior of tetrahedron with bottom face	$[U(4)/T^4] \times G_4$	$\mathbf{F}_{123}^{(4)}$	$15 = 12 + 3$
	$\mathcal{M}_{112}$	$\lambda_1 > \lambda_2 > \lambda_3 = \lambda_4$	faces without side edges	$[U(4)/(U(2) \times T^2)] \times G_3$	$\mathbf{F}_{12}^{(4)}$	$12 = 10 + 2$
	$\mathcal{M}_{121}$	$\lambda_1 > \lambda_2 = \lambda_3 > \lambda_4$			$\mathbf{F}_{13}^{(4)}$	
	$\mathcal{M}_{211}$	$\lambda_1 = \lambda_2 > \lambda_3 > \lambda_4$			$\mathbf{F}_{23}^{(4)}$	
	$\mathcal{M}_{13}$	$\lambda_1 > \lambda_2 = \lambda_3 = \lambda_4$	edges with lower corners	$[U(4)/(U(3) \times T)] \times G_2$	$\mathbf{F}_1^{(4)}$	$7 = 6 + 1$
	$\mathcal{M}_{31}$	$\lambda_1 = \lambda_2 = \lambda_3 > \lambda_4$			$\mathbf{F}_3^{(4)}$	
	$\mathcal{M}_{22}$	$\lambda_1 = \lambda_2 > \lambda_3 = \lambda_4$		$[U(4)/(U(2) \times U(2))] \times G_2$	$\mathbf{F}_2^{(4)}$	$9 = 8 + 1$
	$\mathcal{M}_4$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	upper corner	$[U(4)/U(4)] \times G_1 = \{\rho_*\}$	$\mathbf{F}_0^{(4)}$	$0 = 0 + 0$

because it is the base of the Weyl chamber and has dimension $N - 2$. It is clear from Figure 8.6 that the orbit space of the boundary $\partial\mathcal{M}^{(N)}$ is the same as the orbit space of $\mathcal{M}^{(N-1)}$, but the orbits are different because the group that acts is larger. Generically there will be no degeneracies in the spectrum, so except for a set of measure zero the boundary has the structure $(U(N)/T^N) \times G_{N-1}$.

Alternatively, the boundary can be decomposed into sets of matrices with different rank. It is not hard to show that the dimension of the set of states of rank $r = N - k$ is equal to $N^2 - k^2 - 1$.

The main message of this section has been that the Weyl chamber gives a good picture of the set of density matrices, because it represents the space of orbits under the unitary group. It is a very good picture, because the Euclidean distance between two points within a Weyl chamber is equal to the minimal Hilbert–Schmidt distance between the pair of orbits that they represent. In equations, let U and V

denote arbitrary unitary matrices of size N. Then

$$D_{\rm HS}(Ud_1U^\dagger, Vd_2V^\dagger) \;\geq\; D_{\rm HS}(d_1, d_2)\,, \tag{8.52}$$

where d_1 and d_2 are two diagonal matrices with their eigenvalues in decreasing order. The proof of this attractive observation is simple, once we know something about the majorization order for matrices. For this reason its proof is deferred to Problem 12.5.

8.6 An algebraic afterthought

Quantum mechanics is built around the fact that the set of density matrices is a convex set of a very special shape. From the perspective of Chapter 1 it seems a strange set to consider. There are many convex sets to choose from. The simplex is evidently in some sense preferred, and leads to classical probability theory and – ultimately, once the pure states are numerous enough to form a symplectic manifold – to classical mechanics. But why the convex set of density matrices? A standard answer is that we want the vector space that contains the convex set to have further structure, turning it into an *algebra*.[9] (By definition, an algebra is a vector space where vectors can be multiplied as well as added.) At first sight this looks like an odd requirement: the observables may form an algebra, but why should the states sit in one? We get an answer of sorts if we think of the states as linear maps from the algebra to the real numbers, because then we will obtain a vector space that is dual to the algebra and can be identified with the algebra. Some further hints will emerge in Chapter 11. For now, let us simply accept it. The point is that if the algebra has suitable properties then this will give rise to new ways of defining positive cones – more interesting ways than the simple requirement that the components of the vectors be positive.

To obtain an algebra we must define a product $A \circ B$. We may want the algebra to be *real* in the sense that

$$A \circ A + B \circ B = 0 \quad \Rightarrow \quad A = B = 0\,. \tag{8.53}$$

We do not insist on associativity, but we insist that polynomials of operators should be well defined. In effect we require that $A^n \circ A^m = A^{n+m}$ where $A^n \equiv A \circ A^{n-1}$. Call this *power associativity*. With this structure in hand we do have a natural definition of *positive* vectors, as vectors A that can be written as $A = B^2$ for some vector B, and we can define *idempotents* as vectors obeying $A^2 = A$. If we can also define a trace, we can define pure states as idempotents of unit trace. But by now

[9] The algebraic viewpoint was invented by Jordan; key mathematical results were derived by Jordan, von Neumann and Wigner (1934). To see how far it has developed, see Emch (1972) and Alfsen and Shultz (2003).

Table 8.2. *Jordan algebras*

Jordan algebra	Dimension	Norm	Positive cone	Pure states
$J_N^{\mathbb{R}}$	$N(N+1)/2$	$\det M$	Self dual	$\mathbb{R}\mathbf{P}^{N-1}$
$J_N^{\mathbb{C}}$	N^2	$\det M$	Self dual	$\mathbb{C}\mathbf{P}^{N-1}$
$J_N^{\mathbb{H}}$	$N(2N-1)$	$\det M$	Self dual	$\mathbb{H}\mathbf{P}^{N-1}$
$J_2(V_n)$	$n+1$	$\eta_{IJ}X^I X^J$	Self dual	$\mathbf{S}^{n-1}$
$J_3^{\mathbb{O}}$	27	$\det M$	Self dual	$\mathbb{O}\mathbf{P}^2$

there are not that many algebras to choose from. To make the algebra real in the above sense we would like the algebra to consist of Hermitian matrices. Ordinary matrix multiplication will not preserve Hermiticity, and therefore matrix algebras will not do as they stand. However, because we can square our operators we can define the *Jordan product*

$$A \circ B \equiv \frac{1}{4}(A+B)^2 - \frac{1}{4}(A-B)^2 . \tag{8.54}$$

There is no obvious physical interpretation of this composition law, but it does turn the space of Hermitian matrices into a (commutative) *Jordan algebra*. If A and B are elements of a matrix algebra this product is equal to one half of their anti-commutator, but we need not assume this. Jordan algebras have all the properties we want, including power associativity. Moreover all simple Jordan algebras have been classified, and there are only four kinds (and one exceptional case). A complete list is given in Table 8.2.[10]

The case that really concerns us is $J_N^{\mathbb{C}}$. Here the Jordan algebra is the space of complex valued Hermitian $N \times N$ matrices, and the Jordan product is given by one half of the anti-commutator. This is the very algebra that we have – implicitly – been using, and with whose positive cone we are by now reasonably familiar. We can easily define the trace of any element in the algebra, and the pure states in the table are assumed to be of unit trace. One can replace the complex numbers with real or quaternionic numbers, giving two more families of Jordan algebras. The state spaces that result from them occur in special quantum mechanical situations, as we saw in Section 5.5. The fourth family of Jordan algebras are the *spin factors* $J_2(V_n)$. They can also be embedded in matrix algebras, their norm uses a Minkowski space metric η_{IJ}, and their positive cones are the familiar forward light cones in Minkowski spaces of dimension $n+1$. Their state spaces occur in special quantum

[10] For a survey of Jordan algebras, see McCrimmon (1978).

mechanical situations too (Uhlmann, 1996), but this is not the place to go into that. (Finally there is an exceptional case based on octonions, that need not concern us).

So what is the point? One point is that very little in addition to the quantum mechanical formalism turned up in this way. This is to say: once we have committed ourselves to finding a self dual positive cone in a finite-dimensional real algebra, then we are almost (but not quite) restricted to the standard quantum mechanical formalism already. Another point is that the positive cone now acquires an interesting geometry. Not only is it self dual (see Section 1.1), it is also foliated in a natural way by hypersurfaces for which the determinant of the matrix is constant. These hypersurfaces turn out to be symmetric coset spaces $SL(N,\mathbb{C})/SU(N)$, or relatives of that if we consider a Jordan algebra other than $J_N^{\mathbb{C}}$. Given the norm $\mathcal{N}$ on the algebra, there is a natural looking metric

$$g_{ij} \;=\; -\frac{1}{d}\,\partial_i\partial_j \ln \mathcal{N}^{d/N}\;. \tag{8.55}$$

where d is the dimension of the algebra. (Since the norm is homogeneous of order N, the exponent ensures that the argument of the logarithm is homogeneous of order d.) This metric is positive definite for all the Jordan algebras, and it makes the boundary of the cone sit at an infinite distance from any point in the interior. If we specialize to diagonal matrices – which means that the Jordan algebra is no longer simple – we recover the positive orthant used in classical probability theory, and the natural metric turns out to be flat, although it differs from the 'obvious' flat metric on $\mathbb{R}^N$.

We doubt that the reader feels compelled to accept the quantum mechanical formalism only because it looms large in a Jordan algebra framework. Another way of arguing for quantum mechanics from (hopefully) simple assumptions is provided by quantum logic. This language can be translated into convex set theory (Mielnik, 1981), and turns out to be equivalent to setting conditions on the lattice of faces that one wants the underlying convex set to have. From a physical point of view it concerns the kind of 'yes/no' experiments that one expects to be able to perform; choosing one's expectations suitably (Araki, 1980) one can argue along such lines that the state spaces that will emerge are necessarily state spaces of Jordan algebras, and we are back where we started.

But there is a further algebraic structure waiting in the wings, and it is largely this additional structure that Chapter 9 and *passim* are about.

8.7 Summary

Let us try to summarize basic properties of the set of mixed quantum states $\mathcal{M}^{(N)}$. As before Δ_{N-1} is an $(N-1)$-dimensional simplex, $\widetilde{\Delta}_{N-1}$ is a Weyl chamber in Δ_{N-1}, and $\mathbf{F}^{(N)}$ is the complex flag manifold $\mathbf{F}^{(N)}_{1,2,\ldots,N-1}$.

- $\mathcal{M}^{(N)}$ is a convex set of N^2-1 dimensions. It is topologically equivalent to a ball and does not have pieces of lower dimensions ('no hairs').
- The set $\mathcal{M}^{(N)}$ is inscribed in a ball of radius $R^2_{\rm out} = (N-1)/2N$, and contains a maximal ball of radius $r^2_{\rm in} = [2N(N-1)]^{-1}$.
- It is neither a polytope nor a smooth body. Its faces are copies of $\mathcal{M}^{(K)}$ with $K < N$.
- It is partitioned into orbits of the unitary group, and the space of orbits is a Weyl chamber $\widetilde{\Delta}_{N-1}$.
- The full measure of $\mathcal{M}^{(N)}$ has locally the structure of $\mathbf{F}^{(N)} \times \Delta_{N-1}$.
- The boundary $\partial\mathcal{M}^{(N)}$ contains all states of less than maximal rank.
- The boundary has N^2-2 dimensions. Almost everywhere it has the local structure of $\mathbf{F}^{(N)} \times \Delta_{N-2}$.

In this summary we have not mentioned the remarkable way in which composite systems are handled by quantum theory. The discussion of this topic starts in the next chapter and culminates in Chapter 15.

Problems

Problem 8.1 Prove that the polar decomposition of an invertible operator is unique.

Problem 8.2 Consider a square matrix A. Perform an arbitrary permutation of its rows and/or columns. Will its (a) eigenvalues, (b) singular values change?

Problem 8.3 What are the singular values of (a) a Hermitian matrix, (b) a unitary matrix, (c) any normal matrix A (such that $[A, A^\dagger] = 0$)?

Problem 8.4 A unitary similarity transformation does not change the eigenvalues of any matrix. Show that this is true for the singular values as well.

Problem 8.5 Show that $\mathrm{Tr}(AA^\dagger)\,\mathrm{Tr}(BB^\dagger) \geq |\mathrm{Tr}(AB^\dagger)|^2$, always.

Problem 8.6 Show that the diagonal elements of a positive operator are positive.

Problem 8.7 Take a generic vector in $\mathbb{R}^{N^2-1}$. How many of its components can you set to zero, if you are allowed to act only with an $SU(N)$ subgroup of the rotation group?

Problem 8.8 Transform a density matrix ρ of size 2 into $\rho' = U\rho\,U^\dagger$ by a general unitary matrix $U = \begin{bmatrix} \cos\vartheta\,\mathrm{e}^{\mathrm{i}\phi} & \sin\vartheta\,\mathrm{e}^{\mathrm{i}\psi} \\ -\sin\vartheta\,\mathrm{e}^{-\mathrm{i}\psi} & \cos\vartheta\,\mathrm{e}^{-\mathrm{i}\phi} \end{bmatrix}$. What is the orthogonal matrix $O \in SO(3)$ which transforms the Bloch vector $\vec{\tau}$? Find the rotation angle t and the vector $\vec{\Omega}$ determining the orientation of the rotations axis.

9

Purification of mixed quantum states

> In this significant sense quantum theory subscribes to the view that '*the whole is greater than the sum of its parts*'.
>
> *Hermann Weyl*

In quantum mechanics the whole, built from parts, is described using the *tensor product* that defines the composition of an N-dimensional vector space V and an M-dimensional vector space V' as the NM-dimensional vector space $V \otimes V'$. One can go on, using the tensor product to define an infinite-dimensional tensor algebra. The interplay between the tensor algebra and the other algebraic structures is subtle indeed. In this chapter we study the case of two subsystems only. The arena is Hilbert–Schmidt space (real dimension $2N^2$), but now regarded as the Hilbert space of a composite system. We will use a partial trace to take ourselves from Hilbert–Schmidt space to the space of density matrices acting on an N-dimensional Hilbert space. The result is the quantum analogue of a marginal probability distribution. It is also like a projection in a fibre bundle, with Hilbert–Schmidt space as the bundle space and the group $U(N)$ acting on the fibres, while the positive cone serves as the base space (real dimension $2N^2 - N^2 = N^2$). Physically, the important idea is that of *purification*; a density matrix acting on $\mathcal{H}$ is regarded as a pure state in $\mathcal{H} \otimes \mathcal{H}^*$, with some of its details forgotten. We could now start an argument whether all mixed quantum states are really pure states in some larger Hilbert space, but we prefer to focus on the interesting geometry that is created on the space of mixed states by this construction.[1]

[1] For an eloquent defence of the point of view that regards density matrices as primary, see Mermin (1998). With equal eloquence, Penrose (2004) takes the opposite view.

9.1 Tensor products and state reduction

The *tensor product* of two vector spaces is not all that easy to define. The easiest way is to rely on a choice of basis in each of the factors.[2] We are interested in the tensor product of two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, with dimensions N_1 and N_2, respectively. The tensor product space will be denoted $\mathcal{H}_{12}$, and it will have dimension $N_1 N_2$. The statement that the whole is greater than its parts is related to the fact that $N_1 N_2 > N_1 + N_2$ (unless $N_1 = N_2 = 2$).

We expect the reader to be familiar with the basic features of the tensor product, but to fix our notation let us choose the bases $\{|m\rangle\}_{m=1}^{N_1}$ in $\mathcal{H}_1$, and $\{|\mu\rangle\}_{\mu=1}^{N_2}$ in $\mathcal{H}_2$. Then the Hilbert space $\mathcal{H}_{12} \equiv \mathcal{H}_1 \otimes \mathcal{H}_2$ is spanned by the basis formed by the $N_1 N_2$ elements $|m\rangle \otimes |\mu\rangle = |m\rangle|\mu\rangle$, where the sign $\otimes$ will be written explicitly only on festive occasions. The basis vectors are direct products of vectors in the factor Hilbert spaces, but by taking linear combinations we will obtain vectors that cannot be written in such a form – which explains why the composite Hilbert space $\mathcal{H}_{12}$ is so large. Evidently we can go on to define the Hilbert space $\mathcal{H}_{123}$, starting from three factor Hilbert spaces, and indeed the procedure never stops. By taking tensor products of a vector space with itself, we will end up with an infinite-dimensional *tensor algebra*. Our concern, however, is with *bipartite* systems that use only the Hilbert space $\mathcal{H}_{12}$. In many applications of quantum mechanics, a further elaboration of the idea is necessary: it may be that the subsystems are indistinguishable from each other, in which case one must take symmetric or anti-symmetric combinations of $\mathcal{H}_{12}$ and $\mathcal{H}_{21}$, leading to bosonic or fermionic subsystems, or perhaps utilize some less trivial representation of the symmetric group that interchanges the subsystems. But we will not need this elaboration either.

The matrix algebra of operators acting on a given Hilbert space is itself a vector space – the Hilbert–Schmidt vector space $\mathcal{HS}$ studied in Section 8.1. We can take tensor products also of algebras. If A_1 acts on $\mathcal{H}_1$, and A_2 acts on $\mathcal{H}_2$, then their *tensor* or *Kronecker product* $A_1 \otimes A_2$ is defined by its action on the basis elements:

$$(A_1 \otimes A_2)|m\rangle \otimes |\mu\rangle \;\equiv\; A_1|m\rangle \otimes A_2|\mu\rangle \,. \tag{9.1}$$

Again, this is not the most general operator in the tensor product algebra since we can form linear combinations of operators of this type. For a general operator we can form matrix elements according to

$$A^{m\mu}_{\;\;n\nu} \;=\; \langle m| \otimes \langle \mu|A|n\rangle \otimes |\nu\rangle \,. \tag{9.2}$$

On less festive occasions we may write this as $A_{m\mu,n\nu}$.

[2] This is the kind of procedure that mathematicians despise; the basis independent definition can be found in Kobayashi and Nomizu (1963). Since we tend to think of operators as explicit matrices, the simple-minded definition is good enough for us.

Everything works best if the underlying field is that of the complex numbers (Araki, 1980): let the space of observables, that is Hermitian operators, on a Hilbert space $\mathcal{H}$ be denoted $\mathcal{HM}(\mathcal{H})$. The dimensions of the spaces of observables on a pair of complex Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ obey

$$\dim[\mathcal{HM}(\mathcal{H}_1 \otimes \mathcal{H}_2)] \;=\; \dim[\mathcal{HM}(\mathcal{H}_1)]\,\dim[\mathcal{HM}(\mathcal{H}_2)]\,. \tag{9.3}$$

That is, $(N_1 N_2)^2 = N_1^2 N_2^2$. If we work over the real numbers the left-hand side of Eq. (9.3) is larger than the right-hand side, and if we work over quaternions (using a suitable definition of the ternsor product) the right-hand side is the largest. As an argument for why we should choose to work over the complex numbers, this observation may not be completely compelling. But the tensor algebra over the complex numbers has many wonderful properties.

Most of the time we think of vectors as columns of numbers, and of operators as explicit matrices – in finite dimensions nothing is lost and we gain concreteness. We organize vectors and matrices into arrays, using the obvious lexicographical order. At least, the order should be obvious if we write it out for a simple example: if

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

then

$$A \otimes B = \begin{bmatrix} A_{11}B & A_{12}B \\ A_{21}B & A_{22}B \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22} \end{bmatrix}. \tag{9.4}$$

Contemplation of this expression should make it clear what lexicographical ordering that we are using. At first sight one may worry that A and B are treated quite asymmetrically here, but on reflection one sees that this is only a matter of basis changes, and does not affect the spectrum of $A \otimes B$. See Problems 9.1–9.5 for further information about tensor products.

The tensor product is a main theme in quantum mechanics. We will use it to split the world into two parts; a part 1 that we study and another part 2 that we may refer to as the environment. This may be a physical environment that must be taken into account when doing experiments, but not necessarily so. It may also be a mathematical device that enables us to prove interesting theorems about the system under study, with no pretence of realism as far as the environment is concerned. Either way, the split is more subtle than it used to be in classical physics, precisely because the composite Hilbert space $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$ is so large. Most of its vectors are not direct products of vectors in the factor spaces. If not, the subsystems are said to be *entangled*.

Let us view the situation from the Hilbert space $\mathcal{H}_{12}$. To compute the expectation value of an arbitrary observable we need the density matrix ρ_{12}. It is assumed that we know exactly how $\mathcal{H}_{12}$ is defined as a tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$, so the representation (9.2) is available for all its operators. Then we can define *reduced density matrices* ρ_1 and ρ_2, acting on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, by taking *partial traces*. Thus

$$\rho_1 \equiv \mathrm{Tr}_2\, \rho_{12} \quad \text{where} \quad (\rho_1)^m_n = \sum_{\mu=1}^{N_2} (\rho_{12})^{m\mu}_{n\mu} , \tag{9.5}$$

and similarly for ρ_2. This construction is interesting, because it could be that experiments are performed exclusively on the first subsystem, in which case we are only interested in observables of the form

$$A = A_1 \otimes \mathbb{1}_2 \quad \Leftrightarrow \quad A^{m\mu}_{n\nu} = (A_1)^m_n\, \delta^\mu_\nu . \tag{9.6}$$

Then the state ρ_{12} is more than we need; the reduced density matrix ρ_1 acting on $\mathcal{H}_1$ is enough, because

$$\langle A \rangle = \mathrm{Tr}\rho_{12} A = \mathrm{Tr}_1 \rho_1 A_1 . \tag{9.7}$$

Here Tr_1 denotes the trace taken over the first subsystem only. Moreover $\rho_1 = \mathrm{Tr}_2\rho_{12}$ is the only operator that has this property for every operator of the form $A = A_1 \otimes \mathbb{1}_2$; this observation can be used to give a basis independent definition of the partial trace.

Even if ρ_{12} is a pure state, the state ρ_1 will in general be a mixed state. Interestingly, it is possible to obtain any mixed state as a partial trace over a pure state in a suitably enlarged Hilbert space. To make this property transparent, we need some further preparations.

9.2 The Schmidt decomposition

An exceptionally useful fact is the following:[3]

Theorem 9.1 (Schmidt's) *Every pure state in the Hilbert space $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$ can be expressed in the form*

$$|\Psi\rangle = \sum_{i=1}^{N} \sqrt{\lambda_i} |e_i\rangle \otimes |f_i\rangle , \tag{9.8}$$

[3] The original Schmidt's theorem, that appeared in 1907 (Schmidt, 1907), concerns infinite-dimensional spaces. The present formulation was used by Schrödinger (1936) in his analysis of entanglement, by Everett (1957) in his relative state (or many worlds) formulation of quantum mechanics, and in the 1960s by Carlson and Keller (1961) and Coleman (1963), and Coleman and Yukalov (2000). Simple expositions of the Schmidt decomposition are provided by Ekert and Knight (1995) and by Aravind (1996).

where $\{|e_i\rangle\}_{i=1}^{N_1}$ *is an orthonormal basis for* $\mathcal{H}_1$, $\{|f_i\rangle\}_{i=1}^{N_2}$ *is an orthonormal basis for* $\mathcal{H}_2$, *and* $N \leq min\{N_1, N_2\}$.

This is known as the *Schmidt decomposition* or *Schmidt's polar form* of a *bipartite* pure state. It should come as a surprise, because there is only a single sum; what is obvious is only that any pure state can be expressed in the form

$$|\Psi\rangle \;=\; \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} C_{ij}|\hat{e}_i\rangle \otimes |\hat{f}_j\rangle \;, \tag{9.9}$$

where C is some complex-valued matrix and the bases are arbitrary. The Schmidt decomposition becomes more reasonable when it is observed that the theorem concerns a special state $|\Psi\rangle$; changing the state may force us to change the bases used in Eq. (9.8).

To deduce the Schmidt decomposition we assume, without loss of generality, that $N_1 \leq N_2$. Then we observe that we can rewrite Eq. (9.9) by introducing the states $|\hat{\phi}_i\rangle = \sum_j C_{ij}|\hat{f}_j\rangle$; these will not be orthonormal states but they certainly exist, and permit us to write the state in $\mathcal{H}_{12}$ as

$$|\Psi\rangle \;=\; \sum_{i=1}^{N} |\hat{e}_i\rangle|\hat{\phi}_i\rangle \;. \tag{9.10}$$

Taking a partial trace of $\rho_\Psi = |\Psi\rangle\langle\Psi|$ with respect to the second subsystem, we find

$$\rho_1 \;=\; \mathrm{Tr}_2\big(|\Psi\rangle\langle\Psi|\big) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_1}\langle\hat{\phi}_j|\hat{\phi}_i\rangle\; |\hat{e}_i\rangle\langle\hat{e}_j| \;. \tag{9.11}$$

Now comes the trick. We can always perform a unitary transformation to a new basis $|e_i\rangle$ in $\mathcal{H}_1$, so that ρ_1 takes the diagonal form

$$\rho_1 \;=\; \sum_{i=1}^{N_1} \lambda_i |e_i\rangle\langle e_i| \;, \tag{9.12}$$

where the coefficients λ_i are real and non-negative. Finally we go back and repeat the argument, using this basis from the start. Taking the hats away, we find

$$\langle\phi_j|\phi_i\rangle \;=\; \lambda_i\delta_{ij} \;. \tag{9.13}$$

That is to say, we can set $|\phi_i\rangle = \sqrt{\lambda_i}|f_i\rangle$. The result is precisely the Schmidt decomposition.

An alternative way to obtain the Schmidt decomposition is to rely on the singular value decomposition (8.14) of the matrix C in Eq. (9.9). In Section 8.1 we considered square matrices, but since the singular values are really the square roots of the

eigenvalues of the matrix $CC^\dagger$ – which is square in any case – we can lift that restriction here. Let the singular values of C be $\sqrt{\lambda_i}$. There exist two unitary matrices U and V such that

$$C_{ij} \;=\; \sum_{k,l} U_{ik}\sqrt{\lambda_k}\,\delta_{kl}V_{lj}\;. \tag{9.14}$$

Using U and V to effect changes of the bases in $\mathcal{H}_1$ and $\mathcal{H}_2$ we recover the Schmidt decomposition (9.8). Indeed

$$\rho_1 \;\equiv\; \mathrm{Tr}_2\rho = CC^\dagger \quad \text{and} \quad \rho_2 \;\equiv\; \mathrm{Tr}_1\rho = C^T C^*\;. \tag{9.15}$$

In the generic case all the singular values λ_i are different and the Schmidt decomposition is unique up to phases, which are free parameters determined by any specific choice of the eigenvectors of U and V. The bases used in the Schmidt decomposition are distinguished because they are precisely the eigenbases of the reduced density matrices, one of them is given in Eq. (9.12) and the other being

$$\rho_2 = \mathrm{Tr}_1\big(|\Psi\rangle\langle\Psi|\big) = \sum_i \lambda_i\,|f_i\rangle\langle f_i|\;. \tag{9.16}$$

When the spectra of the reduced density matrices are degenerate the bases may be rotated in the corresponding subspace.

At this point we introduce some useful terminology. The real numbers λ_i that occur in the Schmidt decomposition (9.8) are called *Schmidt coefficients*,[4] and they obey

$$\sum_i \lambda_i = 1\;, \qquad \lambda_i \geq 0\;. \tag{9.17}$$

The set of all possible vectors $\vec{\lambda}$ forms an $(N-1)$- dimensional simplex, known as the *Schmidt simplex*. The number r of non-vanishing λ_i is called the *Schmidt rank* of the state $|\Psi\rangle$. It is equal to the rank of the reduced density matrix. The latter describes a pure state if and only if $r = 1$. If $r > 1$ the state $|\Psi\rangle$ is an entangled state of its two subsystems (see Chapter 15).

A warning concerning the Schmidt decomposition is appropriate: there is no similar strong result available for Hilbert spaces that are direct products of more than two factor spaces.[5] This is evident because if there are M factor spaces, all of dimension N, then the number of parameters describing a general state grows like N^M, while the number of unitary transformations one can use to choose basis vectors within the factors grows like $M \times N^2$. But we can look at the Schmidt

[4] We find this definition convenient. Others (Nielsen and Chuang, 2000) use this name for $\sqrt{\lambda_i}$.

[5] A kind of generalization of the Schmidt decomposition for three qubits is provided in (Carteret, Higuchi and Sudbery, 2000; Acín, Andrianov, Costa, Jané, Latorre and Tarrach, 2000).

decomposition through different glasses. The Schmidt coefficients are not changed by *local unitary transformations*, – that is to say, in the Hilbert space $\mathcal{H} \otimes \mathcal{H}$, where both factors have dimension N, transformations belonging to the subgroup $U(N) \otimes U(N)$, acting on each factor separately. When there are many factors, we can ask for invariants under the action of $U(N) \otimes U(N) \otimes \cdots \otimes U(N)$, characterizing the orbits of that group – but this is an active research subject[6] that we do not go into.

9.3 State purification and the Hilbert–Schmidt bundle

With the Schmidt decomposition in hand we can discuss the opposite of state reduction: given any density matrix ρ on a Hilbert space $\mathcal{H}$, we can use Eq. (9.8) to write down a pure state on a larger Hilbert space whose reduction down to $\mathcal{H}$ is ρ. The key statements are the following:

Lemma 9.1 (Reduction) *Let ρ_{12} be a pure state on $\mathcal{H}_{12}$. Then the spectra of the reduced density matrices ρ_1 and ρ_2 are identical, except possibly for the degeneracy of any zero eigenvalue.*

Lemma 9.2 (Purification) *Given a density matrix ρ_1 on a Hilbert space $\mathcal{H}_1$, there exists a Hilbert space $\mathcal{H}_2$ and a pure state ρ_{12} on $\mathcal{H}_1 \otimes \mathcal{H}_2$ such that $\rho_1 = Tr_2 \rho_{12}$.*

These statements follow trivially from Schmidt's theorem, but they have far-reaching consequences. It is notable that any density matrix ρ acting on a Hilbert space $\mathcal{H}$ can be purified in the Hilbert–Schmidt space $\mathcal{HS} = \mathcal{H} \otimes \mathcal{H}^*$, that we introduced in Section 8.1. Any attempt to use a smaller Hilbert space will fail in general, and, mathematically, there is no point in choosing a larger space since the purified density matrices will always belong to a subspace that is isomorphic to the Hilbert–Schmidt space. Hence Hilbert–Schmidt space provides a canonical arena for the purification of density matrices. We will try to regard it as a fibre bundle, along the lines of Chapter 3. Let us see if we can.

The vectors of $\mathcal{HS}$ can be represented as operators A acting on $\mathcal{H}$, and there is a projection down to the cone $\mathcal{P}$ of positive operators defined by

$$\Pi : \quad A \longrightarrow \rho = AA^\dagger . \tag{9.18}$$

The fibres will consist of operators projecting to the same positive operator, and the unitary group acts on the fibres as

$$A \longrightarrow A' = AU . \tag{9.19}$$

[6] To learn about invariants of local operations for three qubits see (Grassl, Rötteler and Beth, 1998; Sudbery, 2001; Barnum and Linden, 2001).

We could have used the projection $A \longrightarrow \rho' = A^\dagger A$ instead. More interestingly, we could have used the projection $A \longrightarrow \rho' = AA^\dagger/\mathrm{Tr}AA^\dagger$. This would take us all the way down to the density matrices (of unit trace), but the projection (9.18) turns out to be more convenient to work with.

Do we have a fibre bundle? Not quite, because the fibres are not isomorphic. We do have a fibre bundle if we restrict the bundle space to be the open set of Hilbert–Schmidt operators with trivial kernel. The boundary of the base manifold is not really lost, since it can be recovered by continuity arguments. And the fibre bundle perspective is really useful, so we will adopt it here.[7] The structure group of the bundle is $U(N)$ and the base manifold is the interior of the positive cone. The bundle projection is given by Eq. (9.18). From a topological point of view this is a trivial bundle, admitting a global section

$$\tau \; : \quad \rho \longrightarrow \sqrt{\rho}\,. \tag{9.20}$$

The map τ is well defined because a positive operator admits a unique positive square root, it is a section because $\Pi\big(\tau(\rho)\big) = (\sqrt{\rho})^2 = \rho$, and it is global because it works everywhere.

What is interesting about our bundle is its geometry. We want to think of Hilbert–Schmidt space as a real vector space, so we adopt the metric

$$X \cdot Y = \frac{1}{2}(\langle X, Y\rangle + \langle Y, X\rangle) = \frac{1}{2}\mathrm{Tr}(X^\dagger Y + Y^\dagger X)\,, \tag{9.21}$$

where X and Y are tangent vectors. (Because we are in a vector space, the tangent spaces can be identified with the space itself.) This is the *Hilbert–Schmidt bundle*. A matrix in the bundle space will project to a properly normalized density matrix if and only if it sits on the unit sphere in $\mathcal{HS}$. The whole setting is quite similar to that encountered for the 3-sphere in Chapter 3. Like the 3-sphere, the Hilbert–Schmidt bundle space has a preferred metric, and therefore there is a preferred connection and a preferred metric on the base manifold.

According to Section 3.6, a connection is equivalent to a decomposition of the bundle tangent space into vertical and horizontal vectors. The vertical tangent vectors pose no problem. By definition they point along the fibres; since any unitary matrix U can be obtained by exponentiating an Hermitian matrix H, a curve along a fibre is given by

$$A\,U(t) \;=\; A\,\mathrm{e}^{\mathrm{i}Ht}\,. \tag{9.22}$$

Therefore every vertical vector takes the form $\mathrm{i}AH$ for some Hermitian matrix H. The horizontal vectors must be defined somehow, and we do so by requiring that

[7] From this point on, this chapter is mostly an account of ideas developed by Armin Uhlmann (Uhlmann, 1992, 1993, 1995) and his collaborators. For this section, see also Dąbrowski and Jadczyk (1989).

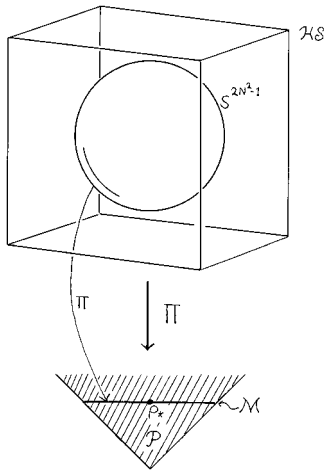


Figure 9.1. The Hilbert–Schmidt bundle. It is the unit sphere in $\mathcal{HS}$ that projects down to density matrices.

they are orthogonal to the vertical vectors under our metric. Thus, for a horizontal vector X, we require

$$\mathrm{Tr}X(\mathrm{i}AH)^\dagger + \mathrm{Tr}(\mathrm{i}AH)X^\dagger \;=\; \mathrm{i}\,\mathrm{Tr}(X^\dagger A - A^\dagger X)H = 0 \tag{9.23}$$

for all Hermitian matrices H. Hence X is a horizontal tangent vector at the point A if and only if

$$X^\dagger A - A^\dagger X \;=\; 0\,. \tag{9.24}$$

Thus equipped, we can lift curves in the base manifold to horizontal curves in the bundle.

In particular, suppose that we have a curve $\rho(s)$ in $\mathcal{M}^{(N)}$. We are looking for a curve $A(s)$ such that $AA^\dagger(s) = \rho(s)$, and such that its tangent vector $\dot{A}$ is horizontal, that is to say that

$$\dot{A}^\dagger A = A^\dagger \dot{A}\,. \tag{9.25}$$

It is easy to see that the latter condition is fulfilled if

$$\dot{A} = GA\,, \tag{9.26}$$

where G is an Hermitian matrix. To find the matrix G, we observe that

$$AA^\dagger(\sigma) = \rho(\sigma) \qquad \Rightarrow \qquad \dot{\rho} = G\rho + \rho G\,. \tag{9.27}$$

As long as ρ is a strictly positive operator this equation determines G uniquely (Sylvester, 1884; Bhatia and Rosenthal, 1997), and it follows that the horizontal

lift of a curve in the base space is uniquely determined. We could go on to define a mixed state generalization of the geometric phase discussed in Section 4.8, but in fact we will turn to somewhat different matters.[8]

9.4 A first look at the Bures metric

Out of our bundle construction comes, not only a connection, but a natural metric on the space of density matrices. It is known as the *Bures metric*, and it lives on the cone of positive operators on $\mathcal{H}$, since this is the base manifold of our bundle. Until further notice then, ρ denotes a positive operator, and we allow $\mathrm{Tr}\rho \neq 1$. The purification of ρ is a matrix A such that $\rho = AA^\dagger$, and A is regarded as a vector in the Hilbert–Schmidt space.

In the bundle space, we have a natural notion of distance, namely the Euclidean distance defined (without any factor $1/2$) by

$$d_B^2(A_1, A_2) = ||A_1 - A_2||_{\mathrm{HS}}^2 = \mathrm{Tr}(A_1A_1^\dagger + A_2A_2^\dagger - A_1A_2^\dagger - A_2A_1^\dagger) \,. \tag{9.28}$$

If A_1, A_2 lie on the unit sphere we have another natural distance, namely the geodesic distance d_A given by

$$\cos d_A \;=\; \frac{1}{2}\,\mathrm{Tr}(A_1A_2^\dagger + A_2A_1^\dagger) \,. \tag{9.29}$$

Unlike the Euclidean distance, which measures the length of a straight chord, the second distance measures the length of a curve that projects down, in its entirety, to density matrices of unit trace. In accordance with the philosophy of Chapter 3, we define the distance between two density matrices ρ_1 and ρ_2 as the length of the shortest path, in the bundle, that connects the two fibres lying over these density matrices. Whether we choose to work with d_A or d_B, the technical task we face is to calculate the *root fidelity*[9]

$$\sqrt{F}(\rho_1, \rho_2) \;\equiv\; \frac{1}{2}\max\,\mathrm{Tr}(A_1A_2^\dagger + A_2A_1^\dagger) = \max|\mathrm{Tr}A_1A_2^\dagger| \,. \tag{9.30}$$

The optimization is with respect to all possible purifications of ρ_1 and ρ_2. Once we have done this, we can define the *Bures distance* D_{B},

$$D_{\mathrm{B}}^2(\rho_1, \rho_2) \;=\; \mathrm{Tr}\rho_1 + \mathrm{Tr}\rho_2 - 2\sqrt{F}(\rho_1, \rho_2) \,, \tag{9.31}$$

[8] Geometric phases were among Uhlmann's motivations for developing the material in this chapter (Uhlmann, 1992; Uhlmann, 1995). Other approaches to geometric phases for mixed states exist (Ericsson, Sjöqvist, Brännlund, Oi and Pati, 2003); for a recent review see Chruściński and Jamiołkowski (2004).

[9] Its square was called *fidelity* by Jozsa (1994). Later several authors, including Nielsen and Chuang (2000), began to refer to our root fidelity as fidelity. We have chosen to stick with the original names, partly to avoid confusion, partly because experimentalists prefer a fidelity to be some kind of a probability – and fidelity is a kind of transition probability, as we will see.

and the *Bures angle* $D_{\rm A}$,

$$\cos D_{\rm A}(\rho_1, \rho_2) \;=\; \sqrt{F}(\rho_1, \rho_2)\;. \tag{9.32}$$

The Bures angle is a measure of the length of a curve within $\mathcal{M}^{(N)}$, while the Bures distance measures the length of a curve within the positive cone. By construction, they are Riemannian distances – and indeed they are consistent with the same Riemannian metric. Moreover they are both monotoneously decreasing functions of the root fidelity.[10]

Root fidelity is a useful concept in its own right and will be discussed in some detail in Section 13.3. It is so useful that we state its evaluation as a theorem:

Theorem 9.2 (Uhlmann's fidelity) *The root fidelity, defined as the maximum of* $|TrA_1A_2^\dagger|$ *over all possible purifications of two density matrices* ρ_1 *and* ρ_2*, is*

$$\sqrt{F}(\rho_1, \rho_2) \;=\; {\rm Tr}|\sqrt{\rho_2}\,\sqrt{\rho_1}| \;=\; {\rm Tr}\sqrt{\sqrt{\rho_2}\,\rho_1\sqrt{\rho_2}}\;. \tag{9.33}$$

To prove this, we first use the polar decomposition to write

$$A_1 = \sqrt{\rho_1}\,U_1 \quad \text{ and } \quad A_2 = \sqrt{\rho_2}\,U_2\;. \tag{9.34}$$

Here U_1 and U_2 are unitary operators that move us around the fibres. Then

$${\rm Tr}A_1A_2^\dagger = {\rm Tr}(\sqrt{\rho_1}\,U_1U_2^\dagger\sqrt{\rho_2}) = {\rm Tr}(\sqrt{\rho_2}\sqrt{\rho_1}\,U_1U_2^\dagger)\;. \tag{9.35}$$

We perform yet another polar decomposition

$$\sqrt{\rho_2}\sqrt{\rho_1} = |\sqrt{\rho_2}\sqrt{\rho_1}|V\;, \qquad VV^\dagger = 1\;. \tag{9.36}$$

We define a new unitary operator $U \equiv VU_1U_2^\dagger$. The final task is to maximize

$${\rm Tr}(|\sqrt{\rho_2}\sqrt{\rho_1}|U) + \text{complex conjugate} \tag{9.37}$$

over all possible unitary operators U. In the eigenbasis of the positive operator $|\sqrt{\rho_2}\sqrt{\rho_1}|$ it is easy to see that the maximum occurs when $U = 1$. This proves the theorem; the definition of the Bures distance, and of the Bures angle, is thereby complete.

The catch is that root fidelity is difficult to compute. Because of the square roots, we must go through the labourious process of diagonalizing a matrix twice. Indeed, although our construction makes it obvious that $\sqrt{F}(\rho_1, \rho_2)$ is a symmetric function of ρ_1 and ρ_2, not even this property is obvious just by inspection of the formula – although in Section 13.3 we will give an elegant direct proof of this property. To

[10] The Bures distance was introduced, in an infinite-dimensional setting, by Bures (1969), and then shown to be a Riemannian distance by Uhlmann (1992). Our Bures angle was called *Bures length* by Uhlmann (1995), and *angle* by Nielsen and Chuang (2000).

come to grips with root fidelity, we work it out in two simple cases, beginning with the case when $\rho_1 = \mathrm{diag}(p_1, p_2, \ldots, p_N)$ and $\rho_2 = \mathrm{diag}(q_1, q_2, \ldots, q_N)$, that is when both matrices are diagonal. We also assume that they have trace one. This is an easy case: we get

$$\sqrt{F}(\rho_1, \rho_2) \;=\; \sum_{i=1}^{N} \sqrt{p_i q_i}\;. \tag{9.38}$$

It follows that the Bures angle D_{A} equals the classical Bhattacharyya distance, while the Bures distance is given by

$$D_{\mathrm{B}}^2(\rho_1, \rho_2) = 2 - 2\sum_{i=1}^{N} \sqrt{p_i q_i} = \sum_{i=1}^{N} (\sqrt{p_i} - \sqrt{q_i})^2 = D_{\mathrm{H}}^2(P, Q)\;, \tag{9.39}$$

where D_{H} is the Hellinger distance between two classical probability distributions. These distances are familiar from Section 2.5. Both of them are consistent with the Fisher–Rao metric on the space of classical probability distributions, so this is our first hint that what we are doing will have some statistical significance.

The second easy case is that of two pure states. The good thing about a pure density matrix is that it squares to itself and therefore equals its own square root. For a pair of pure states a very short calculation shows that

$$\sqrt{F}\big(|\psi_1\rangle\langle\psi_1|, |\psi_2\rangle\langle\psi_2|\big) = |\langle\psi_1|\psi_2\rangle| = \sqrt{\kappa}\;, \tag{9.40}$$

where κ is the projective cross-ratio, also known as the transition probability. It is therefore customary to refer to *fidelity*, that is the square of root fidelity, also as the *Uhlmann transition probability*, regardless of whether the states are pure or not. Anyway, we can conclude that the Bures angle between two pure states is equal to their Fubini–Study distance.

With some confidence that we are studying an interesting definition, we turn to the Riemannian metric defined by the Bures distance. It admits a compact description that we will derive right away, although we will not use it until Section 14.1. It will be convenient to use an old-fashioned notation for tangent vectors, so that $\mathrm{d}A$ is a tangent vector on the bundle, projecting to $\mathrm{d}\rho$, which is a tangent vector on $\mathcal{M}^{(N)}$. The length squared of $\mathrm{d}\rho$ is then defined by

$$\mathrm{d}s^2 = \min\big[\mathrm{Tr}\,\mathrm{d}A\,\mathrm{d}A^\dagger\big]\;, \tag{9.41}$$

where the minimum is sought among all vectors dA that project to $\mathrm{d}\rho$, and achieved if $\mathrm{d}A$ is a horizontal vector (orthogonal to the fibres). According to Eq. (9.26) this happens if and only if $\mathrm{d}A = GA$, where G is a Hermitian matrix. As we know from Eq. (9.27), as long as ρ is strictly positive, G will be determined uniquely by ,Ç

$$\mathrm{d}\rho = G\rho + \rho\, G\;. \tag{9.42}$$

Pulling the strings together, we find that

$$ds^2 = \mathrm{Tr}\, GAA^\dagger G = \mathrm{Tr} G\rho G = \frac{1}{2}\mathrm{Tr}\, G d\rho\ . \tag{9.43}$$

This is the *Bures metric*. Its definition is somewhat implicit. It is difficult to do better though: explicit expressions in terms of matrix elements tend to become so complicated that they seem useless – except when ρ and $d\rho$ commute, in which case $G = d\rho/(2\rho)$, and except for the special case $N = 2$ to which we now turn.[11] A head on attack on Eq. (9.43) will be made in Section 14.1.

9.5 Bures geometry for $N = 2$

It happens that for the qubit case, $N = 2$, we can get fully explicit results with elementary means. The reason is that every 2×2 matrix M obeys

$$M^2 - M\,\mathrm{Tr}M + \det M = 0\ . \tag{9.44}$$

Hence

$$(\mathrm{Tr}M)^2 = \mathrm{Tr}M^2 + 2\det M\ . \tag{9.45}$$

If we set

$$M \equiv \sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\,, \tag{9.46}$$

we find, as a result of an elementary calculation, that

$$F = (\mathrm{Tr}M)^2 = \mathrm{Tr}\rho_1\rho_2 + 2\sqrt{\det\rho_1 \det\rho_2}\,, \tag{9.47}$$

(where the fidelity F is used for the first time!). The $N = 2$ Bures distance is now given by

$$D_{\mathrm{B}}^2(\rho_1, \rho_2) = \mathrm{Tr}\rho_1 + \mathrm{Tr}\rho_2 - 2\sqrt{\mathrm{Tr}\rho_1\rho_2 + 2\sqrt{\det\rho_1 \det\rho_2}}\,. \tag{9.48}$$

It is pleasing that no square roots of operators appear in this expression.

It is now a matter of straightforward calculation to obtain an explicit expression for the Riemannian metric on the positive cone, for $N = 2$. To do so, we set

$$\rho_1 = \frac{1}{2}\begin{bmatrix} t+z & x-\mathrm{i}y \\ x+\mathrm{i}y & t-z \end{bmatrix}\,, \qquad \rho_2 = \rho_1 + \frac{1}{2}\begin{bmatrix} \mathrm{d}t+\mathrm{d}z & \mathrm{d}x-\mathrm{i}\mathrm{d}y \\ \mathrm{d}x+\mathrm{i}\mathrm{d}y & \mathrm{d}t-\mathrm{d}z \end{bmatrix}\,. \tag{9.49}$$

It is elementary (although admittedly a little labourious) to insert this in Eq. (9.48),

[11] In the $N = 2$ case we follow Hübner (1992). Actually Dittmann (1999a) has provided an expression valid for all N, which is explicit in the sense that it depends only on matrix invariants, and does not require diagonalization of any matrix.

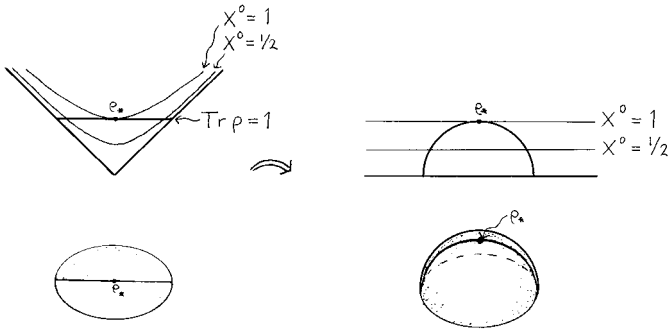


Figure 9.2. Left: a faithful illustration of the Hilbert–Schmidt geometry of a rebit (a flat disc, compare with Figure 8.2). Right: the same for its Bures geometry (a round hemisphere). Above the rebit we show exactly how it sits in the positive cone. On the right the latter appears very distorted, because we have adapted its coordinates to the Bures geometry.

and expand to second order. The final result, for the Bures line element squared, is

$$\mathrm{d}s^2 = \frac{1}{4t}\left(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \frac{(x\mathrm{d}x + y\mathrm{d}y + z\mathrm{d}z - t\mathrm{d}t)^2}{t^2 - x^2 - y^2 - z^2}\right) . \tag{9.50}$$

In the particular case that t is constant, so that we are dealing with matrices of constant trace, this is recognizable as the metric on the upper hemisphere of the 3-sphere, of radius $1/2\sqrt{t}$, in the orthographic coordinates introduced in Eq. (3.2). Indeed we can introduce the coordinates

$$X^0 = \sqrt{t^2 - x^2 - y^2 - z^2}\,, \quad X^1 = x\,, \quad X^2 = y\,, \quad X^3 = z\,. \tag{9.51}$$

Then the Bures metric on the positive cone is

$$\mathrm{d}s^2 = \frac{1}{4t}(\mathrm{d}X^0\mathrm{d}X^0 + \mathrm{d}X^1\mathrm{d}X^1 + \mathrm{d}X^2\mathrm{d}X^2 + \mathrm{d}X^3\mathrm{d}X^3)\,, \tag{9.52}$$

where

$$\mathrm{Tr}\rho = t = \sqrt{(X^0)^2 + (X^1)^2 + (X^2)^2 + (X^3)^2}\,. \tag{9.53}$$

Only the region for which $X^0 \geq 0$ is relevant.

Let us set $t = 1$ for the remainder of this section, so that we deal with matrices of unit trace. We see that, according to the Bures metric, they form a hemisphere of a 3-sphere of radius $1/2$; the pure states sit at its equator, which is a 2-sphere isometric with $\mathbb{C}\mathbf{P}^1$. Unlike a 2-sphere in Euclidean space, the equator of the 3-sphere is a totally geodesic surface – by definition, a surface such that a geodesic within the surface itself is also a geodesic in the embedding space. We can draw a picture (Figure 9.2) that summarizes the Bures geometry of the qubit. Note that the set of

diagonal density matrices appears as a semicircle in this picture, not as the quarter circle that we had in Figure 2.13. Actually, because this set is one-dimensional, the intrinsic geometries on the two circle segments are the same, the length is $\pi/2$ in both cases, and there is no contradiction.

Finally, the qubit case is instructive, but it is also quite misleading in some respects – in particular the case $N = 2$ is especially simple to deal with.

9.6 Further properties of the Bures metric

When $N > 2$ it does not really pay to proceed as directly as we did for the qubit, but the fibre bundle origins of the Bures metric mean that much can be learned about it with indirect means. First, what is a geodesic with respect to the Bures metric? The answer is that it is a projection of a geodesic in the unit sphere embedded in the bundle space $\mathcal{HS}$, with the added condition that the latter geodesic must be chosen to be orthogonal to the fibres of the bundle. We know what a geodesic on a sphere looks like, namely (Section 3.1)

$$A(s) \;=\; A(0)\cos s + \dot{A}(0)\sin s \;, \tag{9.54}$$

where

$$\mathrm{Tr}A(0)A^\dagger(0) = \mathrm{Tr}\dot{A}(0)\dot{A}^\dagger(0) = 1 \;,\quad \mathrm{Tr}\left(A(0)\dot{A}^\dagger(0) + \dot{A}(0)A^\dagger(0)\right) = 0 \;. \tag{9.55}$$

The second equation just says that the tangent vector of the curve is orthogonal to the vector defining the starting point on the sphere. In addition the tangent vector must be horizontal; according to Eq. (9.24) this means that we must have

$$\dot{A}^\dagger(0)A(0) \;=\; A^\dagger(0)\dot{A}(0) \;. \tag{9.56}$$

That is all. An interesting observation – we will see why in a moment – is that if we start the geodesic in a point where A, and hence $\rho = AA^\dagger$, is block diagonal, and if the tangent vector $\dot{A}$ at that point is block diagonal too, then the entire geodesic will consist of block diagonal matrices. The conclusion is that block diagonal density matrices form totally geodesic submanifolds in the space of density matrices.

Now let us consider a geodesic that joins the density matrices ρ_1 and ρ_2, and let them be projections of A_1 and A_2, respectively. The horizontality condition says that $A_1^\dagger A_2$ is a Hermitian operator, and in fact a positive operator if the geodesic does not hit the boundary in between. From this one may deduce that

$$A_2 \;=\; \frac{1}{\sqrt{\rho_1}}\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\,\frac{1}{\sqrt{\rho_1}}\,A_1 \;. \tag{9.57}$$

The operator front of A_1 is known as the *geometric mean* of ρ_1^{-1} and ρ_2; see Section 12.1. It can also be proved that the geodesic will bounce N times from the

boundary of $\mathcal{M}^{(N)}$, before closing on itself (Uhlmann, 1995). The overall conclusion is that we do have control over geodesics and geodesic distances with respect to the Bures metric.

Concerning symmetries, it is known that any bijective transformation of the set of density matrices into itself which conserves the Bures distance (or angle) is implemented by a unitary or an anti-unitary operation (Molnár, 2001). This result is a generalization of Wigner's theorem concerning the transformations of pure states that preserve the transition probabilities (see Section 4.5).

For further insight we turn to a cone of density matrices in $\mathcal{M}^{(3)}$, having a pure state for its apex and a Bloch ball of density matrices with orthogonal support for its base. This can be coordinatized as

$$\rho = \begin{bmatrix} t\rho^{(2)} & 0 \\ 0 & 1-t \end{bmatrix} = \begin{bmatrix} t(1+z)/2 & t(x-\mathrm{i}y)/2 & 0 \\ t(x+\mathrm{i}y)/2 & t(1-z)/2 & 0 \\ 0 & 0 & 1-t \end{bmatrix} . \tag{9.58}$$

This is a submanifold of block diagonal matrices. It is also simple enough so that we can proceed directly, as in the Section 9.5. Doing so, we find that the metric is

$$\mathrm{d}s^2 \;=\; \frac{\mathrm{d}t^2}{4t(1-t)} + \frac{t}{4}\mathrm{d}^2\Omega \,, \tag{9.59}$$

where $\mathrm{d}^2\Omega$ is the metric on the unit 3-sphere (in orthographic coordinates, and only one half of the 3-sphere is relevant). As $t \to 0$, that is as we approach the tip of our cone, the radii of the 3-spheres shrink, and their intrinsic curvature diverges. This does not sound very dramatic, but in fact it is, because by our previous argument about block diagonal matrices these 3-hemispheres are totally geodesic submanifolds of the space of density matrices. Now it is a fact from differential geometry that if the intrinsic curvature of a totally geodesic submanifold diverges, then the curvature of the entire manifold also diverges. (More precisely, the sectional curvatures, evaluated for 2-planes that are tangent to the totally geodesic submanifold, will agree. We hope that this statement sounds plausible, even though we will not explain it further.) The conclusion is that $\mathcal{M}^{(3)}$, equipped with the Bures metric, has conical curvature singularities at the pure states. The general picture is as follows (Dittmann, 1995):

Theorem 9.3 (Dittmann's) *For $N \geq 2$, the Bures metric is everywhere well defined on submanifolds of density matrices with constant rank. However, the sectional curvature of the entire space diverges in the neighbourhood of any submanifold of rank less than $N-1$.*

For $N > 2$, this means that it is impossible to embed $\mathcal{M}^{(N)}$ into a Riemannian manifold of the same dimension, such that the restriction of the embedding to

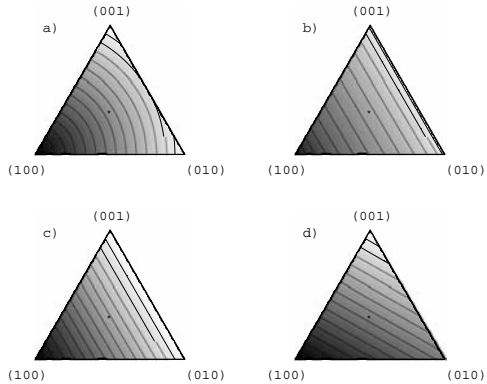


Figure 9.3. The eigenvalue simplex for $N = 3$. The curves consist of points equidistant from the pure state $(1, 0, 0)$ with respect to (a) Hilbert–Schmidt distance, (b) Bures distance, (c) trace distance (Section 13.2) and (d) Monge distance (Section 7.7).

submanifolds of density matrices of constant rank is isometric. The problem does not arise for the special case $N = 2$, and indeed we have seen that $\mathcal{M}^{(2)}$ can be embedded into the 3-sphere. Some further facts are known. Thus, the curvature scalar R assumes its global minimum at the maximally mixed state $\rho_* = \mathbb{1}/N$. It is then natural to conjecture that the scalar curvature is monotone, in the sense that if $\rho_1 \prec \rho_2$, that is if ρ_1 is majorized by ρ_2, then $R(\rho_1) \leq R(\rho_2)$. However, this is not true.[12]

This is perhaps a little disappointing. To recover our spirits, let us look at the Bures distance in two cases where it is very easy to compute. The Bures distance to the maximally mixed state is

$$D_{\mathrm{B}}^2(\rho, \rho_*) \;=\; 1 + \mathrm{Tr}\rho - \frac{2}{\sqrt{N}}\,\mathrm{Tr}\,\sqrt{\rho}\,. \tag{9.60}$$

To compute this it is enough to diagonalize ρ. The distance from an arbitrary density matrix ρ to a pure state is even easier to compute, and is given by a single matrix element of ρ. Figure 9.3 shows where density matrices equidistant to a pure state lie on the probability simplex, for some of the metrics that we have considered. In particular, the distance between a face and its complementary, opposite face – that is, between density matrices of orthogonal support – is constant and maximal, when the Bures metric is used.

[12] The result here is due to Dittmann (1999b), who also found a counter-example to the conjecture (but did not publish it, as far as we know).

We are not done with fidelity and the Bures metric. We will come back to these things in Section 13.3, and place them in a wider context in Section 14.1. This context is – as we have hinted already – that of statistical distinguishability and monotonicity under appropriate stochastic maps. Precisely what is appropriate here will be made clear in the next two chapters.

Problems

Problem 9.1 Check that:

(a) $A \otimes (B + C) = A \otimes B + A \otimes C$;
(b) $(A \otimes B)(C \otimes D) = (AC \otimes BD)$;
(c) $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(B \otimes A) = (\mathrm{Tr}A)(\mathrm{Tr}B)$;
(d) $\det(A \otimes B) = (\det A)^M (\det B)^N$;
(e) $(A \otimes B)^T = A^T \otimes B^T$;

where N and M denote sizes of A and B, respectively (Horn and Johnson, 1985, 1991).

Problem 9.2 Define the Hadamard product $C = A \circ B$ of two matrices as the matrix whose elements are the products of the corresponding elements of A and B, $C_{ij} = A_{ij} B_{ij}$. Show that $(A \otimes B) \circ (C \otimes D) = (A \circ C) \otimes (B \circ D)$.

Problem 9.3 Consider any matrix of size 4 written in standard basis in terms of four 2×2 blocks

$$G = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \tag{9.61}$$

and two local unitary operations $V_1 = \mathbb{1} \otimes U$ and $V_2 = U \otimes \mathbb{1}$, where U is arbitrary unitary matrix of size 2. Compute $G_1 = V_1 G V_1^\dagger$ and $G_2 = V_2 G V_2^\dagger$.

Problem 9.4 Let A and B be square matrices with eigenvalues α_i and β_i, respectively. Find the spectrum of $C = A \otimes B$. Use this to prove that $C' = B \otimes A$ is unitarily similar to C, and also that C is positive definite whenever A and B are positive definite.

Problem 9.5 Show that the singular values of a tensor product satisfy the relation $\{\mathrm{sv}(A \otimes B)\} = \{\mathrm{sv}(A)\} \times \{\mathrm{sv}(B)\}$.

Problem 9.6 Let ρ be a density matrix and A and B denote any matrices of the same size. Show that $|\mathrm{Tr}(\rho AB)|^2 \leq \mathrm{Tr}(\rho AA^\dagger) \times \mathrm{Tr}(\rho BB^\dagger)$.

10

Quantum operations

There is no measurement problem. Bohr cleared that up.
Stig Stenholm

So far we have described the space of quantum states. Now we will allow some action in this space: we shall be concerned with quantum dynamics. At first sight this seems to be two entirely different issues – it is one thing to describe a given space and another to characterize the way you can travel in it – but we will gradually reveal an intricate link between them.

In this chapter we draw on results from the research area known as *open quantum systems*. Our aim is to understand the quantum analogue of the classical stochastic maps, because with their help we reach a better understanding of the structure of the space of states. Stochastic maps can also be used to provide a kind of stroboscopic time evolution; much of the research on open quantum systems is devoted to understanding how continuous time evolution takes place, but for this we have to refer to the literature.[1]

10.1 Measurements and POVMs

Throughout, the system of interest is described by a Hilbert space $\mathcal{H}_N$ of dimension N. All quantum operations can be constructed by composing four kinds of transformations.

The dynamics of an isolated quantum system are given by **i) unitary transformations**. But quantum theory for open systems admits non-unitary processes as well. We can **ii) extend the system** and define a new state in an extended Hilbert

[1] Pioneering results in this direction were obtained by Gorini, Kossakowski and Sudarshan (1976) and by Lindblad (1976). Good books on the subject include Alicki and Lendi (1987), Streater (1995), Ingarden, Kossakowski and Ohya (1997), Breuer and Petruccione (2002) and Alicki and Fannes (2001).

space $\mathcal{H} = \mathcal{H}_N \otimes \mathcal{H}_K$,

$$\rho \;\rightarrow\; \rho' \;=\; \rho \otimes \sigma \;. \tag{10.1}$$

The auxiliary system is described by a Hilbert space $\mathcal{H}_K$ of dimension K (as yet unrelated to N). It represents an environment, and is often referred to as the *ancilla*.[2] The reverse of this operation is given by the **iii) partial trace** and leads to a reduction of the size of the Hilbert space,

$$\rho \;\rightarrow\; \rho' \;=\; \mathrm{Tr}_K \rho \quad \text{so that} \quad \mathrm{Tr}_K(\rho \otimes \sigma) \;=\; \rho \;. \tag{10.2}$$

This corresponds to discarding the redundant information concerning the fate of the ancilla. Transformations that can be achieved by a combination of these three kinds of transformation are known as *deterministic* or *proper quantum operations*.

Finally, we have the **iv) selective measurement**, in which a concrete result of a measurement is specified. This is called a *probabilistic quantum operation*.

Let us see where we get using tranformations of the first three kinds. Let us assume that the ancilla starts out in a pure state $|\nu\rangle$, while the system we are analysing starts out in the state ρ. The entire system including the ancilla remains isolated and evolves in a unitary fashion. Adding the ancilla to the system (ii), evolving the combined system unitarily (i), and tracing out the ancilla at the end (iii), we find that the state ρ is changed to

$$\rho' = \mathrm{Tr_K}\Big[U\big(\rho \otimes |\nu\rangle\langle\nu|\big)U^\dagger\Big] = \sum_{\mu=1}^{K} \langle\mu|U|\nu\rangle \rho \langle\nu|U^\dagger|\mu\rangle \;. \tag{10.3}$$

where $\{|\mu\rangle\}_{\mu=1}^{K}$ is a basis in the Hilbert space of the ancilla – and we use Greek letters to denote its states. We can then define a set of operators in the Hilbert space of the original system through

$$A_\mu \;\equiv\; \langle\mu|U|\nu\rangle \;. \tag{10.4}$$

We observe that

$$\sum_{\mu=1}^{K} A_\mu^\dagger A_\mu = \sum_{\mu} \langle\nu|U^\dagger|\mu\rangle\langle\mu|U|\nu\rangle = \langle\nu|U^\dagger U|\nu\rangle = \mathbb{1}_N \;, \tag{10.5}$$

where $\mathbb{1}_N$ denotes the unit operator in the Hilbert space of the system of interest.

In conclusion, first we assumed that an isolated quantum system evolves through a unitary transformation,

$$\rho \;\rightarrow\; \rho' \;=\; U \rho\, U^\dagger \;, \qquad U^\dagger U = \mathbb{1} \;. \tag{10.6}$$

[2] In Latin an *ancilla* is a maidservant. This not 100 per cent politically correct expression was imported to quantum mechanics by Helstrom (1976) and has become widely accepted.

By allowing ourselves to add an ancilla, later removed by a partial trace, we were led to admit operations of the form

$$\rho \;\to\; \rho' = \sum_{i=1}^{K} A_i \rho A_i^\dagger \,, \qquad \sum_{i=1}^{K} A_i^\dagger A_i = \mathbb{1} \,, \tag{10.7}$$

where we dropped the subscript on the unit operator and switched to Latin indices, since we are not interested in the environment per se. Formally, this is the *operator sum representation* of a *completely positive map*. Although a rather special assumption was slipped in – a kind of *Stoßzahlansatz* whereby the combined system started out in a product state – we will eventually adopt this expression as the most general quantum operation that we are willing to consider.

The process of *quantum measurement* remains somewhat enigmatic. Here we simply accept without proof a postulate concerning the *collapse of the wave function*. It has the virtue of generality, not of preciseness:

Measurement postulate. Let the space of possible measurement outcomes consist of k elements, related to k *measurement operators* A_i, which satisfy the *completeness relation*

$$\sum_{i=1}^{k} A_i^\dagger A_i \;=\; \mathbb{1} \,. \tag{10.8}$$

The quantum measurement performed on the initial state ρ produces the ith outcome with probability p_i and transforms ρ into ρ_i according to

$$\rho \;\to\; \rho_i \;=\; \frac{A_i \rho A_i^\dagger}{\mathrm{Tr}(A_i \rho A_i^\dagger)} \quad \text{with} \quad p_i = \mathrm{Tr}\big(A_i \rho A_i^\dagger\big) \,. \tag{10.9}$$

The probabilities are positive and sum to unity due to the completeness relation. Such measurements, called *selective* since concrete results labelled by i are recorded, cannot be obtained by the transformations (i)–(iii) and form another class of transformations (iv) on their own. If no selection is made based on the outcome of the measurement, the initial state is transformed into a convex combination of all possible outcomes – namely that given by Eq. (10.7).

Note that the 'collapse' happens in the statistical description that we are using. Similar 'collapses' occur also in classical probability theory. Suppose that we know that either Alice or Bob is in jail, but not both. Let the probability that Bob is in jail be p. If this statement is accepted as meaningful, we find that there is a collapse of the probability distribution associated to Bob as soon as we run into Alice in the cafeteria – even though nothing happened to Bob. This is not a real philosophical difficulty, but the quantum case is subtler. Classically the pure states

are safe from collapse, but in quantum mechanics there are no safe havens. Also, a classical probability distribution $P(X)$ can collapse to a conditional probability distribution $P(X|Y_i)$, but if no selection according to the outcomes Y_i is made classical probability theory informs us that

$$\sum_i P(X|Y_i)\, P(Y_i) \;=\; P(X)\,. \tag{10.10}$$

Thus nothing happens to the probability distribution in a *non-selective* measurement, while the quantum state is severely affected also in this case. A non-selective quantum measurement is described by Eq. (10.7), and this is a mixed state even if the initial state ρ is pure. In general one cannot receive any information about the fate of a quantum system without performing a measurement that perturbs its unitary time evolution.

In a *projective measurement* the measurement operators are orthogonal projectors, so $A_i = P_i = A_i^\dagger$, and $P_i P_j = \delta_{ij} P_i$ for $i, j = 1, \ldots, N$. A projective measurement is described by an observable – an Hermitian operator O. Possible outcomes of the measurement are labelled by the eigenvalues of O, which for now we assume to be non-degenerate. Using the spectral decomposition $O = \sum_i \lambda_i P_i$ we obtain a set of orthogonal measurement operators $P_i = |e_i\rangle\langle e_i|$, satisfying the completeness relation (10.8). In a non-selective projective measurement, the initial state is transformed into the mixture

$$\rho \;\to\; \rho' \;=\; \sum_{i=1}^{N} P_i \rho P_i\,. \tag{10.11}$$

The state has been forced to commute with O.

In a selective projective measurement the outcome labelled by λ_i occurs with probability p_i; the initial state is transformed as

$$\rho \to \rho_i = \frac{P_i \rho P_i}{\mathrm{Tr}(P_i \rho P_i)}\,, \quad \text{where} \quad p_i = \mathrm{Tr}\big(P_i \rho P_i\big) = \mathrm{Tr}\big(P_i \rho\big)\,. \tag{10.12}$$

The expectation value of the observable reads

$$\langle O \rangle = \sum_{i=1}^{N} p_i \lambda_i = \sum_{i=1}^{N} \lambda_i \mathrm{Tr} P_i \rho = \mathrm{Tr}(O\rho)\,. \tag{10.13}$$

A key feature of projective measurements is that they are repeatable, in the sense that the state in Eq. (10.12) remains the same – and gives the same outcome – if the measurement is repeated.[3]

[3] Projective measurements are also called *Lüders–von Neumann* measurements (of the first kind), because of the contributions by von Neumann (1955) and Lüders (1951).

Most measurements are not repeatable. The formalism deals with this by relaxing the orthogonality constraint on the measurement operators. This leads to *Positive Operator Valued Measures* (POVM), which are defined by any partition of the identity operator into a set of k positive operators E_i acting on an N-dimensional Hilbert space $\mathcal{H}_N$. They satisfy

$$\sum_{i=1}^{k} E_i = \mathbb{1} \quad \text{and} \quad E_i = E_i^\dagger, \quad E_i \geq 0\,, \quad i = 1, \ldots, k. \tag{10.14}$$

A POVM measurement applied to the state ρ produces the ith outcome with probability $p_i = \mathrm{Tr} E_i \rho$. Note that the elements of the POVM – the operators E_i – need not commute. The name POVM refers to any set of operators satisfying (10.14), and suggests correctly that the discrete sum may be replaced by an integral over a continuous index set, thus defining a measure in the space of positive operators. Indeed the coherent states resolution of unity (6.7) is the paradigmatic example, yielding the Husimi Q-function as the resulting probability distribution. POVMs fit into the general framework of the measurement postulate, since one may choose $E_i = A_i^\dagger A_i$. Note however that the POVM does not determine the measurement operators A_i uniquely (except in the special case of a projective measurement). Exactly what happens to the state when a measurement is made depends on how the POVM is implemented in the laboratory.[4]

The definition of the POVM ensures that the probabilities $p_i = \mathrm{Tr} E_i \rho$ sum to unity, but the probability distribution that one obtains is a constrained one. (We came across this phenomenon in Section 6.2, when we observed that the Q-function is a very special probability distribution.) This is so because the POVM defines an affine map from the set of density matrices $\mathcal{M}^{(N)}$ to the probability simplex Δ_{k-1}. To see this, use the Bloch vector parametrization

$$\rho = \frac{1}{N}\mathbb{1} + \sum_a \tau_a \sigma_a \quad \text{and} \quad E_i = e_{i0}\mathbb{1} + \sum_a e_{ia}\sigma_a\,. \tag{10.15}$$

Then an easy calculation yields

$$p_i = \mathrm{Tr} E_i \rho = 2\sum_a e_{ia}\tau_a + e_{i0}\,. \tag{10.16}$$

This is an affine map. Conversely, any affine map from $\mathcal{M}^{(N)}$ to Δ_{k-1} defines a POVM. We know from Section 1.1 that an affine map preserves convexity. Therefore the resulting probability vector $\vec{p}$ must belong to a convex subset of the probability simplex. For qubits $\mathcal{M}^{(2)}$ is a ball. Therefore its image is an ellipsoid, degenerating

[4] POVMs were introduced by Jauch and Piron (1967) and they were explored in depth in the books by Davies (1976) and Holevo (1982). Holevo's book is the best source of knowledge that one can imagine. For a more recent discussion, see Peres and Terno (1998).

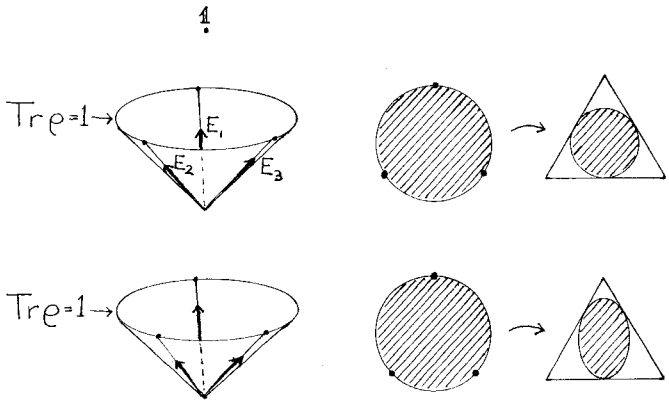


Figure 10.1. Two informationally complete POVMs for a rebit; we show first a realistic picture in the space of real Hermitian matrices, and then the affine map from the rebit to the probability simplex.

to a line segment if the measurement is projective. Figure 10.1 illustrates the case of real density matrices, for which we can draw the positive cone. For $N > 2$ illustration is no longer an easy matter.

A POVM is called *informationally complete* if the statistics of the POVM uniquely determine the density matrix. This requires that the POVM has N^2 elements – a projective measurement will not do. A POVM is called *pure* if each operator E_i is of rank one, so there exists a pure state $|\phi_i\rangle$ such that E_i is proportional to $|\phi_i\rangle\langle\phi_i|$. An impure POVM can always be turned into a pure POVM by replacing each operator E_i by its spectral decomposition. Observe that a set of k pure states $|\phi_i\rangle$ defines a pure POVM if and only if the maximally mixed state $\rho_* = \mathbb{1}/N$ may be decomposed as $\rho_* = \sum_{i=1}^{k} p_i|\phi_i\rangle\langle\phi_i|$, where $\{p_i\}$ form a suitable set of positive coefficients. Indeed any ensemble of pure or mixed states representing ρ_* defines a POVM (Hughston et al., 1993). For any set of operators E_i defining a POVM we take quantum states $\rho_i = E_i/\mathrm{Tr}E_i$ and mix them with probabilities $p_i = \mathrm{Tr}E_i/N$ to obtain the maximally mixed state:

$$\sum_{i=1}^{k} p_i\rho_i = \sum_{i=1}^{k} \frac{1}{N} E_i = \frac{1}{N}\mathbb{1} = \rho_* \,. \tag{10.17}$$

Conversely, any such ensemble of density matrices defines a POVM (see Figure 10.2).

Arguably the most famous of all POVMs is the one based on coherent states. Assume that a classical phase space Ω has been used to construct a family of coherent states, $x \in \Omega \to |x\rangle \in \mathcal{H}$. The POVM is given by the resolution of the

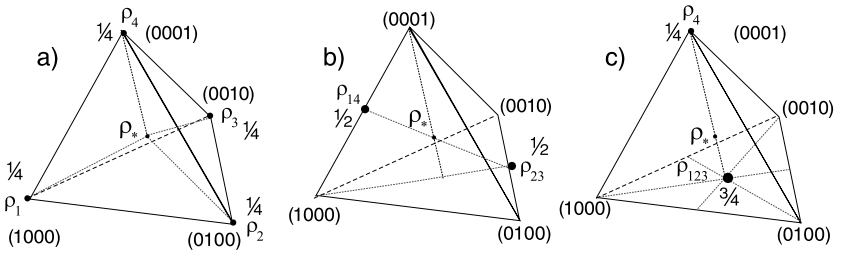


Figure 10.2. Any POVM is equivalent to an ensemble representing the maximally mixed state. For $N = 4$ ρ_* is situated in the centre of the tetrahedron of diagonal density matrices; (a) a pure POVM – in the picture a combination of four projectors with weights 1/4, (b) and (c) unpure POVMs.

identity

$$\int_\Omega |x\rangle\langle x| \, \mathrm{d}x = \mathbb{1} \, , \tag{10.18}$$

where $\mathrm{d}x$ is a natural measure on Ω. Examples of this construction were given in Chapter 6, and include the 'canonical' phase space where $x = (q, p)$. Any POVM can be regarded as an affine map from the set of quantum states to a set of classical probability distributions; in this case the resulting probability distributions are precisely those given by the Q-function. A discrete POVM can be obtained by introducing a partition of phase space into cells,

$$\Omega = \Omega_1 \cup \cdots \cup \Omega_k \, . \tag{10.19}$$

This partition splits the integral into k terms and defines k positive operators E_i that sum to unity. They are not projection operators since the coherent states overlap, and thus

$$E_i \equiv \int_{\Omega_i} \mathrm{d}x \, |x\rangle\langle x| \; \neq \; \int_{\Omega_i} \mathrm{d}x \int_{\Omega_i} \mathrm{d}y \, |x\rangle\langle x|y\rangle\langle y| = E_i^2 \, . \tag{10.20}$$

Nevertheless they do provide a notion of localization in phase space; if the state is ρ the particle will be registered in cell Ω_i with probability

$$p_i = \mathrm{Tr}(E_i \rho) = \int_{\Omega_i} \langle x|\rho|x\rangle \, \mathrm{d}x = \int_{\Omega_i} Q_\rho(x) \, \mathrm{d}x \, . \tag{10.21}$$

These ideas can be developed much further, so that one can indeed perform approximate but simultaneous measurements of position and momentum.[5]

[5] A pioneering result is due to Arthurs and Kelly, Jr (1965); for more, see the books by Holevo (1982), Busch, Lahti and Mittelstaedt (1991) and Leonhardt (1997).

The final twist of the story is that POVM measurements are not only more general than projective measurements, they are a special case of the latter too. Given any pure POVM with k elements and a state ρ in a Hilbert space of dimension N, we can find a state $\rho \otimes \rho_0$ in a Hilbert space $\mathcal{H} \otimes \mathcal{H}'$ such that the statistics of the original POVM measurement is exactly reproduced by a projective measurement of $\rho \otimes \rho_0$. This statement is a consequence of Naimark's theorem:

Theorem 10.1 (Naimark's) *Any POVM $\{E_i\}$ in the Hilbert space $\mathcal{H}$ can be dilated to an orthogonal resolution of identity $\{P_i\}$ in a larger Hilbert space in such a way that $E_i = \Pi P_i \Pi$, where Π projects down to $\mathcal{H}$.*

For a proof, see Problem 10.1. The next idea is to choose a pure state ρ_0 such that, in a basis in which Π is diagonal,

$$\rho \otimes \rho_0 = \left[\begin{array}{c|c} \rho & 0 \\ \hline 0 & 0 \end{array}\right] = \Pi\rho \otimes \rho_0 \Pi \, . \tag{10.22}$$

It follows that $\mathrm{Tr} P_i \rho \otimes \rho_0 = \mathrm{Tr} P_i \Pi \rho \otimes \rho_0 \Pi = \mathrm{Tr} E_i \rho$.

We are left somewhat at a loss to say which notion of measurement is the fundamental one. Let us just observe that classical statistics contains the notion of *randomized experiments*: equip an experimenter in an laboratory with a random number generator and surround the laboratory with a black box. The experimeter has a choice between different experiments, and will perform them with different probabilites p_i. It may not sound like a useful notion, but it is. We can view a POVM measurement as a randomized experiment in which the source of randomness is a quantum mechanical ancilla. Again the quantum case is more subtle than its classical counterpart; the set of all possible POVMs forms a convex set whose extreme points include the projective measurements, but there are other extreme points as well. The symmetric POVM shown in the upper panel in Figure 10.1, reinterpreted as a POVM for a qubit, may serve as an example.

10.2 Algebraic detour: matrix reshaping and reshuffling

Before proceeding with our analysis of quantum operations, we will discuss some simple algebraic transformations that one can perform on matrices. We also introduce a notation that we sometimes find convenient for work in the composite Hilbert space $\mathcal{H}_N \otimes \mathcal{H}_M$, or in the Hilbert–Schmidt (HS) space of linear operators $\mathcal{H}_{\mathrm{HS}}$.

Consider a rectangular matrix A_{ij}, $i = 1, \ldots, M$ and $j = 1, \ldots, N$. The matrix may be *reshaped* by putting its elements in lexicographical order (row after

row)[6] into a vector $\vec{a}_k$ of size MN,

$$\vec{a}_k = A_{ij} \quad \text{where} \quad k = (i-1)N + j, \quad i = 1, \ldots, M, \quad j = 1, \ldots N \; . \tag{10.23}$$

Conversely, any vector of length MN may be reshaped into a rectangular matrix. The simplest example of such a vectorial notation for matrices reads

$$A \; = \; \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad \longleftrightarrow \quad \vec{a} \; = \; (A_{11}, A_{12}, A_{21}, A_{22}) \; . \tag{10.24}$$

The scalar product in Hilbert–Schmidt space (matrices of size N) now looks like an ordinary scalar product between two vectors of size N^2,

$$\langle A|B\rangle \; \equiv \; \mathrm{Tr} A^\dagger B \; = \; \vec{a}^* \cdot \vec{b} \; = \; \langle a|b\rangle \; . \tag{10.25}$$

Thus the Hilbert–Schmidt norm of a matrix is equal to the norm of the associated vector, $||A||^2_{\mathrm{HS}} = \mathrm{Tr} A^\dagger A = |\vec{a}|^2$.

Sometimes we will label a component of $\vec{a}$ by a_{ij}. This vector of length MN may be linearly transformed into $a' = Ca$ by a matrix C of size $MN \times MN$. Its elements may be denoted by $C_{kk'}$ with $k, k' = 1, \ldots, MN$, but it is also convenient to use a four index notation, $C_{\substack{m\mu \\ n\nu}}$ where $m, n = 1, \ldots, N$ while $\mu, \nu = 1, \ldots, M$. In this notation the elements of the transposed matrix are $C^T_{\substack{m\mu \\ n\nu}} = C_{\substack{n\nu \\ m\mu}}$, since the upper pair of indices determines the row of the matrix, while the lower pair determines its column. The matrix C may represent an operator acting in a composite space $\mathcal{H} = \mathcal{H}_N \otimes \mathcal{H}_M$. The tensor product of any two bases in both factors provides a basis in $\mathcal{H}$, so that

$$C_{\substack{m\mu \\ n\nu}} = \langle e_m \otimes f_\mu | C | e_n \otimes f_\nu \rangle \; , \tag{10.26}$$

where Latin indices refer to the first subsystem, $\mathcal{H}_A = \mathcal{H}_N$, and Greek indices to the second, $\mathcal{H}_B = \mathcal{H}_M$. For instance the elements of the identity operator $\mathbb{1}_{NM} \equiv \mathbb{1}_N \otimes \mathbb{1}_M$ are $\mathbb{1}_{\substack{m\mu \\ n\nu}} = \delta_{mn}\delta_{\mu\nu}$. The trace of a matrix reads $\mathrm{Tr} C = C_{\substack{m\mu \\ m\mu}}$, where summation over repeating indices is understood. The operation of partial trace over the second subsystem produces the matrix $C^A \equiv \mathrm{Tr}_B C$ of size N, while tracing over the first subsystem leads to an $M \times M$ matrix $C^B \equiv \mathrm{Tr}_A C$,

$$C^A_{mn} = C_{\substack{m\mu \\ n\mu}}, \quad \text{and} \quad C^B_{\mu\nu} = C_{\substack{m\mu \\ m\nu}} \; . \tag{10.27}$$

If $C = A \otimes B$, then $C_{\substack{m\mu \\ n\nu}} = A_{mn} B_{\mu\nu}$. This form should not be confused with a product of two matrices $C = AB$, the elements of which are given by a double sum

[6] Some programs like MATLAB offer a built-in matrix command *reshape*, which performs such a task. Storing matrix elements column after column leads to the anti-lexicographical order.

over repeating indices, $C_{\substack{m\mu\\ n\nu}} = A_{\substack{m\mu\\ l\lambda}} B_{\substack{l\lambda\\ n\nu}}$. Observe that the standard product of three matrices may be rewritten by means of an object Φ,

$$ABC \;=\; \Phi B \quad \text{where} \quad \Phi \;=\; A \otimes C^T . \tag{10.28}$$

This is a telegraphic notation; since Φ is a linear map acting on B we might write $\Phi(B)$ on the right-hand side, and the left-hand side could be written as $A \cdot B \cdot C$ to emphasize that matrix multiplication is being used there. Equation (10.28) is a concise way of saying all this. It is unambiguous once we know the nature of the objects that appear in it.

Consider a unitary matrix U of size N^2. Unitarity of U implies that its N^2 columns $\vec{a}_k = U_{ik}\ k, i = 1, \ldots N^2$ reshaped into square $N \times N$ matrices A_k as in (10.24), form an orthonormal basis in $\mathcal{H}_{\rm HS}$, since $\langle A_k | A_j \rangle := {\rm Tr} A_k^\dagger A_j = \delta_{kj}$. Alternatively, in a double index notation with $k = (m-1)N + \mu$ and $j = (n-1)N + \nu$ this orthogonality relation reads $\langle A^{m\mu} | A^{n\nu} \rangle = \delta_{mn}\delta_{\mu\nu}$. Note that in general the matrices A_k are not unitary.

Let X denote an arbitrary matrix of size N^2. It may be represented as a double (quadruple) sum,

$$|X\rangle \;=\; \sum_{k=1}^{N^2} \sum_{j=1}^{N^2} C_{kj} |A_k\rangle \otimes |A_j\rangle = C_{\substack{m\mu\\ n\nu}} |A^{m\mu}\rangle \otimes |A^{n\nu}\rangle \,, \tag{10.29}$$

where $C_{kj} = {\rm Tr}((A_k \otimes A_j)^\dagger X)$. The matrix X may be considered as a vector in the composite Hilbert–Schmidt space $\mathcal{H}_{\rm HS} \otimes \mathcal{H}_{\rm HS}$, so applying its Schmidt decomposition (9.8) we arrive at

$$|X\rangle \;=\; \sum_{k=1}^{N^2} \sqrt{\lambda_k}\, |A'_k\rangle \otimes |A''_k\rangle \,, \tag{10.30}$$

where $\sqrt{\lambda_k}$ are the singular values of C, that is the square roots of the non-negative eigenvalues of $CC^\dagger$. The sum of their squares is determined by the norm of the operator, $\sum_{k=1}^{N^2} \lambda_k = {\rm Tr}(XX^\dagger) = ||X||^2_{\rm HS}$.

Since the Schmidt coefficients do *not* depend on the initial basis, let us choose the basis in $\mathcal{H}_{\rm HS}$ obtained from the identity matrix, $U = \mathbb{1}$ of size N^2, by reshaping its columns. Then each of the N^2 basis matrices of size N consists of only one non-zero element which equals unity, $A_k = A^{m\mu} = |m\rangle\langle\mu|$, where $k = N(m-1) + \mu$. Their tensor products form an orthonormal basis in $\mathcal{H}_{HS} \otimes \mathcal{H}_{HS}$ and allow us to represent an arbitrary matrix X in the form (10.29). In this case the matrix of the coefficients C has a particularly simple form, $C_{\substack{m\mu\\ n\nu}} = {\rm Tr}[(A^{m\mu} \otimes A^{n\nu})X] = X_{\substack{mn\\ \mu\nu}}$.

This particular reordering of a matrix deserves a name, so we shall write $X^R \equiv C(X)$ and call it *reshuffling*.[7] Using this notion our findings may be summarized in the following lemma:

Lemma 10.1 (Operator Schmidt decomposition) *The Schmidt coefficients of an operator X acting on a bipartite Hilbert space are equal to the squared singular values of the reshuffled matrix, X^R.*

More precisely, the Schmidt decomposition (10.30) of an operator X of size MN may be supplemented by a set of three equations

$$\left\{\begin{array}{rcl} \{\lambda_k\}_{k=1}^{N^2} & = & \left\{\mathrm{SV}(X^R)\right\}^2 \;:\; \text{eigenvalues of } (X^R)^\dagger X^R \\ |A'_k\rangle & : & \text{reshaped eigenvectors of } (X^R)^\dagger X^R \\ |A''_k\rangle & : & \text{reshaped eigenvectors of } X^R(X^R)^\dagger \end{array}\right. , \tag{10.31}$$

where SV denotes singular values and we have assumed that $N \leq M$. The initial basis is transformed by a local unitary transformation $W_a \otimes W_b$, where W_a and W_b are matrices of eigenvectors of matrices $(X^R)^\dagger X^R$ and $X^R(X^R)^\dagger$, respectively. If and only if the rank of $X^R(X^R)^\dagger$ equals one, the operator can be factorized into a product form, $X = X_1 \otimes X_2$, where $X_1 = \mathrm{Tr}_2 X$ and $X_2 = \mathrm{Tr}_1 X$.

To get a better feeling for the reshuffling transformation, observe that reshaping each row of an initially square matrix X of size MN according to Eq. (10.23) into a rectangular $M \times N$ submatrix, and placing it in lexicographical order block after block, one produces the reshuffled matrix X^R. Let us illustrate this procedure for the simplest case $N = M = 2$, in which any row of the matrix X is reshaped into a 2×2 matrix

$$C_{kj} \;=\; X^R_{kj} \;\equiv\; \left[\begin{array}{cc|cc} \mathbf{X_{11}} & \mathbf{X_{12}} & X_{21} & X_{22} \\ X_{13} & X_{14} & \mathbf{X_{23}} & \mathbf{X_{24}} \\ \hline \mathbf{X_{31}} & \mathbf{X_{32}} & X_{41} & X_{42} \\ X_{33} & X_{34} & \mathbf{X_{43}} & \mathbf{X_{44}} \end{array}\right]. \tag{10.32}$$

In the symmetric case with $M = N$, N^3 elements of X (typeset **boldface**) do not change position during reshuffling, while the remaining $N^4 - N^3$ elements do. Thus the space of complex matrices with the reshuffling symmetry $X = X^R$ is $2N^4 - 2(N^4 - N^3) = 2N^3$ dimensional.

The operation of reshuffling can be defined in an alternative way, say the reshaping of the matrix A from (10.23) could be performed column after column

[7] In general one may reshuffle square matrices, if their size K is not prime. The symbol X^R has a unique meaning if a concrete decomposition of the size $K = MN$ is specified. If $M \neq N$ the matrix X^R is a $N^2 \times M^2$ rectangular matrix. Since $(X^R)^R = X$ we see that one may also reshuffle rectangular matrices, provided both dimensions are squares of natural numbers. Similar reorderings of matrices were considered by Oxenrider and Hill (1985) and Yopp and Hill (2000).

into a vector $\vec{a}'$. In the four indices notation introduced above the two reshuffling operations take the form

$$X^{R}_{\substack{m\mu\\ n\nu}} \equiv X_{\substack{mn\\ \mu\nu}} \quad \text{and} \quad X^{R'}_{\substack{m\mu\\ n\nu}} \equiv X_{\substack{\nu\mu\\ nm}} . \tag{10.33}$$

Two reshuffled matrices are equivalent up to permutation of rows and columns and transposition, so the singular values of $X^{R'}$ and X^{R} are equal.

For comparison we provide analogous formulae showing the action of *partial transposition*: with respect to the first subsystem, $T_A \equiv T \otimes \mathbb{1}$ and with respect to the second, $T_B \equiv \mathbb{1} \otimes T$,

$$X^{T_A}_{\substack{m\mu\\ n\nu}} = X_{\substack{n\mu\\ m\nu}} \quad \text{and} \quad X^{T_B}_{\substack{m\mu\\ n\nu}} = X_{\substack{m\nu\\ n\mu}} . \tag{10.34}$$

Note that all these operations consist of exchanging a given pair of indices. However, while partial transposition (10.34) preserves Hermiticity, the reshuffling (10.33) does not. There is a related *swap* transformation among the two subsystems, $X^{S}_{\substack{m\mu\\ n\nu}} \equiv X_{\substack{\mu m\\ \nu n}}$, the action of which consists in relabelling certain rows (and columns) of the matrix, so its spectrum remains preserved. Note that for a tensor product $X = Y \otimes Z$ one has $X^S = Z \otimes Y$. Alternatively, define a SWAP operator

$$S \equiv \sum_{i,j=1}^{N} |i, j\rangle\langle j, i| \quad \text{so that} \quad S_{\substack{m\mu\\ n\nu}} = \delta_{m\nu}\delta_{n\mu} . \tag{10.35}$$

Observe that S is symmetric, Hermitian, and unitary and the identity $X^S = SXS$ holds. In full analogy to partial transposition we use also two operations of *partial swap*, $X^{S_1} = SX$ and $X^{S_2} = XS$.

All the transformations listed in Table 10.1 are involutions, since performed twice they are equal to identity. It is not difficult to find relations between them, for example $X^{S_1} = [(X^{R'})^{T_A}]^{R'} = [(X^R)^{T_B}]^R$. Since $X^{R'} = [(X^R)^S]^T = [(X^R)^T]^S$, while $X^{T_B} = (X^{T_A})^T$ and $X^{S_1} = (X^{S_2})^S$, thus the spectra and singular values of the reshuffled (partially transposed, partially swapped) matrices do not depend on the way, each operation has been performed, that is $\mathrm{eig}(X^R) = \mathrm{eig}(X^{R'})$ and $\mathrm{SV}(X^R) = \mathrm{SV}(X^{R'})$, while $\mathrm{eig}(X^{S_1}) = \mathrm{eig}(X^{S_2})$ and $\mathrm{SV}(X^{S_1}) = \mathrm{SV}(X^{S_2})$.

10.3 Positive and completely positive maps

Thus equipped, we return to physics. We will use the notation of Section 10.2 freely, so an alternative title for this section is 'Complete positivity as an exercise in index juggling'. Let $\rho \in \mathcal{M}^{(N)}$ be a density matrix acting on an N-dimensional Hilbert space. What conditions need to be fulfilled by a map $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$, if it is

Table 10.1. *Reorderings of a matrix X representing an operator which acts on a composite Hilbert space. The arrows denote the indices exchanged.*

Transformation	Definition	Symbol	Preserves Hermiticity	Preserves spectrum
transposition	$X^{T}_{\substack{m\mu\\ n\nu}} = X_{\substack{n\nu\\ m\mu}}$	$\updownarrow\updownarrow$	yes	yes
partial	$X^{T_A}_{\substack{m\mu\\ n\nu}} = X_{\substack{n\mu\\ m\nu}}$	$\updownarrow$.	yes	no
transpositions	$X^{T_B}_{\substack{m\mu\\ n\nu}} = X_{\substack{m\nu\\ n\mu}}$	. $\updownarrow$	yes	no
reshuffling	$X^{R}_{\substack{m\mu\\ n\nu}} = X_{\substack{mn\\ \mu\nu}}$	$\swarrow\!\!\nearrow$	no	no
reshuffling $'$	$X^{R'}_{\substack{m\mu\\ n\nu}} = X_{\substack{\nu\mu\\ nm}}$	$\nwarrow\!\!\searrow$	no	no
swap	$X^{S}_{\substack{m\mu\\ n\nu}} = X_{\substack{\mu m\\ \nu n}}$	$\leftrightarrow$ $\leftrightarrow$	yes	yes
partial	$X^{S_1}_{\substack{m\mu\\ n\nu}} = X_{\substack{\mu m\\ n\nu}}$	$\leftrightarrow$.	no	no
swaps	$X^{S_2}_{\substack{m\mu\\ n\nu}} = X_{\substack{m\mu\\ \nu n}}$	. $\leftrightarrow$	no	no

to represent a physical operation? One class of maps that we will admit are those given in Eq. (10.7). We will now argue that nothing else is needed.

Our first requirement is that the map should be a linear one. It is always hard to argue for linearity, but at this level linearity is also hard to avoid, since we do not want the image of ρ to depend on the way in which ρ is presented as a mixture of pure states – the entire probabilistic structure of the theory is at stake here.[8] We are thus led to postulate the existence of a *linear superoperator* Φ,

$$\rho' \;=\; \Phi\rho \quad \text{or} \quad \rho'_{m\mu} = \Phi_{\substack{m\mu\\ n\nu}}\,\rho_{n\nu}\;. \tag{10.36}$$

Summation over repeated indices is understood throughout this section. Inhomogeneous maps $\rho' = \Phi\rho + \sigma$ are automatically included, since

$$\Phi_{\substack{m\mu\\ n\nu}}\,\rho_{n\nu} + \sigma_{m\mu} = \big(\Phi_{\substack{m\mu\\ n\nu}} + \sigma_{m\mu}\delta_{n\nu}\big)\rho_{n\nu} = \Phi'_{\substack{m\mu\\ n\nu}}\,\rho_{n\nu} \tag{10.37}$$

due to $\mathrm{Tr}\rho = 1$. We deal with affine maps of density matrices.

The map should take density matrices to density matrices. This means that whenever ρ is (i) Hermitian, (ii) of unit trace, and (iii) positive, its image ρ' must share

[8] Non-linear quantum mechanics is actually a lively field of research; see Mielnik (2001) and references therein.

these properties.[9] These three conditions impose three constraints on the matrix Φ:

$$\text{(i)} \qquad \rho' = (\rho')^\dagger \quad \Leftrightarrow \quad \Phi_{\substack{m\mu \\ n\nu}} = \Phi^*_{\substack{\mu m \\ \nu n}} \quad \text{so} \quad \Phi^* = \Phi^S, \tag{10.38}$$

$$\text{(ii)} \qquad \mathrm{Tr}\rho' = 1 \quad \Leftrightarrow \quad \Phi_{\substack{mm \\ n\nu}} = \delta_{n\nu} \tag{10.39}$$

$$\text{(iii)} \qquad \rho' \geq 0 \quad \Leftrightarrow \quad \Phi_{\substack{m\mu \\ n\nu}} \rho_{n\nu} \geq 0 \quad \text{when } \rho > 0\,. \tag{10.40}$$

As they stand, these conditions are not very illuminating.

The meaning of our three conditions becomes much clearer if we reshuffle Φ according to (10.33) and define the *dynamical matrix* [10]

$$D_\Phi \;\equiv\; \Phi^R \quad \text{so that} \quad D_{\substack{mn \\ \mu\nu}} = \Phi_{\substack{m\mu \\ n\nu}}\,. \tag{10.41}$$

The dynamical matrix D_Φ uniquely determines the map Φ. It obeys

$$D_{a\Phi+b\Psi} = aD_\Phi + bD_\Psi\,, \tag{10.42}$$

that is to say it is a linear function of the map.

In terms of the dynamical matrix our three conditions become

$$\text{(i)} \qquad \rho' = (\rho')^\dagger \quad \Leftrightarrow \quad D_{\substack{mn \\ \mu\nu}} = D^\dagger_{\substack{mn \\ \mu\nu}} \quad \text{so} \quad D_\Phi = D^\dagger_\Phi\,, \tag{10.43}$$

$$\text{(ii)} \qquad \mathrm{Tr}\rho' = 1 \quad \Leftrightarrow \quad D_{\substack{mn \\ m\nu}} = \delta_{n\nu} \tag{10.44}$$

$$\text{(iii)} \qquad \rho' \geq 0 \quad \Leftrightarrow \quad D_{\substack{mn \\ \mu\nu}} \rho_{n\nu} \geq 0 \quad \text{when } \rho > 0\,. \tag{10.45}$$

Condition (i) holds if and only if D_Φ is Hermitian. Condition (ii) also takes a familiar form – the partial trace with respect to the first subsystem is the unit operator for the second subsystem:

$$D_{\substack{mn \\ m\nu}} = \delta_{n\nu} \qquad \Leftrightarrow \qquad \mathrm{Tr}_A D = \mathbb{1} \tag{10.46}$$

Only condition (iii), for positivity, requires further unravelling.

The map is said to be a *positive map* if it takes positive matrices to positive matrices. To see if a map is positive, we must test if condition (iii) holds. Let us first assume that the original density matrix is a pure state, so that $\rho_{n\nu} = z_n z^*_\nu$. Then its image will be positive if and only if, for all vectors x_m,

$$x_m \rho'_{m\mu} x^*_\mu \;=\; x_m z_n \, D_{\substack{mn \\ \mu\nu}} \, x^*_\mu z^*_\nu \;\geq\; 0\,. \tag{10.47}$$

This means that the dynamical matrix itself must be positive when it acts on product states in $\mathcal{H}_{N^2}$. This property is called *block-positivity*. We have arrived at the following (Jamiołkowski, 1972):

[9] Any map $\Phi\rho$ can be normalized according to $\rho \to \Phi\rho/\mathrm{Tr}\big[\Phi\rho\big]$. It is sometimes convenient to work with unnormalized maps, but the a-posteriori normalization procedure may spoil linearity.

[10] This concept was introduced by Sudarshan, Mathews and Rau (1961), and even earlier (in the mathematics literature) by Schatten (1950).

Theorem 10.2 (Jamiołkowski's theorem) *A linear map Φ is positive if and only if the corresponding dynamical matrix D_Φ is block-positive.*

The converse holds since condition (10.47) is strong enough to ensure that condition (10.45) holds for all mixed states ρ as well.

Interestingly, the condition for positivity has not only one but two drawbacks. First, it is difficult to work with. Second, it is not enough from a physical point of view. Any quantum state ρ may be extended by an ancilla to a state $\rho \otimes \sigma$ of a larger composite system. The mere possibility that an ancilla may be added requires us to check that the map $\Phi \otimes \mathbb{1}$ is positive as well. Since the map leaves the ancilla unaffected this may seem like a foregone conclusion. Classically it is so, but quantum mechanically it is not. Let us state this condition precisely: a map Φ is said to be *completely positive* if and only if for an arbitrary K-dimensional extension

$$\mathcal{H}_N \to \mathcal{H}_N \otimes \mathcal{H}_K \quad \text{the map} \quad \Phi \otimes \mathbb{1}_K \quad \text{is positive.} \tag{10.48}$$

This is our final condition on a physical map.[11]

In order to see what the condition of complete positivity says about the dynamical matrix we will backtrack a little, and introduce a canonical form for the latter. Since the dynamical matrix is an Hermitian matrix acting on $\mathcal{H}_{N^2}$, it admits a spectral decomposition

$$D_\Phi = \sum_{i=1}^{r} d_i \, |\chi_i\rangle \, \langle \chi_i| \quad \text{so that} \quad D_{\substack{mn \\ \mu\nu}} = \sum_{i=1}^{N^2} d_i \; \chi^i_{mn} \bar{\chi}^i_{\mu\nu} \; . \tag{10.49}$$

The eigenvalues d_i are real, and the notation emphasizes that the matrices χ^i_{mn} are (reshaped) vectors in $\mathcal{H}_{N^2}$.

Now we are in a position to investigate the conditions that ensure that the map $\Phi \otimes \mathbb{1}$ preserves positivity when it acts on matrices in $\mathcal{HS} = \mathcal{H} \otimes \mathcal{H}'$. We pick an arbitrary vector $z_{nn'}$ in $\mathcal{HS}$, and act with our map on the corresponding pure state:

$$\rho'_{mm'\mu\mu'} = \Phi_{\substack{m\mu \\ n\nu}} \, \delta_{\substack{m'\mu' \\ n'\nu'}} \, z_{nn'} \, z^*_{\nu\nu'} = D_{\substack{mn \\ \mu\nu}} \, z_{nm'} \, z^*_{\nu\mu'} = \sum_i d_i \, \chi^i_{mn} z_{nm'} (\chi^i_{\mu\nu} z_{\nu\mu'})^* \; . \tag{10.50}$$

Then we pick another arbitrary vector $x_{mm'}$, and test whether ρ' is a positive operator:

$$x_{mm'} \, \rho'_{mm'\mu\mu'} \, x^*_{\mu\mu'} = \sum_i d_i \, |\chi^i_{mn} x_{mn'} z_{nm'}|^2 \; \geq \; 0 \; . \tag{10.51}$$

This must hold for arbitrary $x_{mm'}$ and $z_{mm'}$, and therefore all the eigenvalues d_i must be positive (or zero). In this way we have arrived at Choi's theorem:

[11] The mathematical importance of complete positivity was first noted by Stinespring (1955); its importance in quantum theory was emphasized by Kraus (1971), Accardi (1976) and Lindblad (1976).

Theorem 10.3 (Choi's) *A linear map Φ is completely positive if and only if the corresponding dynamical matrix D_Φ is positive.*

There is some fine print. If condition (10.48) holds for a fixed K only, the map is said to be *K-positive*. The map will be completely positive if and only if it is N-positive – which is the condition that we actually investigated.[12]

It is striking that we obtain such a simple result when we strengthen the condition on the map from positivity to complete positivity. The set of completely positive maps is isomorphic to the set of positive matrices D_Φ of size N^2. When the map is also trace preserving we add the extra condition (10.46), which implies that $\mathrm{Tr} D_\Phi = N$. We can therefore think of the set of trace preserving completely positive maps as a subset of the set of density matrices in $\mathcal{H}_{N^2}$, albeit with an unusual normalization. This analogy will be further pursued in Chapter 11, where (for reasons that will become clear later) we will also occupy ourselves with understanding the way in which the set of completely positive maps forms a proper subset of the set of all positive maps.

The dynamical matrix is positive if and only if it can be written in the form

$$D_\Phi = \sum_i |A_i\rangle\langle A_i| \quad \text{so that} \quad D_{\substack{mn \\ \mu\nu}} = \sum_i A^i_{mn} \bar{A}^i_{\mu\nu} , \tag{10.52}$$

where the vectors A_i are arbitrary to an extent given by Schrödinger's mixture theorem (see Section 8.4). In this way we obtain an alternative characterization of completely positive maps. They are the maps that can be written in the *operator sum representation*:

Theorem 10.4 (Operator sum representation) *A linear map Φ is completely positive if and only if it is of the form*

$$\rho \to \rho' = \sum_i A_i \rho A_i^\dagger . \tag{10.53}$$

This is also known as the *Kraus* or *Stinespring form*, since its existence follows from the *Stinespring dilation theorem*.[13] The operators A_i are known as *Kraus operators*. The map will be *trace preserving* if and only if condition (10.44) holds, which translates itself to

$$\sum_i A_i^\dagger A_i = \mathbb{1}_N . \tag{10.54}$$

[12] 'Choi's theorem' is theorem 2 in Choi (1975a). Theorem 1 (the existence of the operator sum representation) and theorem 5 follow below. The paper contains no theorems 3 or 4.

[13] In physics the operator sum representation was introduced by Kraus (1971), based on an earlier (somewhat more abstract) theorem by Stinespring (1955), and independently by Sudarshan et al. (1961). See also Kraus (1983) and Evans (1984).

We have recovered the class of operations that were introduced in Eq. (10.7), but our new point of view has led us to the conclusion that this is the most general class that we need to consider. Trace preserving completely positive maps go under various names: *deterministic* or *proper quantum operations*, *quantum channels*, or *stochastic maps*. They are the sought for analogue of classical stochastic maps.

The convex set of proper quantum operations is denoted $\mathcal{CP}_N$. To find its dimension we note that the dynamical matrices belong to the positive cone in the space of Hermitian matrices of size N^2, which has dimension N^4; the dynamical matrix corresponds to a trace preserving map if only if its partial trace (10.46) is the unit operator, so it is subject to N^2 conditions. Hence the dimension of $\mathcal{CP}_N$ equals $N^4 - N^2$.

Since the operator sum representation does not determine the Kraus operators uniquely we would like to bring it to a canonical form. The problem is quite similar to that of introducing a canonical form for a density matrix – in both cases, the solution is to present an Hermitian matrix as a mixture of its eigenstates. Such a decomposition of the dynamical matrix was given in Eq. (10.49). A set of canonical Kraus operators can be obtained by setting $A_i = \sqrt{d_i}\,\chi_i$. The following results:

Theorem 10.5 (Canonical Kraus form) *A completely positive map* $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$ *can be represented as*

$$\rho \;\to\; \rho' \;=\; \sum_{i=1}^{r\le N^2} d_i\;\chi_i \rho\;\chi_i^\dagger \;=\; \sum_{i=1}^{r} A_i \rho A_i^\dagger\;, \tag{10.55}$$

where

$$\mathrm{Tr} A_i^\dagger A_j \;=\; \sqrt{d_i d_j}\,\langle \chi_i | \chi_j \rangle \;=\; d_i\,\delta_{ij}\;. \tag{10.56}$$

If the map is also trace preserving then

$$\sum_i A_i^\dagger A_i \;=\; \mathbb{1}_N \quad\Rightarrow\quad \sum_{i=1}^{r} d_i \;=\; N\;. \tag{10.57}$$

If D_Φ is non-degenerate the canonical form is unique up to phase choices for the Kraus operators. The *Kraus rank* of the map is the number of Kraus operators that appear in the canonical form, and equals the rank r of the dynamical matrix.

The operator sum representation can be written

$$\Phi \;=\; \sum_{i=1}^{N^2} A_i \otimes \bar{A}_i \;=\; \sum_{i=1}^{N^2} d_i\;\chi^i \otimes \bar{\chi}^i\,. \tag{10.58}$$

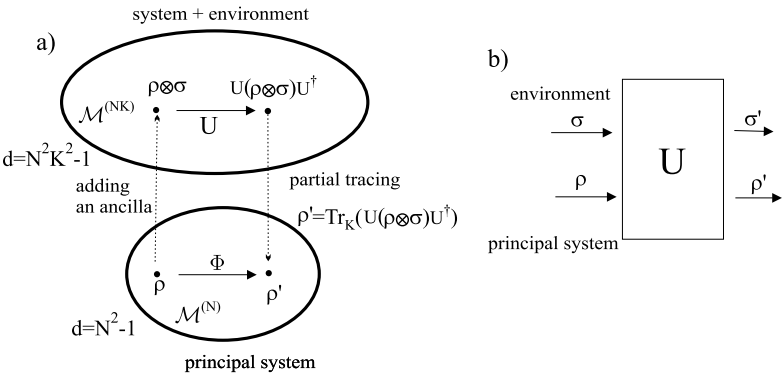


Figure 10.3. Quantum operations represented by (a) unitary operator U of size NK in an enlarged system including the environment, (b) black box picture.

The CP map can be described in the notation of Eq. (10.28), and the operator sum representation may be considered as a Schmidt decomposition (10.30) of Φ, with Schmidt coefficients $\lambda_i = d_i^2$.

10.4 Environmental representations

We began this chapter by adding an ancilla (in the state σ) to the system, evolving the composite system unitarily, and then removing the ancilla through a partial trace at the end. This led us to the *environmental representation* of the map Φ, that is to

$$\rho \;\to\; \rho' \;=\; \mathrm{Tr}_{\mathrm{env}}\Big[U\big(\rho\otimes\sigma\big)U^\dagger\Big]\,; \tag{10.59}$$

see Figure 10.3. We showed that the resulting map can be written in the Kraus form, and now we know that this means that it is a completely positive map. What was missing from the argument was a proof that any CP map admits an environmental representation, and indeed one in which the ancilla starts out in a pure state $\sigma = |\nu\rangle\langle\nu|$. This we will now supply.[14]

We are given a set of K Kraus operators A_μ (equipped with Greek indices because we use such letters to denote states of the ancilla). Due to the completeness relation (10.54) we can regard them as defining N orthogonal columns in a matrix U with NK rows,

$$A^\mu_{mn} = \langle m,\mu|U|n,\nu\rangle = U_{\substack{m\mu\\ n\nu}} \qquad\Leftrightarrow\qquad A_\mu = \langle\mu|U|\nu\rangle\ . \tag{10.60}$$

[14] Originally this was noted by Arveson (1969) and Lindblad (1975).

Here ν is fixed, but we can always find an additional set of columns that turns U into a unitary matrix of size NK. By construction then, for an ancilla of dimension K,

$$\rho' = \mathrm{Tr}_{\mathrm{env}}\Big[U\big(\rho \otimes |\nu\rangle\langle\nu|\big)U^\dagger\Big] = \sum_{\mu=1}^{K}\langle\mu|U|\nu\rangle\rho\langle\nu|U^\dagger|\mu\rangle = \sum_{\mu=1}^{K} A_\mu \rho A_\mu^\dagger \,. \tag{10.61}$$

This is the Kraus form, since the operators A_μ satisfy the completeness relation

$$\sum_{\mu=1}^{K} A_\mu^\dagger A_\mu = \sum_{\mu=1}^{K}\langle\nu|U^\dagger|\mu\rangle\langle\mu|U|\nu\rangle = \langle\nu|U^\dagger U|\nu\rangle = \mathbb{1}_N \,. \tag{10.62}$$

Note that the 'extra' columns that we added to the matrix U do not influence the quantum operation in any way.

Although we may choose an ancilla that starts out in a pure state, we do not have to do it. If the initial state of the environment in the representation (10.59) is a mixture $\sigma = \sum_{\nu=1}^{r} q_\nu |\nu\rangle\langle\nu|$, we obtain an operator sum representation with rK terms,

$$\rho \to \rho' = \mathrm{Tr}_{\mathrm{env}}\Big[U\Big(\rho \otimes \sum_{\nu=1}^{r} q_\nu |\nu\rangle\langle\nu|\Big)U^\dagger\Big] = \sum_{l=1}^{rK} A_l \rho A_l^\dagger \tag{10.63}$$

where $A_l = \sqrt{q_\nu}\langle\mu|U|\nu\rangle$ and $l = \mu + \nu(K-1)$.

If the initial state of the ancilla is pure the dimension of its Hilbert space needs not exceed N^2, the maximal number of Kraus operators required. More precisely its dimension may be set equal to the Kraus rank of the map. If the environment is initially in a mixed state, its weights q_ν are needed to specify the operation. Counting the number of parameters one could thus speculate that the action of any quantum operation may be simulated by a coupling with a mixed state of an environment of size N. However, this is not the case: already for $N = 2$ there exist operations which have to be simulated with a three-dimensional environment (Terhal, Chuang, DiVincenzo, Grassl and Smolin, 1999; Zalka and Rieffel, 2002). The general question of the minimal size of $\mathcal{H}_{\mathrm{env}}$ remains open.

It is illuminating to discuss the special case in which the initial state of the N-dimensional environment is maximally mixed, $\sigma = \mathbb{1}_N/N$. The unitary matrix U of size N^2, defining the map, may be treated as a vector in the composite Hilbert–Schmidt space $\mathcal{H}_{\mathrm{HS}}^A \otimes \mathcal{H}_{\mathrm{HS}}^B$ and represented in its Schmidt form $U = \sum_{i=1}^{N^2}\sqrt{\lambda_i}|\tilde{A}_i\rangle \otimes |\tilde{A}_i'\rangle$, where λ_i are eigenvalues of $(U^R)^\dagger U^R$. Since the operators $\tilde{A}_i'$ (reshaped eigenvectors of $(U^R)^\dagger U^R$) form an orthonormal basis in $\mathcal{H}_{\mathrm{HS}}$,

the procedure of partial tracing leads to a Kraus form with N^2 terms:

$$\begin{aligned}\rho' = \Phi_U \rho &= \mathrm{Tr}_{\mathrm{env}}\Big[U(\rho \otimes \frac{1}{N}\mathbb{1}_N)\,U^\dagger\Big] \\ &= \mathrm{Tr}_{\mathrm{env}}\Big[\sum_{i=1}^{N^2}\sum_{j=1}^{N^2}\sqrt{\lambda_i\lambda_j}\big(\tilde{A}_i\rho\tilde{A}_j^\dagger\big)\otimes\big(\frac{1}{N}\tilde{A}_i'\tilde{A}_j{}'^\dagger\big)\Big] \\ &= \frac{1}{N}\sum_{i=1}^{N^2}\lambda_i\tilde{A}_i\rho\tilde{A}_i^\dagger\,.\end{aligned} \tag{10.64}$$

The standard Kraus form is obtained by rescaling the operators, $A_i = \sqrt{\lambda_i/N}\tilde{A}_i$. Operations for which there exist a unitary matrix U providing a representation in the above form, we shall call *unistochastic channels*.[15] Note that the matrix U is determined up to a local unitary matrix V of size N, in the sense that U and $U' = U(\mathbb{1}_N \otimes V)$ generate the same unistochastic map, $\Phi_U = \Phi_{U'}$.

One may consider analogous maps with an arbitrary size of the environment. Their physical motivation is simple: not knowing anything about the environment (apart from its dimensionality), one assumes that it is initially in the maximally mixed state. In particular we define generalized,*K-unistochastic maps*[16], determined by a unitary matrix $U(N^{K+1})$, in which the environment of size N^K is initially in the state $\mathbb{1}_{N^K}/N^K$.

A debatable point remains, namely that the combined system started out in the product state. This may look like a very special intitial condition. However, in general it not so easy to present a well-defined procedure for how to assign a state of the composite system, given only a state of the system of interest to start with. Suppose $\rho \to \omega$ is such an assignment, where ω acts on the composite Hilbert space. Ideally one wants the assignment map to obey three conditions: (i) it preserves mixtures, (ii) $\mathrm{Tr}_{\mathrm{env}}\omega = \rho$, and (iii) ω is positive for all positive ρ. But it is known[17] that these conditions are so stringent that the only solution is of the form $\omega = \rho \otimes \sigma$.

10.5 Some spectral properties

A quantum operation Φ is uniquely characterized by its dynamical matrix, but the spectra of these matrices are quite different. The dynamical matrix is Hermitian,

[15] In analogy to classical transformations given by unistochastic matrices, $\vec{p}' = T\vec{p}$, where $T_{ij} = |U_{ij}|^2$.

[16] Such operations were analysed in the context of quantum information processing (Knill and Laflamme, 1998; Poulin, Blume–Kohout, Laflamme and Olivier, 2004), and, under the name 'noisy maps', when studying reversible transformations from pure to mixed states (Horodecki, Horodecki and Oppenheim, 2003a). By definition, 1–unistochastic maps are unistochastic.

[17] See the exchange between Pechukas (1994) and Alicki (1995).

Table 10.2. *Quantum operations $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$: properties of the superoperator Φ and the dynamical matrix $D_\Phi = \Phi^R$. For the coarse graining map consult Eq. (12.77) while for the entropy S see Section 12.6.*

Matrices	Superoperator $\Phi = (D_\Phi)^R$	Dynamical matrix D_Φ
Hermiticity	No	Yes
Trace	spectrum is symmetric $\Rightarrow$ $\mathrm{Tr}\,\Phi \in \mathbb{R}$	$\mathrm{Tr}\,D_\Phi = N$
Eigenvectors (right)	invariant states or transient corrections	Kraus operators
Eigenvalues	$\lvert z_i\rvert \le 1$, $-\ln\lvert z_i\rvert$ = decay rates	weights of Kraus operators, $d_i \ge 0$
Unitary evolution $D_U = (U \otimes U^*)^R$	$\lvert\lvert\Phi_U\rvert\rvert^2_{\mathrm{HS}} = N^2$	$S(D_U) = 0$
Coarse graining	$\lvert\lvert\Phi_{\mathrm{CG}}\rvert\rvert^2_{\mathrm{HS}} = N$	$S(D_{\mathrm{CG}}) = \ln N$
Complete depolarization	$\lvert\lvert\Phi_*\rvert\rvert^2_{\mathrm{HS}} = 1$	$S(D_*) = 2\ln N$

but Φ is not and its eigenvalues z_i are complex. Let us order them according to their moduli, $|z_1| \ge |z_2| \ge \cdots \ge |z_{N^2}| \ge 0$. The operation Φ sends the convex compact set $\mathcal{M}^{(N)}$ into itself. Therefore, due to the fixed-point theorem, the transformation has a fixed point – an invariant state σ_1 such that $\Phi\sigma_1 = \sigma_1$. Thus $z_1 = 1$ and all eigenvalues fulfil $|z_i| \le 1$, since otherwise the assumption that Φ is positive would be violated. These spectral properties are similar to those enjoyed by classical stochastic matrices (Section 2.1).

The trace preserving condition, applied to the equation $\Phi\sigma_i = z_i\sigma_i$, implies that if $z_i \neq 1$ then $\mathrm{Tr}\sigma_i = 0$. If $R = |z_2| < 1$, then the matrix Φ is *primitive* (Marshall and Olkin, 1979); under repeated applications of the map all states converge to the invariant state σ_1. If Φ is diagonalizable (its Jordan decomposition has no non-trivial blocks, so that the number of right eigenvectors σ_i is equal to the size of the matrix), then any initial state ρ_0 may be expanded in the eigenbasis of Φ,

$$\rho_0 = \sum_{i=1}^{N^2} c_i\sigma_i \quad \text{while} \quad \rho_t = \Phi^t\rho_0 = \sum_{i=1}^{N^2} c_i z_i^t\sigma_i \ . \tag{10.65}$$

Therefore ρ_0 converges exponentially fast to the invariant state σ_1 with a decay rate not smaller than $-\ln R$ and the right eigenstates σ_i for $i \ge 2$ play the role of the transient traceless corrections to ρ_0. The superoperator Φ sends Hermitian operators to Hermitian operators, $\rho_1^\dagger = \rho_1 = \Phi\rho_0 = \Phi\rho_0^\dagger$, so

$$\text{if} \quad \Phi\chi = z\chi \quad \text{then} \quad \Phi\chi^\dagger = z^*\chi^\dagger \ , \tag{10.66}$$

and the spectrum of Φ (contained in the unit circle) is symmetric with respect to the real axis. Thus the trace of Φ is real, as follows also from the hermiticity of $D_\Phi = \Phi^R$. Using (10.58) we obtain[18]

$$\mathrm{Tr}\,\Phi \;=\; \sum_{i=1}^{r} (\mathrm{Tr}A_i)(\mathrm{Tr}\bar{A}_i) \;=\; \sum_{i=1}^{r} |\mathrm{Tr}A_i|^2 \,, \tag{10.67}$$

equal to N^2 for the identity map and to unity for a map given by the rescaled identity matrix, $D_* = \mathbb{1}_{N^2}/N$. The latter map describes the *completely depolarizing channel* Φ_*, which transforms any initial state ρ into the maximally mixed state, $\Phi_*\rho = \rho_* = \mathbb{1}_N/N$.

Given a set of Kraus operators A_i for a quantum operation Φ, and any two unitary matrices V and W of size N, the operators $A'_i = V A_i W$ will satisfy the relation (10.54) and define the operation

$$\rho \to \rho' = \Phi_{VW}\rho = \sum_{i=1}^{k} A'_i\, \rho A_i'^{\dagger} = V\Big(\sum_{i=1}^{k} A_i (W\rho W^\dagger) A_i^\dagger\Big) V^\dagger \,. \tag{10.68}$$

The operations Φ and Φ_{VW} are in general different, but *unitarily similar*, in the sense that their dynamical matrices have the same spectra. The equality $||\Phi||_{\mathrm{HS}} = ||\Phi_{VW}||_{\mathrm{HS}}$ follows from the transformation law

$$\Phi_{VW} = (V \otimes V^*)\,\Phi\,(W \otimes W^*) \,, \tag{10.69}$$

which is a consequence of (10.58). This implies that the dynamical matrix transforms by a local unitary, $D_{WV} = (U \otimes V^T)D(U \otimes V^T)^\dagger$.

10.6 Unital and bistochastic maps

A trace preserving completely positive map is called a *bistochastic map* if it is also *unital*, that is to say if it leaves the maximally mixed state invariant.[19] Evidently this is the quantum analogue of a bistochastic matrix – a stochastic matrix that leaves the uniform probability vector invariant. The composition of two bistochastic maps is bistochastic. In the operator sum representation the condition that the map be bistochastic reads

$$\rho \;\to\; \rho' = \sum_i A_i \rho A_i^\dagger \,, \qquad \sum_i A^\dagger A_i = \mathbb{1} \,, \qquad \sum_i A_i A_i^\dagger = \mathbb{1} \,. \tag{10.70}$$

For the dynamical matrix this means that $\mathrm{Tr}_A D = \mathrm{Tr}_B D = \mathbb{1}$.

[18] This trace determines the *mean operation fidelity* $\langle F\big(\rho_\psi, \Phi\rho_\psi\big)\rangle_\psi$ averaged over random pure states ρ_ψ (Nielsen, 2002; Zanardi and Lidar, 2004).

[19] See the book by Alberti and Uhlmann (1982).

The channel is bistochastic if all the Kraus operators obey $[A_i, A_i^\dagger] = 0$. Indeed the simplest example is a unitary transformation. A more general class of bistochastic channels is given by convex combinations of unitary operations, also called *random external fields* (REF),

$$\rho' = \Phi_{\rm REF}\rho = \sum_{i=1}^{k} p_i\, V_i \rho V_i^\dagger, \quad \text{with} \quad p_i > 0 \quad \text{and} \quad \sum_{i=1}^{k} p_i = 1, \qquad (10.71)$$

where each operator V_i is unitary. The Kraus form (10.53) can be reproduced by setting $A_i = \sqrt{p_i}\, V_i$.

The set of all bistochastic CP maps, denoted $\mathcal{B}_N$, is a convex set in itself. The set of all bistochastic matrices is, as we learned in Section 2.1, a convex polytope with permutation matrices as its extreme points. Reasoning by analogy one would guess that the extreme points of $\mathcal{B}_N$ are unitary transformations, in which case $\mathcal{B}_N$ would coincide with the set of random external fields. This happens to be true for qubits, as we will see in detail in Section 10.7, but it fails for all $N > 2$.

There is a theorem that characterizes the extreme points the set of stochastic maps (Choi, 1975a):

Lemma 10.2 (Choi's) *A stochastic map Φ is extreme in $\mathcal{CP}_N$ if and only if it admits a canonical Kraus form for which the matrices $A_i^\dagger A_j$ are linearly independent.*

We prove that the condition is sufficient: assume that $\Phi = p\Psi_1 + (1-p)\Psi_2$. If so it will be true that

$$\begin{aligned} \Phi\rho &= \sum_i A_i \rho A_i^\dagger = p\Psi_1\rho + (1-p)\Psi_2\rho \\ &= p\sum_i B_i \rho B_i^\dagger + (1-p)\sum_i C_i \rho\, C_i^\dagger\,. \end{aligned} \qquad (10.72)$$

The right-hand side is not in the canonical form, but we assume that the left-hand side is. Therefore there is a unique way of writing

$$B_i \;=\; \sum_j m_{ij}\, A_j\,. \qquad (10.73)$$

In fact this is Schrödinger's mixture theorem in slight disguise. Next we observe that

$$\sum_i B_i^\dagger B_i = \mathbb{1} = \sum_i A_i^\dagger A_i \quad \Rightarrow \quad \sum_{i,j} \big((m^\dagger m)_{ij} - \delta_{ij}\big) A_i^\dagger A_j = 0\,. \qquad (10.74)$$

Because of the linear independence condition this means that $(m^\dagger m)_{ij} = \delta_{ij}$. This is what we need in order to show that that $\Phi = \Psi_1$, and it follows that the map Φ is indeed pure.

The matrices $A_i A_j^\dagger$ are of size N, so there can be at most N^2 linearly independent ones. This means that there can be at most N Kraus operators occurring in the canonical form of an extreme stochastic map.

It remains to find an example of an extreme bistochastic map which is not unitary. Using the $N = (2j+1)$-dimensional representation of $SU(2)$, we take the three Hermitian angular momentum operators J_i and define the map

$$\rho \to \rho' = \frac{1}{j(j+1)} \sum_{i=1}^{3} J_i \rho J_i \ , \qquad J_1^2 + J_2^2 + J_3^2 = j(j+1) \ . \tag{10.75}$$

Choi's condition for an extreme map is that the set of matrices $J_i J_j^\dagger = J_i J_j$ must be linearly independent. By angular momentum addition our set spans a $9 = 5 + 3 + 1$ dimensional representation of $SO(3)$, and all the matrices will be linearly independent provided that they are non-zero. The example fails for $N = 2$ only – in that case $J_i J_j + J_j J_i = 0$, the J_i are both Hermitian and unitary, and we do not get an extreme point.[20]

For any quantum channel Φ one defines its *dual channel* $\tilde{\Phi}$, such that the Hilbert–Schmidt scalar product satisfies $\langle \Phi\sigma | \rho \rangle = \langle \sigma | \tilde{\Phi}\rho \rangle$ for any states σ and ρ. If a CP map is given by the Kraus form $\Phi\rho = \sum_i A_i \rho A_i^\dagger$, the dual channel reads $\tilde{\Phi}\rho = \sum_i A_i^\dagger \rho A_i$. This gives a link between the dynamical matrices representing dual channels,

$$\tilde{\Phi} = (\Phi^T)^S = (\Phi^S)^T \quad \text{and} \quad D_{\tilde{\Phi}} = (D_\Phi^T)^S = (D_\Phi^S)^T = \overline{D_\Phi^S}. \tag{10.76}$$

Since neither the transposition nor the swap modify the spectrum of a matrix, the spectra of the dynamical matrices for dual channels are the same.

If channel Φ is trace preserving, its dual $\tilde{\Phi}$ is unital, and conversely, if Φ is unital then $\tilde{\Phi}$ is trace preserving. Thus the channel dual to a bistochastic one is bistochastic.

Let us analyse in some detail the set $\mathcal{BU}_N$ of all *unistochastic operations*, for which the representation (10.64) exists. The initial state of the environment is maximally mixed, $\sigma = \mathbb{1}/N$, so the map Ψ_U is determined by a unitary matrix U of size N^2. The Kraus operators A_i are eigenvectors of the dynamical matrix D_{Ψ_U}. On the other hand, they enter also the Schmidt decomposition (10.30) of U as shown in (10.64), and are proportional to the eigenvectors of $(U^R)^\dagger U^R$. Therefore[21]

$$D_{\Psi_U} = \frac{1}{N} (U^R)^\dagger \, U^R \quad \text{so that} \quad \Psi_U = \frac{1}{N} \big[(U^R)^\dagger \, U^R \big]^R. \tag{10.77}$$

[20] This example is due to Landau and Streater (1993).

[21] The same formula holds for K-unistochastic maps (Section 10.3), but then U^R is a rectangular matrix of size $N^2 \times N^{2K}$.

Table 10.3. *Quantum maps acting on density matrices and given by a positive definite dynamical matrix D versus classical Markov dynamics on probability vectors defined by transition matrix T with non-negative elements*

Quantum	Completely positive maps:	Classical	Markov chains given by:
S_1^Q	Trace preserving, $\mathrm{Tr}_A D = \mathbb{1}$	S_1^{Cl}	Stochastic matrices T
S_2^Q	Unital, $\mathrm{Tr}_B D = \mathbb{1}$	S_2^{Cl}	T^T is stochastic
S_3^Q	Unital & trace preserving maps	S_3^{Cl}	Bistochastic matrices B
S_4^Q	Maps with $A_i = A_i^\dagger$ $\Rightarrow D = D^T$	S_4^{Cl}	Symmetric stochastic matrices, $B = B^T$
S_5^Q	Unistochastic operations, $D = U^R (U^R)^\dagger$	S_5^{Cl}	Unistochastic matrices, $B_{ij} = \lvert U_{ij}\rvert^2$
S_6^Q	Unitary transformations	S_6^{Cl}	Permutations

We have thus arrived at an important result: for any unistochastic map the spectrum of the dynamical matrix is given by the Schmidt coefficients, $d_i = \lambda_i / N$, of the unitary matrix U treated as an element of the composite HS space. For any local operation, $U = U_1 \otimes U_2$ the superoperator is unitary, $\Psi_U = U_1 \otimes U_1^*$ so $||\Psi_U||^2_{\mathrm{HS}} = \mathrm{Tr}\Psi_U \Psi_U^\dagger = N^2$. The resulting unitary operation is an isometry, and can be compared with a permutation S_6^{Cl} acting on classical probability vectors. The spaces listed in Table 10.3 satisfy the relations $S_1 \cap S_2 = S_3$ and $S_3 \supset S_5 \supset S_6$ in both the classical and the quantum set-up. However, the analogy is not exact since the inclusion $S_3^{\mathrm{Cl}} \supset S_4^{\mathrm{Cl}}$ does not have a quantum counterpart.

10.7 One qubit maps

When $N = 2$ the quantum operations are called *binary channels*. In general, the space $\mathcal{CP}_N$ is $(N^4 - N^2)$-dimensional. Hence the set of binary channels has 12 dimensions. To parametrize it we begin with the observation that a binary channel is an affine map of the Bloch ball into itself – subject to restrictions that we will deal with later.[22] If we describe density matrices through their Bloch vectors, as in Eq. (5.9), this means that the map $\rho' = \Phi\rho$ can be written in the form

$$\vec{\tau}\,' \;=\; t\vec{\tau} + \vec{\kappa} \;=\; O_1 \eta\, O_2^{\mathrm{T}} \vec{\tau} + \vec{\kappa}\;, \tag{10.78}$$

[22] The explicit description given below is due to Fujiwara and Algoet (1999) and to King and Ruskai (2001). Geometric properties of the set of positive one-qubit maps were also studied in (Oi, n.d.; Wódkiewicz, 2001). A relation with Lorentz transformations is explained in Arrighi and Patricot (2003).

where t denotes a real matrix of size 3 which we diagonalize by orthogonal transformations O_1 and O_2. Actually we permit only rotations belonging to the $SO(3)$ group, which means that some of the elements of the diagonal matrix η may be negative – the restriction is natural because it corresponds to using unitarily similar quantum operations, cf. Eq. (10.68). The elements of the diagonal matrix η are collected into a vector $\vec{\eta} = (\eta_x, \eta_y, \eta_z)$, called the *distortion vector* because the transformation $\vec{\tau}\,' = \eta\,\vec{\tau}$ takes the Bloch ball to an ellipsoid given by

$$\frac{1}{4} = \tau_x^2 + \tau_y^2 + \tau_z^2 = \Big(\frac{\tau_x{}'}{\eta_x}\Big)^2 + \Big(\frac{\tau_y{}'}{\eta_y}\Big)^2 + \Big(\frac{\tau_z{}'}{\eta_z}\Big)^2 . \tag{10.79}$$

Finally the vector $\vec{\kappa} = (\kappa_x, \kappa_y, \kappa_z)$ is called the *translation vector*, because it moves the centre of mass of the ellipsoid. We can now see where the 12 dimensions come from: there are three parameters $\vec{\eta}$ that determine the shape of the ellipsoid, three parameters $\vec{\kappa}$ that determine its centre of mass, three parameters to determine its orientation, and three parameters needed to rotate the Bloch ball to a standard position relative to the ellipsoid (before it is subject to the map described by $\vec{\eta}$).

The map is positive whenever the ellipsoid lies within the Bloch ball. The map is unital if the centre of the Bloch ball is a fixed point of the map, which means that $\vec{\kappa} = 0$. But the map is completely positive only if the dynamical matrix is positive definite, which means that not every ellipsoid inside the Bloch ball can be the image of a completely positive map. We are not so interested in the orientation of the ellipsoid, so as our canonical form of the affine map we choose

$$\vec{\tau}\,' \;=\; \eta\vec{\tau} + \vec{\kappa}\;. \tag{10.80}$$

It is straightforward to work out the superoperator Φ of the map. Reshuffling this according to (10.41) we obtain the dynamical matrix

$$D = \frac{1}{2}\begin{bmatrix} 1+\eta_z+\kappa_z & 0 & \kappa_x+\mathrm{i}\kappa_y & \eta_x+\eta_y \\ 0 & 1-\eta_z+\kappa_z & \eta_x-\eta_y & \kappa_x+\mathrm{i}\kappa_y \\ \kappa_x-\mathrm{i}\kappa_y & \eta_x-\eta_y & 1-\eta_z-\kappa_z & 0 \\ \eta_x+\eta_y & \kappa_x-\mathrm{i}\kappa_y & 0 & 1+\eta_z-\kappa_z \end{bmatrix}. \tag{10.81}$$

Note that $\mathrm{Tr}_A D = \mathbb{1}$, as required. But the parameters $\vec{\eta}$ and $\vec{\kappa}$ must now be chosen so that D is positive definite, otherwise the transformation is not completely positive.

We will study the simple case when $\vec{\kappa} = 0$, in which case the totally mixed state is invariant and the map is unital (bistochastic). Then the matrix D splits into two blocks and its eigenvalues are

$$d_{0,3} = \frac{1}{2}[1+\eta_z \pm (\eta_x+\eta_y)] \quad \text{and} \quad d_{1,2} = \frac{1}{2}[1-\eta_z \pm (\eta_x-\eta_y)]\,. \tag{10.82}$$

Hence, if the *Fujiwara–Algoet conditions*

$$(1 \pm \eta_z)^2 \geq (\eta_x \pm \eta_y)^2 \tag{10.83}$$

hold, the dynamical matrix is positive definite and the corresponding positive map $\Phi_{\vec{\eta}}$ is CP. There are four inequalities: they define a convex polytope, and indeed a regular tetrahedron whose extreme points are $\vec{\eta} = (1, 1, 1)$, $(1, -1, -1)$, $(-1, 1, -1)$, $(-1, -1, 1)$. All maps within the cube defined by $|\eta_i| \leq 1$ are positive, so the tetrahedron of completely positive unital maps is a proper subset $\mathcal{B}_2$ of the set of all positive unital maps.

Note that dynamical matrices of unital maps of the form (10.81) commute. In effect, if we think of dynamical matrices as rescaled density matrices, our tetrahedron can be regarded as an eigenvalue simplex in $\mathcal{M}^{(4)}$. The eigenvectors consist of the identity $\sigma_0 = \mathbb{1}_2$ and the three Pauli matrices. Our conclusion is that any map $\Phi \in \mathcal{B}_2$ can be brought by means of unitary rotations (10.68) into the **canonical form of one-qubit bistochastic maps**:

$$\rho \;\to\; \rho' \;=\; \frac{1}{2} \sum_{i=0}^{3} d_i \, \sigma_i \rho \, \sigma_i \qquad \text{with} \qquad \sum_{i=0}^{3} d_i = 2 \,. \tag{10.84}$$

This explains the name *Pauli channels*. The factor of $1/2$ compensates the normalization of the Pauli matrices, $\mathrm{Tr}\, \sigma_i^2 = 2$. The Kraus operators are $A_i = \sqrt{d_i/2}\, \sigma_i$. For the Pauli matrices $\sigma_j = -\mathrm{i} \exp(\mathrm{i}\pi \sigma_j/2)$ and the overall phase is not relevant, so the extreme points that they represent are rotations of the Bloch ball around the corresponding axis by the angle π. This confirms that the set of binary bistochastic channels is the convex hull of the unitary operations, which is no longer true when $N > 2$.

For concreteness let us distinguish some one-qubit channels; the following list should be read in conjunction with Table 10.4, which gives the distortion vector $\vec{\eta}$ and the Kraus spectrum $\vec{d}$ for each map. Figure 10.4 illustrates the action of the maps.

unital channels (with $\vec{\kappa} = 0$)

- *Identity* which is our canonical form of a unitary rotation.
- *Phase flip* (or *phase-damping channel*), $\vec{\eta} = (1 - 2p, 1 - 2p, 1)$. This channel turns the Bloch ball into an ellipsoid touching the Bloch sphere at the north and south poles. When $p = 1/2$ the image degenerates to a line. The analogous channel with $\vec{\eta} = (1, 1 - 2p, 1 - 2p)$ is called a *bit flip*, while the channel with $\vec{\eta} = (1 - 2p, 1, 1 - 2p)$ is called a *bit–phase flip*. To understand these names we observe that, with probability p, a bit flip exchanges the states $|0\rangle$ and $|1\rangle$.
- *Linear channel*, $\vec{\eta} = (0, 0, q)$. It sends the entire Bloch ball into a line segment of length $2q$. For $q = 0$ and $q = 1$ we arrive at the *completely depolarizing channel* Ψ_* and the *coarse graining* operation, $\Psi_{\mathrm{CG}}(\rho) = \mathrm{diag}(\rho)$, respectively.

Table 10.4. *Some one-qubit channels: distortion vector $\vec{\eta}$, translation vector $\vec{\kappa}$ equal to zero for unital channels, Kraus spectrum $\vec{d}$, and Kraus rank r.*

Channels	$\vec{\eta}$	$\vec{\kappa}$	unital	$\vec{d}$	r
rotation	$(1, 1, 1)$	$(0, 0, 0)$	yes	$(2, 0, 0, 0)$	1
phase flip	$(1-p, 1-p, 1)$	$(0, 0, 0)$	yes	$(2-p, p, 0, 0)$	2
decaying	$(\sqrt{1-p}, \sqrt{1-p}, 1-p)$	$(0, 0, p)$	no	$(2-p, p, 0, 0)$	2
depolarizing	$[1-x](1, 1, 1)$	$(0, 0, 0)$	yes	$\frac{1}{2}(4-3x, x, x, x)$	4
linear	$(0, 0, q)$	$(0, 0, 0)$	yes	$\frac{1}{2}(1+q, 1-q,$ $1-q, 1+q)$	4
planar	$(0, s, q)$	$(0, 0, 0)$	yes	$\frac{1}{2}(1+q+s, 1-q-s,$ $1-q+s, 1+q-s)$	4

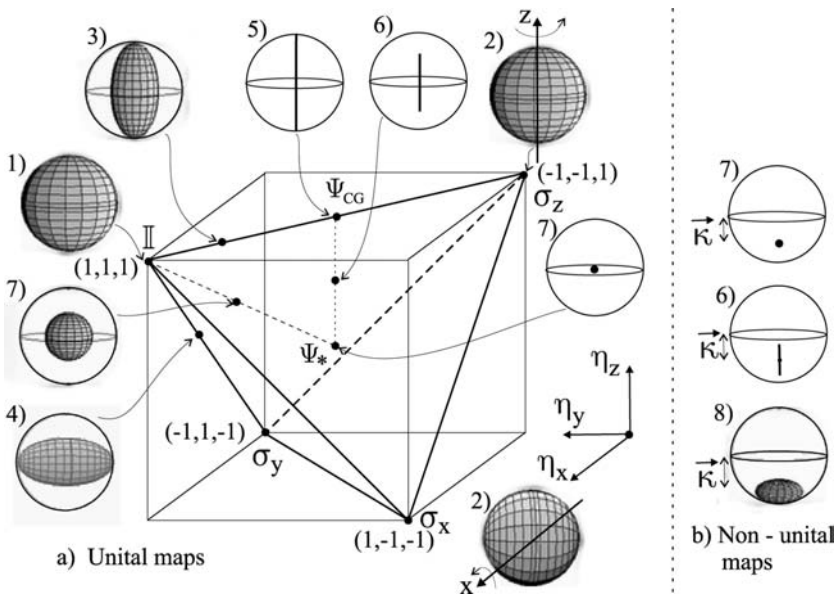


Figure 10.4. One-qubit maps: (a) unital Pauli channels: 1) identity, 2) rotation, 3) phase flip, 4) bit-phase flip, 5) coarse graining, 6) linear channel, 7) completely depolarizing channel; (b) non-unital maps ($\kappa \neq 0$): 8) decaying channel.

- *Planar channel*, $\vec{\eta} = (0, s, q)$, sends the Bloch ball into an ellipse with semi-axis s and q. Complete positivity requires $s \leq 1 - q$.
- *Depolarizing channel*, $\vec{\eta} = [1 - x](1, 1, 1)$. This simply shrinks the Bloch ball. When $x = 1$ we again arrive at the centre of the tetrahedron, that is at the completely depolarizing channel Φ_*. Note that the Kraus spectrum has a triple degeneracy, so there is an extra freedom in choosing the canonical Kraus operators.

If we drop the condition that the map be unital our canonical form gives a six-dimensional set; it is again a convex set but considerably more difficult to analyse.[23] A map of $\mathcal{CP}_2$, the canonical form of which consists of two Kraus operators, is either extremal or bistochastic (if $A_1^\dagger A_1 \sim A_2^\dagger A_2 \sim \mathbb{1}$). Table 10.4 gives one example of a non-unital channel, namely:

- *Decaying channel* (also called *amplitude-damping channel*), defined by $\vec{\eta} = (\sqrt{1-p}, \sqrt{1-p}, 1-p)$ and $\vec{\kappa} = (0, 0, p)$. The Kraus operators are

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix}. \tag{10.85}$$

Physically this is an important channel – and it exemplifies that a quantum operation can take a mixed state to a pure state.

Problems

Problem 10.1 Prove Naimark's theorem.

Problem 10.2 A map Φ is called *diagonal* if all Kraus operators A_i mutually commute, so in a certain basis they are diagonal, $\vec{d}^i = \mathrm{diag}(U^\dagger A_i U)$. Show that such a dynamics is given by the Hadamard product, from Problem 9.2, $\Phi\tilde{\rho} = H \circ \tilde{\rho}$, where $H_{mn} = \sum_{i=1}^k d_m^i \bar{d}_n^i$ with $m, n = 1, ...N$ while $\tilde{\rho} = U^\dagger \rho\, U$ (Landau and Streater, 1993; Havel, Sharf, Viola and Cory, 2001).

Problem 10.3 Let ρ be an arbitrary density operator acting on $\mathcal{H}_N$ and $\{A_i\}_{i=1}^{N^2}$ be a set of mutually orthogonal Kraus operators representing a given CP map Φ in its canonical Kraus form. Prove that the matrix

$$\sigma_{ij} \equiv \langle A_i | \rho | A_j \rangle = \mathrm{Tr} \rho A_j A_i^\dagger \tag{10.86}$$

forms a state acting on an extended Hilbert space, $\mathcal{H}_{N^2}$. Show that in particular, if $\rho = \mathbb{1}/N$, then this state is proportional to the dynamical matrix represented in its eigenbasis.

[23] A complete treatment of the qubit case was given by Ruskai, Szarek and Werner (2002).

Problem 10.4 Show that a binary, unital map $\Phi_{\eta_x,\eta_y,\eta_z}$ defined by (10.81) transforms any density matrix ρ in the following way (King and Ruskai, 2001)

$$\Phi_{\eta_x,\eta_y,\eta_z}\begin{bmatrix} a & z \\ \bar{z} & 1-a \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1+(2a-1)\eta_z & (z+\bar{z})\eta_x + (z-\bar{z})\eta_y \\ (z+\bar{z})\eta_x - (z-\bar{z})\eta_y & 1-(2a-1)\eta_z \end{bmatrix}.$$

Problem 10.5 What qubit channel is described by the Kraus operators

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{bmatrix} ? \tag{10.87}$$

Problem 10.6 Let $\rho \in \mathcal{M}^{(N)}$. Show that the operation Φ_ρ defined by $D_\Phi \equiv \rho \otimes \mathbb{1}_N$ acts as a complete one-step contraction, $\Phi_\rho \sigma = \rho$ for any $\sigma \in \mathcal{M}^{(N)}$.

11

Duality: maps versus states

Good mathematicians see analogies.
Great mathematicians see analogies between analogies.

Stefan Banach

We have already discussed the static structure of our 'Quantum Town' – the set of density matrices – on the one hand, and the set of all physically realizable processes which may occur in it on the other hand. Now we are going to reveal a quite remarkable property: the set of all possible ways to travel in the 'Quantum Town' is equivalent to a 'Quantum Country' – an appropriately magnified copy of the initial 'Quantum Town'! More precisely, the set of all transformations which map the set of density matrices of size N into itself (dynamics) is identical to a subset of the set of density matrices of size N^2 (kinematics). From a mathematical point of view this relation is based on the *Jamiołkowski isomorphism*, analysed later in this chapter. Before discussing this intriguing duality, let us leave the friendly set of quantum operations and pay a short visit to a neighbouring land of maps, as yet unexplored, which are positive but not completely positive.

11.1 Positive and decomposable maps

Quantum transformations which describe physical processes are represented by completely positive (CP) maps. Why should we care about maps which are not CP? On the one hand it is instructive to realize that seemingly innocent transformations are not CP, and thus do not correspond to any physical process. On the other hand, as discussed in Chapter 15, positive but not completely positive maps provide a crucial tool in the investigation of quantum entanglement.

Consider the transposition of a density matrix in a fixed basis, $T : \rho \to \rho^T$. (Since ρ is Hermitian this is equivalent to complex conjugation.) The superoperator entering (10.36) is the SWAP operator, $T_{\substack{m\mu \\ n\nu}} = \delta_{m\nu}\delta_{n\mu} = S_{\substack{m\mu \\ n\nu}}$. Hence it

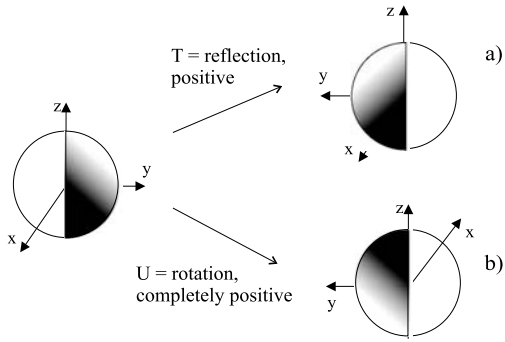


Figure 11.1. Non-contracting transformations of the Bloch ball: (a) transposition (reflection with respect to the x–z plane) – not completely positive; (b) rotation by π around z-axis – completely positive.

is symmetric with respect to reshuffling, $T = T^R = D_T$. This permutation matrix contains N diagonal entries equal to unity and $N(N-1)/2$ blocks of size two. Thus its spectrum consists of $N(N+1)/2$ eigenvalues equal to unity and $N(N-1)/2$ eigenvalues equal to -1, consistent with the constraint $\mathrm{Tr}D = N$. The matrix D_T is not positive, so the transposition T is not CP. Another way to reach this conclusion is to act with the extended map of partial transposition on the maximally entangled state (11.21) and to check that $[T \otimes \mathbb{1}](|\psi\rangle\langle\psi|)$ has negative eigenvalues.

The transposition of an N-dimensional Hermitian matrix changes the signs of the imaginary part of the elements D_{ij}. This is a reflection in an $N(N+1)/2 - 1$ dimensional hyperplane. As shown in Figure 11.1 this is simple to visualize for $N = 2$: when we use the representation (5.8) the transposition reflects the Bloch ball in the (x, z) plane. Note that a unitary rotation of the Bloch ball around the z-axis by the angle π also exchanges the 'western' and the 'eastern' hemispheres, but is completely positive.

As discussed in Section 10.3 a map fails to be CP if its dynamical matrix D contains at least one negative eigenvalue. Let $m \geq 1$ denote the number of the negative eigenvalues (in short, the *neg rank* [1] of D). Ordering the spectrum of D decreasingly allows us to rewrite its spectral decomposition

$$D = \sum_{i=1}^{N^2-m} d_i |\chi_i\rangle\langle\chi_i| \; - \sum_{i=N^2-m+1}^{N^2} |d_i| \, |\chi_i\rangle\langle\chi_i| \; . \tag{11.1}$$

[1] For transposition, the neg rank of D_T is $m = N(N-1)/2$.

Thus a not completely positive map has the canonical form

$$\rho' = \sum_{i=1}^{N^2-m} d_i \chi^i \rho (\chi^i)^\dagger - \sum_{i=N^2-m+1}^{N^2} |d_i| \chi^i \rho (\chi^i)^\dagger , \tag{11.2}$$

where the Kraus operators $A_i = \sqrt{|d_i|}\, \chi^i$ form an orthogonal basis. This is analogous to the canonical form (10.55) of a CP map, and it shows that a positive map may be represented as a *difference* of two completely positive maps (Sudarshan and Shaji, 2003). While this is true, it does not solve the problem: taking any two CP maps and constructing a quasi-mixture[2] $\Phi = (1+a)\Psi_1^{\rm CP} - a\Psi_2^{\rm CP}$, we do not know in advance how large the contribution a of the negative part might be to keep the map Φ positive... .[3] In fact the characterization of the set $\mathcal{P}_N$ of positive maps: $\mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$ for $N > 2$ is by far not simple.[4] By definition, $\mathcal{P}_N$ contains the set $\mathcal{CP}_N$ of all CP maps as a proper subset. To learn more about the set of positive maps we will need some other features of the operation of transposition T. For any operation Φ the modifications of the dynamical matrix induced by a composition with T may be described by the transformation of partial transpose (see Table 10.1),

$$T\Phi = \Phi^{S_1} , \quad D_{T\Phi} = D_\Phi^{T_A} , \quad \text{and} \quad \Phi T = \Phi^{S_2} , \quad D_{\Phi T} = D_\Phi^{T_B} . \tag{11.3}$$

To demonstrate this it is enough to use the explicit form of Φ_T and the observation that

$$D_{\Psi\Phi} = [D_\Psi^R D_\Phi^R]^R . \tag{11.4}$$

Positivity of $D_{\Psi\Phi}$ follows also from the fact that the composition of two CP maps is completely positive. This follows directly from the identity $(\Psi\Phi)\otimes\mathbb{1} = (\Psi\otimes\mathbb{1})\cdot(\Phi\otimes\mathbb{1})$ and implies the following:

Lemma 11.1 (Reshuffling) *Consider two Hermitian matrices A and B of the same size KN.*

$$\textit{If} \quad A \geq 0 \quad \textit{and} \quad B \geq 0 \quad \textit{then} \quad (A^R B^R)^R \geq 0 . \tag{11.5}$$

For a proof (Havel, 2003) see Problem 11.1.

[2] The word *quasi* is used here to emphasize that some weights are negative.

[3] Although some criteria for positivity are known (Størmer, 1963; Jamiołkowski, 1975; Majewski, 1975), they do not lead to a practical test of positivity. A recently proposed technique of extending the system (and the map) a certain number of times gives a constructive test for positivity for a large class of maps (Doherty, Parillo and Spedalieri, 2004).

[4] This issue was a subject of mathematical interest many years ago (Størmer, 1963; Choi, 1972; Woronowicz, 1976b; Takesaki and Tomiyama, 1983) and quite recently (Eom and Kye, 2000; Majewski and Marciniak, 2001; Kye, 2003).

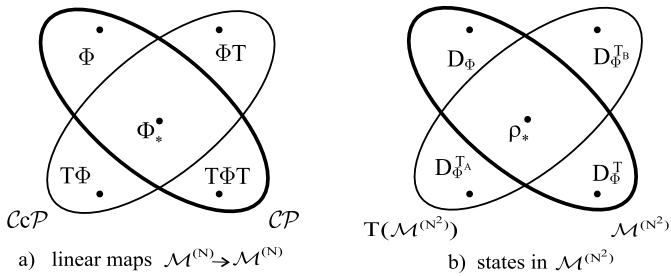


Figure 11.2. (a) The set of CP maps, its image with respect to transposition $\mathcal{C}c\mathcal{P} = T(\mathcal{CP})$, and their intersection $\mathcal{PPTM} \equiv \mathcal{CP} \cap \mathcal{C}c\mathcal{P}$; (b) the isomorphic sets $\mathcal{M}^{(N)}$ of quantum states (dynamical matrices), its image $T_A(\mathcal{M}^{(N)})$ under the action of partial transposition, and the set of PPT states.

Sandwiching Φ between two transpositions does not influence the spectrum of the dynamical matrix, $T\Phi T = \Phi^S = \Phi^*$ and $D_{T\Phi T} = D_\Phi^T = D_\Phi^*$. Thus if Φ is a CP map, so is $T\Phi T$ (if D_Φ is positive so is D_Φ^T). See Figure 11.2.

The not completely positive transposition map T allows one to introduce the following definition (Størmer, 1963; Choi, 1975a; Choi, 1980):

A map Φ *is called* **completely co-positive** *(CcP), if the map* $T\Phi$ *is CP.*

Properties (11.3) of the dynamical matrix imply that the map ΦT could be used instead to define the same set of CcP maps. Thus any CcP map Φ may be written in a Kraus-like form

$$\rho' \;=\; \Phi(\rho) \;=\; \sum_{i=1}^{k} A_i \rho^T A_i^\dagger \,. \tag{11.6}$$

Moreover, as shown in Figure 11.2, the set $\mathcal{C}c\mathcal{P}$ may be understood as the image of $\mathcal{CP}$ with respect to the transposition. Since we have already identified the transposition with a reflection, it is rather intuitive to observe that the set $\mathcal{C}c\mathcal{P}$ is a twin copy of $\mathcal{CP}$ with the same shape and volume. This property is easiest to analyse for the set $\mathcal{B}_2$ of one qubit bistochastic maps (Oi, n.d.), written in the canonical form (10.84). Then the dual set of CcP unital one qubit maps, $T(\mathcal{B}_2)$, forms a tetrahedron spanned by four maps $T\sigma_i$ for $i = 0, 1, 2, 3$. This is the reflection of the set of bistochastic maps with respect to its centre – the completely depolarizing channel Φ_*. See Figure 11.3(b). Observe that the corners of $\mathcal{B}_2$ are formed by proper rotations while the extremal points of the set of CcP maps represent reflections. The intersection of the tetrahedra forms an octahedron of PPT *inducing* maps ($\mathcal{PPTM}$); see Section 15.4.

A positive map Φ is called *decomposable*, if it may be expressed as a convex combination of a CP map and a CcP map,

$$\Phi = a\Phi_{\mathrm{CP}} + (1-a)\Phi_{\mathrm{CcP}} \quad \text{with} \quad a \in [0,1] \,. \tag{11.7}$$

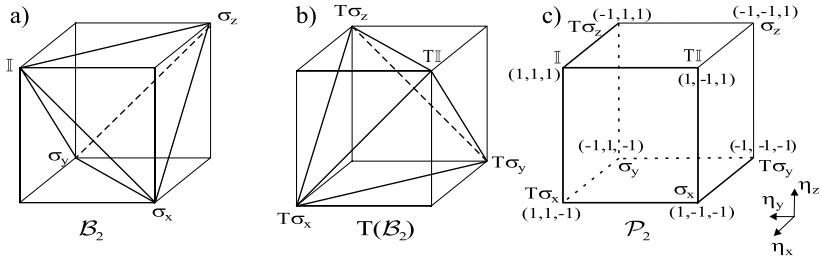


Figure 11.3. Subsets of one-qubit maps: (a) set $\mathcal{B}_2$ of bistochastic maps (unital and CP), (b) set $T(\mathcal{B}_2)$ of unital and CcP maps, (c) set of positive (decomposable) unital maps. The intersection of the two tetrahedra forms an octahedron of super-positive maps.

A relation between CP maps acting on quaternion matrices and the decomposable maps defined on complex matrices was found by Kossakowski (2000). An important characterization of the set $\mathcal{P}_2$ of positive maps acting on (complex) states of one qubit follows from Størmer (1963) and Woronowicz (1976a):

Theorem 11.1 (Størmer–Woronowicz's) *Every one-qubit positive map* $\Psi \in \mathcal{P}_2$ *is decomposable.*

In other words, the set of $N = 2$ positive maps can be represented by the convex hull of the set of CP and CcP maps. This property is illustrated for unital maps (in canonical form) in Figure 11.3, where the cube of positive maps forms the convex hull of the two tetrahedra, and schematically in Figure 11.4(a). It holds also for the maps $\mathcal{M}^{(2)} \to \mathcal{M}^{(3)}$ and $\mathcal{M}^{(3)} \to \mathcal{M}^{(2)}$ (Woronowicz, 1976b), but is not true in higher dimensions, see Figure 11.4(b).

Consider a map defined on $\mathcal{M}^{(3)}$, depending on three non-negative parameters,

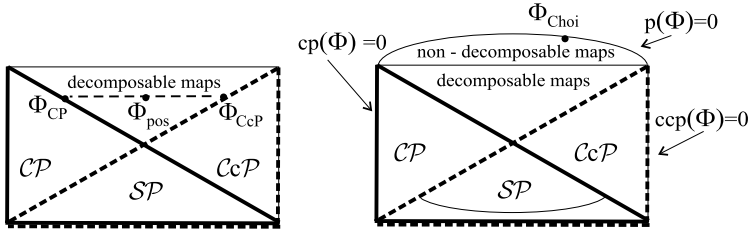
$$\Psi_{a,b,c}(\rho) = \begin{bmatrix} a\rho_{11} + b\rho_{22} + c\rho_{33} & 0 & 0 \\ 0 & c\rho_{11} + a\rho_{22} + b\rho_{33} & 0 \\ 0 & 0 & b\rho_{11} + c\rho_{22} + a\rho_{33} \end{bmatrix} - \rho . \tag{11.8}$$

The map $\Psi_{2,0,2} \in \mathcal{P}_3$ was a first example of a indecomposable map, found by Choi in 1975 (Choi, 1975b). As denoted schematically in Figure 11.4(b) this map is extremal and belongs to the boundary of the convex set $\mathcal{P}_3$. The *Choi map* was generalized in Choi and Lam (1977) and in Cho, Kye and Lee (1992), where it was shown that the map (11.8) is positive if and only if

$$a \geq 1 , \quad a + b + c \geq 3 , \quad 1 \leq a \leq 2 \implies bc \geq (2 - a)^2 , \tag{11.9}$$

while it is decomposable if and only if

$$a \geq 1 , \qquad 1 \leq a \leq 3 \implies bc \geq (3 - a)^2/4 . \tag{11.10}$$



a) $\mathcal{P}_2$ = positive linear maps $\mathcal{M}^{(2)} \to \mathcal{M}^{(2)}$ b) $\mathcal{P}_3$ = positive linear maps $\mathcal{M}^{(3)} \to \mathcal{M}^{(3)}$

Figure 11.4. Sketch of the set positive maps: (a) for $N = 2$ all maps are decomposable so $\mathcal{SP} = \mathcal{CP} \cap \mathcal{CcP} = \mathcal{PPTM}$, (b) for $N > 2$ there exist non-decomposable maps and $\mathcal{SP} \subset \mathcal{CP} \cap \mathcal{CcP}$ – see Section 11.2.

In particular, $\Psi_{2,0,c}$ is positive but not decomposable for $c \geq 1$. All generalized indecomposable Choi maps are known to be *atomic* (Ha, 1998), that is they cannot be written as a convex sum of 2-positive and 2-co-positive maps (Tanahashi and Tomiyama, 1988). Examples of indecomposable maps belonging to $\mathcal{P}_4$ were given in Woronowicz (1976b) and in Robertson (1983). A family of indecomposable maps for an arbitrary finite dimension $N \geq 3$ was recently found by Kossakowski (2003). They consist of an affine contraction of the set $\mathcal{M}^{(N)}$ of density matrices into a ball inscribed in it, followed by a generic rotation from $O(N^2 - 1)$. Although several other methods of construction of indecomposable maps were proposed (Tang, 1986; Tanahashi and Tomiyama, 1988; Osaka, 1991; Kim and Kye, 1994), some of them in the context of quantum entanglement (Terhal, 2000b; Ha, Kye and Park, 2003), the general problem of describing all positive maps remains open. In particular, it is not known if one can find a finite set of K positive maps $\{\Psi_j\}$, such that $\mathcal{P}_N = \text{conv hull}\left(\cup_{j=1}^{K} \Psi_j(\mathcal{CP}_N)\right)$.

Due to the theorem of Størmer and Woronowicz the answer is known for $N = 2$, for which $K = 2$, $\Psi_1 = \mathbb{1}$ and $\Psi_2 = T$. As we shall see in Chapter 15 these properties of the set $\mathcal{P}_N$ are decisive for the separability problem: the separability criterion based on positivity of $(\mathbb{1} \otimes T)\rho$ is conclusive for the system of two qubits, while in the general case of $N \times N$ composite systems it provides a partial solution only (Horodecki, Horodecki and Horodecki, 1996a).

Indecomposable maps are worth investigating, since each such map provides a criterion for separability (see Section 15.4. Conditions for a positive map Φ to be decomposable were found some time ago by Størmer (1982). Since this criterion is not a constructive one, we describe here a simple test. Assume first that the map is not symmetric with respect to the transposition,[5] $\Phi \neq T\Phi$. These two points

[5] If this is the case, one may perform this procedure with a perturbed map, $\Phi' = (1 - \epsilon)\Phi + \epsilon\Psi$, for which $\Phi' \neq T\Phi'$, and study the limit $\epsilon \to 0$.

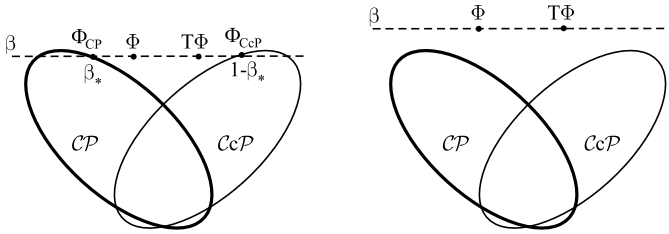


Figure 11.5. Geometric criterion to verify decomposability of a map Φ: (a) if the line passing through Φ and ΦT crosses the set of completely positive maps, a decomposition of Φ is explicitly constructed.

determine a line in the space of maps, parameterized by β, along which we check if the dynamical matrix

$$D_{\beta\Phi+(1-\beta)T\Phi} = \beta D_\Phi + (1-\beta) D_\Phi^{T_A} \tag{11.11}$$

is positive. If it is found to be positive for some $\beta_* < 0$ (or $\beta_* > 1$) then the line (11.11) crosses the set of completely positive maps (see Figure 11.5(a)). Since $D(\beta_*)$ represents a CP map $\Psi_{\rm CP}$, hence $D(1-\beta_*)$ defines a completely co–positive map $\Psi_{\rm CcP}$, and we find an explicit decomposition, $\Phi = [-\beta_* \Psi_{\rm CP} + (1-\beta_*)\Psi_{\rm CcP}]/(1-2\beta_*)$. In this way decomposability of Φ may be established, but one cannot confirm that a given map is indecomposable.

To study the geometry of the set of positive maps one may work with the Hilbert–Schmidt distance $d(\Psi, \Phi) = ||\Psi - \Phi||_{\rm HS}$. Since reshuffling of a matrix does not influence its HS norm, the distance can be measured directly in the space of dynamical matrices, $d(\Psi, \Phi) = D_{HS}(D_\Psi, D_\Phi)$. Note that for unital one qubit maps, (10.81) with $\vec{\kappa} = 0$, one has $d(\Phi_1, \Phi_2) = |\vec{\eta}_1 - \vec{\eta}_2|$, so Figue 11.3 represents correctly the HS geometry of the space of $N = 2$ unital maps.

In order to characterize to what extent a given map is close to the boundary of the set of positive (CP or CcP) maps, let us define the quantities:

- **Complete positivity**

$$cp(\Phi) \equiv \min_{\mathcal{M}^{(N)}} \langle \rho | D_\Phi | \rho \rangle \tag{11.12}$$

- **Complete co-positivity**

$$ccp(\Phi) \equiv \min_{\mathcal{M}^{(N)}} \langle \rho | D_\Phi^{T_A} | \rho \rangle \tag{11.13}$$

- **Positivity**

$$p(\Phi) \equiv \min_{|x\rangle, |y\rangle \in \mathbb{C}P^{N-1}} [\langle x \otimes y | D_\Phi | x \otimes y \rangle]. \tag{11.14}$$

The first two quantities are easily found by diagonalization, since $cp(\Phi) = \min\{\mathrm{eig}(D_\Phi)\}$ and $ccp(\Phi) = \min\{\mathrm{eig}(D_\Phi^{T_A})\}$. Although $p(\Phi) \geq cp(\Phi)$ by construction,[6] the evaluation of positivity is more involved, since one needs to perform the minimization over the space of all product states, that is the Cartesian product $\mathbb{C}\mathbf{P}^{N-1} \times \mathbb{C}\mathbf{P}^{N-1}$. No straightforward method of computing this minimum is known, so one has to rely on numerical minimization.[7]

A given map Φ is completely positive (CcP, positive) if and only if the complete positivity (ccp, positivity) is non-negative. As marked in Figure 11.4(b), the relation $cp(\Phi) = 0$ defines the boundary of the set $\mathcal{CP}_N$, while $ccp(\Phi) = 0$ and $p(\Phi) = 0$ define the boundaries of $\mathcal{CcP}_N$ and $\mathcal{P}_N$. By direct diagonalization of the dynamical matrix we find that $cp(\mathbb{1}) = ccp(T) = 0$ and $ccp(\mathbb{1}) = cp(T) = -1$.

For any not completely positive map Φ_{nCP} one may look for its best approximation with a physically realizable[8] CP map Φ_{CP}, for example by minimizing their HS distance $d(\Phi_{\mathrm{nCP}}, \Phi_{\mathrm{CP}})$ – see Figure 11.8(a). To see a simple application of complete positivity, consider a non-physical positive map with $cp(\Phi_{\mathrm{nCP}}) = -x < 0$. One – in general not optimal – CP approximation may be constructed out of its convex combination with the completely depolarizing channel Ψ_*. Diagonalizing the dynamical matrix representing the map $\Psi_x = a\Phi_{\mathrm{nCP}} + (1-a)\Psi_*$ with $a = 1/(Nx+1)$ we see that its smallest eigenvalue is equal to zero, so Ψ_x belongs to the boundary of $\mathcal{CP}_N$. Hence the distance $d(\Phi_{\mathrm{nCP}}, \Phi_x)$, which is a function of the complete positivity $cp(\Phi_{\mathrm{nCP}})$, gives an upper bound for the distance of Φ_{nCP} from the set $\mathcal{CP}$. In a similar way one may use $ccp(\Phi)$ to obtain an upper bound for the distance of an analysed non-CcP map Φ_{nCcP} from the set $\mathcal{CcP}$ – compare with Problem 11.6. As discussed in further sections and, in more detail, in Chapter 15, the solution of the analogous problem in the space of density matrices allows one to characterize the entanglement of a two-qubit mixed state ρ by its minimal distance to the set of separable states.

11.2 Dual cones and super-positive maps

Since a CP map Φ is represented by a positive dynamical matrix D_Φ, the trace $\mathrm{Tr} P D_\Phi$ is non-negative for any projection operator P. Furthermore, for any two CP maps, the HS scalar product of their dynamical matrices satisfies $\mathrm{Tr} D_\Phi D_\Psi \geq 0$. If such a relation is fulfilled for any CP map Ψ, it implies complete positivity of Φ.

[6] Both quantities are equal if D is a product matrix, which occurs if only one singular value ξ of the superoperator $\Phi = D^R$ is positive.

[7] In certain cases this quantity was estimated analytically by Terhal (2000b) and numerically by Gühne, Hyllus, Bruß, Ekert, Lewenstein, Macchiavello and Sanpera (2002) and Gühne, Hyllus, Bruß, Ekert, Lewenstein, Macchiavello and Sanpera (2003) when characterizing entanglement witnesses.

[8] Such *structural physical approximations* were introduced in Horodecki and Ekert (2002) to propose an experimentally feasible scheme of entanglement detection and later studied in Fiurášek (2002).

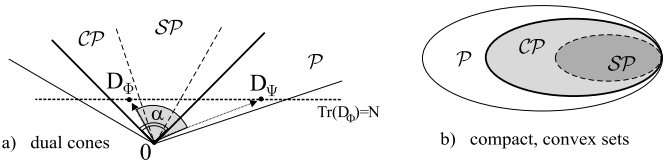


Figure 11.6. (a) Dual cones $\mathcal{P} \leftrightarrow \mathcal{SP}$ and self dual $\mathcal{CP} \leftrightarrow \mathcal{CP}$; (b) the corresponding compact sets of trace preserving positive, completely positive and super-positive maps, $\mathcal{P} \supset \mathcal{CP} \supset \mathcal{SP}$.

More formally, we define a pairing between maps,

$$(\Phi, \Psi) \equiv \langle D_\Phi, D_\Psi \rangle = \mathrm{Tr}\, D_\Phi^\dagger D_\Psi = \mathrm{Tr}\, D_\Phi D_\Psi , \tag{11.15}$$

and obtain the following characterization of the set $\mathcal{CP}$ of CP maps,

$$\{\Phi \in \mathcal{CP}\} \;\Leftrightarrow\; (\Phi, \Psi) \geq 0 \quad \text{for all} \quad \Psi \in \mathcal{CP} . \tag{11.16}$$

This property is illustrated in Figure 11.6(a) – the angle formed by any two positive dynamical matrices at the zero map 0 will not be greater than $\pi/2$. Thus the set of CP maps has a self-dual property and is represented as a right angle cone. All trace preserving maps belong to the horizontal line given by the condition $\mathrm{Tr} D_\Phi = N$.

In a similar way one may define the cone dual to the set $\mathcal{P}$ of positive maps.

A linear map $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$ *is called* **super-positive**[9] (SP) (Ando, 2004), *if*

$$\{\Phi \in \mathcal{SP}\} \;\Leftrightarrow\; (\Phi, \Psi) \geq 0 \quad \text{for all} \quad \Psi \in \mathcal{P} . \tag{11.17}$$

Once $\mathcal{SP}$ is defined as a set containing all SP maps, one may write a dual condition to characterize the set of positive maps,

$$\{\Phi \in \mathcal{P}\} \;\Leftrightarrow\; (\Phi, \Psi) \geq 0 \quad \text{for all} \quad \Psi \in \mathcal{SP} . \tag{11.18}$$

The cones $\mathcal{SP}$ and $\mathcal{P}$ are dual by construction and any boundary line of $\mathcal{P}$ determines the perpendicular (dashed) boundary line of $\mathcal{SP}$. The self-dual set of CP maps is included in the set of positive maps $\mathcal{P}$, and includes the dual set of super-positive maps. All the three sets are convex. See Figure 11.6.

The dynamical matrix of a positive map Ψ is block positive, so it is clear that condition (11.17) implies that a map Φ is super-positive if its dynamical matrix admits a tensor product representation

$$D_\Phi = \sum_i^k A_i \otimes B_i , \quad \text{with} \quad A_i \geq 0, \quad B_i \geq 0; \quad i = 1, \ldots, k . \tag{11.19}$$

[9] SP maps are also called *entanglement breaking channels* (see Section 15.4).

As we shall see in Chapter 15, this very condition is related to separability of the state ρ associated with $D_\Phi \in \mathcal{CP}$. In particular, if the angle α between the vectors pointing to D_Φ and a block positive D_Ψ is obtuse, the state $\rho = D_\Phi/N$ is entangled (compare the notion of entanglement witness in Section 15.4).

In general it is not easy to describe the set $\mathcal{SP}$ explicitly. The situation simplifies for one-qubit maps: due to the theorem by Størmer and Woronowicz any positive map may be represented as a convex combination of a CP map and a CcP map. The sets $\mathcal{CP}$ and $\mathcal{CcP}$ share similar properties and are both self dual. Hence a map Φ is super-positive if it is simultaneously CP and CcP. But the neat relation $\mathcal{SP} = \mathcal{CP} \cap \mathcal{CcP}$ holds for $N = 2$ only. For $N > 2$ the set $\mathcal{P}$ of positive maps is larger (there exist non-decomposable maps), so the dual set becomes smaller, $\mathcal{SP} \subset \mathcal{CP} \cap \mathcal{CcP} \equiv \mathcal{PPTM}$ (see Figure 11.4).

11.3 Jamiołkowski isomorphism

Let $\mathcal{CP}_N$ denote the convex set of all trace preserving, completely positive maps $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$. Any such map may be uniquely represented by its dynamical matrix D_Φ of size N^2. It is a positive, Hermitian matrix and its trace is equal to N. Hence the rescaled matrix $\rho_\Phi \equiv D_\Phi/N$ represents a mixed state in $\mathcal{M}^{(N^2)}$. In fact rescaled dynamical matrices form only a subspace of this set, determined by the trace preserving conditions (10.46), which impose N^2 constraints. Let us denote this $(N^4 - N^2)$-dimensional set by $\mathcal{M}_{\mathbb{1}}^{(N^2)}$. Since any trace preserving CP map has a dynamical matrix, and vice versa, the correspondence between maps in $\mathcal{CP}_N$ and states in $\mathcal{M}_{\mathbb{1}}^{(N^2)}$ is one-to-one. In Table 11.1 this isomorphism is labelled J_{II}.

Let us find the dynamical matrix for the identity operator:

$$\mathbb{1}_{\substack{m\mu \\ n\nu}} = \delta_{mn}\delta_{\mu\nu} \quad \text{so that} \quad D^{\mathbb{1}}_{\substack{m\mu \\ n\nu}} = (\mathbb{1}_{\substack{m\mu \\ n\nu}})^R = \delta_{m\mu}\delta_{n\nu} = N\rho^{\phi}_{\substack{m\mu \\ n\nu}}, \tag{11.20}$$

where $\rho^\phi = |\phi^+\rangle\langle\phi^+|$ represents the operator of projection on a maximally entangled state of the composite system, namely

$$|\phi^+\rangle \;=\; \frac{1}{\sqrt{N}} \sum_{i=1}^{N} |i\rangle \otimes |i\rangle. \tag{11.21}$$

This state is written in its Schmidt form (9.8), and we see that all its Schmidt coefficients are equal, $\lambda_1 = \lambda_i = \lambda_N = 1/N$. Thus we have found that the identity operator corresponds to the maximally entangled pure state $|\phi^+\rangle\langle\phi^+|$ of the composite system. Interestingly, this correspondence may be extended for other operations,

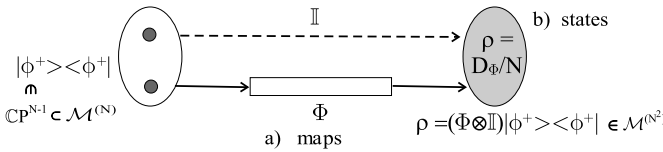


Figure 11.7. Duality (11.22) between a quantum map Φ acting on a part of the maximally entangled state $|\phi^+\rangle$ and the resulting density matrix $\rho = \frac{1}{N} D_\Phi$.

or in general, for arbitrary linear maps. The **Jamiołkowski isomorphism** [10]

$$\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)} \quad \longleftrightarrow \quad \rho_\Phi \equiv D_\Phi / N = \big[\Phi \otimes \mathbb{1}\big](|\phi^+\rangle\langle\phi^+|) \qquad (11.22)$$

allows us to associate a linear map Φ acting on the space of mixed states $\mathcal{M}^{(N)}$ with an operator acting in the enlarged Hilbert state $\mathcal{H}_N \otimes \mathcal{H}_N$. To show this relation write the operator $\Phi \otimes \mathbb{1}$ as an eight-indices matrix[11] and study its action on the state ρ^ϕ expressed by two Kronecker's deltas as in (11.20),

$$\Phi_{\substack{mn \\ m'n'}} \mathbb{1}_{\substack{\mu\nu \\ \mu'\nu'}} \rho^\phi_{\substack{m'\mu' \\ n'\nu'}} = \frac{1}{N} \Phi_{\substack{mn \\ \mu\nu}} = \frac{1}{N} D_{\substack{m\mu \\ n\nu}}. \qquad (11.23)$$

Conversely, for any positive matrix D we find the corresponding map Φ by diagonalization. The reshaped eigenvectors of D, rescaled by the roots of the eigenvalues, give the canonical Kraus form (10.55) of the operation Φ. If $\mathrm{Tr}_A \rho_\Phi = \mathbb{1}/N$ so that $\rho_\Phi \in \mathcal{M}^{(N^2)}_{\mathbb{1}}$, the map Φ is trace preserving.

Consider now a more general case in which ρ denotes a state acting on a composite Hilbert space $\mathcal{H}_N \otimes \mathcal{H}_N$. Let Φ be an arbitrary map which sends $\mathcal{M}^{(N)}$ into itself and let $D_\Phi = \Phi^R$ denote its dynamical matrix (of size N^2). Acting with the extended map on ρ we find its image $\rho' = [\Phi \otimes \mathbb{1}](\rho)$. Writing down the explicit form of the corresponding linear map in analogy to (11.23) and contracting over four indices which represent $\mathbb{1}$ we obtain

$$(\rho')^R \;=\; \Phi\rho^R \quad \text{so that} \quad \rho' \;=\; (D^R_\Phi \rho^R)^R. \qquad (11.24)$$

In the above formula the standard multiplication of square matrices takes place, in contrast to Eq. (10.36) in which the state ρ acts on a simple Hilbert space and is treated as a vector.

Note that Eq. (11.22) may be obtained as a special case of (11.24) if one takes for ρ the maximally entangled state (11.21), for which $(\rho^\phi)^R = \mathbb{1}$. Formula (11.24) provides a useful application of the dynamical matrix corresponding to a map Φ,

[10] This refers to the contribution of Jamiołkowski (1972). Various aspects of the duality between maps and states were recently investigated in (Havel, 2003; Arrighi and Patricot, 2004; Constantinescu and Ramakrishna, 2003).

[11] An analogous operation $\mathbb{1} \otimes \Phi$ acting on ρ^ϕ leads to the matrix D^S with the same spectrum.

which acts on a subsystem. Since the normalization of matrices does not influence positivity, this result implies the reshuffling lemma (11.5).

Formula (11.22) may also be used to find operators D associated with positive maps Φ which are neither trace preserving nor complete positive. The Jamiołkowski isomorphism thus relates the set of positive linear maps with dynamical matrices acting in the composite space and positive on product states. Let us mention explicitly some special cases of this isomorphism, labelled J_I in Table 11.1. The set of completely positive maps Φ is isomorphic to the set of all positive matrices D. The case J_{II} concerns quantum operations which correspond to quantum states $\rho = D/N$ fulfilling an additional constraint,[12] $\mathrm{Tr}_A D = \mathbb{1}$. States satisfying $\mathrm{Tr}_B D = \mathbb{1}$ correspond to unital CP maps.

An important case J_{III} of the isomorphism concerning the super-positive maps which for $N = 2$ are isomorphic with the PPT states (with *positive partial transpose*, $\rho^{T_A} \geq 0$) will be further analysed in Section 15.4, but we are now in position to comment on item E_{III}. If the map Φ is a unitary rotation, $\rho' = \Phi(\rho) = U\rho\, U^\dagger$ then (11.22) results in the pure state $(U \otimes \mathbb{1})|\phi^+\rangle$. The local unitary operation $(U \otimes \mathbb{1})$ preserves the purity of a state and its Schmidt coefficients. As shown in Section 15.2 the set of unitary matrices U of size N – or more precisely $SU(N)/\mathbb{Z}_N$ – is isomorphic to the set of maximally entangled pure states of the composite $N \times N$ system. In particular, vectors obtained by reshaping the Pauli matrices σ_i represent the Bell states in the computational basis, as listed in Table 11.1. Eventually, case J_{IV} consists of a single, distinguished point in both spaces: the completely depolarizing channel $\Phi_* : \mathcal{M}^{(N)} \to \rho_*^{(N)}$ and the corresponding maximally mixed state $\rho_*^{(N^2)}$.

Table 11.1 deserves one more comment: the key word *duality* may be used here in two different meanings. The 'vertical' duality between its both columns describes the isomorphism between maps and states, while the 'horizontal' duality between the rows J_I and J_{III} follows from the dual cones construction. Note the inclusion relations $J_I \supset J_{II} \supset J_{III} \supset J_{IV}$ and $J_{II} \supset E_{III}$, valid for both columns of the Table and visualized in Figure 11.8.

11.4 Quantum maps and quantum states

The relation (11.22) links an arbitrary linear map Φ with the corresponding linear operator given by the dynamical matrix D_Φ. Expressing the maximally entangled state $|\phi^+\rangle$ in (11.22) by its Schmidt form (11.21) we may compute the matrix elements of D_Φ in the product basis consisting of the states $|i \otimes j\rangle$. Due to the factorization of the right-hand side we see that the double sum describing $\rho_\Phi =$

[12] An apparent asymmetry between the role of both subsystems is due to the particular choice of the relation (11.22); if the operator $\mathbb{1} \otimes \Phi$ is used instead, the subsystems A and B need to be interchanged.

Table 11.1. *Jamiołkowski Isomorphism (11.22) between trace-preserving, linear maps Φ on the space of mixed states $\mathcal{M}^{(N)}$ and the normalized Hermitian operators D_Φ acting on the composite space $\mathcal{H}_N \otimes \mathcal{H}_N$.*

Isomorphism	Linear maps $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$	Hermitian operators $D_\Phi : \mathcal{H}_{N^2} \to \mathcal{H}_{N^2}$
J_I	set $\mathcal{P}$ of positive maps Φ	block positive operators D
J_{II}	set $\mathcal{CP}$ of completely positive Φ	positive operators D: subset $\mathcal{M}_{\mathbb{1}}$ of quantum states
J_{III}	set $\mathcal{SP}$ of super-positive Φ	subset of separable quantum states
E_{III}	unitary rotations $\Phi(\rho) = U\rho U^\dagger$, $D_\Phi = (U \otimes U^*)^R$	maximally entangled pure states $(U \otimes \mathbb{1})\|\phi^+\rangle$
$N = 2$ example of E_{III}	$\mathbb{1} \leftrightarrow (1, 0, 0, 1)$	$\|\phi^+\rangle \equiv \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$
Pauli matrices	$\sigma_x \leftrightarrow (0, 1, 1, 0)$	$\|\psi^+\rangle \equiv \frac{1}{\sqrt{2}}(\|01\rangle + \|10\rangle)$
versus	$\sigma_y \leftrightarrow (0, -i, i, 0)$	$\|\psi^-\rangle \equiv \frac{1}{\sqrt{2}}(\|01\rangle - \|10\rangle)$
Bell states $\rho_\phi = \|\phi\rangle\langle\phi\|$	$\sigma_z \leftrightarrow (1, 0, 0, -1)$	$\|\phi^-\rangle \equiv \frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$
J_{IV}	completely depolarizing channel Φ_*	maximally mixed state $\rho_* = \mathbb{1}/N$

D_Φ/N drops out and the result reads

$$\langle k \otimes i | D_\Phi | l \otimes j \rangle \;=\; \langle k \big| \Phi(|i\rangle\langle j|) \big| l \rangle. \tag{11.25}$$

This equation may also be understood as a definition of a map Φ related to the linear operator D_Φ. Its special case, $k = l$ and $i = j$, proves the isomorphism J_I from Table 11.1: if D_Φ is block positive, then the corresponding map Φ sends positive projection operators $|i\rangle\langle i|$ into positive operators (Jamiołkowski, 1972).

As listed in Table 11.1 and shown in Figure 11.8, the Jamiołkowski isomorphism (11.25) may be applied in various setups. Relating linear maps from $\mathcal{P}_N$ with operators acting on an extended space $\mathcal{H}_N \otimes \mathcal{H}_N$ we may compare:

- **(i)** individual objects, e.g. completely depolarizing channel Φ_* and the maximally mixed state ρ_*;
- **(ii)** families of objects, e.g. the depolarizing channels and generalized Werner states;
- **(iii)** entire sets, e.g. the set of $\mathcal{CP} \cap \mathcal{C}c\mathcal{P}$ maps and the set of PPT states;
- **iv)** entire problems, e.g. finding the SP map closest to a given CP map versus finding the separable state closest to a given state; and
- **(v)** their solutions. . . .

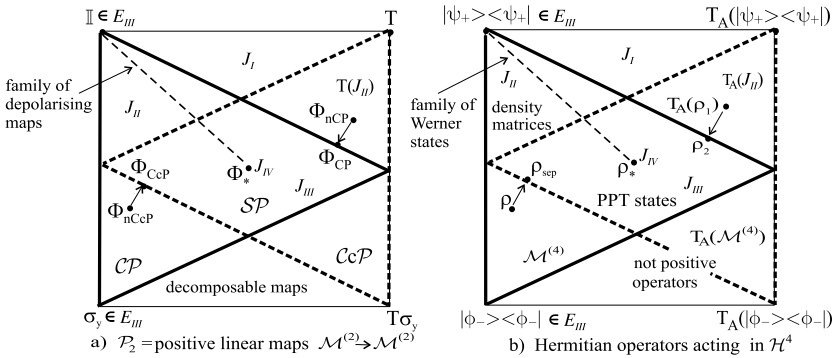


Figure 11.8. Isomorphism between objects, sets, and problems: (a) linear one-qubit maps, (b) linear operators acting in two-qubit Hilbert space $\mathcal{H}_4$. Labels J_i refer to the sets defined in Table 11.1.

For a more comprehensive discussion of the issues related to quantum entanglement see Chapter 15. Some general impression may be gained by comparing both sides of Figure 11.8, in which a drawing of both spaces is presented. Note that this illustration may also be considered as a strict representation of a fragment of the space of one-qubit unital maps (a) or the space of two-qubits density matrices in the HS geometry (b). It is nothing but the cross section of the cube representing the positive maps in Figure 11.3(c) along the plane determined by $\mathbb{1}$, T and Φ_*.

The maps–states duality is particularly fruitful in investigating *quantum gates*: unitary operations U performed on an N-level quantum system. Since the overall phase is not measurable, we may fix the determinant of U to unity, restricting our attention to the group $SU(N)$. For instance, the set of $SU(4)$ matrices may be considered as:

- the space of maximally entangled states of a composite, 4×4 system, $|\psi\rangle \in \mathbb{C}\mathbf{P}^{15} \subset \mathcal{M}^{(16)}$;
- the set of two-qubit unitary gates,[13] acting on $\mathcal{M}^{(4)}$;
- the set $\mathcal{BU}_2$ of one-qubit unistochastic operations (10.64), $\Psi_U \in \mathcal{BU}_2 \subset \mathcal{B}_2$.

There exist a classical analogue of the Jamiołkowski isomorphism. The space of all classical states forms the $(N-1)$-dimensional simplex Δ_{N-1}. A discrete dynamics in this space is given by a stochastic transition matrix $T_N : \Delta_{N-1} \to \Delta_{N-1}$. Its entries are non-negative, and due to stochasticity (2.4.ii), the sum of all its elements is equal to N. Hence the reshaped transition matrix is a vector $\vec{t}$ of

[13] As shown by DiVincenzo (1995) and Lloyd (1995) such gates are *universal* for quantum computing, which means that their suitable composition can produce an arbitrary unitary transformation. Such gates may be realized experimentally (Monroe, Meekhof, King, Itano and Wineland, 1995; DeMarco, Ben-Kish, Leibfried, Meyer, Rowe, Jelenkovic, Itano, Britton, Langer, Rosenband and Wineland, 2002).

length N^2. The rescaled vector $\vec{t}/N$ may be considered as a probability vector. The classical states defined in this way form a measure zero, $N(N-1)$-dimensional, convex subset of Δ_{N^2-1}. Consider, for instance, the set of $N = 2$ stochastic matrices, which can be parameterized as $T_2 = \begin{bmatrix} a & b \\ 1-a & 1-b \end{bmatrix}$ with $a, b \in [0, 1]$. The set of the corresponding probability vectors $\vec{t}/2 = (a, b, 1-a, 1-b)/2$ forms a square of size $1/2$ – the maximal square which may be inscribed into the unit tetrahedron Δ_3 of all $N = 4$ probability vectors.

Classical dynamics may be considered as a (very) special subclass of quantum dynamics, defined on the set of diagonal density matrices. Hence the classical and quantum duality between maps and states may be succinctly summarized in a commutative diagram:

$$\begin{array}{lccc} \text{quantum}: & \big[\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}\big] & \longrightarrow & \frac{1}{N} D_\Phi \in \mathcal{M}^{(N^2)} \\ \downarrow \Psi_{\mathrm{CG}} & \downarrow \ \text{maps} & & \downarrow \ \text{states} \\ \text{classical}: & \big[T : \Delta_{N-1} \to \Delta_{N-1}\big] & \longrightarrow & \frac{1}{N}\vec{t} \ \in \Delta_{N^2-1} \, . \end{array} \tag{11.26}$$

Alternatively, vertical arrows may be interpreted as the action of the coarse graining operation Ψ_{CG} defined in Eq. (12.77). For instance, for the trivial (do nothing) one-qubit quantum map $\Phi_{\mathbb{1}}$, the super-operator $\mathbb{1}_4$ restricted to diagonal matrices gives the identity matrix, $T = \mathbb{1}_2$, and the classical state $\vec{t}/2 = (1, 0, 0, 1)/2 \in \Delta_3$. But this very vector represents the diagonal of the maximally entangled state $\frac{1}{2} D_\Phi = |\phi^+\rangle\langle\phi^+|$. To prove commutativity of the diagram (11.26) in the general case define the stochastic matrix T as a submatrix of the superoperator (10.36), $T_{mn} = \Phi_{\substack{mm \\ nn}}$ (left vertical arrow). Note that the vector $\vec{t}$ obtained by its reshaping satisfies $\vec{t} = \mathrm{diag}(\Phi^R) = \mathrm{diag}(D_\Phi)$. Hence, as denoted by the right vertical arrow, it represents the diagonal of the dynamical matrix, which completes the reasoning.

Problems

Problem 11.1 Prove the reshuffling lemma (11.5).

Problem 11.2 (a) Find the Kraus spectrum of the (non-positive) dynamical matrix representing transposition acting in $\mathcal{M}^{(N)}$. (b) Show that the canonical Kraus representation of the transposition of one qubit is given by a *difference* between two CP maps,

$$\rho^T = (1+a)\Phi_{\mathrm{CP}}(\rho) - a\,\Phi'_{\mathrm{CP}}(\rho) = \frac{1}{2}\big(\sigma_0\rho\sigma_0 + \sigma_x\rho\sigma_x + \sigma_z\rho\sigma_z - \sigma_y\rho\sigma_y\big) \tag{11.27}$$

Problem 11.3 Show that the map $\Psi_r(\rho) \equiv (N\rho_* - \rho)/(N-1)$, acting on $\mathcal{M}^{(N)}$, is not completely positive. Is it positive or completely co-positive?

Problem 11.4 Show that $\Phi_T \equiv \frac{2}{3}\Phi_* + \frac{1}{3}T$ is the best structural physical approximation (SPA) of the non-CP map T of the transposition of a qubit (Horodecki and Ekert, 2002). How does such a SPA look like for the transposition of a quNit?

Problem 11.5 Compute complete positivity (complete co-positivity) of the generalized Choi map (11.8). Find the conditions for $\Psi_{a,b,c}$ to be CP (CcP).

Problem 11.6 Show that the minimal distance of a positive (but not CcP) map Φ_{nCcP} from the set $\mathcal{C}c\mathcal{P}_N$ is smaller than $Nx\sqrt{\text{Tr}(D^{T_A})^2 - 1}/(Nx+1)$, where D_Φ represents the dynamical matrix, and the positive number x is opposite to the negative, minimal eigenvalue of $D_{\Phi T} = D_\Phi^{T_A}$.

12

Density matrices and entropies

> A given object of study cannot always be assigned a unique value, its 'entropy'. It may have many different entropies, each one worthwhile.
>
> *Harold Grad*

In quantum mechanics, the von Neumann entropy

$$S(\rho) = -\mathrm{Tr}\rho \ln \rho \tag{12.1}$$

plays a role analogous to that played by the Shannon entropy in classical probability theory. They are both functionals of the state, they are both monotone under a relevant kind of mapping, and they can be singled out uniquely by natural requirements. In Section 2.2 we recounted the well-known anecdote according to which von Neumann helped to christen Shannon's entropy. Indeed von Neumann's entropy is older than Shannon's, and it reduces to the Shannon entropy for diagonal density matrices. But in general the von Neumann entropy is a subtler object than its classical counterpart. So is the quantum relative entropy, that depends on two density matrices that perhaps cannot be diagonalized at the same time. Quantum theory is a non-commutative probability theory. Nevertheless, as a rule of thumb we can pass between the classical discrete, classical continuous and quantum cases by choosing between sums, integrals and traces. While this rule of thumb has to be used cautiously, it will give us quantum counterparts of most of the concepts introduced in Chapter 2, and conversely we can recover Chapter 2 by restricting the matrices of this chapter to be diagonal.

12.1 Ordering operators

The study of quantum entropy is to a large extent a study in inequalities, and this is where we begin. We will be interested in extending inequalities that are valid for functions defined on $\mathbb{R}$ to functions of operators. This is a large step, but it is

at least straightforward to define *operator functions*, that is functions of matrices, as long as our matrices can be diagonalized by unitary transformations: then, if $A = U\mathrm{diag}(\lambda_i)U^\dagger$, we set $f(A) \equiv U\mathrm{diag}\big(f(\lambda_i)\big)U^\dagger$, where f is any function on $\mathbb{R}$. Our matrices will be Hermitian and therefore they admit a partial order; $B \geq A$ if and only $B - A$ is a positive operator. It is a difficult ordering relation to work with though, ultimately because it does not define a lattice – the set $\{X : X \geq A$ and $X \geq B\}$ has no minimum point in general.

With these observations in hand we can define an *operator monotone function* as a function such that

$$A \leq B \quad \Rightarrow \quad f(A) \leq f(B)\,. \tag{12.2}$$

Also, an *operator convex function* is a function such that

$$f(pA + (1-p)B) \leq pf(A) + (1-p)f(B)\,, \qquad p \in [0,1]\,. \tag{12.3}$$

Finally, an *operator concave function* f is a function such that $-f$ is operator convex. In all three cases it is assumed that the inequalities hold for all matrix sizes (so that an operator monotone function is always monotone in the ordinary sense, but the converse may fail).[1]

The definitions are simple, but we have entered deep water, and we will be submerged by difficulties as soon as we evaluate a function at two operators that do not commute with each other. Quite innocent looking monotone functions fail to be operator monotone. An example is $f(t) = t^2$. Moreover the function $f(t) = \mathrm{e}^t$ is neither operator monotone nor operator convex. To get serious results in this subject some advanced mathematics, including frequent excursions into the complex domain, are needed. We will confine ourselves to stating a few facts. Operator monotone functions are characterized by

Theorem 12.1 (Löwner's) *A function $f(t)$ on an open interval is operator monotone if and only if it can be extended analytically to the upper half plane and transforms the upper half plane into itself.*

Therefore the following functions are operator monotone:

$$\begin{aligned} &f(t) = t^\gamma\,, \quad t \geq 0 \quad \text{if and only if} \quad \gamma \in [0,1] \\ &f(t) = \tfrac{at+b}{ct+d}\,, \quad t \neq -d/c\,, \quad ad - bc > 0 \\ &f(t) = \ln t\,, \quad t > 0\,. \end{aligned} \tag{12.4}$$

This small supply can be enlarged by the observation that the composition of two operator monotone functions is again operator monotone; so is $f(t) = -1/g(t)$

[1] The theory of operator monotone functions was founded by Löwner (1934). An interesting early paper is by Bendat and Sherman (1955). For a survey see Bhatia (1997), and (for matrix means) see Ando (1994).

if $g(t)$ is operator monotone. The set of all operator monotone functions is convex, as a consequence of the fact that the set of positive operators is a convex cone.

A continuous function f mapping $[0, \infty)$ into itself is operator concave if and only if f is operator monotone. Operator convex functions include $f(t) = -\ln t$, and $f(t) = t \ln t$ when $t > 0$; we will use the latter function to construct entropies. More generally $f(t) = tg(t)$ is operator convex if $g(t)$ is operator monotone.

Finally we define the *mean* $A\#B$ of two operators. We require that $A\#A = A$, as well as homogeneity, $\alpha(A\#B) = (\alpha A)\#(\alpha B)$, and monotonicity, $A\#B \geq C\#D$ if $A \geq C$ and $B \geq D$. Moreover we require that $(TAT^\dagger)\#(TBT^\dagger) \geq T(A\#B)T^\dagger$, as well as a suitable continuity property. It turns out (Ando, 1994) that every mean obeying these demands takes the form

$$A \# B \;=\; \sqrt{A}\, f\!\left(\frac{1}{\sqrt{A}}\, B\, \frac{1}{\sqrt{A}}\right) \sqrt{A}\,, \tag{12.5}$$

where $A > 0$ and f is an operator monotone function on $[0, \infty)$ with $f(1) = 1$. The mean will be symmetric in A and B if and only if f is *self inversive*, that is if and only if

$$f(1/t) \;=\; f(t)/t\,. \tag{12.6}$$

Special cases of symmetric means include the *arithmetic mean* for $f(t) = (1 + t)/2$, the *geometric mean* for $f(t) = \sqrt{t}$, and the *harmonic mean* for $f(t) = 2t/(1 + t)$. It can be shown that the arithmetic mean is maximal among symmetric means, while the harmonic mean is minimal (Kubo and Ando, 1980).

We will find use for these results throughout the next three chapters. But to begin with we will get by with inequalites that apply, not to functions of operators directly but to their traces. The subject of convex trace functions is somewhat more manageable than that of operator convex functions. A key result is that the inequality (1.11) for convex functions can be carried over in this way:[2]

Klein's inequality. *If f is a convex function and A and B are Hermitian operators, then*

$$\mathrm{Tr}[f(A) - f(B)] \;\geq\; \mathrm{Tr}[(A - B)f'(B)]\,. \tag{12.7}$$

As a special case

$$\mathrm{Tr}(A \ln A - A \ln B) \;\geq\; \mathrm{Tr}(A - B) \tag{12.8}$$

with equality if and only if $A = B$.

[2] The original statement here is due to Oskar Klein (1931).

To prove this, use the eigenbases:

$$A|e_i\rangle = a_i|e_i\rangle \qquad B|f_i\rangle = b_i|f_i\rangle \qquad \langle e_i|f_j\rangle = c_{ij}\ . \tag{12.9}$$

A calculation then shows that

$$\begin{aligned}\langle e_i|f(A) - f(B) - (A-B)f'(B)|e_i\rangle \\ &= f(a_i) - \sum_j |c_{ij}|^2[f(b_j) - (a_i - b_j)f'(b_j)] \\ &= \sum_j |c_{ij}|^2[f(a_i) - f(b_j) - (a_i - b_j)f'(b_j)]\ .\end{aligned} \tag{12.10}$$

This is positive by Eq. (1.11). The special case follows if we specialize to $f(t) = t\ln t$. The condition for equality requires some extra attention – it is true.

Another useful result is:

Peierl's inequality. *If f is a strictly convex function and A is a Hermitian operator, then*

$$\mathrm{Tr} f(A) \ \geq\ \sum_i f(\langle f_i|A|f_i\rangle)\ , \tag{12.11}$$

where $\{|f_i\rangle\}$ is any complete set of orthonormal vectors, or more generally a resolution of the identity. Equality holds if and only if $|f_i\rangle = |e_i\rangle$ for all i, where $A|e_i\rangle = a_i|e_i\rangle$.

To prove this, observe that for any vector $|f_i\rangle$ we have

$$\langle f_i|A|f_i\rangle = \sum_j |\langle f_i|e_j\rangle|^2 f(a_j) \geq f\left(\sum_j |\langle f_i|e_j\rangle|^2 a_j\right) = f\big(\langle f_i|A|f_i\rangle\big)\ . \tag{12.12}$$

Summing over all i gives the result.

We quote without proofs two further trace inequalities, the *Golden Thompson inequality*

$$\mathrm{Tr} e^A e^B \ \geq\ \mathrm{Tr} e^{A+B}\ , \tag{12.13}$$

with equality if and only if the Hermitian matrices A and B commute, and its more advanced cousin, the *Lieb inequality*

$$\mathrm{Tr} e^{\ln A - \ln C + \ln B} \ \geq\ \mathrm{Tr}\int_0^\infty A\,\frac{1}{C+u\mathbb{1}}\,B\,\frac{1}{C+u\mathbb{1}}\,\mathrm{d}u\ , \tag{12.14}$$

where A, B, C are all positive.[3]

[3] Golden was a person (Golden, 1965). The Lieb inequality was proved in Lieb (1973).

12.2 Von Neumann entropy

Now we can begin. First we establish some notation. In Chapter 2 we used S to denote the Shannon entropy $S(\vec{p})$ of a probability distribution. Now we use S to denote the von Neumann entropy $S(\rho)$ of a density matrix, but we may want to mention the Shannon entropy too. When there is any risk of confusing these entropies, they are distinguished by their arguments. We will also use $S_i \equiv S(\rho_i)$ to denote the von Neumann entropy of a density matrix ρ_i acting on the Hilbert space $\mathcal{H}_i$.

In classical probability theory a state is a probability distribution, and the Shannon entropy is a distinguished function of a probability distribution. In quantum theory a state is a density matrix, and a given density matrix ρ can be associated to many probability distributions because there are many possible POVMs. Also any density matrix can arise as a mixture of pure states in many different ways. From Section 8.4 we recall that if we write our density matrix as a mixture of normalized states,

$$\rho \;=\; \sum_{i=1}^{M} p_i |\psi_i\rangle\langle\psi_i| \;, \tag{12.15}$$

then a large amount of arbitrariness is present, even in the choice of the number M. So if we define the *mixing entropy* of ρ as the Shannon entropy of the probability distribution $\vec{p}$ then this definition inherits a large amount of arbitrariness. But on reflection it is clear that there is one such definition that is more natural than the other. The point is that the density matrix itself singles out one preferred mixture, namely

$$\rho \;=\; \sum_{i=1}^{N} \lambda_i |e_i\rangle\langle e_i| \;, \tag{12.16}$$

where $|e_i\rangle$ are the eigenvectors of ρ and N is the rank of ρ. The *von Neumann entropy* is[4]

$$S(\rho) \;\equiv\; -\mathrm{Tr}\rho\ln\rho \;=\; -\sum_{i=1}^{N} \lambda_i \ln\lambda_i \;. \tag{12.17}$$

Hence the von Neumann entropy is the Shannon entropy of the spectrum of ρ, and varies from zero for pure states to $\ln N$ for the maximally mixed state $\rho_* = \mathbb{1}/N$.

Further reflection shows that the von Neumann entropy has a very distinguished status among the various mixing entropies. While we were proving Schrödinger's

[4] The original reference is von Neumann (1927), whose main concern at the time was with statistical mechanics. His book (von Neumann, 1955), Wehrl (1978) and the (more advanced) book by Ohya and Petz (1993) serve as useful general references for this chapter.

mixture theorem in Section 8.4 we observed that any vector $\vec{p}$ occurring in Eq. (12.15) is related to the spectral vector $\vec{\lambda}$ by $\vec{p} = B\vec{\lambda}$, where B is a bistochastic (and indeed a unistochastic) matrix. Since the Shannon entropy is a Schur concave function, we deduce from the discussion in Section 2.1 that

$$S_{\rm mix} \equiv -\sum_{i=1}^{M} p_i \ln p_i \geq -\sum_{i=1}^{N} \lambda_i \ln \lambda_i = S(\rho) . \tag{12.18}$$

Hence the von Neumann entropy is the smallest possible among all the mixing entropies $S_{\rm mix}$.

The von Neumann entropy is a continuous function of the eigenvalues of ρ, and it can be defined in an axiomatic way as the only such function that satisfies a suitable set of axioms. In the classical case, the key axiom that singles out the Shannon entropy is the recursion property. In the quantum case this becomes a property that concerns *disjoint* states – two density matrices are said to be disjoint if they have orthogonal support, that is if their respective eigenvectors span orthogonal subspaces of the total Hilbert space.

• **Recursion property**. If the density matrices ρ_i have support in orthogonal subspaces $\mathcal{H}_i$ of a Hilbert space $\mathcal{H} = \oplus_{i=1}^{M} \mathcal{H}_i$, then the density matrix $\rho = \sum_i p_i \rho_i$ has the von Neumann entropy

$$S(\rho) = S(\vec{p}) + \sum_{i=1}^{M} p_i S(\rho_i) . \tag{12.19}$$

Here $S(\vec{p})$ is a classical Shannon entropy. It is not hard to see that the von Neumann entropy has this property; if the matrix ρ_i has eigenvalues λ_{ij} then the eigenvalues of ρ are $p_i \lambda_{ij}$, and the result follows from the recursion property of the Shannon entropy (Section 2.2).

As with the Shannon entropy, the von Neumann entropy is interesting because of the list of the properties that it has, and the theorems that can be proved using this list. So, instead of presenting a list of axioms we present a selection of these properties in the form of Table 12.1, where we also compare the von Neumann entropy to the Shannon and Boltzmann entropies. Note that most of the entries concern a situation where ρ_1 is defined on a Hilbert space $\mathcal{H}_1$, ρ_2 on another Hilbert space $\mathcal{H}_2$, and ρ_{12} on their tensor product $\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$, or even more involved situations involving the tensor product of three Hilbert spaces. Moreover ρ_1 is always the reduced density matrix obtained by taking a partial trace of ρ_{12}, thus

$$S_1 \equiv S(\rho_1) , \qquad S_{12} \equiv S(\rho_{12}) , \qquad \rho_1 \equiv {\rm Tr}_2 \rho_{12} , \tag{12.20}$$

and so on (with the obvious modifications in the classical cases). As we will see, even relations that involve one Hilbert space only are conveniently proved through

Table 12.1. *Properties of entropies*

Property	Equation	von Neumann	Shannon	Boltzmann
Positivity	$S \geq 0$	Yes	Yes	No
Concavity	(12.22)	Yes	Yes	Yes
Monotonicity	$S_{12} \geq S_1$	No	Yes	No
Subadditivity	$S_{12} \leq S_1 + S_2$	Yes	Yes	Yes
Araki–Lieb inequality	$\|S_1 - S_2\| \leq S_{12}$	Yes	Yes	No
Strong subadditivity	$S_{123} + S_2 \leq S_{12} + S_{23}$	Yes	Yes	Yes

a detour into a larger Hilbert space. We can say more. In Section 9.3 we proved a purification lemma, saying that the ancilla can always be chosen so that, for any ρ_1, it is true that $\rho_1 = \mathrm{Tr}_2 \rho_{12}$ where ρ_{12} is a pure state. Moreover we proved that in this situation the reduced density matrices ρ_1 and ρ_2 have the same spectra (up to vanishing eigenvalues). This means that

$$\rho_{12}\rho_{12} = \rho_{12} \quad \Rightarrow \quad S_1 = S_2 \,. \tag{12.21}$$

If the ancilla purifies the system, the entropy of the ancilla is equal to the entropy of the original system.

Let us begin by taking note of the property (monotonicity) that the von Neumann entropy does not have. As we know very well from Chapter 9 a composite system can be in a pure state, so that $S_{12} = 0$, while its subsystems are mixed, so that $S_1 > 0$. In principle, although it might be a very contrived Universe, your own entropy can increase without limit while the entropy of the world remains zero.

It is clear that the von Neumann entropy is positive. To convince ourselves of the truth of the rest of the entries in the table[5] we must rely on Section 12.1. Concavity and subadditivity are direct consequences of Klein's inequality, for the special case that A and B are density matrices, so that the right-hand side of Eq. (12.8) vanishes. First out is concavity, where all the density matrices live in the same Hilbert space:

• **Concavity**. If $\rho = p\sigma + (1-p)\omega$, $0 \leq p \leq 1$, then

$$S(\rho) \;\geq\; pS(\sigma) + (1-p)S(\omega) \,. \tag{12.22}$$

In the proof we use Klein's inequality, with $A = \sigma$ or ω and $B = \rho$:

$$\mathrm{Tr}\rho \ln \rho = p\mathrm{Tr}\sigma \ln \rho + (1-p)\mathrm{Tr}\omega \ln \rho \leq p\mathrm{Tr}\sigma \ln \sigma + (1-p)\mathrm{Tr}\omega \ln \omega \,. \tag{12.23}$$

[5] These entries have a long history. Concavity and subadditivity were first proved by Delbrück and Molière (1936). Lanford and Robinson (1968) observed that strong subadditivity is used in classical statistical mechanics, and conjectured that it holds in the quantum case as well. Araki and Lieb (1970) were unable to prove this, but found other inequalities that were enough to complete the work of Lanford and Robinson. Eventually strong subadditivity was proved by Lieb and Ruskai (1973).

Sign reversion gives the result, which is a lower bound on $S(\rho)$.

Using Peierl's inequality we can prove a much stronger result. Let f be any convex function. With $0 \le p \le 1$ and A and B any Hermitian operators, it will be true that

$$\mathrm{Tr} f\big(pA + (1-p)B\big) \;\le\; p\mathrm{Tr} f(A) + (1-p)\mathrm{Tr} B\ . \tag{12.24}$$

Namely, let $|e_i\rangle$ be the eigenvectors of $pA + (1-p)B$. Then

$$\begin{aligned}
\mathrm{Tr} f\big(pA + (1-p)B\big) &= \sum_i \langle e_i | f\big(pA + (1-p)B\big) | e_i\rangle \\
&= \sum_i f\big(\langle e_i | pA + (1-p)B | e_i\rangle\big) \\
&\le p \sum_i f\big(\langle e_i | A | e_i\rangle\big) + (1-p) \sum_i f\big(\langle e_i | B | e_i\rangle\big) \\
&\le p\mathrm{Tr} f(A) + (1-p)\mathrm{Tr} f(B)
\end{aligned} \tag{12.25}$$

where Peierl's inequality was used in the last step.

The recursion property (12.19) for disjoint states can be turned into an inequality that, together with concavity, neatly brackets the von Neumann entropy:

$$\rho = \sum_{a=1}^{K} p_a \rho_a \quad \Rightarrow \quad \sum_{a=1}^{K} p_a S(\rho_a) \le S(\rho) \le S(\vec{p}) + \sum_{a=1}^{K} p_a S(\rho_a)\ . \tag{12.26}$$

(The index a is used because, in this chapter, i labels different Hilbert spaces.) To prove this, one first observes that for a pair of positive operators A, B one has the trace inequality

$$\mathrm{Tr} A[\ln(A+B) - \ln A] \ge 0\ . \tag{12.27}$$

This is true because $\ln t$ is operator monotone, and the trace of the product of two positive operators is positive. When $K = 2$ the upper bound in (12.26) follows if we first set $A = p_1\rho_1$ and $B = p_2\rho_2$, then $A = p_2\rho_2$, $B = p_1\rho_1$, add the resulting inequalities, and reverse the sign at the end. The result for arbitrary K follows if we use the recursion property (2.19) for the Shannon entropy $S(\vec{p})$.

The remaining entries in Table 12.1 concern density matrices defined on different Hilbert spaces, and the label on the density matrix tells us which Hilbert space.

• **Subadditivity**.

$$S(\rho_{12}) \;\le\; S(\rho_1) + S(\rho_2)\ , \tag{12.28}$$

with equality if and only if $\rho_{12} = \rho_1 \otimes \rho_2$.

To prove this, use Klein's inequality with $B = \rho_1 \otimes \rho_2 = (\rho_1 \otimes \mathbb{1})(\mathbb{1} \otimes \rho_2)$ and

$A = \rho_{12}$. Then

$$\begin{aligned} \mathrm{Tr}_{12}\rho_{12}\ln\rho_{12} &\geq \mathrm{Tr}_{12}\rho_{12}\ln\rho_1\otimes\rho_2 \\ &= \mathrm{Tr}_{12}\rho_{12}(\ln\rho_1\otimes\mathbb{1} + \ln\mathbb{1}\otimes\rho_2) \\ &= \mathrm{Tr}_1\rho_1\ln\rho_1 + \mathrm{Tr}_2\rho_2\ln\rho_2\ , \end{aligned} \tag{12.29}$$

which becomes subadditivity when we reverse the sign. It is not hard to see that equality holds if and only if $\rho_{12} = \rho_1 \otimes \rho_2$.

We can now give a third proof of concavity, since it is in fact a consequence of subadditivity. The trick is to use a two level system, with orthogonal basis vectors $|a\rangle$ and $|b\rangle$, as an ancilla. The original density matrix will be written as the mixture $\rho_1 = p\rho_a + (1-p)\rho_b$. Then we define

$$\rho_{12} = p\rho_a \otimes |a\rangle\langle a| + (1-p)\rho_b \otimes |b\rangle\langle b|\ . \tag{12.30}$$

By the recursion property

$$S_{12}(\rho_{12}) = S(p, 1-p) + pS_1(\rho_a) + (1-p)S_1(\rho_b)\ . \tag{12.31}$$

But $S_2 = S(p, 1-p)$, so that subadditivity implies that S_1 is concave, as advertized.

Next on the list:

• **The Araki–Lieb triangle inequality**.

$$|S(\rho_1) - S(\rho_2)| \ \leq\ S(\rho_{12})\ . \tag{12.32}$$

This becomes a triangle inequality when combined with subadditivity.

The proof is a clever application of the fact that if a bipartite system is in a pure state (with zero von Neumann entropy) then the von Neumann entropies of the factors are equal. Of course the inequality itself quantifies how much the entropies of the factors can differ if this is not the case. But we can consider a purification of the state ρ_{12} using a Hilbert space $\mathcal{H}_{123}$. From subadditivity we know that $S_3 + S_1 \geq S_{13}$. By construction $S_{123} = 0$, so that $S_{13} = S_2$ and $S_3 = S_{12}$. A little rearrangement gives the result.

The final entry on our list is:

• **Strong subadditivity**.

$$S(\rho_{123}) + S(\rho_2) \ \leq\ S(\rho_{12}) + S(\rho_{23})\ . \tag{12.33}$$

This is a deep result, and we will not prove it – although it follows fairly easily from Lieb's inequality (12.14).[6] Let us investigate what it says, however. First, it is

[6] For a proof – in fact several proofs – and more information on entropy inequalities generally, we recommend two reviews written by experts, Lieb (1975) and Ruskai (2002).

equivalent to the inequality

$$S(\rho_1) + S(\rho_2) \;\leq\; S(\rho_{13}) + S(\rho_{23})\,. \tag{12.34}$$

To see this, purify the state ρ_{123} by factoring with a fourth Hilbert space. Then we have

$$S_{1234} = 0 \quad \Rightarrow \quad S_{123} = S_4 \quad \text{and} \quad S_{12} = S_{34}\,. \tag{12.35}$$

Inserting this in (12.33) yields (12.34), and conversely. This shows that strong subadditivity of the Shannon entropy is a rather trivial thing, since in that case monotonicity implies that $S_1 \leq S_{13}$ and $S_2 \leq S_{23}$. In the quantum case these inequalities do not hold separately, but their sum does!

The second observation is that strong subadditivity implies subadditivity – to see this, let the Hilbert space $\mathcal{H}_2$ be one dimensional, so that $S_2 = 0$. It implies much more, though. It is tempting to say that every deep result follows from it; we will meet with an example in the next section. Meanwhile we can ask if this is the end of the story? Suppose we have a state acting on the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$. Taking partial traces in all possible ways we get a set of $2^n - 1$ non-trivial density matrices, and hence $2^n - 1$ possible entropies constrained by the inequalities in Table 12.1. These inequalities define a convex cone in an $(2^n - 1)$-dimensional space, and we ask if the possible entropies fill out this cone. The answer is no. There are points on the boundary of the cone that cannot be reached in this way, and there may be further inequalities waiting to be discovered (Linden and Winter, n.d.).

To end on a somewhat different note, we recall that the operational significance of the Shannon entropy was made crystal clear by Shannon's noiseless coding theorem. There is a corresponding quantum noiseless coding theorem. To state it, we imagine that Alice has a string of pure states $|\psi_i\rangle$, generated with the probabilities p_i. She codes her states in qubit states, using a *channel system* C. The qubits are sent to Bob, who decodes them and produces a string of output states ρ_i. The question is: how many qubits must be sent over the channel if the message is to go through undistorted? More precisely, we want to know the *average fidelity*

$$\bar{F} \;=\; \sum_i p_i \,\langle\psi_i|\rho_i|\psi_i\rangle \tag{12.36}$$

that can be achieved. The quantum problem is made more subtle by the fact that generic pure states cannot be distinguished with certainty, but the answer is given by

Theorem 12.2 (Schumacher's noiseless coding theorem) *Let*

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \quad \textit{and} \quad S(\rho) = -\mathrm{Tr}\rho\log_2\rho\,. \tag{12.37}$$

Also let $\epsilon, \delta > 0$ and let $S(\rho) + \delta$ qubits be available in the channel per input state. Then for large $\mathcal{N}$, it is possible to transmit blocks of $\mathcal{N}$ states with average fidelity $\bar{F} > 1 - \epsilon$.

This theorem marks the beginning of quantum information theory.[7]

12.3 Quantum relative entropy

In the classical case the relative entropy of two probability distributions played a key role, notably as a measure of how different two probability distributions are from each other. There is a quantum relative entropy too, and for roughly similar reasons it plays a key role in the description of the quantum state space. In some ways it is a deeper concept than the von Neumann entropy itself and we will see several uses of it as we proceed. The definition looks deceptively simple: for any pair of quantum states ρ and σ their *relative entropy* is[8]

$$S(\rho||\sigma) \equiv \mathrm{Tr}[\rho(\ln \rho - \ln \sigma)] \,. \tag{12.38}$$

If σ has zero eigenvalues this may diverge, otherwise it is is a finite and continuous function. The quantum relative entropy reduces to the classical Kullback–Leibler relative entropy for diagonal matrices, but is not as easy to handle in general. Using the result of Problem 12.1, it can be rewritten as

$$S(\rho||\sigma) = \int_0^\infty \mathrm{Tr}\rho \, \frac{1}{\sigma + u\mathbb{1}}(\rho - \sigma)\frac{1}{\rho + u\mathbb{1}} \, \mathrm{d}u \,. \tag{12.39}$$

This is convenient for some manipulations that one may want to perform.

Two of the properties of relative entropy are immediate:

- **Unitary invariance**. $S(\rho_1||\rho_2) = S(U\rho_1 U^\dagger||U\rho_2 U^\dagger)$.
- **Positivity**. $S(\rho||\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.

The second property is immediate only because it is precisely the content of Klein's inequality – and we proved that in Section 12.1. More is true:

$$S(\rho||\sigma) \geq \frac{1}{2}\mathrm{Tr}(\rho - \sigma)^2 = D_2^2(\rho, \sigma) \,. \tag{12.40}$$

[7] The quantum noiseless coding theorem is due to Schumacher (1995) and Jozsa and Schumacher (1994); a forerunner is due to Holevo (1973). For Shannon's theorem formulated in the same language, see section 3.2 of Cover and Thomas (1991).

[8] The relative entropy was introduced into quantum mechanics by Umegaki (1962) and resurfaced in a paper by Lindblad (1973). A general reference for this section is Ohya and Petz (1993); for recent reviews see Schumacher and Westmoreland (n.d.) and Vedral (2002).

This is as in the classical case, Eq. (2.30); in both cases a stronger statement can be made, and we will come to it in Chapter 13. In general $S(\rho||\sigma) \neq S(\sigma||\rho)$; also as in the classical case.

Three deep properties of relative entropy are as follows:

- **Joint convexity**. For any $p \in [0, 1]$ and any four states

$$S\big(p\rho_a + (1-p)\rho_b||p\sigma_c + (1-p)\sigma_d\big) \leq pS(\rho_a||\sigma_c) + (1-p)S(\rho_b||\sigma_d)\,. \quad (12.41)$$

- **Monotonicity under partial trace**.

$$S(\mathrm{Tr}_2\rho_{12}||\mathrm{Tr}_2\sigma_{12}) \;\leq\; S(\rho_{12}||\sigma_{12})\,. \quad (12.42)$$

- **Monotonicity under CP-maps**. For any completely positive map Φ

$$S\big(\Phi\rho||\Phi\sigma\big) \;\leq\; S\big(\rho||\sigma\big)\,. \quad (12.43)$$

Any of these properties imply the other two, and each is equivalent to strong subadditivity of the von Neumann entropy.[9] The importance of monotonicity is obvious – it implies everything that monotonicity under stochastic maps implies for the classical Kullback–Leibler relative entropy.

It is clear that monotonicity under CP maps implies monotonicity under partial trace – taking a partial trace is a special case of a CP map. To see the converse, use the environmental representation of a CP map given in Eq. (10.61); we can always find a larger Hilbert space in which the CP-map is represented as

$$\rho' = \Phi(\rho) = \mathrm{Tr}_2(U\rho \otimes P_\nu U^\dagger)\,, \quad (12.44)$$

where P_ν is a projector onto a pure state of the ancilla. A simple calculation ensures:

$$\begin{aligned} S\left(\mathrm{Tr}_2(U\rho \otimes P_\nu U^\dagger)||\mathrm{Tr}_2(U\sigma \otimes P_\nu U^\dagger)\right) &\leq S\left(U\rho \otimes P_\nu U^\dagger||U\sigma \otimes P_\nu U^\dagger\right) \\ &= S(\rho \otimes P_\nu||\sigma \otimes P_\nu) \\ &= S(\rho||\sigma)\,, \end{aligned} \quad (12.45)$$

where we used monotonicity under partial trace, unitary invariance, and the easily proved additivity property that

$$S(\rho_1 \otimes \rho_2||\sigma_1 \otimes \sigma_2) = S(\rho_1||\rho_2) + S(\rho_2||\sigma_2)\,. \quad (12.46)$$

To see that monotonicity under partial trace implies strong subadditivity, we introduce a third Hilbert space and consider

$$S(\rho_{23}||\rho_2 \otimes \mathbb{1}) \;\leq\; S(\rho_{123}||\rho_{12} \otimes \mathbb{1})\,. \quad (12.47)$$

[9] Again the history is intricate. Monotonicity of relative entropy was proved from strong subadditivity by Lindblad (1975). A proof from first principles is due to Uhlmann (1977).

Now we just apply the definition of relative entropy, and rearrange terms to arrive at Eq. (12.33). The converse statement, that strong subadditivity implies monotonicity under partial trace, is true as well. One proof proceeds via the Lieb inequality (12.14).

The close link between the relative entropy and the von Neumann entropy can be unearthed as follows: the relative entropy between ρ and the maximally mixed state ρ_* is

$$S(\rho||\rho_*) \;=\; \ln N - S(\rho)\,. \tag{12.48}$$

This is the quantum analogue of (2.37), and shows that the von Neumann entropy $S(\rho)$ is implicit in the definition of the relative entropy. In a sense the link goes the other way too. Form the one parameter family of states

$$\rho_p = p\rho + (1-p)\sigma\,, \qquad p \in [0,1]\,. \tag{12.49}$$

Then define the function

$$f(p) \equiv S(\rho_p) - pS(\rho) - (1-p)S(\sigma)\,. \tag{12.50}$$

With elementary manipulations this can be rewritten as

$$f(p) = pS(\rho||\rho_p) + (1-p)S(\sigma||\rho_p) = pS(\rho||\sigma) - S(\rho_p||\sigma)\,. \tag{12.51}$$

From the strict concavity of the von Neumann entropy we conclude that $f(p) \geq 0$, with equality if and only if $p = 0, 1$. This further implies that the derivative of f is positive at $p = 0$. We are now in position to prove that

$$\lim_{p\to 0} \frac{1}{p} f(p) \;=\; S(\rho||\sigma)\,. \tag{12.52}$$

This is so because the limit exists (Lindblad, 1973) and because Eqs. (12.51) imply, first, that the limit is greater than or equal to $S(\rho||\sigma)$, and, second, that it is smaller than or equal to $S(\rho||\sigma)$. In this sense the definition of relative entropy is implicit in the definition of the von Neumann entropy. If we recall Eq. (1.11) for convex functions – and reverse the sign because $f(p)$ is concave – we can also express the conclusion as

$$S(\rho||\sigma) = \sup_p \frac{1}{p}\Big(S(\rho_p) - pS(\rho) - (1-p)S(\sigma)\Big)\,. \tag{12.53}$$

The same argument applies in the classical case, and in Section 13.1 we will see that the symmetric special case $f(1/2)$ deserves attention for its own sake.

For $N = 2$, Cortese (n.d.) found an explicit formula for the relative entropy between any two mixed states,

$$S(\rho_a||\rho_b) = \frac{1}{2}\ln\Big(\frac{1-\tau_a^2}{1-\tau_b^2}\Big) + \frac{\tau_a}{2}\ln\Big(\frac{1+\tau_a}{1-\tau_a}\Big) - \frac{\tau_a\cos\theta}{2}\ln\Big(\frac{1+\tau_b}{1-\tau_b}\Big)\,, \tag{12.54}$$

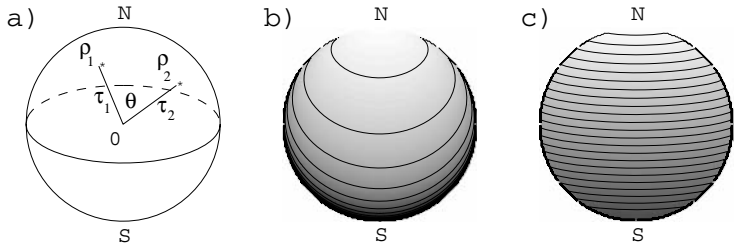


Figure 12.1. (a) Relative entropy between $N = 2$ mixed states depends on the lengths of their Bloch vectors and the angle θ between them. Relative entropies with respect to the north pole ρ_N: (b) $S(\rho||\rho_N)$ and (c) $S(\rho_N||\rho)$.

where the states are represented by their Bloch vectors, for example $\rho_a = \frac{1}{2}(\mathbb{1} + \vec{\tau}_a \cdot \vec{\sigma})$, τ_a is the length of a Bloch vector, and θ is the angle between the two. See Figure 12.1; the relative entropy with respect to a pure state is shown there. Note that the data along the polar axis, representing diagonal matrices, coincide with these plotted at the vertical axis in Figure 2.8(c) and (d) for the classical case.

Now we would like to argue that the relative entropy, as defined above, is indeed a quantum analogue of the classical Kullback–Leibler relative entropy. We could have tried a different way of defining quantum relative entropy, starting from the observation that there are many probability distributions associated to every density matrix, in fact one for every POVM $\{E\}$. Since we expect relative entropy to serve as an information divergence, that is to say that it should express 'how far' from each other two states are in the sense of statistical hypothesis testing, this suggests that we should define a relative entropy by taking the supremum over all possible POVMs:

$$S_1(\rho||\sigma) = \sup_E \sum_i p_i \ln \frac{p_i}{q_i}\,, \quad \text{where} \quad p_i = \mathrm{Tr} E_i \rho \ \text{ and } \ q_i = \mathrm{Tr} E_i \sigma\,. \tag{12.55}$$

Now it can be shown (Lindblad, 1974) (using monotonicity of relative entropy) that

$$S_1(\rho||\sigma) \ \le \ S(\rho||\sigma)\,. \tag{12.56}$$

We can go on to assume that we have several independent and identically distributed systems that we can make observations on, that is to say that we can make measurements on states of the form $\rho^{\mathcal{N}} \equiv \rho \otimes \rho \otimes \cdots \otimes \rho$ (with $\mathcal{N}$ identical factors altogether, and with a similar definition for $\sigma^{\mathcal{N}}$). We optimize over all POVMs $\{\tilde{E}\}$ on the tensor product Hilbert space, and define

$$S_{\mathcal{N}}(\rho||\sigma) = \sup_{\tilde{E}} \frac{1}{\mathcal{N}} \sum_i p_i \ln \frac{p_i}{q_i}\,, \qquad p_i = \mathrm{Tr}\, \tilde{E}_i \rho^{\mathcal{N}}\,, \quad q_i = \mathrm{Tr}\, \tilde{E}_i \sigma^{\mathcal{N}}\,. \tag{12.57}$$

In the large Hilbert space we have many more options for making collective measurements, so this ought to be larger than $S_1(\rho||\sigma)$. Nevertheless we have the bound (Donald, 1987)

$$S_{\mathcal{N}}(\rho||\sigma) \;\leq\; S(\rho||\sigma)\,. \tag{12.58}$$

Even more is true. In the limit when the number of copies of our states go to infinity, it turns out that

$$\lim_{\mathcal{N}\to\infty} S_{\mathcal{N}} \;=\; S(\rho||\sigma)\,. \tag{12.59}$$

This limit can be achieved by projective measurements. We do not intend to prove these results here, we only quote them in order to inspire some confidence in Umegaki's definition of quantum relative entropy.[10]

12.4 Other entropies

In the classical case we presented a wide selection of alternative definitions of entropy, some of which are quite useful. When we apply our rule of thumb – turn sums into traces – to Section 2.7, we obtain (among others) the *quantum Rényi entropy*, labelled by a non-negative parameter q,

$$S_q(\rho) \;\equiv\; \frac{1}{1-q}\ln[\mathrm{Tr}\rho^q] \;=\; \frac{1}{1-q}\ln\Big[\sum_{i=1}^{N}\lambda_i^q\Big]\,. \tag{12.60}$$

It is a function of the L_q-norm of the density matrix, $||\rho||_q = \left(\frac{1}{2}\mathrm{Tr}\rho^q\right)^{1/q}$. It is non-negative and, in the limit $q \to 1$, it tends to the von Neumann entropy $S(\rho)$. The logarithm is used in the definition to ensure additivity for product states:

$$S_q(\rho_1 \otimes \rho_2) \;=\; S_q(\rho_1) + S_q(\rho_2) \tag{12.61}$$

for any real q. This is immediate, given the spectrum of a product state (see Problem 9.4). The quantum Rényi entropies fulfil properties already discussed for their classical versions. In particular, for any value of the coefficient q the generalized entropy S_q equals zero for pure states, and achieves its maximum $\ln N$ for the maximally mixed state ρ_*. In analogy to (2.80), the Rényi entropy is a continuous, non-increasing function of its parameter q.

Some special cases of S_q are often encountered. The quantity $\mathrm{Tr}\rho^2$, called the *purity* of the quantum state, is frequently used since it is easy to compute. The larger the purity, the more pure the state (or more precisely, the larger is its Hilbert–Schmidt

[10] The final result here is due to Hiai and Petz (1991). And the reader should be warned that our treatment is somewhat simplified.

distance from the maximally mixed state). Obviously one has $S_2(\rho) = -\ln[\mathrm{Tr}\rho^2]$. The *Hartley entropy* S_0 is a function of the rank r of ρ; $S_0(\rho) = \ln r$. In the other limiting case the entropy depends on the largest eigenvalue of ρ; $S_\infty = -\ln \lambda_{\max}$. For any positive, finite value of q the generalized entropy is a continuous function of the state ρ. The Hartley entropy is not continuous at all. The concavity relation (12.22) holds at least for $q \in (0, 1]$, and the quantum Rényi entropies for different values of q are correlated in the same way as their classical counterparts (see Section 2.7). They are additive for product states, but not subadditive. A weak version of subadditivity holds (van Dam and Hayden, n.d.):

$$S_q(\rho_1) - S_0(\rho_2) \;\le\; S_q(\rho_{12}) \;\le\; S_q(\rho_1) + S_0(\rho_2)\,, \tag{12.62}$$

where S_0 denotes the Hartley entropy – the largest of the Rényi entropies.

The entropies considered so far have been unitarily invariant, and they take the value zero for any pure state. This is not always an advantage. An interesting alternative is the Wehrl entropy, that is the classical Boltzmann entropy of the Husimi function $Q(z) = \langle z|\rho|z\rangle$. It is not unitarily invariant because it depends on the choice of a special set of coherent states $|z\rangle$ (see Sections 6.2 and 7.4). The Wehrl entropy is important in situations where this set is physically singled out, say as 'classical states'. A key property is (Wehrl, 1979):

Wehrl's inequality. *For any state ρ the Wehrl entropy is bounded from below by the von Neumann entropy,*

$$S_W(\rho) \;\ge\; S(\rho) \tag{12.63}$$

To prove this it is sufficient to use the continuous version of Peierls' inequality (12.11): for any convex function f convexity implies

$$\mathrm{Tr} f(\rho) = \int_\Omega \langle z|f(\rho)|z\rangle \,\mathrm{d}^2 z \;\ge \int_\Omega f\big(\langle z|\rho|z\rangle\big)\,\mathrm{d}^2 z = \int_\Omega f\big(Q(z)\big)\,\mathrm{d}^2 z\,. \tag{12.64}$$

Setting $f(t) = t\ln t$ and reverting the sign of the inequality we get Wehrl's result. Rényi–Wehrl entropies can be defined similarly, and the argument applies to them as well, so that for any $q \ge 0$ and any state ρ the inequality $S_q^{RW}(\rho) \ge S_q(\rho)$ holds.

For composite systems we can define a Husimi function by

$$Q(z_1, z_2) \;=\; \langle z_1|\langle z_2|\rho_{12}|z_2\rangle|z_1\rangle \tag{12.65}$$

and analyse its Wehrl entropy (see Problem 12.4). For a pure product state the Husimi function factorizes and its Wehrl entropy is equal to the sum of the Wehrl entropies of both subsystems. There are two possible definitions of the marginal Husimi distribution, and happily they agree, in the sense that

$$Q(z_1) \;\equiv\; \int_{\Omega_2} Q(z_1, z_2)\,\mathrm{d}^2 z_2 = \langle z_1|\mathrm{Tr}_2\rho_{12}|z_1\rangle \equiv Q(z_1)\,. \tag{12.66}$$

The Wehrl entropy can then be shown to be very well behaved, in particular it is monotone in the sense that $S_{12} \geq S_1$. Like the Shannon entropy, but unlike the Boltzmann entropy when the latter is defined over arbitrary distributions, the Wehrl entropy obeys all the inequalities in Table 12.1.

Turning to relative entropy we find many alternatives to Umegaki's definition. Many of them reproduce the classical relative entropy (2.25) when their two arguments commute. An example is the *Belavkin–Staszewski relative entropy* (Belavkin and Staszewski, 1982)

$$S_{\mathrm{BS}}(\rho||\sigma) \;=\; \mathrm{Tr}\Big(\rho \ln \rho^{1/2}\sigma^{-1}\rho^{1/2}\Big)\,. \tag{12.67}$$

It is monotone, and it can be shown that $S_{\mathrm{BS}}(\rho||\sigma) \geq S(\rho||\sigma)$ (Hiai and Petz, 1991).

The classical relative entropy itself is highly non-unique. We gave a very general class of monotone classical relative entropies in Eq. (2.74). In the quantum case we insist on monotonicity under completely positive stochastic maps, but not really on much else besides. A straightforward attempt to generalize the classical definition to the quantum case encounters the difficulty that the operator $\frac{\rho}{\sigma}$ is ambiguous in the non-commutative case. There are various ways of circumventing this difficulty, and then one can define a large class of monotone relative entropies. Just to be specific, let us mention a one-parameter family of monotone and jointly convex relative entropies:

$$S_\alpha(\rho,\sigma) = \frac{4}{1-\alpha^2}\mathrm{Tr}\big(\mathbb{1} - \sigma^{(\alpha+1)/2}\rho^{(\alpha-1)/2}\big)\rho\,, \qquad -1 < \alpha < 1\,. \tag{12.68}$$

Umegaki's definition is recovered in a limiting case. In fact in the limit $\alpha \to -1$ we obtain $S(\rho||\sigma)$, while we get $S(\sigma||\rho)$ when $\alpha \to 1$. Many more monotone relative entropies exist.[11]

12.5 Majorization of density matrices

The von Neumann entropy (like the Rényi entropies, but unlike the Wehrl entropy) provides a measure of the 'degree of mixing' of a given quantum state. A more sophisticated ordering of quantum states, with respect to the degree of mixing, is provided by the theory of majorization (Section 2.1). In the classical case the majorization order is really between orbits of probability vectors under permutations of their components – a fact that is easily missed since in discussing majorization one tends to represent these orbits with a representative $\vec{p}$, whose components appear in non-increasing order. When we go from probability vectors to density matrices, majorization will provide an ordering of the orbits under the unitary group. By

[11] See Petz (1998) and Lesniewski and Ruskai (1999) for the full story here.

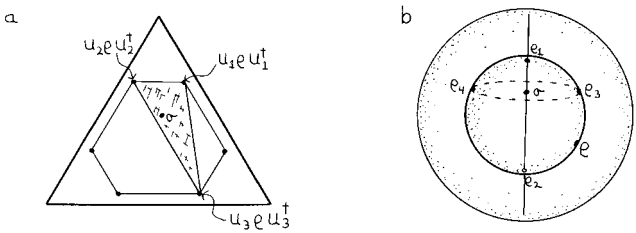


Figure 12.2. Left: $N = 3$, and we show majorization in the eigenvalue simplex. Right: $N = 2$, and we show two different ways of expressing a given σ as a convex sum of points on the orbit (itself a sphere!) of a majorizing ρ.

definition the state σ is majorized by the state ρ if and only if the eigenvalue vector of σ is majorized by the eigenvalue vector of ρ,

$$\sigma \prec \rho \quad \Leftrightarrow \quad \vec{\lambda}(\sigma) \prec \vec{\lambda}(\rho)\,. \tag{12.69}$$

This ordering relation between matrices has many advantages; in particular it does form a lattice.[12]

The first key fact to record is that if $\sigma \prec \rho$ then σ lies in the convex hull of the unitary orbit to which ρ belongs. We state this as a theorem:

Theorem 12.3 (Uhlmann's majorization) *If two density matrices of size N are related by $\sigma \prec \rho$, then there exists a probability vector $\vec{p}$ and unitary matrices U_I such that*

$$\sigma \;=\; \sum_I p_I\, U_I \rho\, U_I^\dagger\,. \tag{12.70}$$

Despite its importance, this theorem is easily proved. Suppose σ is given in diagonal form. We can find a unitary matrix U_I such that $U_I \rho U_I^\dagger$ is diagonal too; in fact we can find altogether $N!$ such unitary matrices since we can permute the eigenvalues. But now all matrices are diagonal and we are back to the classical case. From the classical theory we know that the eigenvalue vector of σ lies in the convex hull of the $N!$ different eigenvalue vectors of ρ. This provides one way of realizing Eq. (12.70).

There are many ways of realizing σ as a convex sum of states on the orbit of ρ, as we can see from Figure 12.1. In fact it is known that we can arrange things so that all components of p_I become equal. The related but weaker statement that any density matrix can be realized as a uniform mixture of pure states is very easy to prove (Bengtsson and Ericsson, 2003). Let $\sigma = \mathrm{diag}(\lambda_i)$. For $N = 3$, say, form a

[12] In physics, the subject of this section was begun by Uhlmann (1971); the work by him and his school is summarized by Alberti and Uhlmann (1982). A more recent survey is due to Ando (1994).

closed curve of pure state vectors by

$$Z^{\alpha}(\tau) = \begin{bmatrix} e^{in_1\tau} & 0 & 0 \\ 0 & e^{in_2\tau} & 0 \\ 0 & 0 & e^{in_3\tau} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \\ \sqrt{\lambda_2} \\ \sqrt{\lambda_3} \end{bmatrix} , \tag{12.71}$$

where the n_i are integers. Provided that the n_i are chosen so that $n_i - n_j$ is non-zero when $i \neq j$, it is easy to show that

$$\sigma \;=\; \frac{1}{2\pi} \int_0^{2\pi} d\tau \; Z^{\alpha} \bar{Z}_{\beta}(\tau) \, . \tag{12.72}$$

The off-diagonal terms are killed by the integration, so that σ is realized by a mixture of pure states distributed uniformly on the circle. The argument works for any N. Moreover a finite set of points on the curve will do as well, but we need at least N points since then we must ensure that $n_i - n_j \neq 0$ modulo N. When $N > 2$ these results are somewhat surprising – it was not obvious that one could find such a curve consisting only of pure states, since the set of pure states is a small subset of the outsphere.

Return to Uhlmann's theorem: in the classical case bistochastic matrices made their appearance at this point. This is true in the quantum case also; the theorem explicitly tells us that σ can be obtained from ρ by a bistochastic completely positive map, of the special kind known from Eq. (10.71) as random external fields. The converse holds:

Lemma 12.1 (Quantum HLP) *There exists a completely positive bistochastic map transforming ρ into σ if and only if $\sigma \prec \rho$,*

$$\rho \overset{\text{bistochastic}}{\longrightarrow} \sigma \quad \Leftrightarrow \quad \sigma \prec \rho. \tag{12.73}$$

To prove 'only if', introduce unitary matrices such that $\mathrm{diag}(\vec{\lambda}(\sigma)) = U\sigma U^{\dagger}$ and $\mathrm{diag}(\vec{\lambda}(\rho)) = V\rho V^{\dagger}$. Given a bistochastic map such that $\Phi\rho = \sigma$ we construct a new bistochastic map Ψ according to

$$\Psi X \;\equiv\; U[\Phi(V^{\dagger} X V)]U^{\dagger} \, . \tag{12.74}$$

By construction $\Psi\left(\mathrm{diag}(\vec{\lambda}(\rho))\right) = \mathrm{diag}(\vec{\lambda}(\sigma))$. Next we introduce a complete set of projectors P_i onto the basis vectors, in the basis we are in. We use them to construct a matrix whose elements are given by

$$B_{ij} \equiv \mathrm{Tr} P_i \Psi P_j \, . \tag{12.75}$$

We recognize that B is a bistochastic matrix, and finally we observe that

$$\lambda_i(\sigma) = \mathrm{Tr} P_i \, \mathrm{diag}(\vec{\lambda}(\sigma)) = \mathrm{Tr} P_i \Psi \left(\sum_j P_j \lambda_j(\rho) \right) = \sum_j B_{ij} \lambda_j(\rho) \, , \tag{12.76}$$

where we used linearity in the last step. An appeal to the classical HLP lemma concludes the proof.

Inspired by the original Horn's lemma (Section 2.1) one may ask if the word *bistochastic* in the quantum HLP lemma might be replaced by *unistochastic*. This we do not know. However, a concrete result may be obtained if one allows the size of ancilla to be large (Horodecki et al., 2003a).

Weak version of quantum Horn's lemma. *If two quantum states of size N satisfy $\rho' \prec \rho$, then there exists a K-unistochastic map transforming ρ into ρ' up to an arbitrary precision controlled by K.*

To prove this one uses the HLP lemma to find a bistochastic matrix B of size N which relates the spectra, $\lambda' = B\lambda$, of the states ρ' and ρ. Then using the Birkhoff theorem one represents the matrix by a convex sum of permutations, $B = \sum_{m=1}^{j} \alpha_m P_m$ with $j \leq N^2 - 2N + 2$. The next step consists in setting the size of the ancilla to $M = KN$ and a decomposition $M = \sum_{m=1}^{j} M_m$ such that the fractions M_m/M approximate the weights α_m. The initial state ρ can be rotated unitarily, so we may assume it is diagonal and commutes with the target ρ'. The spectrum of the extended state $\rho \otimes \mathbb{1}_M$ consists of N degenerate blocks, each containing M copies of the same eigenvalue λ_i. Let us split each block into j groups of M_m elements each and allow every permutation P_m to act M_m times, permuting elements from the mth group of each block. This procedure determines the unitary matrix U of size KN^2 which defines the K-unistochastic operation (see Eq. (10.64)). The partial trace over an M-dimensional environment produces state ρ'' with the spectrum $\lambda'' = B_a\lambda$, where $B_a = \sum_{m=1}^{j}(M_m/M)P_m$. The larger K, the better the matrix B_a approximates B, so one may produce an output state ρ'' arbitrarily close to the target ρ'.

An interesting example of a completely positive and bistochastic map is the operation of coarse graining with respect to a given Hermitian operator H (e.g. a Hamiltonian). We denote it by Φ_{CG}^{H}, and define it by

$$\rho \;\to\; \Phi_{CG}^{H}(\rho) \;=\; \sum_{i=1}^{N} P_i \rho P_i = \sum_{i=1}^{N} p_i \, |h_i\rangle\langle h_i| \,, \tag{12.77}$$

where the P_i project onto the eigenvectors $|h_i\rangle$ of H (assumed non-degenerate for simplicity). In more mundane terms, this is the map that deletes all off-diagonal elements from a density matrix. It obeys Schur–Horn's theorem:

Theorem 12.4 (Schur–Horn's) *Let ρ be an Hermitian matrix, $\vec{\lambda}$ its spectrum, and $\vec{p}$ its diagonal elements in a given basis. Then*

$$\vec{p} \;\prec\; \vec{\lambda} \,. \tag{12.78}$$

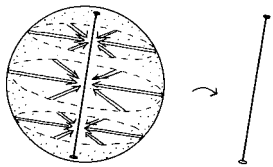


Figure 12.3. Coarse graining a density matrix.

Conversely, if this equation holds then there exists an Hermitian matrix with spectrum $\vec{\lambda}$ whose diagonal elements are given by $\vec{p}$.

We prove this one way. There exists a unitary matrix that diagonalizes the matrix, so we can write

$$p_i = \rho_{ii} = \sum_{j,k} U_{ij}\lambda_j\delta_{jk}U^\dagger_{ki} = \sum_j |U_{ij}|^2\lambda_j \ . \tag{12.79}$$

The vector $\vec{p}$ is obtained by acting on $\vec{\lambda}$ with a unistochastic, hence bistochastic, matrix, and the result follows from Horn's lemma (Section 2.1).[13]

The Schur–Horn theorem has weighty consequences. It is clearly of interest when one tries to quantify decoherence, since the entropy of the coarse grained density matrix Φ^H_{CG} will be greater than that of ρ. It also leads to an interesting definition of the von Neumann entropy, that again brings out the distinguished status of the latter. Although we did not bring it up in Section 12.2, we could have defined the entropy of a density matrix relative to any POVM $\{E\}$, as the Shannon entropy of the probability distribution defined cooperatively by the POVM and the density matrix. That is, $S(\rho) \equiv S(\vec{p})$, where $p_i = \mathrm{Tr}E_i\rho$. To make the definition independent of the POVM, we could then minimize the resulting entropy over all possible POVMs, so a possible definition that depends only on ρ itself would be

$$S(\rho) \equiv \min_{\mathrm{POVM}} S(\vec{p}) \ , \qquad p_i = \mathrm{Tr}E_i\rho \ . \tag{12.80}$$

But the entropy defined in this way is equal to the von Neumann entropy. The Schur–Horn theorem shows this for the special case that we minimize only over projective measurements, and the argument can be extended to cover the general case. Note that Wehrl's inequality (12.63) is really a special case of this observation, since the Wehrl entropy is the entropy that we get from the POVM defined by the coherent states.

From a mathematical point of view, the Schur–Horn theorem is much more interesting than it appears to be at first sight. To begin to see why, we can restate it: consider the map that takes an Hermitian matrix to its diagonal entries. Then the

[13] This part of the theorem is due to Schur (1923). Horn (1954) proved the converse.

theorem says that the image of the space of Hermitian matrices, under this map, is a convex polytope whose corners are the $N!$ fixed points of the map. Already it sounds more interesting! Starting from this example, mathematicians have developed a theory that deals with maps from connected symplectic manifolds, and conditions under which the image will be the convex hull of the image of the set of fixed points of a group acting on the manifold.[14]

12.6 Entropy dynamics

What we have not discussed so far is the role of entropy as an arrow of time – which is how entropy has been regarded ever since the word was coined by Clausius. If this question is turned into the question how the von Neumann entropy of a state changes under the action of some quantum operation $\Phi : \rho \to \rho'$, it does have a clear cut answer. Because of Eq. (12.48), it follows from monotonicity of relative entropy that a CP map increases the von Neumann entropy of every state if and only if it is unital (bistochastic), that is if it transforms the maximally mixed state into itself. For quantum operations that are stochastic, but not bistochastic, this is no longer true – for such quantum channels the von Neumann entropy may decrease. Consider for instance the decaying or amplitude damping channel (Section 10.7), which describes the effect of spontaneous emission on a qubit. It sends any mixed state towards the pure ground state, for which the entropy is zero. But then this is not an isolated system, so this would not worry Clausius.

Even for bistochastic maps, when the von Neumann entropy does serve as an arrow of time, it does not point very accurately to the future (see Figure 12.4). Relative to any given state, the state space splits into three parts, the 'future' F that consists of all states that can be reached from the given state by bistochastic maps, the 'past' P that consists of all states from which the given state can be reached by such maps, and a set of incomparable states that we denote by C in the figure. This is reminiscent of the causal structure in special relativity, where the light cone divides Minkowski space into the future, the past, and the set of points that cannot communicate in either direction with a point sitting at the vertex of the light cone. There is also the obvious difference that the actual shape of the 'future' depends somewhat on the position of the given state, and very much so when its eigenvalues degenerate. The isoentropy curves of the von Neumann entropy do not do justice to this picture. To do better one would have to bring in a complete set of Schur concave functions such as the Rényi entropies (see Figure 2.14).

[14] For an overview of this theory, and its connections to symplectic geometry and to interesting problems of linear algebra, see Knutson (2000). A related problem of finding constraints between spectra of composite systems and their partial traces was recently solved by Bravyi (2004) and Klyachko (n.d.).

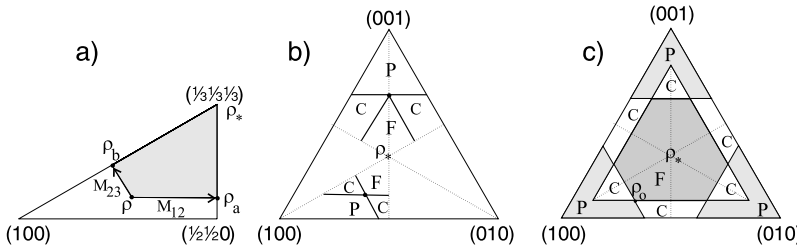


Figure 12.4. The eigenvalue simplex for $N = 3$: (a) a Weyl chamber; the shaded region is accessible from ρ with bistochastic maps. (b) The shape of the 'light cone' depends on the degeneracy of the spectrum. F denotes Future, P Past, and C the noncomparable states. (c) Splitting the simplex into Weyl chambers.

Naturally, the majorization order may not be the last word on the future. Depending on the physics, it may well be that majorization provides a necessary but not sufficient condition for singling it out.[15]

We turn from the future to a more modest subject, namely the *entropy of an operation* Φ. This can be conveniently defined as the von Neumann entropy of the state that corresponds to the operation via the Jamiolkowski isomorphism, namely as

$$S(\Phi) \equiv S\left(\frac{1}{N} D_\Phi\right) \quad \in \quad [0, \ln N^2] \,. \tag{12.81}$$

where D_Φ is the dynamical matrix defined in Section 10.3. Generalized entropies may be defined similarly. The entropy of an operation vanishes if D_Φ is of rank one, that is to say if the operation is a unitary transformation. The larger the entropy S of an operation, the more terms enter effectively into the canonical Kraus form, and the larger are the effects of decoherence in the system. The maximum is attained for the completely depolarizing channel Φ_*. The entropy for an operation of the special form (10.71), that is for random external fields, is bounded from above by the Shannon entropy $S(\vec{p})$. The norm $\sqrt{\mathrm{Tr}\Phi\Phi^\dagger} = ||\Phi||_{\mathrm{HS}}$ may also be used to characterize the decoherence induced by the map. It varies from unity for Φ_* (total decoherence) to N for a unitary operation (no decoherence) – see Table 10.2.

A different way to characterize a quantum operation is to compute the amount of entropy it creates when acting on an initially pure state. In Section 2.3 we defined the entropy of a stochastic matrix with respect to a fixed probability distribution. This definition, and the bound (2.39), has a quantum analogue due to Lindblad

[15] For a lovely example from thermodynamics, see Alberti and Uhlmann (1981).

(1991), and it will lead us to our present goal. Consider a CP map Φ represented in the canonical Kraus form $\rho' = \sum_{i=1}^{r} A_i \rho A_i^\dagger$. Define an operator acting on an auxiliary r-dimensional space $\mathcal{H}_r$ by

$$\sigma_{ij} \;=\; \mathrm{Tr}\rho A_j^\dagger A_i \;. \tag{12.82}$$

In Problem 10.3 we show that σ is a density operator in its own right. The von Neumann entropy of σ depends on ρ, and equals $S(\Phi)$, as defined above, when ρ is the maximally mixed state. Next we define a density matrix in the composite Hilbert space $\mathcal{H}_N \otimes \mathcal{H}_r$,

$$\omega \;=\; \sum_{i=1}^{r}\sum_{j=1}^{r} A_i \rho A_j^\dagger \otimes |i\rangle\langle j| = W\rho\, W^\dagger \;, \tag{12.83}$$

where $|i\rangle$ is an orthonormal basis in $\mathcal{H}_r$. The operator W maps a state $|\phi\rangle$ in $\mathcal{H}_N$ into $\sum_{j=1}^{r} A_j|\phi\rangle \otimes |j\rangle$, and the completeness of the Kraus operators implies that $W^\dagger W = \mathbb{1}_N$. It follows that $S(\omega) = S(\rho)$. Since it is easy to see that $\mathrm{Tr}_N\omega = \sigma$ and $\mathrm{Tr}_r\omega = \rho$, we may use the triangle inequalities (12.28) and (12.32), and some slight rearrangement, to deduce that

$$|S(\rho) - S(\sigma)| \;\le\; S(\rho') \;\le\; S(\sigma) + S(\rho) \;, \tag{12.84}$$

in exact analogy to the classical bound (2.39). If the intial state is pure, that is if $S(\rho) = 0$, we find that the final state has entropy $S(\sigma)$. For this reason $S(\sigma)$ is sometimes referred to as the *entropy exchange* of the operation.

Finally, and in order to give a taste of a subject that we omit, let us define the capacity of a quantum channel Φ. The capacity for a given state is

$$C_\Phi(\rho) \equiv \max_{\mathcal{E}} \sum_i p_i S(\Phi\sigma_i || \Phi\rho) = \min_{\mathcal{E}} \Big[S(\Phi\rho) - \sum_i p_i S(\Phi\sigma_i) \Big] \;. \tag{12.85}$$

The quantity that is being optimized will be discussed, under the name *Jensen–Shannon divergence*, in Section 13.1. The optimization is performed over all ensembles $\mathcal{E} = \{\sigma_i; p_i\}$ such that $\rho = \sum_i p_i\sigma_i$. It is not an easy one to carry out. In the next step the *channel capacity* is defined by optimizing over the set of all states:

$$C_{\mathrm{Hol}}(\Phi) \;\equiv\; \max_{\rho} \big[C_\Phi(\rho) \big] \;. \tag{12.86}$$

There is a theorem due to Holevo (1973) which employs these definitions to give an upper bound on the information carrying capacity of a noisy quantum channel. Together with the quantum noiseless coding theorem, this result brings quantum information theory up to the level achieved for classical information theory in Shannon's classical work (Shannon, 1948). But these matters are beyond the

scope of our book. Let us just mention that there are many things that are not known. Notably there is an additivity conjecture stating that $C_{\rm Hol}(\Phi_1 \otimes \Phi_2) = C_{\rm Hol}(\Phi_1) + C_{\rm Hol}(\Phi_2)$. In one respect its status is similar to that of strong subadditivity before the latter was proven – it is equivalent to many other outstanding conjectures.[16]

Problems

Problem 12.1 Show, for any positive operator A, that

$$\ln(A + xB) - \ln A = \int_0^\infty \frac{1}{A+u} x B \frac{1}{A+xB+u}\, {\rm d}u \,. \tag{12.87}$$

Problem 12.2 Compute the two contour integrals

$$S(\rho) = -\frac{1}{2\pi i} \oint (\ln z) {\rm Tr}\Big(\mathbb{1} - \frac{\rho}{z}\Big)^{-1} {\rm d}z \tag{12.88}$$

and

$$S_Q(\rho) = -\frac{1}{2\pi i} \oint (\ln z) {\rm det}\Big(\mathbb{1} - \frac{\rho}{z}\Big)^{-1} {\rm d}z \,, \tag{12.89}$$

with a contour that encloses all the eigenvalues of ρ. The second quantity is known as *subentropy* (Jozsa, Robb and Wootters, 1994).

Problem 12.3 Prove *Donald's identity* (Donald, 1987): for any mixed state $\rho = \sum_k p_k \rho_k$ and another state σ

$$\sum_k p_k\, S(\rho_k||\sigma) \;=\; \sum_k p_k\, S(\rho_k||\rho) + S(\rho||\sigma) \,. \tag{12.90}$$

Problem 12.4 Compute the Wehrl entropy for the Husimi function (12.65) of a two qubit pure state written in its Schmidt decomposition.

Problem 12.5 Prove that Euclidean distances between orbits can be read off from a picture of the Weyl chamber (i.e. prove Eq. (8.52)).

Problem 12.6 Prove that $\big({\rm det}(A+B)\big)^{1/N} \geq ({\rm det}A)^{1/N} + ({\rm det}B)^{1/N}$, where A and B are positive matrices of size N.

[16] See the review by Amosov, Holevo and Werner (n.d.) and the work by Shor (n.d.).

Problem 12.7 For any operation Φ given by its canonical Kraus form (10.55) one defines its *purity Hamiltonian*

$$\Omega \equiv \sum_{i=1}^{r} \sum_{j=1}^{r} A_j^\dagger A_i \otimes A_i^\dagger A_j, \tag{12.91}$$

the trace of which characterizes an average decoherence induced by Φ (Zanardi and Lidar, 2004). Show that $\mathrm{Tr}\Omega = ||\Phi||^2_{\mathrm{HS}} = \mathrm{TrD}^2$, hence it is proportional to the purity $\mathrm{Tr}\,\rho_\Phi^2$ of the state corresponding to Φ in the isomorphism (11.22).

13

Distinguishability measures

> Niels Bohr supposedly said that if quantum mechanics did not make you dizzy then you did not understand it. I think that the same can be said about statistical inference.
>
> *Robert D. Cousins*

In this chapter we quantify how easy it may be too distinguish probability distributions from each other (a discussion that was started in Chapter 2). The issue is a very practical one and arises whenever one is called upon to make a decision based on imperfect data. There is no unique answer because everything depends on the data – the l_1-distance appears if there has been just one sampling of the distributions, the relative entropy governs the approach to the 'true' distribution as the number of samplings goes to infinity, and so on.

The quantum case is even subtler. A quantum state always stands ready to produce a large variety of classical probability distributions, depending on the choice of measurement procedure. It is no longer possible to distinguish pure states from each other with certainty, unless they are orthogonal. The basic idea behind the quantum distinguishability measures is the same as that which allowed us, in Section 5.3, to relate the Fubini–Study metric to the Fisher–Rao metric. We will optimize over all possible measurements.

13.1 Classical distinguishability measures

If a distance function has an operational significance as a measure of statistical distinguishability, then we expect it to be monotone (and decreasing) under general stochastic maps. Coarse graining means that information is being discarded, and this cannot increase distinguishability. From Čencov's theorem (Section 2.5) we know that the Fisher–Rao metric is the only Riemannian metric on the probability simplex that is monotone under general stochastic maps. But there is another simple distance measure that does have the desirable property, namely the l_1-distance from

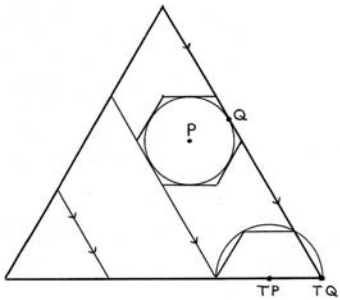


Figure 13.1. Coarse graining, according to Eq. (13.3), collapses the probability simplex to an edge. The l_1-distance never increases (the hexagon is unchanged), but the l_2-distance sometimes does (the circle grows).

Eq. (1.55). The proof of monotonicity uses the observation that the difference of two probability vectors can be written in the form

$$p_i - q_i \;=\; N_i^+ - N_i^- \;, \tag{13.1}$$

where N^+ and N^- are two positive vectors with orthogonal support, meaning that for each component i at least one of N_i^+ and N_i^- is zero. We follow this up with the triangle inequality, and condition (ii) from Eq. (2.4) that defines a stochastic matrix T:

$$\begin{aligned} ||Tp - Tq||_1 &= ||TN^+ - TN^-||_1 \leq ||TN^+||_1 + ||TN^-||_1 \\ &= \frac{1}{2}\sum_{i,j} T_{ij} N_j^+ + \frac{1}{2}\sum_{i,j} T_{ij} N_j^- \\ &= \frac{1}{2}\sum_{j} (N_j^+ + N_j^-) \\ &= ||p - q||_1 \;. \end{aligned} \tag{13.2}$$

By contrast, the Euclidean l_2-distance is not monotone. To see this, consider a coarse graining stochastic matrix such as

$$T \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \;. \tag{13.3}$$

Applying this transformation has the effect of collapsing the entire simplex onto one of its edges. If we draw a picture of this, as in Figure 13.1, it becomes evident why $p = 1$ is the only value of p for which the l_p-distance is monotone under this map. The picture should also make it clear that it is coarse graining maps like (13.3) that may cause problems with monotonicity – monotonicity under bistochastic maps, that cause a contraction of the probability simplex towards its centre, is

much easier to ensure. In fact the flat l_2-distance is monotone under bistochastic maps. Incidentally, it is clear from the picture that the l_1-distance succeeds in being monotone (under general stochastic maps) only because the distance between probability distributions with orthogonal support is constant (and maximal). This is a property that the l_1-distance shares with the monotone Fisher–Rao distance – and if a distance is to quantify how easily two probability distributions can be distinguished, then it must be monotone.

The question remains to what extent, and in what sense, our various monotone notions of distance – the Bhattacharyya and Hellinger distances, and the l_1-distance – have any clear-cut operational significance. For the latter, an answer is known. Consider two probability distributions P and Q over N events, and mix them, so that the probability for event i is

$$r_i = \pi_0 p_i + \pi_1 q_i \ . \tag{13.4}$$

A possible interpretation here is that Alice sends Bob a message in the form of an event drawn from one of two possible probability distributions. Bob is ignorant of which particular distribution Alice uses, and his ignorance is expressed by the distribution (π_0, π_1). Having sampled once, Bob is called upon to guess what distribution was used by Alice. It is clear – and this answer is given stature with technical terms like 'Bayes' decision rule' – that his best guess, given that event i occurs, is to guess P if $p_i > q_i$, and Q if $q_i > p_i$. (If equality holds the two guesses are equally good.) Given this strategy, the probability that Bob's guess is right is

$$P_{\mathrm{R}}(P, Q) \ = \ \sum_{i=1}^{N} \max\{\pi_0 p_i, \pi_1 q_i\} \tag{13.5}$$

and the probability of error is

$$P_{\mathrm{E}}(P, Q) \ = \ \sum_{i=1}^{N} \min\{\pi_0 p_i, \pi_1 q_i\} \ . \tag{13.6}$$

Now consider the case $\pi_0 = \pi_1 = 1/2$. Then Bob has no reason to prefer any distribution in advance. In this situation it is easily shown that $P_{\mathrm{R}} - P_{\mathrm{E}} = D_1$, or equivalently

$$D_1(P, Q) = \frac{1}{2} \sum_{i=1}^{N} |p_i - q_i| = 1 - 2P_{\mathrm{E}}(P, Q) \ , \tag{13.7}$$

that is, the l_1-distance grows as the probability of error goes down. In this sense the l_1-distance has a precise meaning, as a measure of how reliably two probability distributions can be distinguished by means of a single sampling.

Of course it is not clear why we should restrict ourselves to one sampling; the probability of error goes down as the number of samplings $\mathcal{N}$ increases. There is a theorem that governs how it does so:

Theorem 13.1 (Chernoff's) *Let $P_{\mathrm{E}}^{(\mathcal{N})}(P, Q)$ be the probability of error after $\mathcal{N}$ samplings of two probability distributions. Then*

$$P_{\mathrm{E}}^{(\mathcal{N})}(P, Q) \;\leq\; \left(\min_{\alpha \in [0,1]} \sum_{i=1}^{N} p_i^{\alpha}\, q_i^{1-\alpha} \right)^{\mathcal{N}} . \tag{13.8}$$

The bound is approached asymptotically when $\mathcal{N}$ goes to infinity.

Unfortunately it is not easy to obtain an analytic expression for the Chernoff bound (the one that is approached asymptotically), but we do not have to find the minimum in order to obtain useful upper bounds. The non-minimal bounds are of interest in themselves. They are related to the *relative Rényi entropy*

$$I_q(P, Q) \;=\; \frac{1}{1-q} \ln\Big[\sum_{i=1}^{N} p_i^{q}\, q_i^{1-q} \Big] . \tag{13.9}$$

When $q = 1/2$ the relative Rényi entropy is symmetric, and it is a monotone function of the geodesic Bhattacharyya distance D_{Bhatt} from Eq. (2.56).

In the limit $q \to 1$, the relative Rényi entropy tends to the usual relative entropy $S(P||Q)$, which figured in a different calculation of the probability of error in Section 2.3. The setting there was that we made a choice between the distributions P and Q, using a large number of samplings, in a situation where it happened to be the case that the statistics were governed by Q. The probability of erroneously concluding that the true distribution is P was shown to be

$$P_{\mathrm{E}}(P, Q) \;\sim\; \mathrm{e}^{-\mathcal{N} S(P||Q)} . \tag{13.10}$$

The asymmetry of the relative entropy reflects the asymmetry of the situation. In fact, suppose the choice is between a fair coin and a biased coin that only shows heads. Using Eq. (2.32) we find that

$$P_{\mathrm{E}}(\text{fair}||\text{biased}) = \mathrm{e}^{-\mathcal{N}\cdot\infty} = 0 \;\text{ and }\; P_{\mathrm{E}}(\text{biased}||\text{fair}) = \mathrm{e}^{-\mathcal{N}\ln 2} = \frac{1}{2^{\mathcal{N}}} . \tag{13.11}$$

This is exactly what intuition dictates; the fair coin can produce the frequencies expected from the biased coin, but not the other way around.

But sometimes we insist on true distance functions. Relative entropy cannot be turned into a true distance just by symmetrization, because the triangle inequality will still be violated. However, there is a simple modification that does lead to a proper distance function. Given two probability distributions P and Q, let us define

their mean R by

$$R = \frac{1}{2}P + \frac{1}{2}Q \quad \Leftrightarrow \quad r_i = \frac{1}{2}p_i + \frac{1}{2}q_i \,. \tag{13.12}$$

Then the *Jensen–Shannon divergence* is defined by

$$J(P, Q) \equiv 2S(R) - S(P) - S(Q) \,, \tag{13.13}$$

where S is the Shannon entropy. An easy calculation shows that this is related to relative entropy:

$$J(P, Q) = \sum_{i=1}^{N} \left(p_i \ln \frac{2p_i}{p_i + q_i} + q_i \ln \frac{2q_i}{p_i + q_i} \right) = S(P||R) + S(Q||R) \,. \tag{13.14}$$

Interestingly, the function $D(P, Q) = \sqrt{J(P, Q)}$ is not only symmetric but obeys the triangle inequality as well, and hence it qualifies as a true distance function – moreover as a distance function that is consistent with the Fisher–Rao metric.[1]

The Jensen–Shannon divergence can be generalized to a measure of the divergence between an arbitrary number of M probability distributions $P_{(m)}$, weighted by some probability distribution π over M events:

$$J(P_{(1)}, P_{(2)}, \ldots, P_{(M)}) \equiv S\left(\sum_{m=1}^{M} \pi_m P_{(m)} \right) - \sum_{m=1}^{M} \pi_m S(P_{(m)}) \,. \tag{13.15}$$

It has been used, in this form, in the study of DNA sequences – and in the definition (12.85) of the capacity of a quantum channel. Its interpretation as a distinguishability measure emerges when we sample from a statistical mixture of probability distributions. Given that the Shannon entropy measures the information gained when sampling a distribution, the Jensen–Shannon divergence measures the average gain of information about how that mixture was made (that is about π), since we subtract that part of the information that concerns the sampling of each individual distribution in the mixture.

The reader may now have begun to suspect that there are many measures of distinguishability available, some of them more useful, and some of them easier to compute, than others. Fortunately there are inequalities that relate different measures of distinguishability. An example is the *Pinsker inequality* that relates the l_1-distance to the relative entropy:

$$S(P||Q) \;\geq\; \frac{1}{2}\left(\sum_{i=1}^{N} |p_i - q_i| \right)^2 \;=\; 2D_1^2(P, Q) \,. \tag{13.16}$$

[1] This is a recent result; see Endres and Schindelin (2003). For a survey of the Jensen–Shannon divergence and its properties, see Lin (1991).

This is a stronger bound than (2.30) since $D_1 \geq D_2$. The proof is quite interesting. First one uses brute force to establish that

$$2(p-q)^2 \;\leq\; p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q} \tag{13.17}$$

wherever $0 \leq q \leq p \leq 1$. This is the Pinsker inequality for $N = 2$. We are going to reduce the general case to this. Without loss of generality we assume that $p_i \geq q_i$ for $1 \leq i \leq K$, and $p_i < q_i$ otherwise. Then we define a stochastic matrix T by

$$\begin{bmatrix} T_{11} & \dots & T_{1K} & T_{1K+1} & \dots & T_{1N} \\ T_{21} & \dots & T_{2K} & T_{2K+1} & \dots & T_{2N} \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \end{bmatrix} . \tag{13.18}$$

We get two binomial distributions TP and TQ, and define

$$p \equiv \sum_{i=1}^{K} p_i = \sum_{i=1}^{N} T_{1i} p_i \;, \qquad q \equiv \sum_{i=1}^{K} q_i = \sum_{i=1}^{N} T_{1i} q_i \;. \tag{13.19}$$

It is easy to see that $D_1(P, Q) = p - q$. Using this and Eq. (13.17), we get

$$2D_1^2(P, Q) \;\leq\; S(TP||TQ) \;\leq\; S(P||Q) \,. \tag{13.20}$$

Thus monotonicity of relative entropy was used in the punchline.

The Pinsker inequality is not sharp; it has been improved to[2]

$$S(P||Q) \;\geq\; 2D_1^2 + \frac{4}{9} D_1^4 + \frac{32}{135} D_1^6 + \frac{7072}{42525} D_1^8 \,. \tag{13.21}$$

Relative entropy is unbounded from above. But it can be shown (Lin, 1991) that

$$2D_1(P, Q) \geq J(P, Q) \,. \tag{13.22}$$

Hence the l_1-distance bounds the Jensen–Shannon divergence from above.

13.2 Quantum distinguishability measures

We now turn to the quantum case. When density matrices rather than probability distributions are sampled we face new problems, since the probability distribution $P(E, \rho)$ that governs the sampling will depend, not only on the density matrix ρ, but on the POVM that represents the measurement as well. The probabilities that we actually have to play with are given by

$$p_i(E, \rho) = \mathrm{Tr} E_i \rho \;, \tag{13.23}$$

where $\{E_i\}_{i=1}^K$ is some POVM. The quantum distinguishability measures will be defined by varying over all possible POVMs until the classical distinguishability

[2] Inequality (13.16) is due to Pinsker (1964), while (13.21) is due to Topsøe (2001).

of the resulting probability distributions is maximal. In this way any classical distinguishability measure will have a quantum counterpart – except that for some of them, notably for the Jensen–Shannon divergence, the optimization over all POVMs is very difficult to carry out, and we will have to fall back on bounds and inequalities.[3]

Before we begin, let us define the L_p-norm of an operator A by

$$||A||_p \equiv \left(\frac{1}{2}\mathrm{Tr}|A|^p\right)^{1/p}, \tag{13.24}$$

where the absolute value of the operator was defined in Eq. (8.12). The factor of $1/2$ is included in the definition because it is convenient when we restrict ourselves to density matrices. In obvious analogy with Section 1.4 we can now define the L_p-*distance* between two operators as

$$D_p(A, B) \equiv ||A - B||_p . \tag{13.25}$$

Like all distances based on a norm, these distances are useful because convex mixtures will appear as (segments of) metric lines. The factor $1/2$ in the definition ensures that all the L_p-distances coincide when $N = 2$. For 2×2 matrices, an L_p ball looks like an ordinary ball. Although the story becomes more involved when $N > 2$, it will always be true that all the L_p-distances coincide for a pair of pure states, simply because a pair of pure states taken in isolation always span a two-dimensional Hilbert space. We may also observe that, given two density matrices ρ and σ, the operator $\rho - \sigma$ can be diagonalized, and the distance $D_p(\rho, \sigma)$ becomes the l_p–distance expressed in terms of the eigenvalues of that operator.

For $p = 2$ the L_p-distance is Euclidean. It has the virtue of simplicity, and we have already used it extensively. For $p = 1$ we have the *trace distance*[4]

$$D_{\mathrm{tr}}(A, B) = \frac{1}{2}\,\mathrm{Tr}|A - B| = \frac{1}{2}\,D_{\mathrm{Tr}}(A, B) . \tag{13.26}$$

It will come as no surprise to learn that the trace distance will play a role similar to that of the l_1-distance in the classical case. It is interesting to get some understanding of the shape of its unit ball. All L_p-distances can be computed from the eigenvalues of the operator $\rho - \sigma$, and therefore Eq. (1.56) for the radii of its in- and outspheres can be directly taken over to the quantum case. But there is a difference between the trace and the l_1-distances, and we see it as soon as we look at a set of density matrices that cannot be diagonalized at the same time (Figure 13.2).

[3] The subject of quantum decision theory, which we are entering here, was founded by Helstrom (1976) and Holevo (1982). A very useful (and more modern) account is due to Fuchs (1996); see also Fuchs and van de Graaf (1999). Here we give some glimpses only.

[4] For convenience we are going to use, in parallel, two symbols, $D_{\mathrm{Tr}} = 2\,D_{\mathrm{tr}}$.

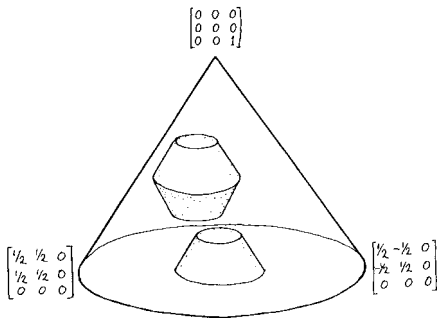


Figure 13.2. Some balls of constant radius, as measured by the trace distance, inside a three-dimensional slice of the space of density matrices (obtained by rotating the eigenvalue simplex around an axis).

Thus equipped, we take up the task of quantifying the probability of error in choosing between two density matrices ρ and σ, based on a single measurement. Mathematically, the task is to maximize the l_1-distance over all possible POVMs $\{E_i\}$, given ρ and σ. Thus our quantum distinguishability measure D is defined by

$$D(\rho, \sigma) \equiv \max_E D_1\Big(P(E, \rho), P(E, \sigma)\Big) . \tag{13.27}$$

As the reader may suspect already, the answer is the trace distance. We will carry out the maximization for projective measurements only – the generalization to arbitrary POVMs being quite easy – and start with a lemma that contains a key fact about the trace distance:

Lemma 13.1 (Trace distance) *If P is any projector onto a subspace of Hilbert space then*

$$D_{\rm tr}(\rho, \sigma) \geq TrP(\rho - \sigma) = D_1(\rho, \sigma) . \tag{13.28}$$

Equality holds if and only if P projects onto the support of N_+, where $\rho - \sigma = N_+ - N_-$, with N_+ and N_- being positive operators of orthogonal support.

To prove this, observe that by construction $\mathrm{Tr}N_+ = \mathrm{Tr}N_-$ (since their difference is a traceless matrix), so that $D_{\rm tr} = \mathrm{Tr}N_+$. Then

$$\mathrm{Tr}P(\rho - \sigma) = \mathrm{Tr}P(N_+ - N_-) \leq \mathrm{Tr}PN_+ \leq \mathrm{Tr}N_+ = D_{\rm tr}(\rho, \sigma) . \tag{13.29}$$

Clearly, equality holds if and only if $P = P_+$, where $P_+N_- = 0$ and $P_+N_+ = N_+$.

The useful properties of the trace distance now follow suit:

Theorem 13.2 (Helstrom's) *Let $p_i = TrE_i\rho$ and $q_i = TrE_i\sigma$. Then*

$$D_{\rm tr}(\rho,\sigma) \;=\; \max_E D_1(P,Q)\,, \tag{13.30}$$

where we maximize over all POVMs.

The proof (for projective measurements) begins with the observation that

$$\mathrm{Tr}|E_i(\rho-\sigma)| = \mathrm{Tr}|E_i(N_+ - N_-)| \le \mathrm{Tr}E_i(N_+ + N_-) = \mathrm{Tr}E_i|\rho-\sigma|\,. \tag{13.31}$$

For every POVM, and the pair of probability distributions derived from it, this implies that

$$D_1(P,Q) = \frac{1}{2}\sum_i \mathrm{Tr}|E_i(\rho-\sigma)| \le \frac{1}{2}\sum_i \mathrm{Tr}E_i|\rho-\sigma| = D_{\rm tr}(\rho,\sigma)\,. \tag{13.32}$$

The inequality is saturated when we choose a POVM that contains one projector onto the support of N_+ and one projector onto the support of N_-. The interpretation of $D_{\rm tr}(\rho,\sigma)$ as a quantum distinguishability measure for 'one shot samplings' is thereby established.

It is important to check that the trace distance is monotone under trace preserving CP maps $\rho \to \Phi(\rho)$. This is not hard to do if we first use our lemma to find a projector P such that

$$D_{\rm tr}(\Phi(\rho),\Phi(\sigma)) \;=\; \mathrm{Tr}P(\Phi(\rho)-\Phi(\sigma))\,. \tag{13.33}$$

We decompose $\rho-\sigma$ as above. Since the map is trace preserving it is true that $\mathrm{Tr}\Phi(N_+) = \mathrm{Tr}\Phi(N_-)$. Then

$$\begin{aligned} D_{\rm tr}(\rho,\sigma) &= \frac{1}{2}(N_+ + N_-) = \frac{1}{2}(\Phi(N_+)+\Phi(N_-)) = \mathrm{Tr}\Phi(N_+) \\ &\ge \mathrm{Tr}P\,\Phi(N_+) \ge \mathrm{Tr}P(\Phi(N_+)-\Phi(N_-)) = \mathrm{Tr}P(\Phi(\rho)-\Phi(\sigma)) \end{aligned} \tag{13.34}$$

– and the monotonicity of the trace distance follows when we take into account how P was selected. The lemma can also be used to prove a strong convexity result for the trace distance, namely,

$$D_{\rm tr}\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \le D_1(P,Q) + \sum_i p_i D_{\rm tr}(\rho_i,\sigma_i)\,. \tag{13.35}$$

We omit the proof (Nielsen and Chuang, 2000). Joint convexity follows if we set $P = Q$.

The trace distance sets limits on how much the von Neumann entropy of a given state may change under a small perturbation. To be precise, we have Fannes' lemma:

Lemma 13.2 (Fannes') *Let the quantum states ρ and σ act on an N-dimensional Hilbert space, and be close enough in the sense of the trace metric so that $D_{\rm tr}(\rho,\sigma) \leq 1/(2e)$. Then*

$$|S(\rho) - S(\sigma)| \;\leq\; 2D_{\rm tr}(\rho,\sigma) \ln \frac{N}{2D_{\rm tr}(\rho,\sigma)} \,. \tag{13.36}$$

Again we omit the proof (Fannes, 1973), but we take note of a rather interesting intermediate step: let the eigenvalues of ρ and σ be r_i and s_i, respectively, and assume that they have been arranged in decreasing order (e.g. $r_1 \geq r_2 \geq \cdots \geq r_N$). Then

$$D_{\rm tr}(\rho,\sigma) \;\geq\; \frac{1}{2}\sum_i |r_i - s_i| \,. \tag{13.37}$$

The closeness assumption in Fannes' lemma has to do with the fact that the function $-x\ln x$ is monotone on the interval $(0, 1/e)$. A weaker bound holds if it is not fulfilled.

The relative entropy between any two states is bounded by their trace distance by a quantum analogue of the Pinsker inequality (13.16)

$$S(\rho||\sigma) \geq 2[D_{\rm tr}(\rho,\sigma)]^2 \,. \tag{13.38}$$

The idea of the proof (Hiai, Ohya and Tsukada, 1981) is similar to that used in the classical case, that is, one relies on Eq. (13.17) and on the monotonicity of relative entropy.

What about relative entropy itself? The results of Hiai and Petz (1991), briefly reported in Section 12.3, can be paraphrased as saying that, in certain well-defined circumstances, the probability of error when performing measurements on a large number $\mathcal{N}$ of copies of a quantum system is be given by

$$P_{\rm E}(\rho,\sigma) \;=\; {\rm e}^{-\mathcal{N}S(\rho||\sigma)} \,. \tag{13.39}$$

That is to say, this is the smallest achievable probability of erroneously concluding that the state is ρ, given that the state in fact is σ. Although our account of the story ends here, the story itself does not. Let us just mention that the step from one to many samplings turns into a giant leap in quantum mechanics, because the set of all possible measurements on density operators such as $\rho\otimes\rho\otimes\cdots\otimes\rho$ will include sophisticated measurements performed on the whole ensemble, that cannot be described as measurements on the systems separately.[5]

[5] For further discussion of this interesting point, see Peres and Wootters (1991) and Bennett, DiVincenzo, Mor, Shor, Smolin and Terhal (1999b).

13.3 Fidelity and statistical distance

Among the quantum distinguishability measures, we single out the fidelity function for special attention. It is much used, and it is closely connected to the Bures geometry of quantum states. It was defined in Section 9.4 as

$$F(\rho_1, \rho_2) = \left(\mathrm{Tr}\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\right)^2 = \left(\mathrm{Tr}|\sqrt{\rho_1}\sqrt{\rho_2}|\right)^2 . \tag{13.40}$$

Actually, in Section 9.4 we worked mostly with the *root fidelity*

$$\sqrt{F}(\rho_1, \rho_2) \;=\; \mathrm{Tr}\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}\,. \tag{13.41}$$

But in some contexts fidelity is the more useful notion. If both states are pure it equals the transition probability between the states. A little more generally, suppose that one of the states is pure, $\rho_1 = |\psi\rangle\langle\psi|$. Then ρ_1 equals its own square root and in fact

$$F(\rho_1, \rho_2) \;=\; \langle\psi|\rho_2|\psi\rangle\,. \tag{13.42}$$

In this situation fidelity has a direct interpretation as the probability that the state ρ_2 will pass the yes/no test associated to the pure state ρ_1. It serves as a figure of merit in many statistical estimation problems.

This still does not explain why we use the definition (13.40) of fidelity – for the quantum noiseless coding theorem we used Eq. (13.42) only, and there are many expressions that reduce to this equation when one of the states is pure (such as $\mathrm{Tr}\rho_1\rho_2$). The definition not only looks odd, it has obvious drawbacks too: in order to compute it we have to compute two square roots of positive operators – that is to say that we must go through the labourious process of diagonalizing a Hermitian matrix twice. But on further inspection the virtues of fidelity emerge. The key statement about it is Uhlmann's theorem (proved in Section 9.4). The theorem says that $F(\rho_1, \rho_2)$ equals the maximal transition probability between a pair of purifications of ρ_1 and ρ_2. It also enjoys a number of other interesting properties (Jozsa, 1994):

(1) $0 \le F(\rho_1, \rho_2) \le 1$;
(2) $F(\rho_1, \rho_2) = 1$ if and only if $\rho_1 = \rho_2$ and $F(\rho_1, \rho_2) = 0$ if and only if ρ_1 and ρ_2 have orthogonal supports;
(3) Symmetry, $F(\rho_1, \rho_2) \;=\; F(\rho_2, \rho_1)$;
(4) Concavity, $F(\rho, a\rho_1 + (1-a)\rho_2) \;\ge\; aF(\rho, \rho_1) + (1-a)F(\rho, \rho_2)$;
(5) Multiplicativity, $F(\rho_1 \otimes \rho_2, \rho_3 \otimes \rho_4) = F(\rho_1, \rho_3)\, F(\rho_2, \rho_4)$;
(6) Unitary invariance, $F(\rho_1, \rho_2) = F(U\rho_1 U^\dagger, U\rho_2 U^\dagger)$;
(7) Monotonicity, $F(\Phi(\rho_1), \Phi(\rho_2)) \ge F(\rho_1, \rho_2)$, where Φ is a trace preserving CP map.

Root fidelity enjoys all these properties too, as well as the stronger property of joint concavity in its arguments.

It is interesting to prove property (3) directly. To do so, observe that the trace can be written in terms of the square roots of the non-zero eigenvalues λ_n of a positive operator, as follows:

$$\sqrt{F} = \sum_n \sqrt{\lambda_n}, \text{ where } AA^\dagger|\psi_n\rangle = \lambda_n|\psi_n\rangle\ , \qquad A \equiv \sqrt{\rho_1}\sqrt{\rho_2}\ . \tag{13.43}$$

But an easy argument shows that the non-zero eigenvalues of $AA^\dagger$ are the same as those of $A^\dagger A$:

$$AA^\dagger|\psi_n\rangle = \lambda_n|\psi_n\rangle \quad \Rightarrow \quad A^\dagger AA^\dagger|\psi_n\rangle = \lambda_n A^\dagger|\psi_n\rangle\ . \tag{13.44}$$

Unless $A^\dagger|\psi_n\rangle = 0$ this shows that any eigenvalue of $AA^\dagger$ is an eigenvalue of $A^\dagger A$. Therefore we can equivalently express the fidelity in terms of the square roots of the non-zero eigenvalues of $A^\dagger A$, in which case the roles of ρ_1 and ρ_2 are interchanged.

Property (7) is a key entry: fidelity is a *monotone* function. The proof (Barnum, Caves, Fuchs, Jozsa and Schumacher, 1996) is a simple consequence of Uhlmann's theorem (Section 9.4). We can find a purification of our density matrices, such that $F(\rho_1, \rho_2) = |\langle\psi_1|\psi_2\rangle|^2$. We can also introduce an environment – a rather 'mathematical' environment, but useful for our proof – that starts in the pure state $|0\rangle$, so that the quantum operation is described by a unitary transformation $|\psi\rangle|0\rangle \to U|\psi\rangle|0\rangle$. Then Uhlmann's theorem implies that $F(\Phi(\rho_1), \Phi(\rho_2)) \geq |\langle\psi_1|\langle 0|U^\dagger U|\psi_2\rangle|0\rangle|^2 = |\langle\psi_1|\langle 0|\psi_2\rangle|0\rangle|^2 = F(\rho_1, \rho_2)$. Thus the fidelity is non-decreasing with respect to any physical operation, including measurement.

Finally, we observe that the fidelity may be defined implicitly (Alberti, 1983) by

$$F(\rho_1, \rho_2) \ = \ \inf\Big[\mathrm{Tr}(A\rho_1)\,\mathrm{Tr}(A^{-1}\rho_2)\Big]\,, \tag{13.45}$$

where the infimum is taken over all invertible positive operators A. There is a closely related representation of the root fidelity as an infimum over the same set of operators A (Alberti and Uhlmann, 2000),

$$\sqrt{F}(\rho_1, \rho_2) = \frac{1}{2}\inf\Big[\mathrm{Tr}(A\rho_1) + \mathrm{Tr}(A^{-1}\rho_2)\Big]\,, \tag{13.46}$$

since after squaring this expression only cross terms contribute to (13.45).

In Section 9.4 we introduced the Bures distance as a function of the fidelity. This is also a monotone function, and no physical operation can increase it. It follows that the corresponding metric, the Bures metric, is a *monotone metric* under stochastic maps, and may be a candidate for a quantum version of the Fisher–Rao metric. It is a good candidate.[6] To see this, let us choose a POVM $\{E_i\}$. A given density

[6] The link between Bures and statistical distance was forged by Helstrom (1976), Holevo (1982), and Braunstein and Caves (1994). Our version of the argument follows Fuchs and Caves (1995). Actually the precise sense in

matrix ρ will respond with the probability distribution $P(E, \rho)$. For a pair of density matrices we can define the *quantum Bhattacharyya coefficient*

$$B(\rho, \sigma) \equiv \min_E B\Big(P(E, \rho), P(E, \sigma)\Big) = \min_E \sum_i \sqrt{p_i q_i}\,, \tag{13.47}$$

where

$$p_i = \mathrm{Tr} E_i \rho\,, \qquad q_i = \mathrm{Tr} E_i \sigma\,, \tag{13.48}$$

and the minimization is carried out over all possible POVMs. If we succeed in doing this, we will obtain a quantum analogue of the Fisher–Rao distance as a function of $B(\rho, \sigma)$.

We will assume that both density matrices are invertible. As a preliminary step, we rewrite p_i, using an arbitrary unitary operator U, as

$$p_i = \mathrm{Tr}\left((U\sqrt{\rho}\sqrt{E_i})(U\sqrt{\rho}\sqrt{E_i})^\dagger\right)\,. \tag{13.49}$$

Then we use the Cauchy–Schwarz inequality (for the Hilbert–Schmidt inner product) to set a lower bound:

$$\begin{aligned} p_i q_i &= \mathrm{Tr}\left((U\sqrt{\rho}\sqrt{E_i})(U\sqrt{\rho}\sqrt{E_i})^\dagger\right)\mathrm{Tr}\left((\sqrt{\sigma}\sqrt{E_i})(\sqrt{\sigma}\sqrt{E_i})^\dagger\right) \\ &\geq \left(\mathrm{Tr}\left((U\sqrt{\rho}\sqrt{E_i})(\sqrt{\sigma}\sqrt{E_i})^\dagger\right)\right)^2\,. \end{aligned} \tag{13.50}$$

Equality holds if and only if

$$\sqrt{\sigma}\sqrt{E_i} \;=\; \mu_i\, U\sqrt{\rho}\sqrt{E_i} \tag{13.51}$$

for some real number μ_i. Depending on the choice of U, this equation may or may not have a solution. Anyway, using the linearity of the trace, it is now easy to see that

$$\sum_i \sqrt{p_i q_i} \geq \sum_i \left|\mathrm{Tr}(U\sqrt{\rho}E_i\sqrt{\sigma})\right| \geq \left|\mathrm{Tr}\left(\sum_i U\sqrt{\rho}E_i\sqrt{\sigma}\right)\right| = \mathrm{Tr}(U\sqrt{\rho}\sqrt{\sigma})\,. \tag{13.52}$$

The question is: how should we choose U if we wish to obtain a sharp inequality? We have to make sure that Eq. (13.51) holds, and also that all the terms in (13.52) are positive. A somewhat tricky argument (Fuchs and Caves, 1995) shows that the

which the Bures metric is the analogue of the classical Fisher–Rao metric is quite subtle (Barndorff-Nielsen and Gill, 2000).

answer is

$$U = \sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\,\frac{1}{\sqrt{\sigma}}\,\frac{1}{\sqrt{\rho}}\,. \tag{13.53}$$

This gives $\sum_i \sqrt{p_i q_i} \geq \sqrt{F}(\rho, \sigma)$, where $\sqrt{F}$ is the root fidelity. The optimal POVM turns out to be a projective measurement, associated to the Hermitian observable

$$M = \frac{1}{\sqrt{\sigma}}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\,\frac{1}{\sqrt{\sigma}}\,. \tag{13.54}$$

The end result is that

$$B(\rho, \sigma) \equiv \min_E B\Big(P(E, \rho), P(E, \sigma)\Big) = \mathrm{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}} \equiv \sqrt{F}(\rho, \sigma)\,. \tag{13.55}$$

It follows that the Bures angle distance $D_{\mathrm{A}} = \cos^{-1}\sqrt{F(\rho, \sigma)}$ is precisely the Fisher–Rao distance, maximized over all the probability distributions that one can obtain by varying the POVM.

For the case when the two states to be distinguished are pure we have already seen (in Section 5.3) that the Fubini–Study distance is the answer. These two answers are consistent. But in the pure state case the optimal measurement is not uniquely determined, while here it is: we obtained an explicit expression for the observable that gives optimal distinguishability, namely M. The operator has an ugly look, but it has a name: it is the *geometric mean* of the operators σ^{-1} and ρ. As such it was briefly discussed in Section 12.1, where we observed that the geometric mean is symmetric in its arguments. From this fact it follows that $M(\sigma, \rho) = M^{-1}(\rho, \sigma)$. Therefore $M(\sigma, \rho)$ and $M(\rho, \sigma)$ define the same measurement. The operator M also turned up in our discussion of geodesics with respect to the Bures metric, in Eq. (9.57). When $N = 2$ this fact can be used to determine M in an easy way: draw the unique geodesic that connects the two states, given that we view the Bloch ball as a round hemi-3-sphere. This geodesic will meet the boundary of the Bloch ball in two points, and these points are precisely the eigenstates of M.

We now have a firm link between statistical distance and the Bures metric, but we are not yet done with it – we will come back to it in Chapter 14. Meanwhile, let us compare the three distances that we have brought into play (Table 13.1). The first observation is that the two monotone distances, trace and Bures, have the property that the distance between states of orthogonal support is maximal:

$$\mathrm{supp}(\rho) \perp \mathrm{supp}(\sigma) \quad \Leftrightarrow \quad D_{\mathrm{tr}}(\rho, \sigma) = 1 \quad \Leftrightarrow \quad D_{\mathrm{B}}(\rho, \sigma) = \sqrt{2}\,. \tag{13.56}$$

This is not true for the Hilbert–Schmidt distance. The second observation concerns a bound (Fuchs and van de Graaf, 1999) that relates fidelity (and hence the Bures

Table 13.1. *Metrics in the space of quantum states*

Metric	Bures	Hilbert–Schmidt	Trace
Is it Riemannian ?	Yes	Yes	No
Is it monotone ?	Yes	No	Yes

distance) to the trace distance, namely,

$$1 - \sqrt{F}(\rho, \sigma) \;\leq\; D_{\rm tr}(\rho, \sigma) \;\leq\; \sqrt{1 - F(\rho, \sigma)}\,. \tag{13.57}$$

To prove that the upper bound holds, observe that it becomes an equality for a pair of pure states (which is easy to check, since we can work in the two-dimensional Hilbert space spanned by the two pure states). But Uhlmann's theorem means that we can find a purification such that $F(\rho, \sigma) = |\langle\psi|\phi\rangle|^2$. In the purifying Hilbert space the bound is saturated, and taking a partial trace can only decrease the trace distance (because of its monotonicity), while the fidelity stays constant by definition. For the lower bound, see Problem 13.2.

The Bures and trace distances are both monotone, and the close relation between them means that, for many purposes, they can be used interchangeably. There exist also relations between the Bures and Hilbert–Schmidt distances, but the latter does not have the same fundamental importance. It is evident, from the way that the Bloch sphere is deformed by an orthographic projection from the flat Hilbert–Schmidt Bloch ball to the round Bures hemi-3-sphere, that it may happen that $D_2(\rho_a, \rho_b) > D_2(\rho_c, \rho_d)$ while $D_{\rm B}(\rho_a, \rho_b) < D_{\rm B}(\rho_c, \rho_d)$. To find a concrete example, place $\rho_a = \rho_c$ at the north pole, ρ_b on the surface, and ρ_d on the polar axis through the Bloch ball.

For $N = 2$ we can use the explicit formula (9.48) for the Bures distance to compare it with the flat Hilbert–Schmidt distance. Since for one-qubit states the trace and HS distances agree, we arrive in this way at strict bounds between $D_{\rm B} = D_{\rm B}(\rho_a, \rho_b)$ and $D_{\rm tr} = D_{\rm tr}(\rho_a, \rho_b)$ valid for any $N = 2$ states,

$$\sqrt{2 - 2\sqrt{1 - (D_{\rm tr})^2}} \;\leq\; D_B \;\leq\; \sqrt{2 - 2\sqrt{1 - D_{\rm tr}}}\,. \tag{13.58}$$

The lower bound comes from pure states. The upper bound comes from the family of mixed states situated on an axis through the Bloch ball, and does not hold in higher dimensions. However, making use of the relation (9.31) between Bures distance and fidelity we may translate the general bounds (13.57) into

$$\sqrt{2 - 2\sqrt{1 - (D_{\rm tr})^2}} \;\leq\; D_B \;\leq\; \sqrt{2D_{\rm tr}}. \tag{13.59}$$

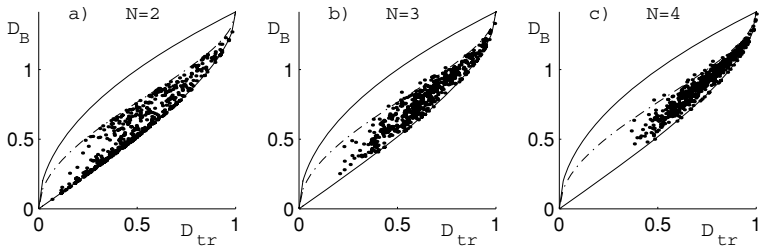


Figure 13.3. Bures distance plotted against trace distance for random density matrices of size (a) $N = 2$, (b) $N = 3$ and (c) $N = 4$. The single dots are randomly drawn density matrices, the solid lines denote the bounds (13.59), and the dotted lines the upper bound in (13.58) which holds for $N = 2$.

This upper bound, valid for an arbitary N, is not strict. Figure 13.3 presents Bures distances plotted as a function of the trace distance for an ensemble of 500 pairs consisting of a random pure state and a random mixed state distributed according to the Hilbert–Schmidt measure (see Chapter 14). The upper bound (13.58) is represented by a dashed curve, and violated for $N > 2$.

Problems

Problem 13.1 Prove that the flat metric on the classical probability simplex is monotone under bistochastic maps.

Problem 13.2 Complete the proof of the inequality (13.57).

Problem 13.3 Derive the inequalities (a): $F(\sigma, \rho) \geq \left(\mathrm{Tr}\sqrt{\sigma}\sqrt{\rho}\right)^2$ and (b): $F(\sigma, \rho) \geq \mathrm{Tr}\sigma\rho$. What are the conditions for equality?

14

Monotone metrics and measures

Probability theory is a measure theory – with a soul.
Mark Kac

Section 2.6 was devoted to *classical ensembles*, that is to say ensembles defined by probability measures on the set of classical probability distributions over N events. In this chapter *quantum ensembles* are defined by choosing probability measures on the set of density matrices of size N. A warning should be issued first: there is no single, naturally distinguished measure in $\mathcal{M}^{(N)}$, so we have to analyse several measures, each of them with different physical motivations, advantages and drawbacks. This is in contrast to the set of pure quantum states, where the Fubini–Study measure is the only natural choice for a measure that defines 'random states'.

A simple way to define a probability measure goes through a metric. Hence we will start this chapter with a review of the metrics defined on the set of mixed quantum states.

14.1 Monotone metrics

In Section 2.5 we explained how the Fisher metric holds its distinguished position due to the theorem of Čencov, which states that the Fisher metric is the unique monotone metric on the probability simplex Δ_{N-1}. Now that the latter has been replaced by the space of quantum states $\mathcal{M}^{(N)}$ we must look again at the question of metrics. Since the uniqueness in the classical case came from the behaviour under stochastic maps, we turn our attention to stochastic quantum maps – the completely positive, trace preserving maps discussed in Chapter 10. A distance D in the space of quantum states $\mathcal{M}^{(N)}$ is called *monotone* if it does not grow under the action of a stochastic map Φ,

$$D_{\text{mon}}\big(\Phi\rho, \Phi\sigma\big) \leq D_{\text{mon}}\big(\rho, \sigma\big) \,. \tag{14.1}$$

If a monotone distance is geodesic the corresponding metric on $\mathcal{M}^{(N)}$ is called monotone. However, in contrast to the classical case it turns out that there exist infinitely many monotone Riemannian metrics on the space of quantum states.

The appropriate generalization of Čencov's classical theorem is as follows:[1]

Theorem 14.1 (Morozova–Čencov–Petz's) *At a point where the density matrix is diagonal, $\rho = diag(\lambda_1, \lambda_2, \ldots, \lambda_N)$, every monotone metric on $\mathcal{M}^{(N)}$ assigns the length squared*

$$||A||^2 = \frac{1}{4}\left[C \sum_{i=1}^{N} \frac{A_{ii}^2}{\lambda_i} + 2 \sum_{i<j}^{N} c(\lambda_i, \lambda_j)\, |A_{ij}|^2 \right] \tag{14.2}$$

to any tangent vector A, where C is an arbitrary constant, the function $c(x, y)$ is symmetric,

$$c(x, y) = c(y, x)\,, \quad \text{and obeys} \quad c(sx, sy) = s^{-1} c(x, y)\,, \tag{14.3}$$

and the function $f(t) \equiv \frac{1}{c(t,1)}$ is operator monotone.

The tangent vector A is a traceless Hermitian matrix; if A is diagonal then the second term vanishes and the first term gives the familiar Fisher–Rao metric on the simplex. But the second term is new. And the result falls far short of providing uniqueness. Any function $f(t) : \mathbb{R}_+ \to \mathbb{R}_+$ will be called a *Morozova–Chentsov* (MC) function, if it fulfils three restrictions:

$$\begin{aligned} &\textbf{i)} && f \quad \text{is operator monotone,} \\ &\textbf{ii)} && f \quad \text{is self inversive:} \quad f(1/t) = f(t)/t\,, \\ &\textbf{iii)} && f(1) = 1\,. \end{aligned} \tag{14.4}$$

The meaning of condition (i) was discussed at some length in Section 12.1. Condition (ii) was also encountered there, as the condition for an operator mean to be symmetric. Here it ensures that the function $c(x, y) = 1/[yf(x/y)]$ satisfies (14.3). Condition (iii) is a normalization with consequences – it forces us to set $C = 1$ in order to avoid a conical singularity at the maximally mixed state, as we will soon see explicitly in the case $N = 2$. The metric is now said to be *Fisher adjusted*. The Morozova–Čencov–Petz (MCP) theorem can now be rephrased: there is a one-to-one correspondence between MC functions and monotone Riemannian metrics.

The infinite-dimensional set $\mathcal{F}$ of all MC functions is convex; the set of operator monotone functions itself is convex, and an explicit calculation shows that any

[1] Here we have merged the theorem by Morozova and Čencov (1990) with a theorem by Petz (1996) that completed it. See also Petz and Sudár (1996), Lesniewski and Ruskai (1999) and, for a guided tour through the garden of monotone metrics, Petz (1998).

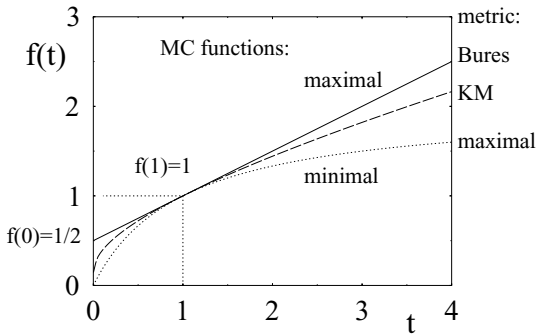


Figure 14.1. Morozova–Chentsov functions $f(t)$: minimal function for the maximal metric (dotted), Kubo–Mori metric (dashed), and maximal function for the minimal Bures metric (solid line). These metrics are Fisher-adjusted, $f(1) = 1$. The Bures metric is also Fubini–Study adjusted, $f(0) = 1/2$.

convex combination of two self-inversive functions is self-inversive. A monotone Riemannian metric will be called *pure*, if the corresponding MC function f is an extreme point of the convex set $\mathcal{F}$. Among all operator monotone functions on $[0, +\infty)$ which are self inversive and obey $f(1) = 1$, there exists a minimal and a maximal function (Kubo and Ando, 1980). In Figure 14.1 we plot three choices:

$$f_{\min}(t) = \frac{2t}{t+1}\,, \qquad f_{\mathrm{KM}}(t) = \frac{t-1}{\ln t}\,, \qquad f_{\max}(t) = \frac{1+t}{2}\,. \tag{14.5}$$

The maximal MC function $f_{\max}$ is a straight line and gives rise to the Bures metric, while the minimal function $f_{\min}$ is a hyperbola. The intermediate case f_{KM} leads to the Kubo–Mori metric used in quantum statistical mechanics. Our MC functions correspond to

$$c_{\min}(x, y) = \frac{x+y}{2xy}, \quad c_{\mathrm{KM}}(x, y) = \frac{\ln x - \ln y}{x-y}, \quad c_{\max}(x, y) = \frac{2}{x+y}, \tag{14.6}$$

so the inverse, $1/c$, is equal to the harmonic, logarithmic and arithmetic mean, respectively. For other interesting choices of $f(t)$, see Problem 14.1.

To familiarize ourselves with the frightening expression (14.2) we can take a look at the case $N = 2$, the Bloch ball. We set $\rho = \frac{1}{2}\mathrm{diag}(1+r, 1-r)$ and find, after a minor calculation, that the metric is

$$\mathrm{d}s^2 = \frac{1}{4}\left[\frac{\mathrm{d}r^2}{1-r^2} + \frac{1}{f\left(\frac{1-r}{1+r}\right)}\frac{r^2}{1+r}\,\mathrm{d}\Omega^2\right], \qquad 0 < r < 1\,. \tag{14.7}$$

Here $\mathrm{d}\Omega^2$ is the metric on the unit 2-sphere – the second term corresponds to the second, tangential, term in Eq. (14.2), and we used spherical symmetry to remove

the restriction to diagonal density matrices. We can now see, given $C = 1$, that the condition $f(1) = 1$ means that the metric is regular at the origin. We also see that $f(0) = 0$ means that the area of the boundary is infinite. If $f(0)$ is finite the boundary is a round 2-sphere. It will have radius $1/2$ if $f(0) = 1/2$; such a metric is said to be *Fubini–Study adjusted*.

Because the Morozova–Chentsov function f appears in the denominator, the larger the (normalized) function, the smaller the area of a sphere at constant r. So, slightly confusingly, the metric that uses $f_{\min}$ will be called the *maximal metric* and the metric using $f_{\max}$ will be called the *minimal metric*.

Let us now go through the geometries corresponding to our three choices of the function $f(t)$. If $f = f_{\max}$ then

$$\mathrm{d}s^2_{\min} = \frac{1}{4}\left[\frac{\mathrm{d}r^2}{1-r^2} + r^2\mathrm{d}\Omega^2\right] . \tag{14.8}$$

This is the metric on a round 3-sphere of radius $1/2$; the scalar curvature is $R_{\max} = 4 \cdot 6$ and the boundary of the Bloch ball is a round 2-sphere corresponding to the equator of the 3-sphere. This is a Fubini–Study adjusted metric. In fact it is the Bures metric that we have encountered in Chapters 9 and 13. Since it is given by the *maximal* function $f_{\max}(t)$, it is distinguished by being the *minimal* Fisher-adjusted, monotone Riemannian metric. If $f = f_{\min}$ then

$$\mathrm{d}s^2_{\max} = \frac{1}{4}\left[\frac{1}{1-r^2}(\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2)\right] . \tag{14.9}$$

The curvature is everywhere negative and diverges to minus infinity at $r = 1$. Indeed the area of the boundary – the space of pure states – is infinite. For the intermediate choice $f = f_{\mathrm{KM}}$ we get

$$\mathrm{d}s^2_{\mathrm{KM}} = \frac{1}{4}\left[\frac{\mathrm{d}r^2}{1-r^2} + \frac{r}{2}\ln\left(\frac{1+r}{1-r}\right)\mathrm{d}\Omega^2\right] . \tag{14.10}$$

The curvature is zero at the origin and decreases (very slowly at first) to minus infinity at the boundary, which again has a diverging area although the divergence is only logarithmic.

Now where does Eq. (14.2) come from? Recall from Section 2.5 that there were two natural definitions of the (uniquely monotone) Fisher–Rao metric. We can define it as the Hessian of the relative entropy function, or in terms of expectation values of score vectors (logarithmic derivatives of the probability distribution). In the classical case these definitions led to the same metric. But in the quantum case this is no longer so. The logarithmic derivative of a density matrix is an ambiguous notion because we have entered a non-commutative probability theory. To see this let us assume that we are looking at a set of density matrices ρ_θ parametrized by

some parameters θ^a. For simplicity we consider an affine parametrization of $\mathcal{M}^{(N)}$ itself,

$$\rho_\theta = \frac{1}{N}\mathbb{1} + \sum_{a=1}^{N} \theta^a A_a \,, \tag{14.11}$$

where A_a are a set of traceless Hermitian matrices, that is to say that they are tangent vectors of $\mathcal{M}^{(N)}$. Evidently

$$\partial_a \rho_\theta = A_a \,. \tag{14.12}$$

We may define the logarithmic derivative, that is a *quantum score vector* L_a, by

$$A_a = \rho_\theta L_a \quad \Rightarrow \quad L_a = \rho_\theta^{-1} A_a \,. \tag{14.13}$$

In analogy with the classical equation (2.65), we define the squared length of a tangent vector as

$$||A||^2 = \frac{1}{4}\mathrm{Tr}\rho_\theta LL = \frac{1}{4}\mathrm{Tr}AL = \frac{1}{4}\mathrm{Tr}\rho_\theta^{-1}A^2 \,. \tag{14.14}$$

This defines a metric. If we set $\rho = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ and perform a pleasant calculation we find that it is exactly the maximal metric, that is the metric (14.2) for the choice $f = f_{\min}$.

But other definitions of the logarithmic derivative suggest themselves, such as the *symmetrical logarithmic derivative* L_a occurring in the equation

$$A_a = \rho_\theta \circ L_a = \frac{1}{2}(\rho_\theta L_a + L_a \rho_\theta) \,, \tag{14.15}$$

where we found a use for the Jordan product from Section 8.6.[2] This equation for L_a was first encountered in Eq. (9.27), and there we claimed that there is a unique solution if ρ_θ is invertible. To find it, we first choose a basis where ρ_θ is diagonal with eigenvalues λ_i. The equation for the matrix elements of $L_a \equiv L$ becomes

$$A_{ij} = \frac{1}{2}(\lambda_i L_{ij} + L_{ij}\lambda_j) = \frac{1}{2}(\lambda_i + \lambda_j)L_{ij} \tag{14.16}$$

(where no summation is involved). The solution is immediate:

$$L_{ij} = \frac{2}{\lambda_i + \lambda_j} A_{ij} \,. \tag{14.17}$$

The length squared of the vector A becomes

$$||A||^2 \equiv \frac{1}{4}\mathrm{Tr}AL = \frac{1}{2}\sum_{i,j} \frac{A_{ij}A_{ji}}{\lambda_i + \lambda_j} = \frac{1}{4}\sum_i \frac{A_{ii}^2}{\lambda_i} + \sum_{i<j} \frac{|A_{ij}|^2}{\lambda_i + \lambda_j} \,. \tag{14.18}$$

[2] The symmetric logarithmic derivative was first introduced by Helstrom (1976) and Holevo (1982).

Again a pleasant calculation confirms that this agrees with Eq. (14.2), this time for the choice $f = f_{\max}$.

The Kubo–Mori metric is a different kettle of fish.[3] It arises when we try to define a metric as the Hessian of the relative entropy. It takes a little effort to do this explicitly. We need to know that, for any positive operator A,

$$\ln(A + xB) - \ln A = \int_0^\infty \frac{1}{A+u} xB \frac{1}{A+xB+u}\, \mathrm{d}u \,. \tag{14.19}$$

To prove this is an exercise (namely Problem 12.1). It follows that

$$\partial_x \ln(\rho + xA)_{|x=0} = \int_0^\infty \frac{1}{\rho+u} A \frac{1}{\rho+u}\, \mathrm{d}u \,. \tag{14.20}$$

With this result in hand it is straightforward to compute

$$-\partial_\alpha \partial_\beta S(\rho + \alpha A || \rho + \beta B)_{|\alpha=\beta=0} = \int_0^\infty \mathrm{Tr} A(\rho+u)^{-1} B(\rho+u)^{-1}\, \mathrm{d}u \,. \tag{14.21}$$

This is precisely the expression that defines the *Kubo–Mori* scalar product $g(A, B)(\rho)$, in analogy with the classical equation (2.59). If we evaluate this expression for a diagonal density matrix ρ we recover Eq. (14.2), with the function $c = c_{\mathrm{KM}}$ given in Eq. (14.6).

The Kubo–Mori metric does have a uniqueness property: it is the only monotone metric for which the two flat affine connections, mixture and exponential, are mutually dual (Grasseli and Streater, 2001). Therefore it allows us to play a quantum version of the classical game played in Section 3.2. There is also a conjecture by Petz (1994), that says that the Kubo–Mori metric is the only metric having the property that the scalar curvature is monotone, in the sense that if $\rho_1 \prec \rho_2$, that is if ρ_1 is majorized by ρ_2, then $R(\rho_1) \leq R(\rho_2)$ (which is not true for the Bures metric, as mentioned in Section 9.6). On the other hand the fact that the Kubo–Mori metric is the Hessian of the relative entropy is not really a uniqueness property, because it can be shown that every monotone metric is the Hessian of a certain monotone relative entropy – not Umegaki's relative entropy, but one of the larger family of monotone relative entropies[4] whose existence we hinted at in Section 12.4.

14.2 Product measures and flag manifolds

It will not have escaped the reader that the price of putting a Riemannian geometry on $\mathcal{M}^{(N)}$ is rather high, in the sense that the monotone metrics that appear are quite difficult to work with when $N > 2$. Fortunately the measures that come from our

[3] Studies of this metric include those by Ingarden (1981), Balian, Alhassid and Reinhardt (1986) and Petz (1994).
[4] Lesniewski and Ruskai (1999) explored this matter in depth; see also Jenčova (2004).

monotone metrics are not that difficult to handle, so that calculations of volumes and the like are quite doable. The basic trick that we will use is the same as that one uses in flat space, when the Euclidean measure is decomposed into a product of a measure on the orbits of the rotation group and a measure in the radial direction (in other words, when one uses spherical polar coordinates). The set of density matrices that can be written in the form $\rho = U\Lambda U^\dagger$, for a fixed diagonal matrix Λ with strictly positive eigenvalues, is a flag manifold $\mathbf{F}^{(N)} = U(N)/[U(1)]^N$ (see Section 8.5). A natural assumption concerning a probability distribution in $\mathcal{M}^{(N)}$ is to require invariance with respect to unitary rotations, $P(\varrho) = P(W\varrho W^\dagger)$. This is the case if (a) the choice of eigenvalues and eigenvectors is independent, and (b) the eigenvectors are drawn according to the Haar measure, $\mathrm{d}\nu_H(W) = \mathrm{d}\nu_H(UW)$.

Such a *product measure*, $\mathrm{d}V = \mathrm{d}\nu_H(U) \times \mathrm{d}\mu(\vec{\lambda})$, defined on the Cartesian product $\mathbf{F}^{(N)} \times \Delta_{N-1}$, leads to the probability distribution,

$$P(\rho) = P_H(\mathbf{F}^{(N)}) \times P(\vec{\lambda}) \tag{14.22}$$

in which the first factor denotes the natural, unitarily invariant distribution on the flag manifold $\mathbf{F}^{(N)}$, induced by the Haar measure on $U(N)$.

We are going to compute the volume of $\mathbf{F}^{(N)}$ with respect to this measure. Let us rewrite the complex flag manifold as a Cartesian product of complex projective spaces,

$$\begin{aligned}\mathbf{F}^{(N)} = \frac{U(N)}{[U(1)]^N} &\simeq \frac{U(N)}{U(N-1)\times U(1)}\frac{U(N-1)}{U(N-2)\times U(1)}\cdots\frac{U(2)}{U(1)\times U(1)} \\ &\simeq \mathbb{C}\mathbf{P}^{N-1}\times\mathbb{C}\mathbf{P}^{N-2}\times\cdots\times\mathbb{C}\mathbf{P}^1,\end{aligned} \tag{14.23}$$

where $\simeq$ means 'equal, if the volumes are concerned'. This ignores a number of topological complications, but our previous experience with fibre bundles, say in Section 4.7, makes it at least plausible that we can proceed like this. (And the result is correct!) Making use of Eq. (4.87), $\mathrm{Vol}(\mathbb{C}\mathbf{P}^k) = \pi^k/k!$, we find for $N \geq 2$

$$\mathrm{Vol}(\mathbf{F}^{(N)}) = \prod_{k=1}^{N-1}\mathrm{Vol}(\mathbb{C}\mathbf{P}^k) = \frac{\pi^{N(N-1)/2}}{\Xi_N}\,, \tag{14.24}$$

with

$$\Xi_N \equiv 0!\,1!\,2!\ldots(N-1)! = \prod_{k=1}^{N}\Gamma(k)\,. \tag{14.25}$$

The result for $\mathrm{Vol}(\mathbf{F}^{(N)})$ still depends on a free multiplicative factor which sets the scale, just as the volume of a sphere depends on its radius. In Eq. (14.24) we have implicitly fixed the scale by the requirement that a closed geodesic on $\mathbb{C}\mathbf{P}^k$ has circumference π. In this way we have adjusted the metric with Eq. (3.137), which

Table 14.1. *Volumes of flag manifolds and unitary groups;* $\Xi_N = \prod_{k=1}^{N} \Gamma(k)$.

Manifold	Dimension	$\mathrm{Vol}[X]$, $a = 1/2$	$\mathrm{Vol}'[X]$, $a = 1$
$\mathbb{C}P^N$	$2N$	$\frac{\pi^N}{N!}$	$\frac{(2\pi)^N}{N!}$
$\mathbf{F}^{(N)} = \frac{U(N)}{[U(1)]^N}$	$N(N-1)$	$\frac{\pi^{N(N-1)/2}}{\Xi_N}$	$\frac{(2\pi)^{N(N-1)/2}}{\Xi_N}$
$U(N)$	N^2	$2^{N/2}\frac{\pi^{N(N+1)/2}}{\Xi_N}$	$\frac{(2\pi)^{N(N+1)/2}}{\Xi_N}$
$SU(N)$	N^2-1	$2^{(N-1)/2}\sqrt{N}\frac{\pi^{(N+2)(N-1)/2}}{\Xi_N}$	$\sqrt{N}\frac{(2\pi)^{(N+2)(N-1)/2}}{\Xi_N}$

corresponds to the following normalization of the measure on the unitary group

$$\mathrm{d}s^2 \equiv -a\,\mathrm{Tr}(U^{-1}\mathrm{d}U)^2 \quad \text{with} \quad a = \frac{1}{2}. \tag{14.26}$$

Direct integration over the circle $U(1)$ gives $\mathrm{Vol}[U(1)] = 2\pi\sqrt{a} = \sqrt{2}\pi$. This result combined with (14.24) allows us to write the volume of the unitary group[5]

$$\mathrm{Vol}[U(N)] = \mathrm{Vol}(\mathbf{F}^{(N)})\big(\mathrm{Vol}[U(1)]\big)^N = 2^{N/2}\frac{\pi^{N(N+1)/2}}{\Xi_N}. \tag{14.27}$$

For completeness let us state an analogous results for the special unitary group,

$$\mathrm{Vol}[SU(N)] = 2^{(N-1)/2}\,\sqrt{N}\,\frac{\pi^{(N+2)(N-1)/2}}{\Xi_N}. \tag{14.28}$$

This is again the quotient of the volumes of $U(N)$ and $U(1)$, but it is a different $U(1)$ than the one above, hence the presence of the stretching factor $\sqrt{N}$.

Let us return to the discussion of the product measures (14.22). The second factor $P(\vec{\lambda})$ may in principle be an arbitrary measure on the simplex Δ_{N-1}. To obtain a measure supported on the set of mixed states one may use the Dirichlet distribution (2.73), for example with the parameter s equal to 1 and $1/2$, for the flat and round measures on the simplex. These measures were called *unitary* and *orthogonal product measures* (Życzkowski, 1999), respectively, since they represent the distribution of the squared components of a random complex (real) vector and are induced by the Haar measures on the unitary (orthogonal) groups. The unitary product measure P_u is induced by the coarse graining map (see Eq. (14.72)) but, as discussed in the following sections of this chapter, other product measures seem to be better motivated by physical assumptions.

[5] This result was derived by Hua (1963) with the normalization $a = 1$, for which the volume of the circle becomes $\mathrm{Vol}'[U(1)] = 2\pi$. Different normalizations give proportional volumes, $\mathrm{Vol}'(X) = (\sqrt{2})^{\dim(X)}\mathrm{Vol}(X)$. For the reader's convenience we have collected the results for both normalizations in Table 14.1. For orthogonal groups, see Problem 14.2. For more details consult (Boya, Sudarshan and Tilma, 2003; Życzkowski and Sommers, 2003) and references therein.

14.3 Hilbert–Schmidt measure

A metric always generates a measure. For Riemannian metrics this was explained in Section 1.4, but we can also use more exotic metrics like the L_p-metrics defined in section 13.2. All L_p-metrics, including the trace metric for $p = 1$, will generate the same *Lebesque measure*. Here we concentrate on the $p = 2$ case, the Hilbert–Schmidt metric, and when deriving the measure we treat it in the same way that we would treat any Riemannian metric. The Hilbert–Schmidt metric is defined by the line element squared,

$$\mathrm{d}s^2_{\mathrm{HS}} = \frac{1}{2}\mathrm{Tr}[(\mathrm{d}\rho)^2]\,, \tag{14.29}$$

valid for any dimension N.[6] Making use of the diagonal form $\rho = U\Lambda U^{-1}$ and of the differentiation rule of Leibnitz, we write

$$\mathrm{d}\rho \;=\; U\,[\mathrm{d}\Lambda + U^{-1}\mathrm{d}U\Lambda - \Lambda U^{-1}\mathrm{d}U]\,U^{-1}. \tag{14.30}$$

Hence (14.29) can be rewritten as

$$\mathrm{d}s^2_{\mathrm{HS}} \;=\; \frac{1}{2}\sum_{i=1}^{N}(\mathrm{d}\lambda_i)^2 + \sum_{i<j}^{N}(\lambda_i-\lambda_j)^2\,|(U^{-1}\mathrm{d}U)_{ij}|^2. \tag{14.31}$$

Due to the normalization condition $\sum_{i=1}^N \lambda_i = 1$, the sum of differentials vanishes, $\sum_{i=1}^N \mathrm{d}\lambda_i = 0$. Thus we may consider the variation of the Nth eigenvalue as a dependent one, $\mathrm{d}\lambda_N = -\sum_{i=1}^{N-1}\mathrm{d}\lambda_i$, which implies

$$\frac{1}{2}\sum_{i=1}^{N}(\mathrm{d}\lambda_i)^2 = \frac{1}{2}\sum_{i=1}^{N-1}(\mathrm{d}\lambda_i)^2 + \frac{1}{2}\Big(\sum_{i=1}^{N-1}\mathrm{d}\lambda_i\Big)^2 = \sum_{i,j=1}^{N-1}\mathrm{d}\lambda_i\, g_{ij}\,\mathrm{d}\lambda_j. \tag{14.32}$$

The corresponding volume element gains a factor $\sqrt{\det g}$, where g is the metric in the $(N-1)$-dimensional simplex Δ_{N-1} of eigenvalues. From (14.32) one may read out the explicit form of the metric

$$g = \frac{1}{2}\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}. \tag{14.33}$$

It is easy to check that the $(N-1)$-dimensional matrix g has one eigenvalue equal to $N/2$ and $N-2$ eigenvalues equal to $1/2$, so that $\det g = N/2^{N-1}$. Thus the Hilbert–Schmidt volume element has the product form

$$\mathrm{d}V_{\mathrm{HS}} = \frac{\sqrt{N}}{2^{(N-1)/2}}\prod_{j=1}^{N-1}\mathrm{d}\lambda_j \prod_{j<k}^{1\ldots N}(\lambda_j-\lambda_k)^2 \left|\prod_{j<k}^{1\ldots N}\mathrm{Re}(U^{-1}\mathrm{d}U)_{jk}\mathrm{Im}(U^{-1}\mathrm{d}U)_{jk}\right|. \tag{14.34}$$

[6] As in Chapter 8, a factor $1/2$ is included here to ensure that the length of a closed geodesic in $\mathbb{C}\mathbf{P}^n$ equals π.

The first factor, depending only on the spectrum $\vec{\lambda}$, induces the *Hilbert–Schmidt probability distribution* in the simplex of eigenvalues (Braunstein, 1996; Hall, 1998),

$$P_{\rm HS}(\vec{\lambda}) \;=\; C_N^{\rm HS}\,\delta(1-\sum_{j=1}^{N}\lambda_j)\prod_{j<k}^{N}(\lambda_j-\lambda_k)^2, \tag{14.35}$$

where the normalization constant

$$C_N^{\rm HS} \;=\; \frac{\Gamma(N^2)}{\prod_{k=1}^{N}\Gamma(k)\Gamma(k+1)} = \frac{\Gamma(N^2)}{\Xi_N\,\Xi_{N+1}} \tag{14.36}$$

ensures that the integral over the simplex Δ_{N-1} is equal to unity (Życzkowski and Sommers, 2001).

Let us now analyse the second factor in (14.34). It arises from the off-diagonal elements of the invariant metric (14.26) on the unitary group

$$ds^2 \equiv -\frac{1}{2}{\rm Tr}(U^{-1}dU)^2 = \frac{1}{2}\sum_{j=1}^{N}|(U^{-1}dU)_{jj}|^2 + \sum_{j<k=1}^{N}|(U^{-1}dU)_{jk}|^2\,. \tag{14.37}$$

Hence the Hilbert–Schmidt measure in the space of density matrices belongs to the class of product measures (14.22).

Since the diagonal elements of (14.37) do not contribute to the volume element (14.34), integrating it over the unitary group we obtain the volume of the complex flag manifold (14.24). To compute the Hilbert–Schmidt volume of the set of mixed states $\mathcal{M}^{(N)}$ we need also to integrate (14.34) over the simplex of eigenvalues Δ_{N-1}. Normalization (14.35) implies that the latter integral is equal to $1/C_N^{\rm HS}$. Taking into account the prefactor $\sqrt{\det g} = \sqrt{N/2^{N-1}}$, present in (14.34), we obtain[7]

$${\rm Vol}_{\rm HS}\big(\mathcal{M}^{(N)}\big) = \sqrt{\frac{N}{2^{N-1}}}\;\frac{{\rm Vol}\big(\mathbf{F}^{(N)}\big)}{N!\;C_N^{\rm HS}} = \sqrt{N}\;\frac{\pi^{N(N-1)/2}}{2^{(N-1)/2}}\;\frac{\Gamma(1)\cdots\Gamma(N)}{\Gamma(N^2)}. \tag{14.38}$$

The factor $N!$ in the denominator compensates for the fact that we should really integrate over only one Weyl chamber in the eigenvalue simplex. In other words, different permutations of the vector $\vec{\lambda}$ of N generically different eigenvalues belong to the same unitary orbit, so we may restrict the order of the eigenvalues to, say, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. Substituting $N=2$ it is satisfying to receive that $V_2 = \pi/6$ – exactly the volume of the Bloch ball of radius $1/2$.

The next result $V_3 = \pi^3/(13\,440\sqrt{3})$ allows us to characterize the difference between the set $\mathcal{M}^{(3)} \subset \mathbb{R}^8$ and the ball $\mathbf{B}^8$. The set of mixed states is inscribed

[7] This differs from the result in (Życzkowski and Sommers, 2003) by a factor $2^{(N^2-1)/2}$, since in that work the length of closed geodesic in $\mathbb{C}\mathbf{P}^n$ was set to 2π.

Table 14.2. *Scaling properties of convex bodies in $\mathbb{R}^D$ with volume V, area A. Radii of the inspheres and outspheres are denoted by r and R.*

Body	X	dim	$\zeta = \frac{R}{r}$	$\eta = R\frac{A}{V}$
Round ball	$\bigcirc_D$	D	D^0	D^1
Cube	$\boxtimes_D$	D	$D^{1/2}$	$D^{3/2}$
Quantum states	$\mathcal{M}^{(N)}$	N^2-1	$D^{1/2}$	$D^{3/2}$
Cross-polytope	$\diamond_D$	D	$D^{1/2}$	$D^{3/2}$
Simplex	Δ_D	D	D^1	D^2

into a sphere of radius $R_3 = \sqrt{1/3} \approx 0.577$, while the maximal ball contained inside has the radius $r_3 = R_3/2 \approx 0.289$. These numbers may be compared with the *volume radius* $\bar{r}_3$ of $\mathcal{M}^{(3)}$, that is the radius of an 8-ball of the same volume. Using Eq. (1.22) we find $\bar{r}_3 \approx 0.368$. The distance from the centre of $\mathcal{M}^{(3)}$ to its boundary varies with the direction in $\mathbb{R}^8$ from r_3 to R_3.

The volume V_N tends to zero if $N \to \infty$, but there is no reason to worry about it. The same is true for the volume (1.22) of the N-balls, as it is a consequence of the choice of the units. To get some more information on the properties of the set $\mathcal{M}^{(N)}$, let us compute the (hyper)area of its boundary. Although the boundary $\partial\mathcal{M}^{(N)}$ is far from being trivial and contains orbits of different dimensionality, its full measure is formed of density matrices with exactly one eigenvalue equal to zero (see Section 8.5). Hence the hyperarea A of the boundary may be computed by integrating (14.34) over $\Delta_{N-2} \times \mathbf{F}^{(N)}$. The result is (Życzkowski and Sommers, 2003)

$$\mathrm{Vol}_{\mathrm{HS}}\big(\partial\mathcal{M}^{(N)}\big) \;=\; \sqrt{N-1}\,\frac{\pi^{N(N-1)/2}}{2^{(N-2)/2}}\,\frac{\Gamma(1)\ldots\Gamma(N+1)}{\Gamma(N)\,\Gamma(N^2-1)}. \tag{14.39}$$

For $N = 2$ we get π, just the area of the Bloch sphere of radius $R_2 = 1/2$.

Knowing the HS volume of the set of mixed states and the area of its boundary, we may compute their ratio. As in Section 1.2 we fix the scale by multiplying the ratio with the radius of the outsphere, $R_N = \sqrt{(N-1)/2N}$, compare Eq. (8.35). The result is

$$\eta_N^{\mathrm{HS}} \;\equiv\; R_N\,\frac{\mathrm{Vol}_{\mathrm{HS}}\big(\partial\mathcal{M}^{(N)}\big)}{\mathrm{Vol}_{\mathrm{HS}}\big(\mathcal{M}^{(N)}\big)} \;=\; (N-1)(N^2-1)\,. \tag{14.40}$$

It grows fast with N. We can compare this with the results obtained for balls, simplices, and cubes – if we remember that the dimension of $\mathcal{M}^{(N)}$ is $D = N^2 - 1$. Then we see that the area/volume ratio for the mixed states scales with the dimensionality[8] as $\eta \sim D^{3/2}$, just as for D-cubes (1.29): see Table 14.2. It is

[8] The same scaling also describes the set of real density matrices (Życzkowski and Sommers, 2003).

remarkable that the area to volume ratio for $\mathcal{M}^{(N)}$ behaves just like that for cubes of the same dimension. Note that $\eta = DR/r$ for all the entries, where r is the radius of the insphere. This means that $r\,\mathrm{vol}(\partial X)/\mathrm{vol}(X) = D$ for all of them. At the end of Section 1.2, we explained exactly what such a result implies about a convex body.

14.4 Bures measure

We can repeat the same analysis for the monotone metrics given by the MCP theorem; they may be much more complicated to deal with than the Hilbert–Schmidt metric, but they are unitarily invariant, so all of them will lead to product measures of the form (14.22). To see how this works, we first observe that in a coordinate system where ρ is diagonal Eq. (14.30) simplifies to

$$\mathrm{d}\rho_{ij} = \mathrm{d}\lambda_i \delta_{ij} + (\lambda_j - \lambda_i)(U^{-1}\mathrm{d}U)_{ij} \,. \tag{14.41}$$

Substituting this into the expression that defines the metric, Eq. (14.2), we obtain

$$\mathrm{d}s_f^2 = \frac{1}{4}\left[\sum_{j=1}^{N}\frac{\mathrm{d}\lambda_j^2}{\lambda_j} + 2\sum_{i<j} c(\lambda_i,\lambda_j)(\lambda_i-\lambda_j)^2 |(U^{-1}\mathrm{d}U)_{ij}|^2\right] , \tag{14.42}$$

where f is an arbitrary MC function and $c(x, y) = 1/[yf(x/y)]$ enters the definition (14.2). For $c(x, y) = c_{\max}(x, y) = 2/(x + y)$ this becomes the Bures metric, our topic in this section:

$$\mathrm{d}s_{\mathrm{B}}^2 \;=\; \frac{1}{4}\sum_{i=1}^{N}\frac{(\mathrm{d}\lambda_i)^2}{\lambda_i} + \sum_{i<j}\frac{(\lambda_i-\lambda_j)^2}{\lambda_i+\lambda_j}\,|(U^{-1}\mathrm{d}U)_{ij}|^2 \,. \tag{14.43}$$

Since $\mathrm{Tr}\rho = 1$ not all $\mathrm{d}\lambda_j$ are independent. Eliminating $\mathrm{d}\lambda_N$ we obtain

$$\frac{1}{4}\sum_{j=1}^{N}\frac{(\mathrm{d}\lambda_j)^2}{\lambda_j} = \frac{1}{4}\sum_{j=1}^{N-1}\frac{(\mathrm{d}\lambda_j)^2}{\lambda_j} + \frac{1}{4}\Big(\sum_{j=1}^{N-1}\mathrm{d}\lambda_j\Big)^2\frac{1}{\lambda_N} \,. \tag{14.44}$$

The metric (14.44) in the $(N-1)$-dimensional simplex is of the form $g_{ik} = (\delta_{ik}/4\lambda_i + \mathbb{1}/4\lambda_N)$ with determinant $\det g = (\lambda_1+\lambda_2+\cdots+\lambda_N)/(4^{N-1}\lambda_1\lambda_2\ldots\lambda_N) = 4^{1-N}/\det\rho$, since $\mathrm{Tr}\rho = 1$. Thus the volume element gains a factor $\sqrt{\det g} = 2^{1-N}(\det\rho)^{-1/2}$, so we obtain an expression

$$\mathrm{d}V_{\mathrm{B}} = \frac{1}{2^{N-1}}\frac{1}{\sqrt{\det\rho}}\prod_{j=1}^{N-1}\mathrm{d}\lambda_j\;\prod_{j<k}\frac{(\lambda_j-\lambda_k)^2}{\lambda_j+\lambda_k}\;\Big|\prod_{j<k}^{1\ldots N}\mathrm{Re}(U^{-1}\mathrm{d}U)_{jk}\mathrm{Im}(U^{-1}\mathrm{d}U)_{jk}\Big| \tag{14.45}$$

analogous to (14.34). This volume element gives the *Bures probability distribution* in the eigenvalue simplex (Hall, 1998; Caves, n.d.),

$$P_{\rm B}(\lambda_1,\ldots,\lambda_N) \;=\; C_N^{\rm B}\,\frac{\delta(1-\sum_{i=1}^N\lambda_i)}{(\lambda_1\ldots\lambda_N)^{1/2}}\prod_{j<k}^{N}\frac{(\lambda_j-\lambda_k)^2}{\lambda_j+\lambda_k}. \tag{14.46}$$

The Bures normalization constants[9] read (Sommers and Życzkowski, 2003)

$$C_N^{\rm B} = 2^{N^2-N}\frac{\Gamma(N^2/2)}{\pi^{N/2}\,\Gamma(1)\ldots\Gamma(N+1)} = 2^{N^2-N}\frac{\Gamma(N^2/2)}{\pi^{N/2}\;\Xi_{N+1}}\;. \tag{14.47}$$

Integrating (14.45) over $\mathbf{F}^{(N)}\times\Delta_{N-1}$ we get the Bures volume

$${\rm Vol}_{\rm B}(\mathcal{M}^{(N)}) \;=\; \frac{2^{1-N}}{C_N^{\rm B}}\,\frac{{\rm Vol}(\mathbf{F}^{(N)})}{N!} \;=\; \frac{1}{2^{N^2-1}}\frac{\pi^{N^2/2}}{\Gamma(N^2/2)}\;. \tag{14.48}$$

Observe that the Bures volume of the set of mixed states is equal to the volume of an (N^2-1)-dimensional hemisphere of radius $R_{\rm B}=1/2$. In a similar way one computes the Bures volume of the boundary of the set of mixed states (Sommers and Życzkowski, 2003)

$${\rm Vol}_{\rm B}(\partial\mathcal{M}^{(N)}) \;=\; \frac{N}{2^{N^2-2}}\,\frac{\pi^{(N^2-1)/2}}{\Gamma[(N^2-1)/2]}\;. \tag{14.49}$$

We are pleased to realize that for $N=2$ the above results describe the Uhlmann hemisphere of radius $R_{\rm B}=1/2$,

$${\rm Vol}_{\rm B}(\mathcal{M}^{(2)}) = \frac{1}{2}{\rm Vol}(\mathbf{S}^3)\,R_{\rm B}^3 = \frac{\pi^2}{8}\;,\qquad {\rm Vol}_{\rm B}\big(\partial\mathcal{M}^{(2)}\big) = {\rm Vol}(\mathbf{S}^2)\,R_{\rm B}^2 = \pi\;. \tag{14.50}$$

Although for $N\geq 2$ the Bures geometry is not like that of a hyper-hemisphere (see Section 9.6), we see that the ratio

$$\eta_N^{\rm B} = R_{\rm B}\,\frac{{\rm Vol}_{\rm B}(\partial\mathcal{M}^{(N)})}{{\rm Vol}_{\rm B}(\mathcal{M}^{(N)})} = \frac{N}{\sqrt{\pi}}\,\frac{\Gamma(N^2/2)}{\Gamma(N^2/2-1/2)}\sim D \tag{14.51}$$

asymptotically increases linearly with the dimensionality $D=N^2-1$, which is typical for hemispheres.

14.5 Induced measures

The measures discussed so far belong to the wide class of **S1 measures**:

S1 – Metric-related measures. Any metric generates a measure.

[9] The constant $C_2^{\rm B}=2/\pi$ was computed by Hall (1998), while $C_3^{\rm B}=35/\pi$ and $C_4^{\rm B}=2^{11}35/\pi^2$ were found by Slater (1999a).

Other families of measures may be defined in an operational manner by specifying a recipe to generate a density matrix. Consider a pure state $|\psi\rangle$ of a bipartite $N \times K$ composite quantum system described in $\mathcal{H} = \mathcal{H}_N \otimes \mathcal{H}_K$. It is convenient to work in an arbitrary orthogonal product basis, $|i,k\rangle = |i\rangle \otimes |k\rangle$, where $|i\rangle \in \mathcal{H}_N$ and $|k\rangle \in \mathcal{H}_K$. The pure state $|\psi\rangle$ is then represented by a $N \times K$ rectangular complex matrix A with $A_{ik} = \langle i,k|\psi\rangle$. The normalization condition, $\|\psi\|^2 = \mathrm{Tr}AA^\dagger = 1$, is the only constraint imposed on this matrix. The corresponding density matrix $\sigma = |\psi\rangle\langle\psi|$, acting on the composite Hilbert space $\mathcal{H}$, is represented in this basis by a matrix labelled by four indices, $\sigma^{ik}_{i'k'} = A_{ik}A^*_{i'k'}$. Partial tracing with respect to the subspace $\mathcal{H}_K$ gives the reduced density matrix of size N

$$\rho \;=\; \mathrm{Tr}_K(\sigma) \;=\; AA^\dagger, \tag{14.52}$$

while partial tracing over the first subsystem leads to the reduced density matrix $\rho' = \mathrm{Tr}_N(\sigma) = A^\dagger A$ of size K. Now we are prepared to define a family of measures in the space of mixed states $\mathcal{M}^{(N)}$ labelled by a single parameter – the size K of the ancilla (Lubkin, 1978; Braunstein, 1996; Hall, 1998).

S2 – Measures $P^{\mathrm{trace}}_{N,K}(\rho)$ induced by partial trace over an K-dimensional environment (14.52) of an ensemble of pure states distributed according to the unique, unitarily invariant Fubini-Study measure on the space $\mathbb{C}\mathbf{P}^{KN-1}$ of pure states of the composite system.

There is a simple physical motivation for such measures: they can be used if we do not know anything about the density matrix, apart from the dimensionality K of the environment. When $K = 1$ we get the FS measure on the space of pure states. Since the rank of ρ is limited by K, the induced measure covers the full set of $\mathcal{M}^{(N)}$ for $K \geq N$. When $K < N$ the measure $P^{\mathrm{trace}}_{N,K}$ is supported on the subspace of density matrices of rank K belonging to the boundary $\partial\mathcal{M}^{(N)}$.

Since the pure state $|\psi\rangle$ is drawn according to the FS measure, the induced measure $P^{\mathrm{trace}}_{N,K}$ enjoys the product form (14.22). Hence the distribution of the eigenvectors of ρ is determined by the Haar measure on $U(N)$, and we need to find the joint distribution of the eigenvalues in the simplex Δ_{N-1}.

In the first step we use the relation (14.52) to write down the distribution of matrix elements

$$P(\rho) \;\propto\; \int [\mathrm{d}A]\, \delta(\rho - AA^\dagger)\, \delta(\mathrm{Tr}AA^\dagger - 1)\,, \tag{14.53}$$

in which the first delta function assures the (semi)positive definiteness, while the second delta function provides the normalization. Let us assume that $K \geq N$, so that $\rho = AA^\dagger$ is generically positive definite (the *Wishart case*, since $AA^\dagger$ is called

a *Wishart matrix*[10]). Thus one can perform the transformation

$$A = \sqrt{\rho}\,\tilde{A} \quad \text{and} \quad [\mathrm{d}A] = \det\rho^K\,[\mathrm{d}\tilde{A}]\,. \tag{14.54}$$

Note that $[\mathrm{d}A]$ includes alternating factors $\mathrm{d}A_{ik}$ and $\mathrm{d}A_{ik}{}^*$. The matrix delta function may now be written as

$$\delta\big(\sqrt{\rho}(1-\tilde{A}\tilde{A}^\dagger)\sqrt{\rho}\big) \;=\; \det\rho^{-N}\,\delta\big(1-\tilde{A}\tilde{A}^\dagger\big)\,, \tag{14.55}$$

where the first factor on the right-hand side is the inverse Jacobian of the corresponding transformation. This implies

$$P(\rho) \;\propto\; \theta(\rho)\,\delta(\mathrm{Tr}\rho-1)\,\det\rho^{K-N}, \tag{14.56}$$

in which the step function ensures that ρ is positive definite. The joint probability distribution of eigenvalues is given by (Lubkin, 1978; Lloyd and Pagels, 1988; Page, 1993; Hall, 1998)

$$P^{\mathrm{trace}}_{N,K}(\lambda_1,\ldots,\lambda_N) \;=\; C_{N,K}\,\delta(1-\sum_i\lambda_i)\prod_i\lambda_i^{K-N}\prod_{i<j}(\lambda_i-\lambda_j)^2 \tag{14.57}$$

with the normalization constant (Mehta, 1991; Życzkowski and Sommers, 2001)

$$C_{N,K} = \frac{\Gamma(KN)}{\prod_{j=0}^{N-1}\Gamma(K-j)\Gamma(N-j+1)} = \frac{\Xi_{K-N}\Gamma(KN)}{\Xi_K\,\Xi_{N+1}} \tag{14.58}$$

written here for $K \geq N$ in terms of Ξ_N defined in (14.25) extended by a convention $\Xi_0 = 1$. If the size K of the ancilla equals the size N of the system, the measure $P^{\mathrm{trace}}_{N,N}$ induced by partial tracing of the pure states in $\mathbb{C}\mathbf{P}^{N^2-1}$ coincides with the Hilbert–Schmidt measure (14.35), and $C_{N,N} = C_N^{\mathrm{HS}}$. For instance, the partial trace of $N = 4$ random complex pure states induces the uniform measure in the Bloch ball of $N = 2$ mixed states.

Integrating out all eigenvalues but one from the joint probability distribution (14.57) one receives the density of eigenvalues. This task[11] was performed by Page (1993), who derived the distribution

$$P_{N,K}(x) = \frac{\sqrt{(x-a_-)(a_+-x)}}{2\pi x}, \quad \text{with} \quad a_\pm = 1+\frac{K}{N}\pm 2\sqrt{\frac{K}{N}} \tag{14.59}$$

valid for $K \geq N \gg 1$. Here $x = N\lambda$ stands for the rescaled eigenvalue, so asymptotically the rescaled spectrum is supported on the interval $[a_-, a_+]$.

[10] In the opposite, so called *anti-Wishart*, case (Yu and Zhang, 2002; Janik and Nowak, 2003) the reduced density matrix $\rho = A^\dagger A$ has $N-K$ zero eigenvalues, but the reduced matrix ρ' of size K is positive definite and has the same positive eigenvalues. The formulae (14.57)–(14.59) hold with both parameters exchanged, $N \leftrightarrow K$.

[11] Up to an overall normalization factor this problem is equivalent to finding the density of the spectrum of random Wishart matrices $H = AA^\dagger$ studied by Marchenko and Pastur (1967) and later by Sengupta and Mitra (1999).

A random mixed state may also be obtained as a convex sum of K pure states from $\mathbb{C}\mathbf{P}^{N-1}$. The probability distribution of the weights may be arbitrary, but for simplicity we will consider the uniform distribution, $p_i = 1/K$. The number of pure states K, which governs the rank of ρ, may be treated as a free parameter labelling the measure

S3 – Mixtures of random pure states obtained as a combination of K independent random pure states $|\psi_i^{\rm rand}\rangle$ drawn according to the FS measure,

$$\rho \;=\; \frac{1}{K}\sum_{j=1}^{K} |\psi_i^{\rm rand}\rangle\langle\psi_i^{\rm rand}| \,. \tag{14.60}$$

The measure $P^{\rm mixt}_{N,K}(\rho)$ defined in this way has the product form (14.22). For $K < N$ the measure is supported on a subspace of lower rank included in the boundary $\partial\mathcal{M}^{(N)}$. By construction $P^{\rm mixt}_{N,1} = P^{\rm trace}_{N,1}$ is equivalent to the FS measure on the manifold of pure states. However, for larger K both ways of generating mixed states do not coincide.

In general, one may distinguish the symmetric case, $P^{\rm mixt}_{N,N}(\rho)$, since the number $K = N$ is the minimal one, which typically gives mixed states of the full rank. For instance, analysing the position of the barycentre of two independent random points placed on the surface of the Bloch sphere we infer that $P^{\rm mixt}_{2,2}(\lambda_1, \lambda_2) \propto \delta(1-\lambda_1-\lambda_2)|\lambda_1-\lambda_2|$. In larger dimensions the distributions $P^{\rm mixt}_{N,N}(\vec{\lambda})$ get more complicated and differ from $P^{\rm trace}_{N,N}(\vec{\lambda}) = P_{\rm HS}(\vec{\lambda})$. The larger number K of the states in the mixture, the larger the entropy of the resulting mixed state ρ. In the limit $K \to \infty$ it tends to the maximal value $\ln N$.

14.6 Random density matrices

The measures that we have discussed allow us to generate random density matrices. The picture is particularly transparent for $N = 2$, in which the product form (14.22) assures the rotational symmetry inside the Bloch ball, and the only thing to settle is the radial distribution $P(r)$, where the radius $r = |\lambda_1 - 1/2|$ is equal to the length of the Bloch vector, $r = |\vec{\tau}|$.

For $N = 2$ the HS measure (14.35) gives $P_{\rm HS}(r) = 24r^2$ (for $r \in [0, 1/2]$). This is one possible way of saying that the distribution is uniform inside the Bloch ball. The Bures measure (14.46) induces the distribution

$$P_{\rm B}(r) \;=\; \frac{32\, r^2}{\pi\sqrt{1-4r^2}} \,. \tag{14.61}$$

This is the uniform distribution on the Uhlmann hemisphere. Comparing the HS and Bures measures we realize that the latter is more concentrated on states of high purity

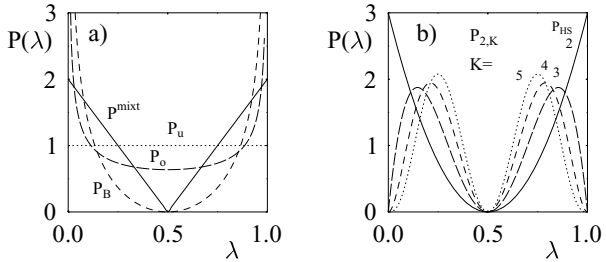


Figure 14.2. Distribution of an eigenvalue λ of density matrices of size $N = 2$: (a) P_u (dotted line); P_o – 'cosine distribution' (dashed line); $P_{2,2}^{\text{mixed}}$ (solid line) and P_B – Bures measure (dash–dotted line); (b) measures $P_{2,K}^{\text{trace}}$ induced by partial tracing; $K = 2$ (i.e. P_{HS}) (solid line); $K = 3$ (dashed line); $K = 4$ (dash–dotted) line, and $K = 5$ (dotted line).

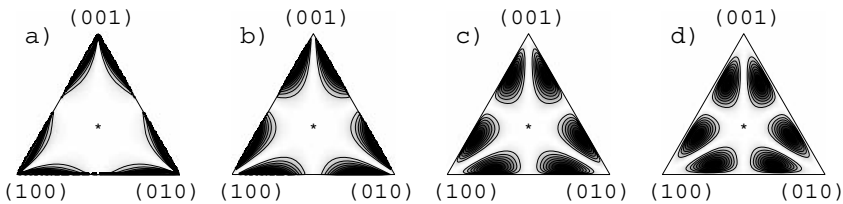


Figure 14.3. Probability distributions in the simplex of eigenvalues for $N = 3$ (a) Bures measure, (b) Hilbert–Schmidt measure equal to $P_{3,3}^{\text{trace}}$; other measures induced by partial tracing: (c) $P_{3,4}^{\text{trace}}$ and (d) $P_{3,5}^{\text{trace}}$.

(with large r). For $N = 2$ one has $\lambda_1 + \lambda_2 = 1$, the denominator in (14.46) equals unity, and the Bures measure coincides with the induced measure $P_{\text{B}}(\lambda_1, \lambda_2) = P_{2,3/2}^{\text{trace}}(\lambda_1, \lambda_2)$. Even though there is no subsystem of the dimensionality $K = 3/2$ and such an induced measure has no physical interpretation, this relation is useful to compute some averages over the Bures measure by an analytical continuation in the parameter K.

Some exemplary radial distributions inside the Bloch ball, sketched in Figure 14.2, include the $s = 1$ Dirichlet distribution (2.73), generated by the unitary product measure, $P_u(r) = 2$, and the $s = 1/2$ Dirichlet distribution, related to the orthogonal product measure, $P_o(r) = 4/(\pi\sqrt{1 - 4r^2})$. Figure 14.2 also presents a family of distributions implied by the induced measures (14.57). The larger the dimension K of the auxiliary space $\mathcal{H}_B$, the more is the induced distribution $P_{N,K}^{\text{trace}}$ concentrated in the centre of the Bloch ball (Hall, 1998).

A similar effect is shown in Figure 14.3 for $N = 3$, in which probability distributions are plotted in the eigenvalue simplex. Again the Bures measure is more

localized on states of high purity, as compared to the HS measure. Due to the factor $(\lambda_i - \lambda_j)^2$ in (14.57) the degeneracies in spectrum are avoided, which is reflected by a low probability (white colour) along all three bisectrices of the triangle. On the other hand, the distribution $P^{\text{trace}}_{3,2}(\vec{\lambda})$ is singular, and located at the edges of the triangle, which represent density matrices of rank 2. It is equal to $P^{\text{trace}}_{2,3}(\vec{\lambda})$ represented by dashed line in Figure 14.2

In order to characterize the average degree of mixing of random states we compute the mean moments $\langle \mathrm{Tr}\rho^k \rangle$, averaged with respect to the measures introduced in this chapter. Averaging over the Hilbert–Schmidt measure (14.35) gives the exact results (Lubkin, 1978; Hall, 1998; Życzkowski and Sommers, 2001) for $k = 2, 3$

$$\langle \mathrm{Tr}\rho^2 \rangle_{\mathrm{HS}} = \frac{2N}{N^2+1} \quad \text{and} \quad \langle \mathrm{Tr}\rho^3 \rangle_{\mathrm{HS}} = \frac{5N^2+1}{(N^2+1)(N^2+2)} \tag{14.62}$$

and an asymptotics for large N

$$\langle \mathrm{Tr}\rho^k \rangle_{\mathrm{HS}} = N^{1-k} \frac{\Gamma(1+2k)}{\Gamma(1+k)\,\Gamma(2+k)} \Big(1 + O\big(\frac{1}{N}\big)\Big). \tag{14.63}$$

Analogous averages with respect to the Bures measure (14.46) read (Sommers and Życzkowski, 2004)

$$\langle \mathrm{Tr}\rho^2 \rangle_{\mathrm{B}} = \frac{5N^2+1}{2N(N^2+2)} \quad \text{and} \quad \langle \mathrm{Tr}\rho^3 \rangle_{\mathrm{B}} = \frac{N(14N^2+10)}{(5N^2+1)(N^2+3)}, \tag{14.64}$$

and give asymptotically

$$\langle \mathrm{Tr}\rho^k \rangle_{\mathrm{B}} = N^{1-k} \frac{2^k \Gamma[(3k+1)/2]}{\Gamma[(1+k)/2]\,\Gamma(2+k)} \Big(1 + O\big(\frac{1}{N}\big)\Big). \tag{14.65}$$

Observe that $\langle \mathrm{Tr}\rho^2 \rangle_{\mathrm{HS}} < \langle \mathrm{Tr}\rho^2 \rangle_{\mathrm{B}}$ which shows that among these two measures the Bures measure is concentrated at the states of higher purity. Allowing the parameter k to be real and performing the limit $k \to 1$ we obtain the asymptotics for the average von Neumann entropy $\langle S \rangle = -\lim_{k\to 1} \partial \langle \mathrm{Tr}\rho^k \rangle / \partial k$,

$$\langle S(\rho) \rangle_{\mathrm{HS}} = \ln N - \frac{1}{2} + O\big(\frac{\ln N}{N}\big) \tag{14.66}$$

and

$$\langle S(\rho) \rangle_{\mathrm{B}} = \ln N - \ln 2 + O\big(\frac{\ln N}{N}\big). \tag{14.67}$$

For comparison note that the mean entropy of the components of complex random vectors of size N, or their average Wehrl entropy (7.70) behaves as $\ln N + \gamma - 1 \approx \ln N - 0.4228$.

The above analysis may be extended for the measures (14.57) induced by partial tracing. The average purity[12] was derived early on by Lubkin (1978):

$$\langle \mathrm{Tr}\rho^2 \rangle_{N,K} = \frac{N+K}{NK+1} . \tag{14.68}$$

For $K = N$ this reduces to (14.62). An exact formula for the average entropy

$$\langle S(\rho) \rangle_{N,K} = \Big(\sum_{j=K+1}^{KN} \frac{1}{j} \Big) - \frac{N-1}{2K} = \Psi(NK+1) - \Psi(K+1) - \frac{N-1}{2K} \tag{14.69}$$

was conjectured by Page (1993) and later proved in Foong and Kanno (1994), Sánchez-Ruiz (1995) and Sen (1996). The above formula implies that the average entropy is close to the maximal value $\ln N$, if the size K of the environment is sufficiently large. Taking, for instance, $K = MN$ and making use the asymptotic properties of the digamma function, $\Psi(x + 1/2) \approx \ln(x)$ (Spanier and Oldham, 1987) we infer[13] that $\langle S \rangle_{N,MN} \sim \ln N - 1/2M$. Hence a typical pure state of a $N \times N$ system is almost maximally entangled, (compare Eq. (15.25)), while the probability to find a state with the entropy S smaller by one than the average $\langle S \rangle_{N,N}$ is exponentially small (Hayden, Leung and Winter, n.d.a).

To wind up this section on random density matrices let us discuss practical methods to generate them. It is rather simple to draw random mixed states with respect to the induced measures: as discussed in Chapter 7 we generate K independent complex random pure state according to the natural measure on N-dimensional Hilbert space and prepare their mixture (14.60). To draw a random mixed state according to $P^{\mathrm{trace}}_{N,K}(\rho)$ we generate a complex random pure state on NK-dimensional Hilbert space, and then perform the partial trace (14.52) over the K-dimensional subsystem (Braunstein, 1996; Hall, 1998; Życzkowski and Sommers, 2001).

Alternatively, to obtain a random state according to this measure we generate a rectangular, $K \times N$ random matrix A with all entries being independent complex Gaussian numbers, and compute $\rho = AA^\dagger/(\mathrm{Tr}AA^\dagger)$. By definition such a matrix is normalized and positive definite, while random matrix theory allows one to show that the joint probability distribution of spectrum coincides with (14.57). In particular, for $K = 1$ the density matrices represent random pure states (see Section 7.6), while for $K = N$ we deal with non-Hermitian square random matrices characteristic of the *Ginibre ensemble* (Ginibre, 1965; Mehta, 1991) and obtain the Hilbert–Schmidt measure (Życzkowski and Sommers, 2001; Tucci, n.d.a).

[12] Average values of the higher moments are computed in (Malacarne, Mandes and Lenzi, 2002). See also problem 14.5.

[13] This generalization of (14.66) was derived first by Lubkin (1978).

To obtain random matrices with respect to unitary (orthogonal) product measures one needs to generate a random unitary (orthogonal) matrix with respect to the Haar measure on $U(N)$ (or $O(N)$, respectively) (Poźniak, Życzkowski and Kuś, 1998). The squared moduli of its components sum to unity and provide the diagonal matrix Λ with spectrum of the density matrix. Taking another random unitary matrix W, one obtains the random state by $\rho = W\Lambda W^\dagger$.

Random mixed states according to the Bures measure may easily be generated in the $N = 2$ case (Hall, 1998). One picks a point on the Uhlmann hemisphere at random. When $N \geq 3$ we are not aware of any clever technique to get such random states, apart from the brute force method: generate a random spectrum $\vec{\lambda}$ according to the distribution (14.46), for example using the Monte Carlo method, pick a random unitary matrix W and compute $W\Lambda W^\dagger$ as in the case discussed above.

14.7 Random operations

Analysing properties of random density matrices we should also discuss random operations, for which the Kraus operators A_i entering the Kraus form (10.53) are taken at random, provided the condition (10.54) is fulfilled. There are several reasons to do that; for instance, due to the Jamiolkowski isomorphism (Section 11.3), measures in the sets of quantum states and quantum maps are closely related. Hence we should not expect that there exists a unique way to generate random operations. Indeed, there exist several relevant measures in the space of quantum maps, and we start our short review with

M1. Random maps induced by random states. Any measure in space $\mathcal{M}^{(N^2)}$ of mixed states of increased dimensionality induces by (11.22) a measure in the space of trace preserving, completely positive maps $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$.

For instance, the HS measure in $\mathcal{M}^{(4)}$ generates in this way a measure supported on the entire space of one-qubit operations. Another interesting class of random operations arise considering the environmental representation (10.59). Let us treat the size k of the environment as a free parameter and assume that the unitary matrix U is distributed according to the Haar measure on $U(kN)$. This natural assumption characterizes eigenvectors of the initial state σ of the environment, so to specify a random operation we need to characterize its spectrum. In principle one may take for this purpose any probability distribution on the simplex Δ_{k-1}, but we shall discuss only two extremal cases.

M2. Random operations with pure k-dimensional environment, $\rho' = \Phi_{\rm p}(\rho) = {\rm Tr}_k[U(\rho \otimes |\nu\rangle\langle\nu|)U^\dagger]$.

Such a random operation Φ_p is thus specified by a random unitary matrix of size kN, since the choice of the pure state $|\nu\rangle = |1\rangle$ does not influence the measure. Looking again at (10.61) we see that the Kraus operators representing Φ_p arise as blocks of size N of the random matrix U,

$$(A_i)_{lm} \equiv U_{l\mu} \quad \text{with} \quad \mu = N(i-1)+m;\ \ l,m = 1,\ldots,N;\ i = 1,\ldots,k\ . \tag{14.70}$$

Due to unitarity of U the set of Kraus operators $\{A_i\}_{i=1}^k$ satisfies the completeness relation (10.54) and defines a quantum operation (10.53). For large dimensionality N the unitarity constraints become (relatively) weaker and the the non-Hermitian random matrices A are described by the Ginibre ensemble (Ginibre, 1965; Mehta, 1991): their spectra cover uniformly the disc of radius $1/\sqrt{N}$ in the complex plane of (Życzkowski and Sommers, 2000).

Note that the random operations $\Phi_{\rm p}$ need not be bistochastic. To get a generic random operation described by a dynamical matrix of the full rank N^2 one needs to choose the dimension $k \geq N^2$. These random operations correspond to a physically motivated situation, if one knows that the state of an environment of a fixed size is initially in a pure state, but not more.

M3. Random operations with mixed k-dimensional environment, $\rho' = \Phi_{\rm m}(\rho) = {\rm Tr}_k[U(\rho \otimes \mathbb{1}/k)U^\dagger]$, where U is a random unitary of size kN.

These random operations are by construction bistochastic. The Haar measure on $U(Nk)$ generates a definite measure in the space of bistochastic maps parametrized by the size k of ancilla. The case $k = N$ is equivalent to the unistochastic map (10.64), while K-unistochastic maps are obtained for $k = KN$. Random operations $\Phi_{\rm m}$ describe the situation in which only the size k of the ancilla is known, and were called *noisy maps* in (Horodecki, Horodecki and Oppenheim, 2003a).

M4. Random external fields defined as convex combination of k unitary transformations (Alicki and Lendi, 1987), $\rho' = \Phi_{\rm REF}(\rho) = \sum_{i=1}^k p_i V_i \rho V_i^\dagger$, where unitary matrices V_i are drawn according to the Haar measure on $U(N)$. Probability vector $\vec{p}$ may be drawn with an arbitrary probability distribution on Δ_{k-1}, but one asumes often that $p_i = 1/k$. As discussed in Section 10.6, for $N \geq 3$ REF's form a proper subset of the set of bistochastic maps.

M5. Maps generated by a measure in the space of stochastic matrices. Take a random stochastic matrix S of size N. Out of its rows construct N diagonal matrices, $E_{kl}^{(i)} = \sqrt{S_{ki}}\,\delta_{kl}$, and define the set operators, $A_i = U E^{(i)} V$. Here U and V are random unitary matrices generated according to the Haar measure on $U(N)$. Since S is stochastic and satisfies (2.4), the Kraus operators A_i fulfil the completeness

relation (10.54),

$$\sum_{i=1}^{N} A_i^\dagger A_i = \sum_{i=1}^{N} V^\dagger E^{(i)} U^\dagger U E^{(i)} V = V^\dagger \Big[\sum_{i=1}^{N} {E^{(i)}}^2 \Big] V = \mathbb{1}, \tag{14.71}$$

hence define a quantum operation. In the same way one demonstrates that the dual condition (10.70) is fulfilled, so the random operation is bistochastic. Instead of using random stochastic matrices one may also take a concrete matrix S, for example related to some specific physically interesting map Φ, and then randomise it[14] by introducing random unitary rotations as in (14.71).

After discussing the methods to generate random maps, let us stress that any measure in the space of quantum operations $\mathcal{CP}_N$ induces a measure in the space of density matrices. Alternatively, random mixed states may be generated by any individual operation assuming that an initial pure state $|\psi\rangle$ is drawn according to the natural Fubini–Study measure.

S4. Operation induced random mixed states defined be a certain operation Φ,

$$\rho' \;=\; U \left[\Phi\big(V|1\rangle\langle 1|V^\dagger \big) \right] U^\dagger \,, \tag{14.72}$$

where U and V are independent random unitary matrices distributed according to the Haar measure on $U(N)$.

Hence any CP map Φ determines the *operation induced measure* μ_Φ. In the trivial case, the identity map, $\Phi = \mathbb{1}$, induces the unitarily invariant FS measure on the space of pure states, $\mu_{\mathbb{1}} = \mu_{\mathrm{FS}}$.

It is instructive to study measures induced in the $N = 2$ case by planar or linear maps defined in Table 10.4 (see Problem 14.9). The latter case may be generalized for arbitrary N. As follows from properties of the FS measure (see Eq. (7.66) and Problem 7.3), the distribution of the diagonal elements of the density matrix $|\psi\rangle\langle\psi|$ of a random pure state is uniform in the simplex of eigenvalues. In this way we arrive at an important result: the coarse graining map, $\Psi_{\mathrm{CG}}(\rho) = \mathrm{diag}(\rho)$, induces by (14.72) the unitary measure $P_u(\vec{\lambda}) = \mathrm{const}$, uniform in Δ_{N-1}.

Allowing operations which reduce the dimensionality of the system, we see that the measures (14.52) also belong to this class, since they are induced by the operation of partial trace, $\Phi(\rho) = \mathrm{Tr}_K \rho$. A more general class of measures is induced via (14.72) by a family of operations Φ_a and a concrete probability distribution $P(a)$. For instance, an interesting measure supported on the subspace of degenerated states arises by usage of the depolarizing channels, $\Phi_a = a\Phi_* + (1-a)\mathbb{1}$, with a uniform distribution of the noise level, $P(a) = 1$ for $a \in [0, 1]$.

[14] Such random maps obtained for $S = \mathbb{1}$ were used in Alicki, Loziński, Pakoński and Życzkowski (2004) to describe the influence of measurement in a random basis on the time evolution of quantum baker maps.

Let us conclude by emphasizing again that there is no single, naturally distinguished probability measure in the set of quantum states. Guessing at random what the state may be, we can use any available additional information. For instance, if a mixed state has arisen by the partial tracing over a K-dimensional environment, the induced measure (14.57) should be used. More generally, if a mixed state has arisen as an image of an initially pure state under the action of a known operation Φ the operation-related measure (14.72) may be applied. Without any prior information whatsoever, it will be legitimate to use the Bures measure (14.46), related to Jeffreys' prior, statistical distance, fidelity and quantum distinguishability.

Problems

Problem 14.1 Show that the following functions $f(t)$ generate monotone Riemannian metrics: $\sqrt{t}$, $(t-1)/\ln t$, $2[(t-1)/\ln t]^2/(1+t)$ (non-informative metric), $2t^{\alpha+1/2}/(1+t^{2\alpha})$ for $\alpha \in [0, 1/2]$ (Petz and Sudár, 1996), $t^{t/(t-1)}/\mathrm{e}$ (quasi-Bures metric) (Slater, 1999b), and $f_{\mathrm{WY}} = (\sqrt{t}+1)^2/4$ (Wigner–Yanase metric) (Gibilisco and Isola, 2003). Is the latter metric pure?

Problem 14.2 Compute the volume of the orthogonal group with respect to the measure $(\mathrm{d}s)^2 \equiv -\frac{1}{2}\,\mathrm{Tr}(O^{-1}\mathrm{d}O)^2$ analogous to (14.26). Show that $\mathrm{Vol}(\mathbb{R}\mathbf{P}^N) = \frac{1}{2}\mathrm{Vol}(\mathbf{S}^N) = \pi^{(N+1)/2}/\Gamma[(N+1)/2] = \mathrm{Vol}[O(N)]/\big(\mathrm{Vol}[O(N-1)]\mathrm{Vol}[O(1)]\big)$; $\mathrm{Vol}[O(2)] = 2$ and $\mathrm{Vol}(\mathbf{F}_{\mathbb{R}}^{(N)}) = \mathrm{Vol}[O(N)]/2^N$ (Życzkowski and Sommers, 2003).

Problem 14.3 Show that the volume radius of the set $\mathcal{M}^{(N)}$ of mixed states behaves for large N as $\mathrm{e}^{-1/4}/\sqrt{N}$ (Szarek, 2005).

Problem 14.4 Compute the mean von Neumann entropy, of $N = 2$ random mixed states averaged over Hilbert–Schmidt, Bures, orthogonal and unitary measures, respectively.

Problem 14.5 Find the mean moments $\langle \mathrm{Tr}\rho^3 \rangle$ and $\langle \mathrm{Tr}\rho^4 \rangle$ averaged over the induced measures (14.57).

Problem 14.6 Calculate the probability distribution of a rescaled eigenvalue $x = N\lambda$ of a random density matrix of size $N \gg 1$ generated according to the Hilbert–Schmidt measure. What has this distribution in common with a circle?

Problem 14.7 Derive the distribution of fidelity $P(F)$ between two random N-dimensional complex pure states.

Problem 14.8 Compute the mean fidelity between two $N = 2$ independent random states distributed according to (a) HS measure; (b) Bures measure.

Problem 14.9 Analyse operation induced measures (14.72) defined for $N = 2$ by planar or linear channels (for definitions see Table 10.4) and show that they are isotropic inside the Bloch ball with the radial distributions $P_{\text{plan}}(r) = 4r/\sqrt{1-4r^2}$ and $P_{\text{lin}}(r) = 2 = P_u(r)$, respectively, where $r \in [0, 1/2]$.

15

Quantum entanglement

> Entanglement is not *one* but rather *the* characteristic trait of quantum mechanics.
>
> *Erwin Schrödinger*

15.1 Introducing entanglement

So far, when working in a Hilbert space that is a tensor product of the form $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, we were really interested in only one of the factors; the other factor played the role of an ancilla describing an environment outside our control. Now the perspective changes: we are interested in a situation where there are two masters. The fate of both subsystems are of equal importance, although they may be sitting in two different laboratories.

The operations performed independently in the two laboratories are described using operators of the form $\Phi_A \otimes \mathbb{1}$ and $\mathbb{1} \otimes \Phi_B$, respectively, but due perhaps to past history, the global state of the system may not be a product state. In general, it may be described by an arbitrary density operator ρ acting on the composite Hilbert space $\mathcal{H}$.

The peculiarities of this situation were highlighted in 1935 by Einstein, Podolsky and Rosen (1935). Their basic observation was that if the global state of the system is chosen suitably then it is possible to change, and to some extent to choose, the state assignment in laboratory A by performing operations in laboratory B. The physicists in laboratory A will be unaware of this until they are told, but they can check in retrospect that the experiments they performed were consistent with the state assignment arrived at from afar – even though there was an element of choice in arriving at that state assignment. Einstein's considered opinion was that 'on one supposition we should ... absolutely hold fast: the real factual situation of the system $\mathcal{S}_2$ is independent of what is done with the system $\mathcal{S}_1$, which is spatially separated from the former' (Einstein, 1949). Then we seem to be forced to the

conclusion that quantum mechanics is an incomplete theory in the sense that its state assignment does not fully describe the factual situation in laboratory A.

In his reply to Einstein, Podolsky and Rosen (EPR), Schrödinger argued that in quantum mechanics 'the best possible knowledge of a *whole* does not include the best possible knowledge of all its *parts*, even though they may be entirely separated and therefore virtually capable of being "best possibly known" '.[1] Schrödinger introduced the word *Verschränkung* to describe this phenomenon, personally translated it into English as *entanglement*, and made some striking observations about it. The subject then lay dormant for many years.

To make the concept of entanglement concrete, we recall that the state of the subsystem in laboratory A is given by the partially traced density matrix $\rho_A = \mathrm{Tr}_B \rho$. This need not be a pure state, even if ρ itself is pure. In the simplest possible case, namely when both $\mathcal{H}_A$ and $\mathcal{H}_B$ are two dimensional, we find an orthogonal basis of four states that exhibit this property in an extreme form. This is the *Bell basis*, already mentioned in Table 11.1

$$|\psi^{\pm}\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|1\rangle \pm |1\rangle|0\rangle\big) \qquad |\phi^{\pm}\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle|0\rangle \pm |1\rangle|1\rangle\big)\,. \tag{15.1}$$

The Bell states all have the property that $\rho_A = \frac{1}{2}\mathbb{1}$, which means that we know nothing at all about the state of the subsystems, even though we have maximal knowledge of the whole. At the opposite extreme we have product states such as $|0\rangle|0\rangle$ and so on; if the global state of the system is in a product state then ρ_A is a projector and the two subsystems are in pure states of their own. Such pure product states are called *separable*, while all other pure states are *entangled*.

Now the point is that if a projective measurement is performed in laboratory B, corresponding to an operator of the form $\mathbb{1} \otimes \Phi_B$, then the global state will collapse to a product state. Indeed, depending on what measurement B chooses to perform, and depending on its outcome, the state in laboratory A can become any pure state in the support of ρ_A. (This conclusion was drawn by Schrödinger from his mixture theorem. He found it 'repugnant'.) Of course, if the global state was one of the Bell states to begin with, then the experimenters in laboratory A still labour under the assumption that their state is $\rho_A = \frac{1}{2}\mathbb{1}$, and it is clear that any measurement results in A will be consistent with this state assignment. Nevertheless it would seem as if the real factual situation in A has been changed from afar.

In the early 1960s John Bell (1964) was able to show that if we hold fast to the locality assumption then there cannot exist a completion of quantum mechanics in the sense of EPR; it is the meaning of the expression 'real factual situation'

[1] Schrödinger's 'general confession' consisted of a series of three papers (1935a, 1935b, 1936).

that is at stake in entangled systems.[2] The idea is that if the quantum mechanical probabilities arise as marginals of a probability distribution over some kind of a set of real factual situations, then the mere existence of the latter gives rise to inequalities for the marginal distributions that, as a matter of fact, are disobeyed by the probabilities predicted by quantum mechanics.

Bell's work caused much excitement in philosophically oriented circles; it seemed to put severe limits on the world view offered by physics.[3] In the early 1990s the emphasis began to shift. Entanglement came to be regarded as a *resource* that allows us to do certain otherwise impossible things. An early and influential example is that of *quantum teleportation*. Let us dwell on this a little. The task is to send information that allows a distant receiver to reconstruct the state of a spin $1/2$ particle – even if the state is unknown to the sender. But since only a single copy of the state is available the sender is unable to figure out what the state to be 'teleported' actually is. So the task appears impossible.[4] A solution is to prepare a composite system in the Bell state $|\phi^+\rangle$, and to share the two entangled subsystems between sender and receiver. Suppose that the state to be sent is $\alpha|0\rangle + \beta|1\rangle$. At the outset the latter is uncorrelated to the former, so the total (unnormalized) state is

$$\begin{aligned} |\Psi\rangle &= \big(\alpha|0\rangle + \beta|1\rangle\big)\big(|0\rangle|0\rangle + |1\rangle|1\rangle\big) \\ &= \alpha|0\rangle|0\rangle|0\rangle + \alpha|0\rangle|1\rangle|1\rangle + \beta|1\rangle|0\rangle|0\rangle + \beta|1\rangle|1\rangle|1\rangle \,. \end{aligned} \tag{15.2}$$

The sender controls the first two factors of the total Hilbert space, and the receiver controls the third. By means of a simple manipulation we rewrite this as

$$\begin{aligned} \sqrt{2}|\Psi\rangle = |\psi^+\rangle\big(\alpha|1\rangle + \beta|0\rangle\big) + |\psi^-\rangle\big(\alpha|1\rangle - \beta|0\rangle\big) \\ + |\phi^+\rangle\big(\alpha|0\rangle + \beta|1\rangle\big) + |\phi^-\rangle\big(\alpha|0\rangle - \beta|1\rangle\big) \,. \end{aligned} \tag{15.3}$$

The sender now performs a projective measurement in the four-dimensional Hilbert space at his disposal, such that the state collapses to one of the four Bell states. If the collapse results in the state $|\phi^+\rangle$ the teleportation is complete. But the other cases are equally likely, so the sender must send two classical bits of information to

[2] At this point opinions diverge; some physicists, including notably David Bohm, have not felt obliged to hold absolutely fast to Einstein's notion of locality. See Bell (1987) for a sympathetic review of Bohm's arguments. Followers of Everett (1957) on the other hand argue that what happened was that the system in A went from being entangled with the system in B to being entangled with the measurement apparatus in B, with no change of the real factual situation in A.

[3] For a thorough discussion of the Bell inequalities consult Clauser and Shimony (1978); experimental tests, notably by Aspect, Dalibard and Roger (1982), show that violation of the Bell inequalities does indeed occur in the laboratory. (Although loopholes still exist; see Gill (2003).)

[4] To send information that allows us to reconstruct a given state elsewhere is referred to as teleportation in the science fiction literature, where it is usually assumed to be trivial for the sender to verify what the state to be sent may be. The idea of teleporting a state that is not known at all is due to Bennett, Brassard, Crépeau, Josza, Peres and Wootters (1993).

the receiver, informing him of the outcome of the measurement. Depending on the result the receiver then performs a unitary transformation (such that $|0\rangle \leftrightarrow |1\rangle$, if the outcome was $|\psi^+\rangle$) and the teleportation of the still unknown qubit is complete.[5]

In the example of teleportation the entangled auxiliary system was used to perform a task that is impossible without it. It will be noted also that the entanglement was used up, in the sense that once the transmission has been achieved no mutual entanglement between sender and receiver remains. In this sense then entanglement is a resource, just as the equally abstract concept of energy is a resource. Moreover it has emerged that there are many interesting tasks for which entanglement can be used, including *quantum cryptography* and *quantum computing* (Preskill, n.d.; Gruska, 1999; Nielsen and Chuang, 2000; Keyl, 2002).

If entanglement is a resource we naturally want to know how much of it we have. As we will see it is by no means easy to answer this question, but it is easy to take a first step in the situation when the global state is a pure one. It is clear that there is no entanglement in a product state, when the subsystems are in pure states too and the von Neumann entropy of the partially traced state vanishes. It is also clear that maximally entangled pure state will lead to a partially traced density matrix that is a maximally mixed state. For the case of two qubits the von Neumann entropy then assumes its maximum value $\ln 2$, and the amount of entanglement in such a state is known as an *e-bit*. States that are neither separable nor maximally entangled require more thought. Let us write a pure state in its Schmidt form $|\Psi\rangle = \cos\chi\,|00\rangle + \sin\chi\,|11\rangle$, (see Section 9.2). Performing the partial trace one obtains

$$\rho_A = \mathrm{Tr}_B |\Psi\rangle\langle\Psi| = \begin{bmatrix} \cos^2\chi & 0 \\ 0 & \sin^2\chi \end{bmatrix}. \tag{15.4}$$

The *Schmidt angle* $\chi \in [0, \pi/4]$ parametrizes the amount of ignorance about the state of the subsystem, that is to say the amount of entanglement. A good thing about it is that its value cannot be changed by local unitary transformations of the form $U(2) \otimes U(2)$. For the general case, when the Hilbert space has dimension $N \times N$, we will have to think more, and for the case when the global state is itself a mixed one much more thought will be required.

At this stage entanglement may appear to be such an abstract notion that the need to quantify it does not seem to be urgent but then, once upon a time, 'energy' must have seemed a very abstract notion indeed, and now there are thriving industries whose role is to deliver it in precisely quantified amounts. Perhaps our governments will eventually have special *Departments of Entanglement* to deal with these things.

[5] This is not a *Gedanken experiment* only; it was first done in Innsbruck (Bouwmeester, Pan, Mattle, Eibl, Weinfurter and Zeilinger, 1997) and in Rome (Boschi, Branca, De Martini, Hardy and Popescu, 1998).

But that is in the far future; here we will concentrate on a geometrical description of entanglement and how it is to be quantified.

15.2 Two qubit pure states: entanglement illustrated

Our first serious move will be to take a look (literally) at entanglement in the two qubit case.[6] Our Hilbert space has four complex dimensions, so the space of pure states is $\mathbb{C}\mathbf{P}^3$. We can make a picture of this space along the lines of Section 4.6. So we draw the positive hyperoctant of a 3-sphere and imagine a 3-torus sitting over each point, using the coordinates

$$(Z^0, Z^1, Z^2, Z^3) = (n_0,\ n_1 e^{i\nu_1},\ n_2 e^{i\nu_2},\ n_3 e^{i\nu_3})\ . \tag{15.5}$$

The four non-negative real numbers n_0, etc. obey

$$n_0^2 + n_1^2 + n_2^2 + n_3^2 = 1\ . \tag{15.6}$$

To draw a picture of this set we use a gnomonic projection of the 3-sphere centred at

$$(n_0, n_1, n_2, n_3) = \frac{1}{2}(1, 1, 1, 1)\ . \tag{15.7}$$

The result is an attractive picture of the hyperoctant, consisting of a tetrahedron centred at the above point, with geodesics on the 3-sphere appearing as straight lines in the picture. The 3-torus sitting above each interior point can be pictured as a rhomboid that is squashed in a position dependent way.

Mathematically, all points in $\mathbb{C}\mathbf{P}^3$ are equal. In physics, points represent states, and some states are more equal than others. In Chapter 6, this happened because we singled out a particular subgroup of the unitary group to generate coherent states. Now it is assumed that the underlying Hilbert space is presented as a product of two factors in a definite way, and this singles out the orbits of $U(N) \times U(N) \subset U(N^2)$ for special attention. More specifically there is a preferred way of using the entries Γ_{ij} of an $N \times N$ matrix as homogeneous coordinates. Thus any (normalized) state vector can be written as

$$|\Psi\rangle\ =\ \frac{1}{\sqrt{N}} \sum_{i=0}^{n} \sum_{j=0}^{n} \Gamma_{ij} |i\rangle |j\rangle\ . \tag{15.8}$$

For two qubit entanglement $N = n + 1 = 2$, and it is agreed that

$$(Z^0, Z^1, Z^2, Z^3) \equiv (\Gamma_{00}, \Gamma_{01}, \Gamma_{10}, \Gamma_{11})\ . \tag{15.9}$$

[6] Such a geometric approach to the problem was initiated by Brody and Hughston (2001) and developed in Kuś and Życzkowski (2001), Mosseri and Dandoloff (2001), Bengtsson et al. (2002) and Lévay (2004).

Let us first take a look at the separable states. For such states

$$|\Psi\rangle = \sum_{i=0}^{n}\sum_{j=0}^{n}(a_i|i\rangle)(b_j|j\rangle) \quad \Leftrightarrow \quad \Gamma_{ij} = a_i b_j \,. \tag{15.10}$$

In terms of coordinates a two qubit case state is separable if and only if

$$Z^0 Z^3 - Z^1 Z^2 = 0 \,. \tag{15.11}$$

We recognize this quadric equation from Section 4.3. It defines the Segre embedding of $\mathbb{C}\mathbf{P}^1 \times \mathbb{C}\mathbf{P}^1$ into $\mathbb{C}\mathbf{P}^3$. Thus the separable states form a four real-dimensional submanifold of the six real-dimensional space of all states. (Had we regarded $\mathbb{C}\mathbf{P}^1$ as a classical phase space, this submanifold would have been enough to describe all the states of the composite system.)

What we did not discuss in Chapter 4 is the fact that the Segre embedding is easily described in the octant picture. Equation (15.11) splits into two real equations:

$$n_0 n_3 - n_1 n_2 = 0 \tag{15.12}$$

$$\nu_1 + \nu_2 - \nu_3 = 0 \,. \tag{15.13}$$

Hence we can draw the space of separable states as a two-dimensional surface in the octant, with a two-dimensional surface in the torus that sits above each separable point in the octant. The surface in the octant has an interesting structure, related to Figure 4.6. In Eq. (15.10) we can keep the state of one of the subsystems fixed; say that b_0/b_1 is some fixed complex number with modulus k. Then

$$\frac{Z^0}{Z^1} = \frac{b_0}{b_1} \quad \Rightarrow \quad n_0 = k n_1 \tag{15.14}$$

$$\frac{Z^2}{Z^3} = \frac{b_0}{b_1} \quad \Rightarrow \quad n_2 = k n_3 \,. \tag{15.15}$$

As we vary the state of the other subsystem we sweep out a curve in the octant that is in fact a geodesic in the hyperoctant (the intersection between the 3-sphere and two hyperplanes through the origin in the embedding space). In the gnomonic coordinates that we are using this curve will appear as a straight line, so what we see when we look at how the separable states sit in the hyperoctant is a surface that is ruled by two families of straight lines.

There is an interesting relation to the Hopf fibration (see Section 3.5) here. Each family of straight lines is also a one parameter family of Hopf circles, and there are two such families because there are two Hopf fibrations, with different twist. We can use our hyperoctant to represent real projective space $\mathbb{R}\mathbf{P}^3$, in analogy with Figure 4.12. The Hopf circles that rule the separable surface are precisely those that get mapped onto each other when we 'fold' the hemisphere into a hyperoctant.

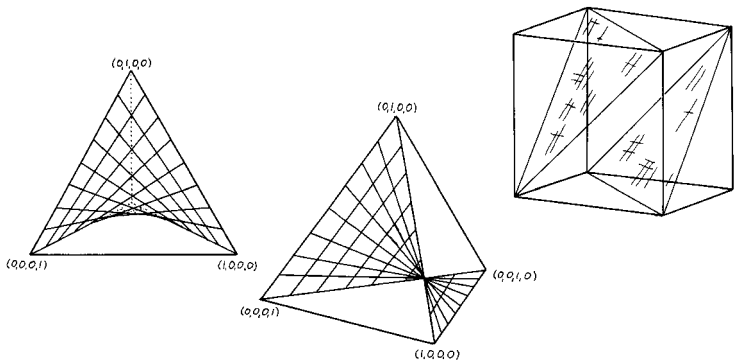


Figure 15.1. The separable states, or the Segre embedding of $\mathbb{C}\mathbf{P}^1 \times \mathbb{C}\mathbf{P}^1$ in $\mathbb{C}\mathbf{P}^3$. Two different perspectives of the tetrahedron are given.

We now turn to the maximally entangled states, for which the reduced density matrix is the maximally mixed state. Using composite indices we write

$$\rho_{\substack{ij\\kl}} = \frac{1}{N}\Gamma_{ij}\Gamma^*_{kl} \quad \Rightarrow \quad \rho^A_{ik} = \sum_{j=0}^{n} \rho_{\substack{ij\\kj}} \,, \tag{15.16}$$

see Eq. (10.27). Thus

$$\rho^A_{ik} = \frac{1}{N}\mathbb{1} \quad \Leftrightarrow \quad \sum_{j=0}^{n} \Gamma_{ij}\Gamma^*_{kj} = \delta_{ik} \,. \tag{15.17}$$

Therefore the state is maximally entangled if and only if the matrix Γ is unitary. Since an overall factor of this matrix is irrelevant for the state we reach the conclusion that the space of maximally entangled states is $SU(N)/\mathbb{Z}_N$. This happens to be an interesting submanifold of $\mathbb{C}\mathbf{P}^{N^2-1}$, because it is at once *Lagrangian* (a submanifold with vanishing symplectic form and half the dimension of the symplectic embedding space) and *minimal* (any attempt to move it will increase its volume).

When $N = 2$ we are looking at $SU(2)/\mathbb{Z}_2 = \mathbb{R}\mathbf{P}^3$. To see what this space looks like in the octant picture we observe that

$$\Gamma_{ij} = \begin{bmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{bmatrix} \quad \Rightarrow \quad Z^\alpha = (\alpha, \beta, -\beta^*, \alpha^*) \,. \tag{15.18}$$

In our coordinates this yields three real equations; the space of maximally entangled states will appear in the picture as a straight line connecting two entangled edges and passing through the centre of the tetrahedron, while there is a two-dimensional surface in the tori. The latter is shifted relative to the separable surface in such a way that the separable and maximally entangled states manage to keep their distance in the torus also when they meet in the octant (at the centre of the tetrahedron where

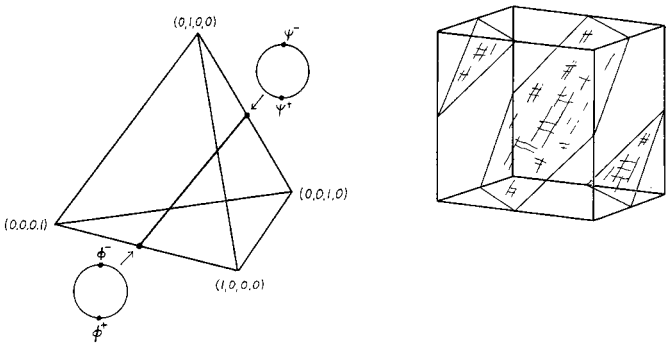


Figure 15.2. The maximally entangled states form an $\mathbb{R}\mathbf{P}^3$, appearing as a straight line in the octant and a surface in the tori. The location of the Bell states is also shown.

the torus is large). Our picture thus displays $\mathbb{R}\mathbf{P}^3$ as a one parameter family of two-dimensional flat tori, degenerating to circles at the ends of the interval. This is similar to our picture of the 3-sphere, except that this time the lengths of the two intersecting shortest circles on the tori stay constant while the angle between them is changing. It is amusing to convince oneself of the validity of this picture, and to verify that it is really a consequence of the way that the 3-tori are being squashed as we move around the octant.

As a further illustration we can consider the collapse of a maximally entangled state, say $|\psi^+\rangle$ for definiteness, when a measurement is performed in laboratory B. The result will be a separable state, and because the global state is maximally entangled all the possible outcomes will be equally likely. It is easily confirmed that the possible outcomes form a 2-sphere's worth of points on the separable surface, distinguished by the fact that they are all lying on the same distance $D_{FS} = \pi/4$ from the original state. This is the minimal Fubini–Study distance between a separable and a maximally entangled state. The collapse is illustrated in Figure 15.3.

A set of states of intermediate entanglement, quantified by some given value of the Schmidt angle χ, is more difficult to draw (although it can be done). For the extreme cases of zero or one e-bit's worth of entanglement we found the submanifolds $\mathbb{C}\mathbf{P}^1 \times \mathbb{C}\mathbf{P}^1$ and $SU(2)/\mathbb{Z}_2$, respectively. There is a simple reason why these spaces turn up, namely that the amount of entanglement must be left invariant under locally unitary transformations belonging to the group $SU(2) \times SU(2)$. In effect therefore we are looking for orbits of this group, and what we have found are the two obvious possibilities. More generally we will get a stratification of $\mathbb{C}\mathbf{P}^3$ into orbits of $SU(2) \times SU(2)$; the problem is rather similar to that discussed in Section 7.2. Of the exceptional orbits, one is a Kähler manifold and one (the maximally entangled

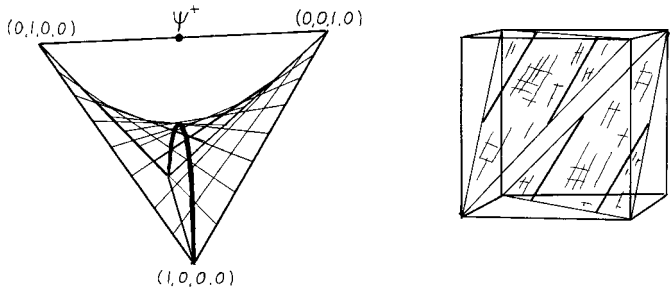


Figure 15.3. A complete measurement on one of the subsystem will collapse the Bell state $|\psi^+\rangle$ to a point on a sphere on the separable surface; it appears as a one parameter family of circles in our picture. All points on this sphere are equally likely.

one) is actually a Lagrangian submanifold of $\mathbb{C}\mathbf{P}^3$, meaning that the symplectic form vanishes on the latter. A generic orbit will be five real-dimensional and the set of such orbits will be labelled by the Schmidt angle χ, which is also the minimal distance from a given orbit to the set of separable states. A generic orbit is rather difficult to describe however. Topologically it is a non-trivial fibre bundle with an $\mathbf{S}^2$ as base space and $\mathbb{R}\mathbf{P}^3$ as fibre.[7] In the octant picture it appears as a three-dimensional volume in the octant and a two-dimensional surface in the torus. And with this observation our tour of the two qubit Hilbert space is at an end.

15.3 Pure states of a bipartite system

Consider a pure state of a composite system $|\psi\rangle \in \mathcal{H}_{NK} = \mathcal{H}_N \otimes \mathcal{H}_K$. The states related by a local unitary transformation,[8]

$$|\psi'\rangle \;=\; U \otimes V\,|\psi\rangle\,, \tag{15.19}$$

where $U \in SU(N)$ and $V \in SU(K)$, are called *locally equivalent*. Sometimes one calls them *interconvertible states*, since they may be reversibly converted by local transformations one into another (Jonathan and Plenio, 1999a). It is clear that not all pure states are locally equivalent, since the product group $SU(N) \times SU(K)$ forms only a measure zero subgroup of $SU(NK)$. How far can one go from a state using local transformations only? In other words, what is the dimensionality and

[7] This can be seen in an elegant way using the Hopf fibration of $\mathbf{S}^7$ – the space of normalized state vectors – as $\mathbf{S}^4 = \mathbf{S}^7/\mathbf{S}^3$; Mosseri and Dandoloff (2001) provide the details.

[8] Non-local properties of a unitary gate may be quantified by its operator Schmidt decomposition (10.30). A canonical form of a two qubit gates was provided in Khaneja, Brockett and Glaser (2001), Kraus and Cirac (2001) and Makhlin (2002), while theory of non-local gates was further developed in Dür and Cirac (2002), Hammerer, Vidal and Cirac (2002) and Nielsen, Dawson, Dodd, Gilchrist, Mortimer, Osborne, Bremner, Harrow and Hines (2003).

topology of the orbit generated by local unitary transformations from a given state $|\psi\rangle$?

To find an answer we are going to rely on the Schmidt decomposition (9.8). It consists of not more than N terms, since without loss of generality we have assumed that $K \geq N$. The normalization condition $\langle\psi|\psi\rangle = 1$ enforces $\sum_{i=1}^{N} \lambda_i = 1$, so the Schmidt vector $\vec{\lambda}$ lives in the $(N-1)$-dimensional simplex Δ_{N-1}. The *Schmidt rank* of a pure state $|\psi\rangle$ is the number of non-zero Schmidt coefficients, equal to the rank of the reduced state. States with maximal Schmidt rank are generic and occupy the interior of the simplex, while states of a lower rank live on its boundary.

The Schmidt vector gives the spectra of the partially reduced states, $\rho_A = \mathrm{Tr}_B(|\psi\rangle\langle\psi|)$ and $\rho_B = \mathrm{Tr}_A(|\psi\rangle\langle\psi|)$, which differ only by $K - N$ zero eigenvalues. The separable states sit at the corners of the simplex. Maximally entangled states are described by the uniform Schmidt vector, $\vec{\lambda}_* = \{1/N, \ldots, 1/N\}$, since the partial trace sends them into the maximally mixed state.

Let $\vec{\lambda} = (0, \ldots, 0, \kappa_1, \ldots, \kappa_1, \kappa_2, \ldots, \kappa_2, \ldots, \kappa_J, \ldots, \kappa_J)$ represent an ordered Schmidt vector, in which each value κ_n occurs m_n times while m_0 is the number of vanishing coefficients. By definition $\sum_{n=0}^{J} m_n = N$, while m_0 might equal to zero. The local orbit $\mathcal{O}_{\mathrm{loc}}$ generated from $|\psi\rangle$ has the structure of a fibre bundle, in which two quotient spaces

$$\frac{U(N)}{U(m_0) \times U(m_1) \times \cdots \times U(m_J)} \quad \text{and} \quad \frac{U(N)}{U(m_0) \times U(1)} \tag{15.20}$$

form the base and the fibre, respectively (Sinołęcka, Życzkowski and Kuś, 2002). In general such a bundle need not be trivial. The dimensionality of the local orbit may be computed from dimensionalities of the coset spaces,

$$\dim(\mathcal{O}_{\mathrm{loc}}) = 2N^2 - 1 - 2m_0^2 - \sum_{n=1}^{J} m_n^2 \,. \tag{15.21}$$

Observe that the base is the set of all unitarily similar mixed states ρ_A of the reduced system A, with spectrum $\vec{\lambda}$ and depends on its degeneracy (see Table 8.1). The fibre characterizes the manifold of pure states which are projected by the partial trace to the same density matrix, and depends on the Schmidt rank equal to $N - m_0$. To understand this structure consider first a generic state of the maximal Schmidt rank, so that $m_0 = 0$. Acting on $|\psi\rangle$ with $U_N \otimes W_N$, where both unitary matrices are diagonal, we see that there exist N redundant phases. Since each pure state is determined up to an overall phase, the generic orbit has the local structure

$$\mathcal{O}_g \;\approx\; \frac{U(N)}{[U(1)]^N} \;\times\; \frac{U(N)}{U(1)} = \mathbf{F}^{(N)} \;\times\; \frac{U(N)}{U(1)}, \tag{15.22}$$

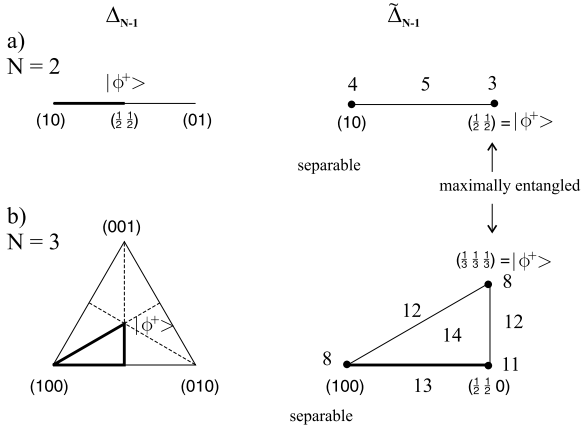


Figure 15.4. Dimensionality of local orbits generated by a given point of the Weyl chamber $\tilde{\Delta}_{N-1}$ – an asymmetric part of the Schmidt simplex Δ_{N-1} – for pure states of $N \times N$ problem with $N = 2, 3$ (compare Table 15.1).

with dimension $\dim(\mathcal{O}_g) = 2N^2 - N - 1$. If some of the coefficients are equal, say $m_J > 1$, then we need to identify all states differing by a block diagonal unitary rotation with $U(m_J)$ in the right lower corner. In the same way one explains the meaning of the factor $U(m_0) \times U(m_1) \times \cdots \times U(m_J)$ which appears in the first quotient space of (15.20). If some Schmidt coefficients are equal to zero the action of the second unitary matrix U_B is trivial in the m_0-dimensional subspace – the second quotient space in (15.20) is $U(N)/[U(m_0) \times U(1)]$.

For separable states there exists only one non-zero coefficient, $\lambda_1 = 1$, so $m_0 = N - 1$. This gives the Segre embedding (4.16),

$$\mathcal{O}_{\text{sep}} = \frac{U(N)}{U(1) \times U(N-1)} \times \frac{U(N)}{U(1) \times U(N-1)} = \mathbb{C}\mathbf{P}^{N-1} \times \mathbb{C}\mathbf{P}^{N-1}, \quad (15.23)$$

of dimensionality $\dim(\mathcal{O}_{\text{sep}}) = 4(N-1)$. For a maximally entangled state one has $\lambda_1 = \lambda_N = 1/N$, hence $m_1 = N$ and $m_0 = 0$. Therefore

$$\mathcal{O}_{\max} = \frac{U(N)}{U(N)} \times \frac{U(N)}{U(1)} = \frac{U(N)}{U(1)} = \frac{SU(N)}{\mathbb{Z}_N}, \quad (15.24)$$

with $\dim(\mathcal{O}_{max}) = N^2 - 1$, which equals *half* the total dimensionality of the space of pure states.

The set of all orbits foliate $\mathbb{C}\mathbf{P}^{N^2-1}$, the space of all pure states of the $N \times N$ system. This foliation is *singular*, since there exist measure zero leaves of various dimensions and topology. The dimensionalities of all local orbits for $N = 2, 3$ are shown in Figure 15.4, and their topologies in Table 15.1.

Table 15.1. *Topological structure of local orbits of the $N \times N$ pure states, D_s denotes the dimension of the subspace of the Schmidt simplex Δ_{N-1}, while D_o represents the dimension of the local orbit.*

N	Schmidt coefficients	D_s	Part of the asymmetric simplex	Local structure: base × fibre	D_o
2	(a, b)	1	line	$\mathbf{F}^{(2)} \times \mathbb{R}\mathbf{P}^3$	5
	$(1, 0)$	0	left edge	$\mathbb{C}\mathbf{P}^2 \times \mathbb{C}\mathbf{P}^2$	4
	$(1/2, 1/2)$	0	right edge	$U(2)/U(1) = \mathbb{R}\mathbf{P}^3$	3
3	(a, b, c)	2	interior of triangle	$\mathbf{F}^{(3)} \times \frac{U(3)}{U(1)}$	14
	$(a, b, 0)$	1	base	$\mathbf{F}^{(3)} \times \frac{U(3)}{[U(1)]^2}$	13
	(a, b, b)	1	2 upper sides	$\frac{U(3)}{U(1)\times U(2)} \times \frac{U(3)}{U(1)}$	12
	$(1/2, 1/2, 0)$	0	right corner	$\frac{U(3)}{U(1)\times U(2)} \times \frac{U(3)}{[U(1)]^2}$	11
	$(1, 0, 0)$	0	left corner	$\mathbb{C}\mathbf{P}^3 \times \mathbb{C}\mathbf{P}^3$	8
	$(1/3, 1/3, 1/3)$	0	upper corner	$U(3)/U(1)$	8

Observe that the local orbit defined by (15.19) contains all purifications of all mixed states acting on $\mathcal{H}_N$ isospectral with $\rho_N = \mathrm{Tr}_K|\psi\rangle\langle\psi|$. Sometimes one modifies (15.19) imposing additional restrictions, $K = N$ and $V = U$. Two states fulfilling this *strong local equivalence* (SLE) relation, $|\psi'\rangle = U \otimes U|\psi\rangle$ are *equal*, up to selection of the reference frame used to describe both subsystems. The basis is determined by a unitary U. Hence the orbit of the strongly locally equivalent states – the base in (15.20) – forms a coset space of all states of the form $U\rho_N U^\dagger$, as discussed in Section 8.5. In particular, for any maximally entangled state, there are no other states satisfying SLE, while for a separable state the orbit of SLE states forms the complex projective space $\mathbb{C}\mathbf{P}^{N-1}$ of all pure states of a single subsystem.

The question, if a given pure state $|\psi\rangle \in \mathcal{H}_N \otimes \mathcal{H}_K$ is separable, is easy to answer: it is enough to compute the partial trace, $\rho_N = \mathrm{Tr}_K(|\psi\rangle\langle\psi|)$, and to check if $\mathrm{Tr}\rho_N^2$ equals unity. If it is so the reduced state is pure, hence the initial pure state is separable. In the opposite case the pure state is entangled. The next question is: to what extent is a given state $|\psi\rangle$ entangled?

There seems not to be a unique answer to this question. Due to the Schmidt decomposition (9.8) one obtains the Schmidt vector $\vec{\lambda}$ of length N (we assume $N \leq K$), and may describe it by entropies analysed in Chapters 2 and 12. For

instance, the *entanglement entropy* is defined as the von Neumann entropy of the reduced state, which is equal to the Shannon entropy of the Schmidt vector,

$$E(|\psi\rangle) \equiv S(\rho_A) = S(\vec{\lambda}) = -\sum_{i=1}^{N} \lambda_i \ln \lambda_i \ . \tag{15.25}$$

It is equal to zero for separable states and $\ln N$ for maximally entangled states. In the similar way to measure entanglement one may also use the Rényi entropies (2.79) of the reduced state, $E_q \equiv S_q(\rho_A)$. We shall need a quantity related to E_2 called *tangle*

$$\tau(|\psi\rangle) \equiv 2(1 - \mathrm{Tr}\rho_A^2) = 2\Big(1 - \sum_{i=1}^{N} \lambda_i^2\Big) = 2\Big(1 - \exp[-E_2(|\psi\rangle)]\Big) \ , \tag{15.26}$$

which runs from 0 to $2(N-1)/N$, and its square root $C = \sqrt{\tau}$, called *concurrence*.[9] Another entropy, $E_\infty = -\ln \lambda_{\max}$, has an elegant geometric interpretation: if the Schmidt vector is ordered decreasingly and $\lambda_1 = \lambda_{\max}$ denotes its largest component then $|1\rangle \otimes |1\rangle$ is the separable pure state closest to $|\psi\rangle$ (Lockhart and Steiner, 2002). Thus the Fubini–Study distance of $|\psi\rangle$ to the set of separable pure states, $D_{\mathrm{FS}}^{\min} = \arccos(\sqrt{\lambda_{\max}})$, is a function of E_∞. Although one uses several different Rényi entropies E_q, the entanglement entropy $E = E_1$ is distinguished among them just as the Shannon entropy is singled out by its operational meaning discussed in Section 2.2.

For the two qubit problem the Schmidt vector has only two components, which sum to unity, so the entropy $E(|\psi\rangle) \in [0, \ln 2]$ characterizes uniquely the entanglement of the pure state $|\psi\rangle$. To analyse its geometry it is convenient to select a three-dimensional section of the space of pure states. The net of the tetrahedron used for the cover picture is shown in Appendix 3 – it presents entanglement at the boundary of the simplex defined by four separable states defining the standard basis. It is defined by Eqs. (4.70) and (4.71) and may be obtained by setting all phases ν_i in (4.67) to zero. Making use of the freedom of choice of the basis vectors we have selected four separable states to define a standard basis.

In general, for an $N \times N$ system the entropy is bounded, $0 \le E \le \ln N$, and to describe the entanglement completely one needs a set of $N-1$ independent quantities. What properties should they fulfil?

Before discussing this issue we need to distinguish certain classes of quantum operations acting on bipartite systems. *Local operations* (LO) arise as the tensor

[9] Concurrence was initially introduced for two qubits by Hill and Wootters (1997). We adopted here the generalization of Rungta, Bužek, Caves, Hillery and Milburn (2001) and Mintert, Kuś and Buchleitner (2004), but there are also other ways to generalize this notion for higher dimensions (Uhlmann, 2000; Wong and Christensen, 2001; Wootters, 2001; Audenaert, Verstraete and Moor, 2001b; Badziąg, Deaur, Horodecki, Horodecki and Horodecki, 2002).

product of two maps, both satisfying the trace preserving condition (10.54),

$$[\Phi_A \otimes \Phi_B](\rho) \;=\; \sum_i \sum_j (A_i \otimes B_j)\, \rho\, (A_i^\dagger \otimes B_j^\dagger)\,. \tag{15.27}$$

Any operation which might be written in the form

$$\Phi_{\rm sep}(\rho) \;=\; \sum_i (A_i \otimes B_i)\, \rho\, (A_i^\dagger \otimes B_i^\dagger)\,, \tag{15.28}$$

is called *separable* (SO). Observe that this form is more general than (15.27), even though the summation goes over one index. The third, important class of maps is called *LOCC*. This name stands for *local operations and classical communication* and means that all quantum operations, including measurements, are allowed, provided they are performed locally in each subsystem. Classical communication allows the two parties to exchange in both ways classical information about their subsystems, and hence to introduce classical correlations between them. One could think, all separable operations may be obtained in this way, but this is not true (Bennett, DiVincenzo, Fuchs, Mor, Rains, Shor, Smolin and Wootters, 1999a), and we have the proper inclusion relations LO $\subset$ LOCC $\subset$ SO.

The concept of local operations leads to the notion of *entanglement monotones*. These are the quantities which are invariant under unitary operations and decrease, on average, under LOCC (Vidal, 2000). The words 'on average' refer to the general case, in which a pure state is transformed by a probabilistic local operation into a mixture,

$$\rho \;\to\; \sum_i p_i \rho_i \quad\Rightarrow\quad \mu(\rho) \;\geq\; \sum_i p_i \mu(\rho_i)\,. \tag{15.29}$$

Note that if μ is a non-decreasing monotone, then $-\mu$ is a non-increasing monotone. Thus we may restrict our attention to the non-increasing monotones, which reflect the key paradigm of any entanglement measure: entanglement cannot increase under the action of local operations. Construction of entanglement monotones can be based on the following theorem (Nielsen, 1999):

Theorem 15.1 (Nielsen's majorization) *A given state $|\psi\rangle$ may be transformed into $|\phi\rangle$ by deterministic LOCC operations if and only if the corresponding vectors of the Schmidt coefficients satisfy the majorization relation (2.1)*

$$|\psi\rangle \;\stackrel{\rm LOCC}{\longrightarrow}\; |\phi\rangle \quad\Longleftrightarrow\quad \vec{\lambda}_\psi \;\prec\; \vec{\lambda}_\phi. \tag{15.30}$$

To prove the forward implication we follow the original proof. Assume that party A performs locally a generalized measurement, which is described by a set of k Kraus operators A_i. By classical communication the result is sent to party B, which

performs a local action Φ_i, conditioned on the result i. Hence

$$\sum_{i=1}^{k} [\mathbb{1} \otimes \Phi_i]\big(A_i|\psi\rangle\langle\psi|A_i^\dagger\big) = |\phi\rangle\langle\phi| \,. \tag{15.31}$$

The result is a pure state so each terms in the sum needs to be proportional to the projector. Tracing out the second subsystem we get

$$A_i \rho_\psi A_i^\dagger = p_i \rho_\phi \,, \qquad i = 1, \ldots, k, \tag{15.32}$$

where $\sum_{i=1}^{k} p_i = 1$ and $\rho_\psi = \mathrm{Tr}_B(|\psi\rangle\langle\psi|)$ and $\rho_\phi = \mathrm{Tr}_B(|\phi\rangle\langle\phi|)$. Due to the polar decomposition of $A_i\sqrt{\rho_\psi}$ we may write

$$A_i \sqrt{\rho_\psi} = \sqrt{A_i \rho_\psi A_i^\dagger}\; V_i = \sqrt{p_i \rho_\phi}\; V_i \tag{15.33}$$

with unitary V_i. Making use of the completeness relation (10.54) we obtain

$$\rho_\psi = \sqrt{\rho_\psi}\, \mathbb{1} \sqrt{\rho_\psi} = \sum_{i=1}^{k} \sqrt{\rho_\psi} A_i^\dagger A_i \sqrt{\rho_\psi} = \sum_{i=1}^{k} p_i V_i^\dagger \rho_\phi V_i, \tag{15.34}$$

and the last equality follows from (15.33) and its adjoint. Hence we arrived at an unexpected conclusion: if a local transformation $|\psi\rangle \to |\phi\rangle$ is possible, then there exists a bistochastic operation (10.71), which acts on the partially traced states with inversed time – it sends ρ_ϕ into ρ_ψ! The quantum HLP lemma (Section 12.5) implies the majorization relation $\vec{\lambda}_\psi \prec \vec{\lambda}_\phi$. The backward implication follows from an explicit conversion protocol proposed by Nielsen, or alternative versions presented in Hardy (1999), Jensen and Schack (2001) and Donald, Horodecki and Rudolph (2002).

The majorization relation (15.30) introduces a partial order into the set of pure states.[10] Hence any pure state $|\psi\rangle$ allows one to split the Schmidt simplex, representing the set of all local orbits, into three regions: the set F (*Future*) contains states which can be produced from $|\psi\rangle$ by LOCC , the set P (*Past*) of states from which $|\psi\rangle$ may be obtained, and eventually the set C of *incomparable* states, which cannot be joined by a local transformation in any direction.[11]

This structure resembles the 'causal structure' defined by the light cone in special relativity. See Figure 15.5, and observe the close similarity to Figure 12.4 showing paths in the simplex of eigenvalues that can be generated by bistochastic operations. The only difference is the *arrow of time*: the 'Past' for the evolution in the space

[10] A similar partial order induced by LOCC into the space of mixed states is analysed in (Hayden, Terhal and Uhlmann, n.d.b).

[11] For $N \geq 4$ there exists an effect of *entanglement catalysis* (Jonathan and Plenio, 1999a; Daftuar and Klimesh, 2001; Bandyopadhyay and Roychowdhury, 2002) that allows one to obtain certain incomparable states in the presence of additional entangled states.

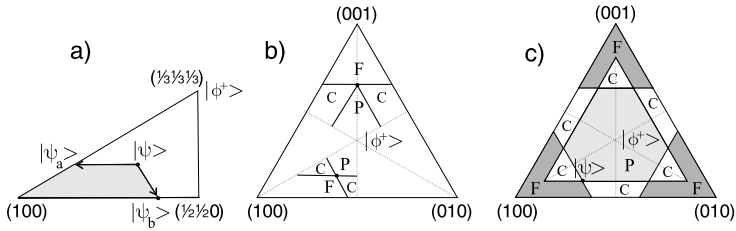


Figure 15.5. Simplex of Schmidt coefficients Δ_2 for 3×3 pure states: the corners represent separable states, centre the maximally entangled state $|\phi^+\rangle$. Panels (a)–(c) show '*Future*' and '*Past*' zones with respect to LOCC and are analogous to those in Figure 12.4, but the direction of the arrow of time is reversed.

of density matrices corresponds to the 'Future' for the local entanglement transformations and vice versa. In both cases the set C of incomparable states contains the same fragments of the simplex Δ_{N-1}. In a typical case C occupies regions close to the boundary of Δ_{N-1}, so one may expect the larger dimensionality N, the larger relative volume of C. This is indeed the case, and in the limit $N \to \infty$ two generic pure states of the $N \times N$ system (or two generic density matrices of size N) are incomparable (Clifton, Hepburn and Wuthrich, 2002).

The majorization relation (15.30) provides another justification for the observation that two pure states are interconvertible (locally equivalent) if and only if the have the same Schmidt vectors. More importantly, this theorem implies that any Schur concave function of the Schmidt vector $\vec{\lambda}$ is an entanglement monotone. In particular, this crucial property is shared by all Rényi entropies of entanglement $E_q(\vec{\lambda})$ including the entanglement entropy (15.25). To ensure a complete description of a pure state of the $N \times N$ problem one may choose $E_1, E_2, \ldots, E_{N-1}$. Other families of entanglement monotones include partial sums of Schmidt coefficients ordered decreasingly, $M_k(\vec{\lambda}) = \sum_{i=1}^{k} \lambda_k$ with $k = 1, \ldots, N-1$ (Vidal, 2000), subentropy (Jozsa et al., 1994; Mintert and Życzkowski, 2004), and symmetric polynomials in Schmidt coefficients (see Problem 15.2).

Since the maximally entangled state is majorized by all pure states, it cannot be reached from other states by any deterministic local transformation. Is it at all possible to create it locally? A possible clue is hidden in the word *average* contained in the majorization theorem.

Let us assume we have at our disposal n copies of a generic pure state $|\psi\rangle$. The majorization theorem does not forbid us to locally create out of them m maximally entangled states $|\psi^+\rangle$, at the expense of the remaining $n - m$ states becoming separable. Such protocols proposed in Bennett, Bernstein, Popescu and Schumacher (1996a) and Lo and Popescu (1998) are called *entanglement concentration*. This local operation is reversible, and the reverse process of transforming m maximally

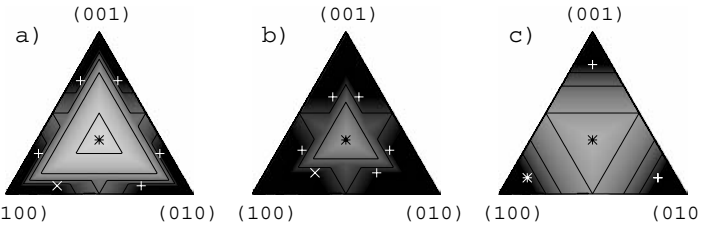


Figure 15.6. Probability of the optimal local conversion of an initial state $|\psi\rangle$ (white $\times$) of the 3×3 problem into a given pure state represented by a point in the Schmidt simplex. Initial Schmidt vector $\vec{\lambda}_\psi$ is (a) $(0.7, 0.25, 0.05)$, (b) $(0.6, 0.27, 0.03)$ and (c) $(0.8, 0.1, 0.1)$. Due to the degeneracy of $\vec{\lambda}$ in the latter case there exist only three interconvertible states in Δ_2, represented by $(+)$.

entangled states and $n - m$ separable states into n entangled states is called *entanglement dilution*. The asymptotic ratio $m/n \leq 1$ obtained by an optimal concentration protocol is called *distillable entanglement* (Bennett et al., 1996a) of the state $|\psi\rangle$ (see Problem 15.3).

Assume now that only one copy of an entangled state $|\psi\rangle$ is at our disposal. To generate maximally entangled state locally we may proceed in a probabilistic way: a local operation produces $|\psi^+\rangle$ with probability p and a separable state otherwise. Hence we allow a pure state $|\psi\rangle$ to be transformed into a mixed state. Consider a probabilistic scheme to convert a pure state $|\psi\rangle$ into a target $|\phi\rangle$ with probability p. Let p_c be the maximal number such that the following majorization[12] relation holds,

$$\lambda_\psi \prec p_c \vec{\lambda}_\phi. \tag{15.35}$$

It is easy to check that the probability p cannot be larger than p_c, since the Nielsen theorem would be violated. The optimal conversion strategy for which $p = p_c$ was explicitly constructed by Vidal (1999). The Schmidt rank cannot increase during any local conversion scheme (Lo and Popescu, 1998). If the rank of the target state $|\phi\rangle$ is larger than the Schmidt rank of $|\psi\rangle$, then $p_c = 0$ and the probabilistic conversion cannot be performed.[13]

This situation is illustrated in Figure 15.6, which shows the probability of accessing different regions of the Schmidt simplex for pure states of a 3×3 system for four different initial states $|\psi\rangle$. The shape of the black figure ($p = 1$ represents deterministic transformations) is identical with the set 'Future' in Figure 15.5. The

[12] If the sum of both vectors is not equal relation (2.1) is sometimes called *submajorization*.

[13] In such a case one may still perform a *faithful conversion* (Jonathan and Plenio, 1999b; Vidal, Jonathan and Nielsen, 2000) transforming the initial state $|\psi\rangle$ into a state $|\phi'\rangle$, for which its fidelity with the target, $|\langle\phi|\phi'\rangle|^2$, is maximal.

more entangled final state $|\phi\rangle$ (closer to the maximally entangled state – black $(*)$ in the centre of the triangle), the smaller probability p of a successful transformation. Observe that the contour lines (plotted at $p = 0.2, 0.4, 0.6$ and 0.8) are constructed from the iso-entropy lines S_q for $q \to 0$ and $q \to \infty$ (compare with Figure 2.14).

Let us close with an envoi: entanglement of a pure state of any bipartite system may be fully characterized by its Schmidt decomposition. In particular, all entanglement monotones are functions of the Schmidt coefficients. However, the Schmidt decomposition cannot be directly applied to the multipartite case (Peres, 1995; Carteret et al., 2000; Acín, Andrianov, Jané and Tarrach, 2001). These systems are still being investigated.[14] Let us just mention that pure states of three qubits can be entangled in two inequivalent ways. There exist three-qubit[15] pure states (Greenberger, Horne and Zeilinger, 1989)

$$|GHZ\rangle = \frac{1}{\sqrt{2}}\big(|000\rangle + |111\rangle\big) \quad \text{and} \quad |W\rangle = \frac{1}{\sqrt{3}}\big(|001\rangle + |010\rangle + |100\rangle\big) \tag{15.36}$$

which cannot be locally converted with a positive probability in any direction (Dür, Vidal and Cirac, 2000b).

15.4 Mixed states and separability

It is a good time to look again at mixed states: in this section we shall analyse bipartite density matrices, acting on a composite Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ of finite dimensionality $d = NK$. A state is called a *product state*, if it has a tensor product structure, $\rho = \rho_A \otimes \rho_B$. A mixed state ρ is called *separable*, if it can be represented as a convex sum of product states (Werner, 1989),

$$\rho_{sep} \;=\; \sum_{j=1}^{M} q_j \, \rho_j^A \otimes \rho_j^B \;, \tag{15.37}$$

where ρ^A acts in $\mathcal{H}_A$ and ρ^B acts in $\mathcal{H}_B$, the weights are positive, $q_j > 0$, and sum to unity, $\sum_{j=1}^{M} q_j = 1$. Such a decomposition is not unique. For any separable ρ, the

[14] Several families of local invariants and entanglement monotones were found (Sudbery, 2001; Brun and Cohen, 2001; Gingrich, 2002), properties of local orbits were analysed (Mosseri and Dandoloff, 2001; Bernevig and Chen, 2003; Miyake, 2003; Lévay, 2004), measures of multipartite entanglement were introduced (Coffman, Kundu and Wootters, 2000; Bennett, Popescu, Rohrlich, Smolin and Thapliyal, 2001; Wong and Christensen, 2001; Meyer and Wallach, 2002; Brennen, 2003; Heydari and Bjork, 2004) and a link between quantum mechanical and topological entanglement including knots and braids (Kauffmann and Lomonaco Jr., 2002; Asoudeh, Karimipour, Memarzadeh and Rezakhani, 2004), rings (O'Connor and Wootters, 2001) and graphs (Plesch and Bužek, 2003; Hein, Eisert and Briegel, 2004) have been discussed.

[15] A curious reader might be pleased to learn that *four* qubits can be entangled in *nine* different ways (Verstraete, Dahaene, DeMoor and Verschelde, 2002b). What is the number of different ways, one may entangle m qubits?

smallest number M of terms is called *cardinality* [16] of the state. By definition the set $\mathcal{M}_S$ of separable mixed states is convex. Separable states can be constructed locally using classical communication, and may exhibit classical correlations only.A mixed state which is not separable, hence may display non-classical correlations, is called *entangled*.[17] It is easy to see that for pure states both definitions are consistent.

Any density matrix ρ acting on d-dimensional Hilbert space may be represented as a sum (8.20) over d^2-1 trace-less generators σ_i of $SU(d)$. However, analysing a composite system for which $d = NK$, it is more advantageous to use the basis of the product group $SU(N) \otimes SU(K)$, which leads us to the **Fano form** (Fano, 1983)

$$\rho = \frac{1}{NK}\Big[\mathbb{1}_{NK} + \sum_{i=1}^{N^2-1} \tau_i^A \sigma_i \otimes \mathbb{1}_K + \sum_{j=1}^{K^2-1} \tau_j^B \mathbb{1}_N \otimes \sigma_j + \sum_{i=1}^{N^2-1}\sum_{j=1}^{K^2-1} \beta_{ij}\sigma_i \otimes \sigma_j\Big]. \tag{15.38}$$

Here $\vec{\tau}^A$ and $\vec{\tau}^B$ are Bloch vectors of the partially reduced states, while a real $(N^2-1)\times(K^2-1)$ matrix β describes the correlation between both subsystems. If $\beta = 0$ then the state is separable, but the reverse is not true.[18] Keeping both Bloch vectors constant and varying β in such a way to preserve positivity of ρ we obtain a $(N^2-1)(K^2-1)$-dimensional family of bipartite mixed states, which are locally indistinguishable.

The definition of separability (15.37) is implicit, so it is in general not easy to see if such a decomposition exists for a given density matrix. Separability criteria found so far may be divided into two disjoint classes: **A)** sufficient and necessary, but not practically usable; and **B)** easy to use, but only necessary (or only sufficient). A simple, albeit amazingly powerful criterion was found by Peres (1996), who analysed the action of partial transposition on an arbitrary separable state,

$$\rho_{\rm sep}^{T_A} \equiv (T \otimes \mathbb{1})(\rho_{\rm sep}) = \sum_j q_j\, (\rho_j^A)^T \otimes \rho_j^B \geq 0. \tag{15.39}$$

Thus any separable state has a positive partial transpose (is PPT), so we obtain directly

B1. PPT criterion. *If* $\rho^{T_A} \not\geq 0$, *the state* ρ *is entangled.*

[16] Due to Carathéodory's theorem the cardinality is not larger then d^2 (Horodecki, 1997). In the two-qubit case it is not larger than $d = 4$ (Sanpera, Tarrach and Vidal, 1998), while for systems of higher dimensions it is typically larger than the rank $r \leq d$ (Lockhart, 2000).

[17] The notion of *entanglement* may also be used in the set-up of classical probability distributions (Tucci, n.d.b), theory of Lie-algebras or convex sets (Barnum, Knill, Ortiz and Viola, 2003) and may be compared with *secret* classical correlations (Collins and Popescu, 2002).

[18] Note that for product states $M_{ij} \equiv \beta_{ij} - \tau_i^A \tau_j^B = 0$, hence the norm $||M||^2$ characterizes to what extent ρ is not a product state (Schlienz and Mahler, 1995).

Is is extremely easy to use: all we need to do is to perform the partial transposition of the density matrix in question, diagonalize, and check if all eigenvalues are non-negative. Although partial transpositions were already defined in (10.34), let us have a look at how both operations act on a block matrix,

$$X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad X^{T_B} \equiv \begin{bmatrix} A^T & B^T \\ C^T & D^T \end{bmatrix}, \quad X^{T_A} \equiv \begin{bmatrix} A & C \\ B & D \end{bmatrix}. \tag{15.40}$$

Note that $X^{T_B} = (X^{T_A})^T$, so the spectra of the two operators are the same and the above criterion may be equivalently formulated with the map $T_B = (\mathbb{1} \otimes T)$. Furthermore, partial transposition applied on a density matrix produces the same spectrum as the transformation of flipping one of both Bloch vectors present in its Fano form (15.38). Alternatively one may change the signs of all generators σ_j of the corresponding group. For instance, flipping the second of the two subsystems of the same size we obtain

$$\rho^{T_B} \equiv \frac{1}{N^2}\Big[\mathbb{1}_{N^2} + \sum_{i=1}^{N^2-1} \tau_i^A \sigma_i \otimes \mathbb{1}_N - \sum_{j=1}^{N^2-1} \tau_j^B \mathbb{1}_N \otimes \sigma_j - \sum_{i,j=1}^{N^2-1} \beta_{ij} \sigma_i \otimes \sigma_j\Big], \tag{15.41}$$

with the same spectrum as ρ^{T_A}. In the two-qubit case, reflection of all three components of the Bloch vector, $\vec{\tau}^{\,B} \to -\vec{\tau}^{\,B}$, is equivalent to changing the sign of its single component τ_y^B (partial transpose), followed by the π-rotation along the y-axis.

To watch the PPT criterion in action consider the family of *generalized Werner states* [19] which interpolate between maximally mixed state ρ_* and the maximally entangled state $P_+ = |\phi^+\rangle\langle\phi^+|$,

$$\rho_W(x) = x|\phi^+\rangle\langle\phi^+| + (1-x)\frac{1}{N}\mathbb{1} \quad \text{with} \quad x \in [0,1]. \tag{15.42}$$

One eigenvalue equals $[1 + (N-1)x]/N$, and the remaining $(N-1)$ eigenvalues are degenerate and equal to $(1-x)/N$. In the $N = 2$ case:

$$\rho_x = \frac{1}{4}\begin{bmatrix} 1+x & 0 & 0 & 2x \\ 0 & 1-x & 0 & 0 \\ 0 & 0 & 1-x & 0 \\ 2x & 0 & 0 & 1+x \end{bmatrix},$$
$$\rho_x^{T_A} = \frac{1}{4}\begin{bmatrix} 1+x & 0 & 0 & 0 \\ 0 & 1-x & 2x & 0 \\ 0 & 2x & 1-x & 0 \\ 0 & 0 & 0 & 1+x \end{bmatrix}. \tag{15.43}$$

[19] For the original *Werner states* (Werner, 1989) the singlet pure state $|\psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ was used instead of $|\phi^+\rangle$.

Diagonalization of the partially transposed matrix $\rho_x^{T_A} = \rho_x^{T_B}$ gives the spectrum $\frac{1}{4}\{1+x, 1+x, 1+x, 1-3x\}$. This matrix is positive definite if $x \leq 1/3$, hence Werner states are entangled for $x > 1/3$. It is interesting to observe that the critical state $\rho_{1/3} \in \partial\mathcal{M}_{\rm sep}$, is localized at the distance $r_{\rm in} = 1/\sqrt{24}$ from the maximally mixed state ρ_*, so it sits on the insphere, the maximal sphere that one can inscribe into the set $\mathcal{M}^{(4)}$ of $N = 4$ mixed states.

As we shall see below the PPT criterion works in both directions only if $\dim(\mathcal{H}) \leq 6$, so there is a need for other separability criteria.[20] Before reviewing the most important of them let us introduce one more notion often used in the physical literature.

An Hermitian operator W is called an *entanglement witness* for a given entangled state ρ if $\mathrm{Tr}\rho W < 0$ and $\mathrm{Tr}\rho\omega \geq 0$ for all separable σ (Horodecki et al., 1996a; Terhal, 2000b). For convenience the normalization $\mathrm{Tr}W = 1$ is assumed. Horodecki, Horodecki and Horodecki (1996a) proved a useful lemma:

Lemma 15.1 (Witness) *For any entangled state ρ there exists an entanglement witness W.*

In fact this is the Hahn–Banach separation theorem (Section 1.1) in slight disguise.

It is instructive to realize there is a direct relation with the dual cones construction discussed in Chapter 11: any witness operator is proportional to a dynamical matrix, $W = D_\Phi/N$, corresponding to a non-completely positive map Φ. Since D_Φ is block positive (positive on product states), the condition $\mathrm{Tr}\, W\sigma \geq 0$ holds for all separable states for which the decomposition (15.37) exists. Conversely, a state ρ is separable if $\mathrm{Tr}\, W\rho \geq 0$ for all block positive W. This is just the definition (11.17) of a super-positive map Ψ. We arrive, therefore, at a key observation: the set $\mathcal{SP}$ of super-positive maps is isomorphic with the set $\mathcal{M}_S$ of separable states by the Jamiołkowski isomorphism, $\rho = D_\Psi/N$.

An intricate link between positive maps and the separability problem is made clear in the

A1. Positive maps criterion (Horodecki et al., 1996a). *A state ρ is separable if and only if $\rho' = (\Phi \otimes \mathbb{1})\rho$ is positive for all positive maps Φ.*

To demonstrate that this condition is necessary, act with an extended map on the separable state (15.37),

$$\Big(\Phi \otimes \mathbb{1}\Big)\Big(\sum_j q_j\, \rho_j^A \otimes \rho_j^B\Big) \;=\; \sum_j q_j\, \Phi(\rho_j^A) \otimes \rho_j^B \;\geq\; 0\,. \qquad (15.44)$$

Due to positivity of Φ the above combination of positive operators is positive. To prove sufficiency, assume that $\rho' = (\Phi \otimes \mathbb{1})\rho$ is positive. Thus $\mathrm{Tr}\rho' P \geq 0$ for any

[20] For recent analysis of the problem consult also Lewenstein, Bruß, Cirac, Kraus, Kuś, Samsonowicz, Sanpera and Tarrach (2000a), Horodecki, Horodecki and Horodecki (2000b), Bruß, Cirac, Horodecki, Hulpke, Kraus, Lewenstein and Sanpera (2002), Terhal (2002) and Bruß (2002).

projector P. Setting $P = P_+ = |\phi^+\rangle\langle\phi^+|$ and making use of the adjoint map we get $\mathrm{Tr}\rho(\Phi \otimes \mathbb{1})P_+ = \frac{1}{N}\mathrm{Tr}\rho \mathrm{D}_\Phi \geq 0$. Since this property holds for all positive maps Φ, it implies separability of ρ due to the witness lemma $\boxtimes$.

The positive maps criterion holds also if the map acts on the second subsystem. However, this criterion is not easy to use: one needs to verify that the required inequality is satisfied for *all* positive maps. The situation becomes simple for the 2×2 and 2×3 systems. In this case any positive map is decomposable due to the Størmer–Woronowicz theorem and may be written as a convex combination of a CP map and a CcP map, which involves the transposition T (see Section 11.1). Hence, to apply the above criterion we need to perform one check working with the partial transposition $T_A = (T \otimes \mathbb{1})$. In this way we become

B1'. Peres–Horodeccy criterion (Peres, 1996; Horodecki et al., 1996a). *A state ρ acting on $\mathcal{H}_2 \otimes \mathcal{H}_2$ (or $\mathcal{H}_2 \otimes \mathcal{H}_3$) composite Hilbert space is separable if and only if $\rho^{T_A} \geq 0$.*

In general, the set of bipartite states may be divided into *PPT states* (positive partial transpose) and *NPPT states* (not PPT). A map Φ is related by the Jamiołkowski isomorphism (11.22) to a PPT state if $\Phi \in \mathcal{CP} \cap \mathcal{CcP}$. Complete co-positivity of Φ implies that $T\Phi$ is completely positive, so $(T\Phi \otimes \mathbb{1})\rho \geq 0$ for any state ρ. Thus $\rho' = (\Phi \otimes \mathbb{1})\rho$ is a PPT state, so such a map may be called *PPT inducing* [21] $\mathcal{PPTM} \equiv \mathcal{CP} \cap \mathcal{CcP}$ (see Figure 11.4).

Similarly, a super-positive map Φ is related by the isomorphism (11.22) with a separable state. Hence $(\Phi \otimes \mathbb{1})$ acting on the maximally entangled state $|\phi^+\rangle\langle\phi^+|$ is separable. It is then not surprising that $\rho' = (\Phi \otimes \mathbb{1})\rho$ becomes separable for an arbitrary state ρ (Horodecki, Shor and Ruskai, 2003b), which explains why SP maps are also called *entanglement breaking channels*. Furthermore, due to the positive maps criterion $(\Psi \otimes \mathbb{1})\rho' \geq 0$ for any positive map Ψ. In this way we have arrived at the first of three duality conditions equivalent to (11.16)–(11.18),

$$\{\Phi \in \mathcal{SP}\} \quad \Leftrightarrow \quad \Psi \cdot \Phi \in \mathcal{CP} \quad \text{for all} \quad \Psi \in \mathcal{P}\,, \tag{15.45}$$

$$\{\Phi \in \mathcal{CP}\} \quad \Leftrightarrow \quad \Psi \cdot \Phi \in \mathcal{CP} \quad \text{for all} \quad \Psi \in \mathcal{CP}\,, \tag{15.46}$$

$$\{\Phi \in \mathcal{P}\} \quad \Leftrightarrow \quad \Psi \cdot \Phi \in \mathcal{CP} \quad \text{for all} \quad \Psi \in \mathcal{SP}\,. \tag{15.47}$$

The second condition reflects the fact that a composition of two CP maps is CP, while the third one is dual to the first.

Due to the Størmer and Woronowicz theorem and the Peres–Horodeccy criterion, all PPT states for 2×2 and 2×3 problems are separable (hence any PPT-inducing map is SP) while all NPPT states are entangled. In higher dimensions there exist

[21] These maps should not be confused with *PPT-preserving* maps (Rains, 2001; Eggeling, Vollbrecht, Werner and Wolf, 2001), which act on bipartite systems and fulfil another property: if $\rho^{T_A} \geq 0$ then $(\Psi(\rho))^{T_A} \geq 0$.

PPT entangled states (PPTES), and this fact motivates investigation of positive, non-decomposable maps and other separability criteria.

B2. Range criterion (Horodecki, 1997). *If a state ρ is separable, then there exists a set of pure product states such that $|\psi_i \otimes \phi_i\rangle$ span the range of ρ and $T_B(|\psi_i \otimes \phi_i\rangle$ span the range of ρ^{T_B}.*

The action of the partial transposition on a product state gives $|\psi_i \otimes \phi_i^*\rangle$, where * denotes complex conjugation in the standard basis. This criterion, proved by P. Horodecki (1997), allowed him to identify the first PPTES in the $2 \otimes 4$ system. Entanglement of ρ was detected by showing that none of the product states from the range of ρ, if partially conjugated, belong to the range of ρ^{T_B}.

The range criterion allows one to construct PPT entangled states related to *unextendible product basis*, (UPB). It is a set of orthogonal product vectors $|u_i\rangle \in \mathcal{H}_N \otimes \mathcal{H}_M$, $i = 1, \ldots, k < MN$, such that there does not exist any product vectors orthogonal to all of them (Bennett, DiVincenzo et al., 1999b; Alon and Lovasz, 2001; DiVincenzo, Mor, Shor, Smolin and Terhal, 2003). We shall recall an example found in Bennett, DiVincenzo, Mor, Shor, Smolin and Terhal (1999b) for 3×3 system,

$$|u_1\rangle = \frac{1}{\sqrt{2}}|0\rangle \otimes |0-1\rangle, \quad |u_2\rangle = \frac{1}{\sqrt{2}}|2\rangle \otimes |1-2\rangle, \quad |u_3\rangle = \frac{1}{\sqrt{2}}|0-1\rangle \otimes |2\rangle,$$
$$|u_4\rangle = \frac{1}{\sqrt{2}}|1-2\rangle \otimes |0\rangle, \quad |u_5\rangle = \frac{1}{3}|0+1+2\rangle \otimes |0+1+2\rangle. \tag{15.48}$$

These five states are mutually orthogonal. However, since they span full three-dimensional spaces in both subsystems, no product state may be orthogonal to all of them.

For a given UPB let $P = \sum_{i=1}^{k} |u_i\rangle\langle u_i|$ denote the projector on the space spanned by these product vectors. Consider the mixed state, uniformly covering the complementary subspace,

$$\rho \equiv \frac{1}{MN-k}(\mathbb{1} - P). \tag{15.49}$$

By construction this subspace does not contain any product vectors, so ρ is entangled due to the range criterion. On the other hand, the projectors $(|u_i\rangle\langle u_i|)^{T_B}$ are mutually orthogonal, so the operator $P^{T_B} = \sum_{i=1}^{k}(|u_i\rangle\langle u_i|)^{T_B}$ is a projector. So is $(\mathbb{1} - P)^{T_B}$, hence ρ^{T_B} is positive. Thus the state (15.49) is a positive partial transpose entangled state.[22]

[22] The UPB method was used to construct PPTES in (Bennett et al., 1999b; Bruß and Peres, 2000; DiVincenzo, Mor, Shor, Smolin and Terhal, 2003; Pittenger, 2003), while not completely positive maps were applied in (Ha et al., 2003; Benatti, Floreanini and Piani, 2004) for this purpose. Conversely, PPTES were used in (Terhal, 2000b) to find non-decomposable positive maps.

B3. Reduction criterion (Cerf, Adami and Gingrich, 1999; Horodecki and Horodecki, 1999). *If a state ρ is separable then the reduced states $\rho_A = \mathrm{Tr}_B \rho$ and $\rho_B = \mathrm{Tr}_A \rho$ satisfy*

$$\rho_A \otimes \mathbb{1} - \rho \geq 0 \quad \text{and} \quad \mathbb{1} \otimes \rho_B - \rho \geq 0 \,. \tag{15.50}$$

This statement follows directly from the positive maps criterion with the map $\Phi(\sigma) = (\mathrm{Tr}\sigma)\mathbb{1} - \sigma$ applied to the first or the second subsystem. Computing the dynamical matrix for this map composed with the transposition, $\Phi' = \Phi T$, we find that $D_{\Phi'} \geq 0$, hence Φ is CcP and (trivially) decomposable. Thus the reduction criterion cannot be stronger[23] than the PPT criterion. There exists, however, a good reason to pay some attention to this criterion: the Horodecki brothers have shown (Horodecki and Horodecki, 1999) that any state ρ violating (15.50) is *distillable*, that is there exists a LOCC protocol which allows one to extract locally maximally entangled states out of ρ or its copies (Bennett, DiVincenzo, Smolin and Wootters, 1996b; Rains, 1999b). Entangled states, which are not distillable are called *bound entangled* (Horodecki, Horodecki and Horodecki, 1998; Horodecki, Horodecki and Horodecki, 1999).

A general question, which mixed state may be distilled,[24] is not solved yet (Bruß et al., 2002). Again the situation is clear for systems with $\dim(\mathcal{H}) \leq 6$: all PPT states are separable, and all NPPT states are entangled and distillable. For larger systems there exist PPT entangled states[25] and all of them are not distillable, hence bound entangled (Horodecki et al., 1998). Conversely, one could think that all NPPT entangled states are distillable, but this seems not to be the case (Dür, Cirac, Lewenstein and Bruß, 2000a; DiVincenzo, Shor, Smolin, Terhal and Thapliyal, 2000a).

B4. Majorization criterion (Nielsen and Kempe, 2001). *If a state ρ is separable, then the reduced states ρ_A and ρ_B satisfy the majorization relations*

$$\rho \prec \rho_A \qquad \text{and} \qquad \rho \prec \rho_B \,. \tag{15.51}$$

In brief, separable states are more disordered globally than locally. To prove this criterion one needs to find a bistochastic matrix B such that the spectra satisfy $\vec{\lambda} = B\vec{\lambda}_A$ (see Problem 15.4). The majorization relation implies that any Schur convex functions satisfies the inequality (2.8). For Schur concave functions the direction of the inequality changes. In particular, the *entropy criterion* follows:

[23] This is the case for the *generalized reduction criterion* proposed in Albeverio, Chen and Fei (2003).

[24] Following literature we use two similar terms: entanglement *concentration* and *distillation*, for local operations performed on pure and mixed states, respectively. While the former operations are reversible, the latter are not.

[25] Interestingly, there are no bound entangled states of rank one or two (Horodecki, Smolin, Terhal and Thapliyal, 2003c).

B5. Entropy criterion. *If a state ρ is separable, then the Rényi entropies fulfil*

$$S_q(\rho) \geq S_q(\rho_A) \quad \text{and} \quad S_q(\rho) \geq S_q(\rho_B) \quad \text{for} \quad q \geq 0\,. \tag{15.52}$$

The entropy criterion was originally formulated for $q = 1$ (Horodecki and Horodecki, 1994). Then this statement may be equivalently expressed in terms of the *conditional entropy*[26] $S(A|B) = S(\rho_{AB}) - S(\rho_A)$: for any separable bipartite state $S(A|B)$ is non-negative. Thus negative conditional entropy of a state ρ_{AB} confirms its entanglement (Horodecki and Horodecki, 1996; Schumacher and Nielsen, 1996; Cerf and Adami, 1999). The entropy criterion was proved for $q = 2$ in (Horodecki, Horodecki and Horodecki, 1996b) and later formulated also for the Havrda–Charvat–Tsallis entropy (2.77) (Abe and Rajagopal, 2001; Tsallis, Lloyd and Baranger, 2001; Rajagopal and Rendell, 2002; Rossignoli and Canosa, 2002). Its combination with the entropic uncertainty relations of Maassen and Uffink (1988) provides yet another interesting family of separability criteria (Gühne and Lewenstein, 2004). However, it is worth emphasizing that in general the spectral properties do not determine separability – there exist pairs of isospectral states, one of which is separable, the other not (compare with Problem 15.5).

A2. Contraction criterion. *A bipartite state ρ is separable if and only if any extended trace preserving positive map act as a (weak) contraction in sense of the trace norm,*

$$||\rho'||_{\mathrm{Tr}} \;=\; ||(\mathbb{1} \otimes \Phi)\rho||_{\mathrm{Tr}} \;\leq\; ||\rho||_{\mathrm{Tr}} \;=\; \mathrm{Tr}\rho \;=\; 1\,. \tag{15.53}$$

This criterion was formulated in (Horodecki, Horodecki and Horodecki, n.d.b) basing on earlier papers (Rudolph, 2003a; Chen and Wu, 2003). To prove it notice that the sufficiency follows from the positive map criterion: since $\mathrm{Tr}\rho' = 1$, hence $||\rho'||_{\mathrm{Tr}} \leq 1$ implies that $\rho' \geq 0$. To show the converse consider a normalized product state $\rho = \rho^A \otimes \rho^B$. Any trace preserving positive map Φ acts as isometry in sense of the trace norm, and the same is true for the extended map,

$$||\rho'||_{\mathrm{Tr}} = \big|\big|(\mathbb{1} \otimes \Phi)(\rho^A \otimes \rho^B)\big|\big|_{\mathrm{Tr}} = ||\rho^A|| \cdot ||\Phi(\rho^B)||_{\mathrm{Tr}} = 1\,. \tag{15.54}$$

Since the trace norm is convex, $||A + B||_{\mathrm{Tr}} \leq ||A||_{\mathrm{Tr}} + ||B||_{\mathrm{Tr}}$, any separable state fulfils

$$\Big|\Big|(\mathbb{1} \otimes \Phi)\Big(\sum_i q_i(\rho_i^A \otimes \rho_i^B)\Big)\Big|\Big|_{\mathrm{Tr}} \;\leq\; \sum_i q_i\, ||\rho_i^A \otimes \Phi(\rho_i^B)||_{\mathrm{Tr}} \;=\; \sum_i q_i = 1\,, \tag{15.55}$$

which ends the proof. ⊠

[26] The opposite quantity, $-S(A|B)$, is called *coherent quantum information* (Schumacher and Nielsen, 1996) and plays an important role in quantum communication (Horodecki, Horodecki, Horodecki and Oppenheim, 2005).

Several particular cases of this criterion could be useful. Note that the celebrated PPT criterion **B1** follows directly, if the transposition T is selected as a trace preserving map Φ, since the norm condition, $||\rho^{T_A}||_{\mathrm{Tr}} \leq 1$, implies positivity, $\rho^{T_A} \geq 0$. Moreover, one may formulate an analogous criterion for global maps Ψ, which act as contractions on any bipartite product states, $||\Psi(\rho_A \otimes \rho_B)||_{\mathrm{Tr}} \leq 1$. As a representative example let us mention

B6. Reshuffling criterion.[27] *If a bipartite state ρ is separable then reshuffling (10.33) does not increase its trace norm,*

$$||\rho^R||_{\mathrm{Tr}} \;\leq\; ||\rho||_{\mathrm{Tr}} = 1 \,. \tag{15.56}$$

We shall start the proof considering an arbitrary product state, $\sigma_A \otimes \sigma_B$. By construction its Schmidt decomposition consists of one term only. This implies

$$||(\sigma_A \otimes \sigma_B)^R||_{\mathrm{Tr}} = 2\,||\sigma_A||_2 \cdot ||\sigma_B||_2 = \sqrt{\mathrm{Tr}\,\sigma_A^2}\,\sqrt{\mathrm{Tr}\,\sigma_B^2} \;\leq\; 1 \,. \tag{15.57}$$

Since the reshuffling transformation is linear, $(A+B)^R = A^R + B^R$, and the trace norm is convex, any separable state satisfies

$$\Big|\Big|\Big(\sum_i q_i(\sigma_i^A \otimes \sigma_i^B)\Big)^R\Big|\Big|_{\mathrm{Tr}} \;\leq\; \sum_i q_i\,||(\sigma_i^A \otimes \sigma_i^B)^R||_{\mathrm{Tr}} \;\leq\; \sum_i q_i = 1 \,, \tag{15.58}$$

which completes the reasoning. ⊠

In the simplest case of two qubits, the latter criterion is weaker than the PPT: examples of NPPT states, the entanglement of which is not detected by reshuffling, were provided by Rudolph (2003b). However, for some larger dimensional problems the reshuffling criterion becomes useful, since it is capable of detecting PPT entangled states, for which $||\rho^R||_{\mathrm{Tr}} > 1$ (Chen and Wu, 2003).

The problem of which separability criterion[28] is the strongest, and what the implication chains among them are, remains a subject of a vivid research (Vollbrecht and Wolf, 2002; Albeverio et al., 2003; Chen and Wu, 2004; Batle, Plastino, Casas and Plastino, 2004). In general, the separability problem is 'hard', since it is known that it belongs to the NP complexity class (Gurvits, 2003). Due to this intriguing mathematical result it is not surprising that all operationally feasible analytic criteria provide partial solutions only. On the other hand, one should appreciate practical

[27] Called also *realignment* criterion (Chen and Wu, 2003) or *computable cross-norm criterion* (Rudolph, 2003b). It is related to the earlier *minimal cross-norm* criterion of Rudolph (2003a), which provides a necessary and sufficient condition for separability, but is in general not practical to use.

[28] There exists several other separability criteria, not discussed here. Let us mention applications of the range criterion for $2 \times N$ systems (Dür et al., 2000a), checks for low rank density matrices (Horodecki, Lewenstein, Vidal and Cirac, 2000c), reduction of the dimensionality of the problem (Woerdeman, 2004), relation between purity of a state and its maximal projection on a pure states (Lewenstein et al., 2000a), or criterion obtained by expanding a mixed state in the Fourier basis (Pittenger and Rubin, 2000).

methods constructed to decide separability numerically. Iterative algorithms based on an extension of the PPT criterion for higher dimensional spaces (Doherty, Parillo and Spedalieri, 2002; Doherty et al., 2004) or non-convex optimization (Eisert, Hyllus, Guhne and Curty, 2004) are able to detect the entanglement in a finite number of steps. Another algorithm provides an explicit decomposition into pure product states (Hulpke and Bruß, 2005), confirming that the given mixed state ρ is separable. A combination of these two approaches terminates after a finite time t and gives an inconclusive answer only if ρ belongs to the ϵ-vicinity of the boundary of the set of separable states. By increasing the computation time t one may make the width ϵ of the 'no man's land' arbitrarily small.

15.5 Geometry of the set of separable states

Equipped with a broad spectrum of separability criteria, we may try to describe the structure of the set $\mathcal{M}_S$ of the separable states. This task becomes easier for the two-qubit system, for which positive partial transpose implies separability. Hence the set of $N = 4$ separable states arises as an intersection of the entire body of mixed states with its reflection induced by partial transpose,

$$\mathcal{M}_S^{(4)} = \mathcal{M}^{(4)} \cap T_A(\mathcal{M}^{(4)}) , \tag{15.59}$$

(see Figure 11.2(b)). This observation suggests that the set of separable states has a positive volume. The maximally mixed state is invariant with respect to partial transpose, $\rho_* = \rho_*^{T_B}$ and occupies the centre of the body $\mathcal{M}^{(4)}$. It is thus natural to ask, what is the radius of the separable ball centred at ρ_*? The answer is very appealing in the simplest, Euclidean geometry: the entire maximal 15-D ball inscribed in $\mathcal{M}^{(4)}$ is separable (Życzkowski, Horodecki, Sanpera and Lewenstein, 1998). Working with the distance D_2 defined in Eq. (8.3), its radius reads $r_{\rm in} = 1/\sqrt{24}$.

The separable ball is sketched in two- or three-dimensional cross sections of $\mathcal{M}^{(4)}$ in Figure 15.7. To prove its separability[29] we shall invoke

Theorem 15.2 (Mehta's) *(Mehta, 1989). Let A be a Hermitian matrix of size D and let $\alpha = {\rm Tr}A/\sqrt{{\rm Tr}A^2}$. If $\alpha \geq \sqrt{D-1}$ then A is positive.*

Its proof begins with an observation that both traces are basis independent, so we may work in the eigenbasis of A. Let $(x_1, \ldots x_D)$ denote the spectrum of A. Assume first that one eigenvalue, say x_1, is negative. Making use of the right-hand side of the standard estimation between the l_1- and l_2-norms (with prefactor 1) of an N-vector,

[29] An explicit separability decomposition (15.37) for any state inside the ball was provided in (Braunstein, Caves, Jozsa, Linden, Popescu and Schack, 1999).

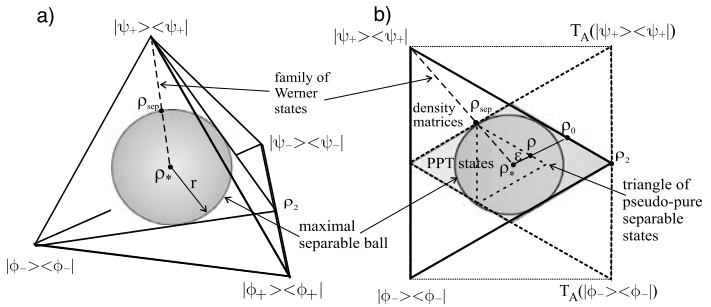


Figure 15.7. Maximal ball inscribed inside the 15-D body $\mathcal{M}^{(4)}$ of mixed states is separable: (a) 3-D cross section containing four Bell states, (b) 2-D cross section defined by two Bell states and ρ_* with the maximal separable triangle of pseudo-pure states.

$||A||_2 \leq ||A||_1 \leq N||A||_2$, we infer

$$\mathrm{Tr}A = \sum_{i=1}^{D} x_i < \sum_{i=2}^{D} x_i \leq \sqrt{D-1}\Big(\sum_{i=2}^{D} x_i^2\Big)^{1/2} < \sqrt{D-1}\sqrt{\mathrm{Tr}A^2}\,. \qquad (15.60)$$

This implies that $\alpha < \sqrt{D-1}$. Hence if the opposite is true and $\alpha \geq \sqrt{D-1}$ then none of the eigenvalues x_i could be negative, so $A \geq 0$. ⊠

The partial transpose preserves the trace and the HS norm of any state, $||\rho^{T_B}||_2^2 = ||\rho||_2^2 = \frac{1}{2}\,\mathrm{Tr}\rho^2$. Taking for A a partially transposed density matrix ρ^{T_B} we see that $\alpha^2 = 1/\mathrm{Tr}\rho^2$. Let us apply the Mehta lemma to an arbitrary mixed state of a $N \times N$ bipartite system, for which the dimension $D = N^2$,

$$1/\mathrm{Tr}\rho^2 \;\geq\; N^2 - 1 \quad \Rightarrow \quad \rho \quad \text{is} \quad \text{PPT}. \qquad (15.61)$$

Since the purity condition $\mathrm{Tr}\rho^2 = 1/(D-1)$ characterizes the insphere of $\mathcal{M}^{(D)}$, we conclude that for any bipartite[30] system the entire maximal ball inscribed inside the set of mixed states consists of PPT states only. This property implies separability for 2×2 systems. Separability of the maximal ball for higher dimensions was established by Gurvits and Barnum (2002), who later estimated the radius of the separable ball for multipartite systems (Gurvits and Barnum, 2003, 2004).

For any $N \times N$ system the volume of the maximal separable ball, $\mathbf{B}^{\mathrm{sep}}_{N^4-1}$ may be compared with the Euclidean volume (14.38) of $\mathcal{M}^{(N^2)}$. The ratio

$$\frac{\mathrm{Vol}(\mathbf{B}^{\mathrm{sep}}_{N^4-1})}{\mathrm{Vol}\big(\mathcal{M}^{(N^2)}\big)} = \frac{\pi^{(N^2-1)/2}\,2^{(N^2-N^4)/2}\,\Gamma(N^4)}{\Gamma[(N^4+1)/2]\;N^{N^4}\,(N^2-1)^{(N^4-1)/2}\;\prod_{k=1}^{N^2}\Gamma(k)} \qquad (15.62)$$

[30] The same is true for multipartite systems (Kendon, Życzkowski and Munro, 2002).

decreases fast with N, which suggests that for higher-dimensional systems the separable states are not typical. The actual probability p to find a separable mixed state is positive for any finite N and depends on the measure used (Życzkowski, 1999; Slater, 1999b; Slater, 2005). However, in the limit $N \to \infty$ the set of separable states is nowhere dense (Clifton and Halvorson, 2000), so the probability p computed with respect to an arbitrary non-singular measure tends to zero.

Another method of exploring the vicinity of the maximally mixed state consists in studying *pseudo-pure states*

$$\rho_\epsilon \equiv \frac{\mathbb{1}}{N^2}(1-\epsilon) + \epsilon |\phi\rangle\langle\phi| \,, \tag{15.63}$$

which are relevant for experiments with nuclear magnetic resonance (NMR) for $\epsilon \ll 1$. The set $\mathcal{M}_\epsilon$ is then defined as the convex hull of all ϵ-pseudo pure states. It forms a smaller copy of the entire set of mixed states of the same shape and is centred at $\rho_* = \mathbb{1}/N^2$.

For instance, since the cross section of the set $\mathcal{M}^{(4)}$ shown in Figure 15.7(b) is a triangle, so is the set $\mathcal{M}_{\epsilon_c}$ – a dashed triangle located inside the dark rhombus of separable states. The rhombus is obtained as a cross section of the separable octahedron,[31] which arises as a common part of the tetrahedron of density matrices spanned by four Bell states and its reflection representing their partial transposition (Horodecki and Horodecki, 1996; Aravind, 1997). An identical octahedron of superpositive maps will be formed by intersecting the tetrahedrons of CP and CcP one-qubit unital maps shown in Figure 11.3(a) and (b).

Making use of the radius (15.61) of the separable ball we obtain that the states $\mathcal{M}_\epsilon$ of a $N \times N$ bipartite[32] system are separable for $\epsilon \leq \epsilon_c = 1/(N^2-1)$.

Usually one considers states separable with respect to a given decomposition of the composed Hilbert space, $\mathcal{H}_{N^2} = \mathcal{H}_A \otimes \mathcal{H}_B$. A state ρ may be separable with respect to a given decomposition and entangled with respect to another one. Consider for instance, two decompositions of $\mathcal{H}_6$: $\mathcal{H}_2 \otimes \mathcal{H}_3$ and $\mathcal{H}_3 \otimes \mathcal{H}_2$ which describe different physical problems. There exist states separable with respect to the former decomposition and entangled with respect to the latter one. On the other hand one may ask, which states are separable with respect to all possible splittings

[31] Properties of a separable octangula obtained for other 3-D cross sections of $\mathcal{M}^{(4)}$ were analysed in Ericsson (2002). Several 2-D cross sections plotted in Jakóbczyk and Siennicki (2001) and Verstraete, Dahaene and DeMoor (2002a) provide further insight into the geometry of the problem.

[32] Bounds for ϵ_c in multipartite systems were obtained in Braunstein et al. (1999), Deuar, Munro and Nemoto (2000), Pittenger and Rubin (2002), Gurvits and Barnum (2003), Szarek (2005) and Gurvits and Barnum (2004). The size of the separable ball is large enough that to generate a genuinely entangled pseudo-pure state in an NMR experiment one would need to deal with at least 34 qubits (Gurvits and Barnum, 2004). Although, to date, experimentalist have gained full control over 7–10 qubits and work with separable states only, the NMR quantum computing does fine (Cory, Fahmy and Havel, 1997; Chuang, Gershenfeld, Kubinec and Leung, 1998; Laflamme, Cory, Negrevergne and Viola, 2002).

of the composed system into subsystems A and B. This is the case if $\rho' = U\rho\, U^\dagger$ is separable for any global unitary U, and states possessing this property are called *absolutely separable* (Kuś and Życzkowski, 2001).

All states belonging to the maximal ball inscribed into the set of mixed states for a bipartite problem are not only separable but also absolutely separable. In the two-qubit case the set of absolutely separable states is larger than the maximal ball: As conjectured in Ishizaka and Hiroshima (2000) and proved in Verstraete, Audenaert and DeMoor (2001a) it contains any mixed state ρ for which

$$C_M(\vec{x}) \;\equiv\; x_1 - x_3 - 2\sqrt{x_2 x_4} \leq 0\,, \tag{15.64}$$

where $\vec{x} = \{x_1 \geq x_2 \geq x_3 \geq x_4\}$ denotes the ordered spectrum of ρ. The problem, whether there exist absolutely separable states outside the maximal ball was solved for 2×3 case (Hildebrand, n.d.), but it remains open in higher dimensions. Numerical investigations suggest that in such a case the set $\mathcal{M}_S$ of separable states, located in central parts of $\mathcal{M}$, is covered by a shell of bound entangled states. However this shell is not perfect, in the sense that the set of NPPT entangled states (occupying certain 'corners' of $\mathcal{M}$) has a common border with the set of separable states.

Some insight into the geometry of the problem may be gained by studying the manifold Σ of mixed products states. To verify whether a given state ρ belongs to Σ one computes the partial traces and checks if $\rho_A \otimes \rho_B$ is equal to ρ. This is the case, for example, for the maximally mixed state, $\rho_* \in \Sigma$. All states tangent to Σ at ρ_* are separable, while the normal subspace contains the maximally entangled states. Furthermore, for any bipartite systems the maximally mixed state ρ_* is the product state closest to any maximally entangled state (with respect to the HS distance) (Lockhart, Steiner and Gerlach, 2002).

Let us return to characterization of the boundary of the set of separable states for a bipartite system. For any entangled state $\sigma_{\rm ent}$ one may define the separable state $\sigma_{\rm sep} \in \partial\mathcal{M}_S$, which is closest to $\sigma_{\rm ent}$ with respect to a given metric. In general it is not easy to find the closest separable state, even in the two qubit case, for which the 14-dim boundary of the set $\mathcal{M}_S^{(4)}$ may be characterized explicitly,

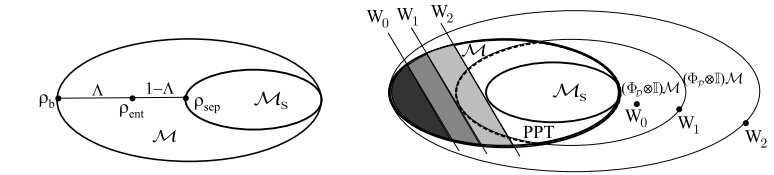
$$(\rho \in \partial\mathcal{M}_S^{(4)}) \;\Rightarrow\; \det\rho = 0 \quad \text{or} \quad \det\rho^{T_A} = 0\,. \tag{15.65}$$

Alternatively, for any entangled state one defines the *best separable approximation* (BSA) [33]

$$\rho_{\rm ent} \;=\; \Lambda\rho_{\rm sep} + (1-\Lambda)\rho_b, \tag{15.66}$$

where the separable state ρ_{sep} and the state ρ_b are chosen in such a way that the positive weight $\Lambda \in [0, 1]$ is maximal. Uniqueness of such a decomposition

[33] Also called *Lewenstein–Sanpera decomposition* (Lewenstein and Sanpera, 1998).



a) optimal decomposition of ρ_{ent} b) entanglement witnesses

Figure 15.8. (a) Best separable approximation of entangled state ρ; (b) a witness W_0 detects entanglement in a subset of entanglement states; W_1 – optimal decomposable witness; W_2 – optimal non-decomposable witness.

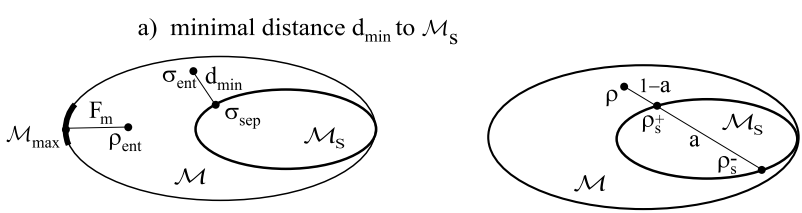


Figure 15.9. (a) Minimal distance D_x to the closest separable state. (b) Maximal fidelity to a maximally entangled state. (c) Robustness, i.e. the minimal ratio of the distance to set $\mathcal{M}_S$ by its width.

was proved in Lewenstein and Sanpera (1998) for two qubits, and in Karnas and Lewenstein (2001) for any bipartite system. In the two-qubit problem the state ρ_b is pure, and is maximally entangled for any full rank state ρ (Karnas and Lewenstein, 2001). An explicit form of the decomposition (15.66) was found in Wellens and Kuś (2001) for a generic two-qubit state and in Akhtarshenas and Jafarizadeh (2004) for some particular cases in higher dimensions. Note the key difference in both approaches: looking for the separable state closest to ρ we probe the boundary $\partial\mathcal{M}_S^{(N)}$ of the set of separable states only, while looking for its best separable approximation we must also take into account the boundary of the entire set of density matrices; compare Figure 15.8(a) and Figure 15.9(a).

The structure of the set of separable states may also be analysed with use of the entanglement witnesses (Pittenger and Rubin, 2003), already defined in the previous section. Any witness W, being a non-positive operator, may be represented as a point located far outside the set $\mathcal{M}$ of density matrices, in its image with respect to an extended positive map, $(\Phi_{\mathcal{P}} \otimes \mathbb{1})$, or as a line perpendicular to the axis OW (see Figure 11.6), which crosses $\mathcal{M}$. The states outside this line satisfy $\mathrm{Tr}\rho W < 0$, hence their entanglement is detected by W. A witness W_1 is called *finer* than W_0 if every entangled state detected by W_0 is also detected by W_1. A witness W_2 is called *optimal* if the corresponding map belongs to the boundary of the set of

positive operators, so the line representing W_2 touches the boundary of the set $\mathcal{M}_S$ of separable states. A witness related to a generic non-CP map $\Phi_{\mathcal{P}} \in \mathcal{P}$ may be *optimized* by sending it toward the boundary of $\mathcal{P}$ (Lewenstein, Kraus, Cirac and Horodecki, 2000b). If a positive map $\Phi_{\mathcal{D}}$ is decomposable, the corresponding witness, $W = D_{\Phi_{\mathcal{D}}}/N$ is called *decomposable*. Any decomposable witness cannot detect PPT bound entangled states (see Figure 15.8(b)).

One might argue that in general a witness $W = D_{\Phi}/N$ is theoretically less useful than the corresponding map Φ, since the criterion $\mathrm{Tr}\rho W < 0$ is not as powerful as $N(W^R \rho^R)^R = (\Phi \otimes \mathbb{1})\rho \geq 0$; see Eq. (11.24). However, a non-CP map Φ cannot be realized in nature, while an observable W may be measured. Suitable witness operators were actually used to detect quantum entanglement experimentally in bipartite (Barbieri, Martini, Nepi, Mataloni, D'Ariano and Macciavello, 2003; Gühne et al., 2003) and multipartite systems (Bourennane, Eibl, Kurtsiefer, Weinfurter, Guehne, Hyllus, Bruß, Lewenstein and Sanpera, 2004). Furthermore, the *Bell inequalities* may be viewed as a kind of separability criterion, related to a particular entanglement witness (Terhal, 2000a; Horodecki et al., 2000b; Hyllus, Gühne, Bruß and Lewenstein, 2005) so evidence of their violation[34] for certain states (Aspect et al., 1982) might be regarded as an experimental detection of quantum entanglement.

15.6 Entanglement measures

We have already learned that the degree of entanglement of any pure state of a $N \times K$ system may be characterized by the entanglement entropy (15.25) or any other Schur concave function f of the Schmidt vector $\vec{\lambda}$. The problem of quantifying entanglement for mixed states becomes complicated (Vedral, Plenio, Rippin and Knight, 1997; Donald et al., 2002; Horodecki, 2001).

Let us first discuss the properties that any potential measure $E(\rho)$ should satisfy. Even in this respect experts seem not to share exactly the same opinions (Bennett et al., 1996b; Popescu and Rohrlich, 1997; Vedral and Plenio, 1998; Vidal, 2000; Horodecki, Horodecki and Horodecki, 2000a). There are three basic axioms,

(E1) Discriminance. $E(\rho) = 0$ *if and only if* ρ *is separable.*

(E2) Monotonicity (15.29) *under* **probabilistic** *LOCC.*

(E3) Convexity, $E\big(a\rho + (1-a)\sigma\big) \leq aE(\rho) + (1-a)E(\sigma)$, *with* $a \in [0, 1]$.

Then there are certain additional requirements,

[34] A state violating the Bell or, in particular, the CHSH inequalities (Clauser, Horne, Shimony and Holt, 1969; Clauser and Shimony, 1978) needs to be entangled (Werner, 1989). The converse is not true: any pure entangled state violates CHSH inequalities (Gisin, 1991; Popescu and Rohrlich, 1992), but this is not always the case for a mixed entangled state (Werner, 1989).

(E4) Asymptotic continuity.[35] *Let ρ_m and σ_m denote sequences of states acting on m copies of the composite Hilbert space, $(\mathcal{H}_N \otimes \mathcal{H}_K)^{\otimes m}$.*

$$\text{If} \quad \lim_{m\to\infty} ||\rho_m - \sigma_m||_1 = 0 \quad \text{then} \quad \lim_{m\to\infty} \frac{E(\rho_m) - E(\sigma_m)}{m \ln NK} = 0 \,, \tag{15.67}$$

(E5) Additivity. $E(\rho \otimes \sigma) = E(\rho) + E(\sigma)$ *for any* $\rho, \sigma \in \mathcal{M}_{NK}$.
(E6) Normalization. $E(|\psi^-\rangle\langle\psi^-|) = 1$.
(E7) Computability. *There exists an efficient method to compute E for any* ρ.
There are also alternative forms of properties **(E1)–(E5)**.
(E1a) Weak discriminance. *If* ρ *is separable then* $E(\rho) = 0$.
(E2a) Monotonicity *under* **deterministic** *LOCC,* $E(\rho) \geq E[\Phi_{\mathrm{LOCC}}(\rho)]$.
(E3a) Pure states convexity. $E(\rho) \leq \sum_i p_i E(\phi_i)$ *where* $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$.
(E4a) Continuity. *If* $||\rho - \sigma||_1 \to 0$ *then* $|E(\rho) - E(\sigma)| \to 0$.
(E5a) Extensivity. $E(\rho^{\otimes n}) = nE(\rho)$.
(E5b) Subadditivity. $E(\rho \otimes \sigma) \leq E(\rho) + E(\sigma)$.
(E5c) Superadditivity. $E(\rho \otimes \sigma) \geq E(\rho) + E(\sigma)$.

The above list of postulates deserves a few comments. The rather natural 'if and only if' condition in **(E1)** is very strong: it cannot be satisfied by any measure quantifying the distillable entanglement, due to the existence of bound entangled states. Hence one often requires the weaker property **(E1a)** instead.

Monotonicity **(E2)** under probabilistic transformations is stronger than monotonicity **(E2a)** under deterministic LOCC. Since local unitary operations are reversible, the latter property implies
(E2b) Invariance with respect to local unitary operations,

$$E(\rho) = E(U_A \otimes U_B \, \rho \, U_A^\dagger \otimes U_B^\dagger) \,. \tag{15.68}$$

Convexity property **(E3)** guarantees that one cannot increase entanglement by mixing.[36] Following Vidal (2000), we will call any quantity satisfying **(E2)** and **(E3)** an *entanglement monotone*.[37] These fundamental postulates reflect the key idea that quantum entanglement cannot be created locally. Or in more economical terms: it is not possible to get any entanglement for free – one needs to invest resources for certain global operations.

The postulate that any two neighbouring states should be characterized by similar entanglement is made precise in **(E4)**. Let us recall here the Fannes continuity lemma (13.36), which estimates the difference between von Neumann entropies

[35] We follow Horodecki (2001) here; slightly different fomulations of this property are used in Horodecki et al. (2000a) and in Donald et al. (2002).
[36] Note that entropy is a *concave* function of its argument: mixing of pure states *increases* their von Neumann entropy, but it *decreases* their entanglement.
[37] Some authors require also continuity **(E4a)**.

of two neighbouring mixed states. Similar bounds may also be obtained for any other Rényi entropy with $q > 0$, but then the bounds for S_q are weaker then for S_1. Although S_q are continuous for $q > 0$, in the asymptotic limit $n \to \infty$ only S_1 remains a continuous function of the state $\rho^{\otimes n}$. In the same way the asymptotic continuity distinguishes the entanglement entropy based on S_1 from other entropy measures related to the generalized entropies S_q (Vidal, 2000; Donald et al., 2002).

Additivity **(E5)** is a most welcome property of an optimal entanglement measure. For certain measures one can show sub- or super-additivity; additivity requires both of them. Unfortunately this is extremely difficult to prove for two arbitrary density matrices, so some authors suggest to require extensivity **(E5a)**. Even this property is not easy to demonstrate. However, for any measure E one may consider the quantity

$$E_R(\rho) \equiv \lim_{n\to\infty} \frac{1}{n} E(\rho^{\otimes n}) \,. \tag{15.69}$$

If such a limit exists, then the *regularized* measure E_R defined in this way satisfies **(E5a)** by construction. The normalization property **(E6)**, useful to compare different quantities, can be achieved by a trivial rescaling.

The complete wish list **(E1)–(E7)** is very demanding, so it is not surprising that instead of one ideal measure of entanglement fulfilling all required properties, the literature contains a plethora of measures (Vedral and Plenio, 1998; Horodecki, 2001; Bruß, 2002), each of them satisfying some axioms only. The pragmatic wish **(E7)** is an especially tough one, since we have learned that even the problem of deciding the separability is a 'hard one' (Gurvits, 2003; Gurvits, 2004), the quantifying of entanglement cannot be easier. Instead of waiting for the discovery of a single, universal measure of entanglement, we have thus no choice but to review some approaches to the problem. In the spirit of this book we commence with

I. Geometric measures

The distance from an analysed state ρ to the set $\mathcal{M}_S$ of separable states satisfies **(E1)** by construction (see Figure 15.9(a)). However, it is not simple to find the separable state σ closest to ρ with respect to a certain metric, necessary to define $D_x(\rho) \equiv D_x(\rho, \sigma)$. There are several distances to choose from, for instance

G1. Bures distance (Vedral and Plenio, 1998) $D_{\mathrm{B}}(\rho) \equiv \min_{\sigma\in\mathcal{M}_S} D_{\mathrm{B}}(\rho, \sigma)$,

G2. Trace distance (Eisert, Audenaert and Plenio, 2003) $D_{\mathrm{Tr}}(\rho) \equiv \min_{\sigma\in\mathcal{M}_S} D_{\mathrm{Tr}}(\rho, \sigma)$,

G3. Hilbert–Schmidt distance (Witte and Trucks, 1999) $D_{\mathrm{HS}}(\rho) \equiv \min_{\sigma\in\mathcal{M}_S} D_{\mathrm{HS}}(\rho, \sigma)$.

The Bures and the trace metrics are monotone (see Sections 13.2 and 14.1), which directly implies **(E2a)**, while D_{B} fulfils also the stronger property **(E2)** (Vedral and

Plenio, 1998). Since the HS metric is not monotone (Ozawa, 2001) it is not at all clear, whether the minimal Hilbert–Schmidt distance is an entanglement monotone (Verstraete et al., 2002a). Since the diameter of the set of mixed states with respect to the above distances is finite, all distance measures cannot satisfy even the partial additivity **(E3a)**.

Although quantum relative entropy is not exactly a distance, but rather a contrast function, it may also be used to characterize entanglement.

G4. Relative entropy of entanglement (Vedral et al., 1997) $D_R(\rho) \equiv \min_{\sigma\in\mathcal{M}_S} S(\rho||\sigma)$.

In view of the discussion in Chapter 13 this measure has an appealing interpretation as distinguishability of ρ from the closest separable state. For pure states it coincides with the entanglement entropy, $D_R(|\phi\rangle) = E_1(|\phi\rangle)$ (Vedral and Plenio, 1998). Analytical formulae for D_R are known in certain cases only (Vedral et al., 1997; Vollbrecht and Werner, 2001; Ishizaka, 2003), but it may be efficiently computed numerically (Řeháček and Hradil, 2003). This measure of entanglement is convex (due to double convexity of relative entropy) and continuous (Donald and Horodecki, 1999), but not additive (Vollbrecht and Werner, 2001). It is thus useful to study the regularized quantity, $\lim_{n\to\infty} D_R(\rho^{\otimes n})/n$. This limit exists due to subadditivity of relative entropy and has been computed in some cases (Audenaert, Eisert, Jané, Plenio, Virmani and Moor, 2001a; Audenaert, Moor, Vollbrecht and Werner, 2002).

G5. Reversed relative entropy of entanglement $D_{RR}(\rho) \equiv \min_{\sigma\in\mathcal{M}_S} S(\sigma||\rho)$.

This quantity with exchanged arguments is not so interesting per se, but its modification D'_{RR} – the minimal entropy with respect to the set $\mathcal{M}_\rho$ of separable states ρ' locally identical to ρ, $\{\rho' \in \mathcal{M}_\rho : \rho'_A = \rho_A \text{ and } \rho'_B = \rho_B\}$, provides a distinctive example of an entanglement measure,[38] which satisfies the additivity condition **(E3)** (Eisert et al., 2003).

G6. Robustness (Vidal and Tarrach, 1999). $R(\rho)$ measures the endurance of entanglement by quantifying the minimal amount of mixing with separable states needed to wipe out the entanglement,

$$R(\rho) \equiv \min_{\rho_s^-\in\mathcal{M}_S} \left(\min_{s\geq 0} s : \rho_s^+ = \frac{1}{1+s}(\rho + s\rho_s^-) \in \mathcal{M}_S \right). \tag{15.70}$$

As shown if Figure 15.9(c) the robustness R may be interpreted as a minimal ratio of the HS distance $1 - a = s/(1+s)$ of ρ to the set $\mathcal{M}_S$ of separable states to the width $a = 1/(1+s)$ of this set. This construction does not depend on the

[38] A similar measure based on modified relative entropy was introduced by Partovi (2004).

boundary of the entire set $\mathcal{M}$, in contrast with the best separable approximation.[39] Robustness is known to be convex and monotone, but is not additive (Vidal and Tarrach, 1999). Robustness for two-qubit states diagonal in the Bell basis was found in Akhtarshenas and Jafarizadeh (2003), while a generalization of this quantity was proposed in Steiner (2003).

G7. Maximal fidelity F_m with respect to the set $\mathcal{M}_{\max}$ of maximally entangled states (Bennett et al., 1996b), $F_m(\rho) \equiv \max_{\phi\in\mathcal{M}_{\max}} F(\rho, |\phi\rangle\langle\phi|)$.

Strictly speaking the maximal fidelity cannot be considered as a measure of entanglement, since it does not satisfy even weak discriminance **(E1a)**. However, it provides a convenient way to characterize, to what extent ρ may approximate a maximally entangled state required for various tasks of quantum information processing, so in the two-qubit case it is called the *maximal singlet fraction*. Invoking (9.31) we see that F_m is a function of the minimal Bures distance from ρ to the set $\mathcal{M}_{\max}$. An explicit formula for the maximal fidelity for a two-qubit state was derived in Badziąg, Horodecki, Horodecki and Horodecki (2000), while relations to other entanglement measures were analysed in Verstraete and Verschelde (2002).

II. Extensions of pure-state measures

Another class of mixed-states entanglement measures can be derived from quantities characterizing entanglement of pure states. There exist at least two different ways of proceeding. The *convex roof* construction (Uhlmann, 1998; Uhlmann, 2003) defines $E(\rho)$ as the minimal average quantity $\langle E(\phi)\rangle$ taken on pure states forming ρ. The most important measure is induced by the entanglement entropy (15.25).

P1. Entanglement of Formation (EoF) (Bennett et al., 1996b)

$$E_F(\rho) \equiv \min_{\mathcal{E}_\rho} \sum_{i=1}^{M} p_i E_1(|\phi_i\rangle)\,, \tag{15.71}$$

where the minimization[40] is performed over an ensemble of all possible decompositions

$$\mathcal{E}_\rho = \{p_i, |\phi_i\rangle\}_{i=1}^{M} \,:\, \rho = \sum_{i=1}^{M} p_i |\phi_i\rangle\langle\phi_i| \quad \text{with} \quad p_i > 0, \quad \sum_{i=1}^{M} p_i = 1\,. \tag{15.72}$$

The ensemble $\mathcal{E}$ for which the minimum (15.71) is realized is called *optimal*. Several optimal ensembles might exist and the minimal ensemble length M is called the *cardinality* of the state ρ. If the state is separable then $E_F(\rho) = 0$, and

[39] In the two-qubit case the entangled state used for BSA (15.66) is pure, $\rho_b = |\phi_b\rangle\langle\phi_b|$, the weight Λ is a monotone (Eisert and Briegel, 2001), so the quantity $(1-\Lambda)E_1(\phi_b)$ works as a measure of entanglement (Lewenstein and Sanpera, 1998; Wellens and Kuś, 2001).

[40] A dual quantity defined by maximization over $\mathcal{E}_\rho$ is called *entanglement of assistance* (DiVincenzo, Fuchs, Mabuchi, Smolin, Thapliyal and Uhlmann, 1999), and both of them are related to relative entropy of entanglement of an extended system (Henderson and Vedral, 2000).

the cardinality coincides with the minimal length of the decomposition (15.37). Due to Carathéodory's theorem the cardinality of $\rho \in \mathcal{M}^{NK}$ does not exceed the squared rank of the state, $r^2 \leq N^2K^2$ (Uhlmann, 1998). In the two-qubit case it is sufficient to take $M = 4$ (Wootters, 1998), and this length is necessary for some states of rank $r = 3$ (Audenaert et al., 2001b). In higher dimensions there exists states for which $M > NK \geq r$ (DiVincenzo, Terhal and Thapliyal, 2000b).

Entanglement of formation enjoys several appealing properties: it may be interpreted as the minimal pure-states entanglement required to build up the mixed state. It satisfies by construction the discriminance property **(E1)** and is convex and monotone (Bennett et al., 1996b). EoF is known to be continuous (Nielsen, 2000), and for pure states it is by construction equal to the entanglement entropy $E_1(|\phi\rangle)$. To be consistent with normalization **(E6)** one often uses a rescaled quantity, $E'_F \equiv E_F / \ln 2$.

Two other properties are still to be desired, if EoF is to be an ideal entanglement measure: we do not know, whether EoF is additive,[41] and EoF is not easy to evaluate.[42] Explicit analytical formulae were derived for the two-qubit system (Wootters, 1998), and a certain class of symmetric states in higher dimensions (Terhal and Vollbrecht, 2000; Vollbrecht and Werner, 2001), while for the $2 \times K$ systems at least lower bounds are known (Chen, Liang, Li and Huang, 2002; Loziński, Buchleitner, Życzkowski and Wellens, 2003; Gerjuoy, 2003).

P2. Generalized Entanglement of Formation (GEoF)

$$E_q(\rho) \equiv \min_{\mathcal{E}_\rho} \sum_{i=1}^{M} p_i \, E_q(|\phi_i\rangle) \,, \tag{15.73}$$

where $E_q(|\phi\rangle) = S_q[\mathrm{Tr}_B(|\phi\rangle\langle\phi|)]$ stands for the Rényi entropy of entanglement. Note that an optimal ensemble for a certain value of q needs not to provide the minimum for $q' \neq q$. GEoF is asymptotically continuous only in the limit $q \to 1$ for which it coincides with EoF. In the very same way, the convex roof construction can be applied to extend any pure states entanglement measure for mixed states. In fact, several measures introduced so far are related to GEoF. For instance, the convex roof extended negativity (Lee, Chi, Oh and Kim, 2003) and *concurrence of formation* (Wootters, 2001; Rungta and Caves, 2003; Mintert et al., 2004) are related to $E_{1/2}$ and E_2, respectively.

[41] Additivity of EoF has been demonstrated in special cases only, if one of the states is a product state (Benatti and Narnhofer, 2001), is separable (Vollbrecht and Werner, 2001) or if it is supported on a specific subspace (Vidal, Dür and Cirac, 2002). At least we can be sure that EoF satisfies subadditivity **(E5b)**, since the tensor product of the optimal decompositions of ρ and σ provides an upper bound for $E(\rho \otimes \sigma)$.

[42] EoF may by computed numerically by minimization over the space of unitary matrices $U(M)$. A search for the optimal ensemble can be based on simulated annealing (Życzkowski, 1999), on a faster conjugate–gradient method (Audenaert et al., 2001b), or on minimizing the conditional mutual information (Tucci, n.d.c).

There is another way to make use of pure state entanglement measures. In analogy to the fidelity between two mixed states, equal to the maximal overlap between their purifications, one may also purify ρ by a pure state $|\psi\rangle \in (\mathcal{H}_N \otimes \mathcal{H}_K)^{\otimes 2}$. Based on the entropy of entanglement (15.25) one defines

P3. Entanglement of purification (Terhal, Horodecki, Leung and DiVincenzo, 2002; Bouda and Bužek, 2002)

$$E_P(\rho) \quad \equiv \quad \min_{|\phi\rangle:\ \rho = \mathrm{Tr}_{KN}(|\phi\rangle\langle\phi|)} E_1(|\phi\rangle)\,, \tag{15.74}$$

but any other measure of pure states entanglement might be used instead.

The entanglement of purification is continuous and monotone under strictly local operations (not under LOCC). It is not convex, but more importantly, it does not satisfy even the weak discriminance **(E1a)**. In fact E_P measures *correlations*[43] between both subsystems, and is positive for any non-product, separable mixed state (Bouda and Bužek, 2002). Hence entanglement of purification is not an entanglement measure, but it happens to be helpful to estimate a variant of the entanglement cost (Terhal et al., 2002). To obtain a reasonable measure one needs to allow for an arbitrary extension of the system size, as assumed by defining

P4. Squashed entanglement (Christandl and Winter, 2004)

$$E_S(\rho^{AB}) \quad \equiv \quad \inf_{\rho^{ABE}} \frac{1}{2}\big[S(\rho^{AE}) + S(\rho^{BE}) - S(\rho^{E}) - S(\rho^{ABE})\big]\,, \tag{15.75}$$

where the infimum is taken over all extensions ρ^{ABE} of an unbounded size such that $\mathrm{Tr}_E(\rho^{ABE}) = \rho^{AB}$. Here ρ^{AE} stands for $\mathrm{Tr}_B(\rho^{ABE})$ while $\rho^E = \mathrm{Tr}_{AB}(\rho^{ABE})$.

Squashed entanglement[44] is convex, monotone and vanishes for every separable state. If ρ^{AB} is pure then $\rho^{ABE} = \rho^E \otimes \rho^{AB}$, hence $E_S = [S(\rho^A) + S(\rho^B)]/2 = S(\rho^A)$ and the squashed entanglement reduces to the entropy of entanglement E_1. It is characterized by asymptotic continuity (Alicki and Fannes, 2004), and additivity **(E5)**, which is a consequence of the strong subadditivity of the von Neumann entropy. Thus E_S would be a perfect measure of entanglement, if we only knew how to compute it!

III. Operational measures

Entanglement may also be quantified in an abstract manner by considering the minimal resources required to generate a given state or the maximal entanglement yield. These measures are defined implicitly, since one deals with an infinite set of copies of the state analysed and assumes an optimization over all possible LOCC protocols.

[43] To quantify them one may use of the operator Schmidt decomposition (10.31) of ρ.

[44] Minimized quantity is proportional to quantum conditional mutual information of ρ^{ABE} (Tucci, n.d.c) and its name refers to 'squashing out' the classical correlations.

O1. Entanglement cost (Bennett et al., 1996b; Rains, 1999a) $E_C(\rho) = \lim_{n\to\infty} \frac{m}{n}$, where m is the number of singlets $|\psi^-\rangle$ needed to produce locally n copies of the analysed state ρ.

Entanglement cost has been calculated for instance for states supported on a subspace such that tracing out one of the parties forms an entanglement breaking channel (super-separable map) (Vidal et al., 2002). Moreover, entanglement cost was shown (Hayden, Horodecki and Terhal, 2001) to be equal to the regularized entanglement of formation, $E_C(\rho) = \lim_{n\to\infty} E_F(\rho^{\otimes n})/n$. Thus, if we knew that EoF is additive, the notions of entanglement cost and entanglement of formation would coincide.

O2. Distillable entanglement (Bennett et al., 1996b; Rains, 1999a). $E_D(\rho) = \lim_{n\to\infty} \frac{m}{n}$, where m is the maximal number of singlets $|\psi^-\rangle$ obtained out of n copies of the state ρ by an optimal LOCC conversion protocol.

Distillable entanglement is a measure of a fundamental importance, since it tells us how much entanglement one may extract out of the state analysed and use, for example for the cryptographic purposes. It is rather difficult to compute, but there exist analytical bounds due to Rains (1999b) and Rains (2001), and an explicit optimization formula was found (Devetak and Winter, 2005). E_D is not likely to be convex (Shor, Smolin and Terhal, 2001), although it satisfies the weaker condition **(E3a)** (Donald et al., 2002).

IV. Algebraic measures

If a partial transpose of a state ρ is not positive then ρ is entangled due to the PPT criterion **B1**. The partial transpose preserves the trace, so if $\rho^{T_A} \geq 0$ then $||\rho^{T_A}||_{\mathrm{Tr}} = \mathrm{Tr}\rho^{T_A} = 1$. Hence we can use the trace norm to characterize the degree, to which the positivity of ρ^{T_A} is violated.

N1. Negativity (Życzkowski et al., 1998; Eisert and Plenio, 1999), $\mathcal{N}_T(\rho) \equiv ||\rho^{T_A}||_{\mathrm{Tr}} - 1$.

Negativity is easy to compute, convex (partial transpose is linear and the trace norm is convex) and monotone (Eisert, n.d.; Vidal and Werner, 2002). It is not additive, but this drawback may be cured by defining the *log - negativity*,[45] $\ln ||\rho^{T_A}||_{\mathrm{Tr}}$. However, the major deficiency of the negativity is its failure to satisfy **(E1)** – by construction $\mathcal{N}_T(\rho)$ cannot detect PPT bound entangled states. In the two-qubit case the spectrum of ρ^{T_A} contains at most a single negative eigenvalue (Sanpera et al., 1998), so $\mathcal{N}'_T \equiv \max\{0, -2\lambda_{\min}\} = \mathcal{N}_T$. This observation explains the name on the one hand, and on the other provides a geometric interpretation: $\mathcal{N}'_T(\rho)$ measures the minimal relative weight of the maximally mixed state ρ_* which needs

[45] This is additive but is not convex. The log–negativity was used by Rains (2001) to obtain bounds on distillable entanglement.

to be mixed with ρ to produce a separable mixture (Verstraete et al., 2002a). In higher dimensions several eigenvalues of the partially transposed state may be negative, so in general $\mathcal{N}_T \neq \mathcal{N}'_T$, and the latter quantity is proportional to complete co-positivity (11.1) of a map Φ associated with the state ρ.

Negativity is not the only application of the trace norm (Vidal and Werner, 2002). Building on the positive maps criterion **A1** for separability one might analyse analogous quantities for any (not completely) positive map, $\mathcal{N}_\Phi(\rho) \equiv ||(\Phi \otimes \mathbb{1})\rho||_{\mathrm{Tr}} - 1$. Furthermore, one may consider another quantity related to the reshuffling criterion **B6**.

N2. Reshuffling negativity (Chen and Wu, 2002; Rudolph, 2003b). $\mathcal{N}_R(\rho) \equiv ||\rho^R||_{\mathrm{Tr}} - 1$.

This quantity is convex due to linearity of reshuffling and non-increasing under local measurements, but may increase under partial trace (Rudolph, 2003b). For certain bound entangled states $\mathcal{N}_R$ is positive; unfortunately not for all of them. A similar quantity with the minimal cross-norm $||\cdot||_\gamma$ was studied by Rudolph (2001), who showed that $\mathcal{N}_\gamma(\rho) = ||\rho||_\gamma - 1$ is convex and monotone under local operations. However, $||\rho^R||_{\mathrm{Tr}}$ is easily computable from the definition (10.33), in contrast to $||\rho||_\gamma$.

We end this short tour through the vast garden of entanglement measures[46] by studying how they behave for pure states. Entanglement of formation and purification coincide by construction with the entanglement entropy E_1. So is the case for both operational measures, since conversion of n copies of an analysed pure state into m maximally entangled states is reversible. The negativities are easy to compute. For any pure state $|\psi\rangle \in \mathcal{H}_N \otimes \mathcal{H}_N$ written in the Schmidt form (9.8), the non-zero entries of the density matrix ρ_ψ of size N^2 are equal to $\sqrt{\lambda_i \lambda_j}$, $i, j = 1, \ldots, N$. The reshuffled matrix ρ_ψ^R becomes diagonal, while partially transposed matrix $\rho_\psi^{T_2}$ has a block structure: it consists of N Schmidt components λ_i at the diagonal which sum to unity and $N(N-1)/2$ off-diagonal blocks of size 2, one for each pair of different indices (i, j). Eigenvalues of each block are $\pm\sqrt{\lambda_i \lambda_j}$, so both traces norms are equal to the sum of all entries. Hence both negativities coincide for pure states,

$$\mathcal{N}_T(\rho_\psi) = \mathcal{N}_R(\rho_\psi) = \sum_{i,j=1}^{N} \sqrt{\lambda_i \lambda_j} - 1 = \Big(\sum_{i}^{N} \sqrt{\lambda_i}\Big)^2 - 1 = \mathrm{e}^{E_{1/2}} - 1 \,, \qquad (15.76)$$

and vary from 0 for separable states to $N-1$ for maximally entangled states. Also maximal fidelity and robustness for pure states become related, $F = \exp(E_{1/2})/N$

[46] There also exist attempts to quantify entanglement by the dynamical properties of a state and the speed of decoherence (Blanchard, Jakóbczyk and Olkiewicz, 2001) or the secure key distillation rate (Horodecki et al., n.d.a; Devetak and Winter, 2005) and several others. For multipartite systems the problem gets even more demanding (Coffman et al., 2000; Eisert and Briegel, 2001; Wei and Goldbart, 2003; Partovi, 2004).

Table 15.2. *Properties of entanglement measures: discriminance E1, monotonicity E2, convexity E3, asymptotic continuity E4, additivity E5, extensivity E5a, computability E7: explicit closed formula C, optimization over a finite space F or an infinite space I; Rényi parameter q for pure states.*

Entanglement measure	E1	E2	E3	E4	E5	E5a	E7	q
G1 Bures distance D_B	Y	Y	Y	N	N	N	F	2
G2 Trace distance D_{Tr}	Y	Y	Y	N	N	N	F	
G3 HS distance D_{HS}	Y	?	?	N	N	N	F	
G4 Relative entropy D_R	Y	Y	Y	Y	N	?	F	1
G5' Reversed RE, D'_{RR}	?	Y	Y	Y	Y	Y	F	1
G6 Robustness R	Y	Y	Y	N	N	N	F	1/2
P1 Entangl. of formation E_F	Y	Y	Y	Y	?	?	C/F	1
P2 Generalized EoF, E_q	Y	Y	?	N	?	?	F	q
P4 Squashed entangl. E_S	?	Y	Y	Y	Y	Y	I	1
O1 Entangl. cost E_C	?	Y	Y	Y	?	Y	I	1
O2 Distillable entangl. E_D	N	Y	N(?)	Y	?	Y	I	1
N1 Negativity $\mathcal{N}_T$	N	Y	Y	N	N	N	C	1/2
N2 Reshuffling negativity $\mathcal{N}_R$	N	N	Y	N	N	N	C	1/2

and $R = \mathcal{N}_T = \exp(E_{1/2}) - 1$ (Vidal and Tarrach, 1999). On the other hand, the minimal Bures distance to the closest separable (mixed) state (Vedral and Plenio, 1998) becomes a function of the Rényi entropy of order two,

$$D_{\mathrm{B}}(\rho_\phi) = \left(2 - 2\sum_{i=1}\lambda_i^2\right)^{1/2} = \sqrt{2 - 2\mathrm{e}^{-E_2/2}}\,, \tag{15.77}$$

and is equal to concurrence (15.26), while the Bures distance to the closest separable pure state $D_{\mathrm{B}}^{\mathrm{pure}}(|\phi\rangle) = [2(1 - \sqrt{1 - \lambda_{max}})]^{1/2}$ is a function of $E_\infty = -\ln \lambda_{\max}$. The Rényi parameters q characterizing behaviour of the discussed measures of entanglement for pure states are collected in Table 15.2.

Knowing that a given state ρ can be locally transformed into ρ' implies that $E(\rho) \geq E(\rho')$ for any measure E, but the converse is not true. Two Rényi entropies of entanglement of different (positive) orders generate different order in the space of pure states. By continuity this is also the case for mixed states, and the relation

$$E_A(\rho_1) \;\leq\; E_A(\rho_2) \;\;\Leftrightarrow\;\; E_B(\rho_1) \;\leq\; E_B(\rho_2) \tag{15.78}$$

does not hold. For a certain pair of mixed states it is thus likely that one state is more entangled with respect to a given measure, while the other one, with respect to another measure of entanglement (Eisert and Plenio, 1999). If two measures E_A and E_B coincide for pure states they are identical or they *do not* generate the same order in the set of mixed states (Virmani and Plenio, 2000). Hence entanglement of

formation and distillable entanglement do not induce the same order. On the other hand, several entanglement measures are correlated and knowing E_A one may try to find lower and upper bounds for E_B.

The set of entanglement measures shrinks, if one imposes even some of the desired properties **(E1)–(E7)**. The asymptotic continuity **(E4)** is particularly restrictive. For instance, among generalized Rényi entropies it is satisfied only by the entropy of entanglement E_1 (Vidal, 2000). If a measure E satisfies additionally monotonicity **(E2a)** under deterministic LOCC and extensivity **(E5a)**, it is bounded by the distillable entanglement and entanglement cost (Horodecki et al., 2000a; Horodecki, 2001),

$$E_D(\rho) \leq E(\rho) \leq E_C(\rho) . \tag{15.79}$$

Interestingly, the two first measures introduced in the pioneering paper by Bennett et al. (1996b) occurred to be the extreme entanglement measures. For pure states both of them coincide, and we arrive at a kind of **uniqueness theorem**: *Any monotone, extensive and asymptotically continuous entanglement measure coincides for pure states with the entropy of formation* E_F (Popescu and Rohrlich, 1997; Vidal, 2000; Donald et al., 2002). This conclusion concerning pure states entanglement of bipartite systems may also be reached by an abstract, thermodynamic approach (Vedral and Kashefi, 2002).

Let us try to recapitulate the similarities and differences between four classes of entanglement measures. For a geometric measure or an extension of a pure states measure it is not simple to check, which of the desired properties are satisfied. Furthermore, to evaluate it for a typical mixed state one needs to perform a cumbersome optimization scheme. One should not expect the remarkable analytical result of Wootters (1998) for entanglement of formation in the 2×2 system, to be extended for the general $N \times N$ problem, since even stating the separability is known to be an algorithmically complex task (Gurvits, 2003).

Operational measures are attractive, especially from the point of view of information science, and extensivity and monotonicity are direct consequence of their definitions. However, they are extremely hard to compute. In contrast, algebraic measures are easy to calculate, but they fail to detect entanglement for all nonseparable states. Summarizing, several different measures of entanglement are thus likely to be still used in future.

15.7 Two-qubit mixed states

Before discussing the entanglement of two-qubit mixed states let us recapitulate, in what sense the case $N = 2$ differs from $N \geq 3$.

A) algebraic properties

- **i)** $SU(N) \times SU(N)$ is homomorphic to $SO(N^2)$ for $N = 2$ only,
- **ii)** $SU(N) \cong SO(N^2 - 1)$ for $N = 2$ only,
- **iii)** All positive maps $\Phi : \mathcal{M}^{(N)} \to \mathcal{M}^{(N)}$ are decomposable for $N = 2$ only,

B) N-level mono-partite systems

- **iv)** Boundary $\partial\mathcal{M}^{(2)}$ consists of pure states only,
- **v)** For any state $\rho_{\vec{\tau}} \in \mathcal{M}^{(N)}$ also its antipode $\rho_{-\vec{\tau}} = 2\rho_* - \rho_{\vec{\tau}}$ forms a state and there exists a universal NOT operation for $N = 2$ only.
- **vi)** $\mathcal{M}^{(N)} \subset \mathbb{R}^{N^2-1}$ forms a ball for $N = 2$ only,

C) $N \times N$ composite systems

- **vii)** For any pure state $|\psi\rangle \in \mathcal{H}_N \otimes \mathcal{H}_N$ there exist $N - 1$ independent Schmidt coefficients λ_i. For $N = 2$ there exists only one independent coefficient λ_1, hence all entanglement measures are equivalent.
- **viii)** The maximally entangled states form the manifold $SU(N)/\mathbb{Z}_N$, which is equivalent to the real projective space $\mathbb{R}\mathbf{P}^{N^2-1}$ only for $N = 2$.
- **ix)** All PPT states of a $N \times N$ system are separable for $N = 2$ only.
- **x)** For any two-qubit mixed state its optimal decomposition consists of pure states of equal concurrence. Thus entanglement of formation becomes a function of concurrence of formation for 2×2 systems.

These features demonstrate why entanglement of two-qubit systems is special (Vollbrecht and Werner, 2000). Several of these issues are closely related. We have already learned that decomposability **iii)** is a consequence of **ii)** and implies the separability **ix)**. We shall see now, how a group-theoretic fact **i)** allows one to to derive a closed formula for EoF of a two-qubit system. We are going to follow the seminal paper of Wootters (1998), who built on Bennett et al. (1996b) and Hill and Wootters (1997).

Consider first a two-qubit pure state $|\psi\rangle$. Due to its normalization the Schmidt components – eigenvalues of the matrix $\Gamma\Gamma^\dagger$ – satisfy $\mu_1 + \mu_2 = 1$. The tangle of $|\psi\rangle$, defined in (15.26), reads

$$\tau = C^2 = 2(1 - \mu_1^2 - \mu_2^2) = 4\mu_1(1 - \mu_1) = 4\mu_1\mu_2 \,, \tag{15.80}$$

and implies that concurrence is proportional to the determinant of (15.8),

$$C = 2\,|\sqrt{\mu_1\mu_2}| = 2\,|\mathrm{det}\Gamma| \,. \tag{15.81}$$

Inverting this relation we find the entropy of entanglement E as a function of concurrence

$$E = S(\mu_1, 1 - \mu_1) \quad \text{where} \quad \mu_1 = \frac{1}{2}\big(1 - \sqrt{1 - C^2}\big) \tag{15.82}$$

and S stands for the Shannon entropy function, $-\sum_i \mu_i \ln \mu_i$.

Let us represent $|\psi\rangle$ in a particular basis consisting of four Bell states

$$|\psi\rangle = \Big[a_1|\phi^+\rangle + a_2\mathrm{i}|\phi^-\rangle + a_3\mathrm{i}|\psi^+\rangle + a_4|\psi^-\rangle\Big] \tag{15.83}$$
$$= \frac{1}{\sqrt{2}}\Big[(a_1 + \mathrm{i}a_2)|00\rangle + (\mathrm{i}a_3 + a_4)|01\rangle + (\mathrm{i}a_3 - a_4)|10\rangle + (a_1 - \mathrm{i}a_2)|11\rangle\Big] .$$

Calculating the determinant in Eq. (15.81) we find that

$$C(|\psi\rangle) \;=\; \Big|\sum_{k=1}^{4} a_k^2\Big| . \tag{15.84}$$

If all its coefficients of $|\psi\rangle$ in the basis (15.84) are real then $C(|\phi\rangle) = 1$ and the state is maximally entangled. This property holds also if we act on $|\psi\rangle$ with an orthogonal gate $O \in SO(4)$ and justifies referring to (15.84) as the *magic basis* (Bennett et al., 1996b). Any two-qubit unitary gate, which is represented in it by a real $SO(4)$ matrix corresponds to a local operation[47] and its action does not influence entanglement. This is how the property **i)** enters the game.

To appreciate another feature of the magic basis consider the transformation $|\psi\rangle \to |\tilde{\psi}\rangle = (\sigma_y \otimes \sigma_y)|\psi^*\rangle$, in which complex conjugation is taken in the standard basis, $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. It represents flipping of both spins of the system. If a state is expressed in the magic basis this transformation is realized just by complex conjugation. Expression (15.84) implies then

$$C(|\phi\rangle) \;=\; |\langle\psi|\tilde{\psi}\rangle| . \tag{15.85}$$

Spin flipping of mixed states is also realized by complex conjugation, if ρ is expressed in magic basis. Working in standard basis this transformation reads

$$\rho \;\to\; \tilde{\rho} \;=\; (\sigma_y \otimes \sigma_y)\rho^*(\sigma_y \otimes \sigma_y) . \tag{15.86}$$

In the Fano form (15.38) flipping corresponds to reversing the signs of both Bloch vectors,

$$\tilde{\rho} = \frac{1}{4}\Big[\mathbb{1}_4 - \sum_{i=1}^{3} \tau_i^A \sigma_i \otimes \mathbb{1}_2 - \sum_{j=1}^{3} \tau_j^B \mathbb{1}_2 \otimes \sigma_j + \sum_{i,j=1}^{3} \beta_{ij}\sigma_i \otimes \sigma_j\Big] . \tag{15.87}$$

Root fidelity between ρ and $\tilde{\rho}$ is given by the trace of the positive matrix

$$\sqrt{F} \;=\; \sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}} . \tag{15.88}$$

[47] A gate represented by an orthogonal matrix with $\det O = -1$, corresponds to SWAP of both qubits. It does not influence the entanglement but is non-local (Vollbrecht and Werner, 2000).

Let us denote by λ_i the decreasingly ordered eigenvalues of $\sqrt{F}$, (singular values of $\sqrt{\rho}\sqrt{\tilde{\rho}}$). The *concurrence* of a two-qubit mixed state is now defined by

$$C(\rho) \equiv \max\{0,\ \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\}\,. \tag{15.89}$$

The number of of positive eigenvalues cannot be greater then rank r of ρ. For a pure state the above definition is thus consistent with $C = |\langle\psi|\tilde{\psi}\rangle|$ and expression (15.84).

Consider a generic mixed state of full rank given by its eigendecomposition $\rho = \sum_{i=1}^{4} |w_i\rangle\langle w_i|$. The eigenstates are subnormalized in a sense that $||w_i\rangle|^2$ is equal to the ith eigenvalue d_i. The flipped states $|\tilde{w}_i\rangle$ are also eigenstates of $\tilde{\rho}$. Defining a symmetric matrix $W_{ij} \equiv \langle w_i|\tilde{w}_j\rangle$ we see that the spectra of $\rho\tilde{\rho}$ and WW^* coincide. Let U be a unitary matrix diagonalizing the Hermitian matrix WW^*.

Other decompositions of ρ may be obtained by the Schrödinger's mixture theorem (8.40). In particular, the unitary matrix U defined above gives a decomposition into four states $|x_i\rangle \equiv \sum_i U^*_{ij}|w_j\rangle$. They fulfil

$$\langle x_i|\tilde{x}_j\rangle = (UWU^T)_{ij} = \lambda_i\delta_{ij}\,. \tag{15.90}$$

Since W is symmetric, an appropriate choice of phases of eigenvectors forming U assures that the diagonal elements of UWU^T are equal to the square roots of the eigenvalues of WW^*, which coincide with the eigenvalues λ_i of $\sqrt{F}$.

We are going to show that a state is separable if $C = 0$. Hence $\lambda_1 < \lambda_2 + \lambda_3 + \lambda_4$ and it is possible to find four phases η_i such that

$$\sum_{j=1}^{4} \mathrm{e}^{2\eta_j}\lambda_j = 0 \tag{15.91}$$

In other words such a chain of four links of length λ_i may be closed to form a polygon,[48] as sketched in Figure 15.10(a), in which the phase η_1 is set to zero. Phases η_i allow us to write four other pure states

$$\begin{aligned}
|z_1\rangle &= \frac{1}{2}\big(\mathrm{e}^{\mathrm{i}\eta_1}|x_1\rangle + \mathrm{e}^{\mathrm{i}\eta_2}|x_2\rangle + \mathrm{e}^{\mathrm{i}\eta_3}|x_3\rangle + \mathrm{e}^{\mathrm{i}\eta_4}|x_4\rangle\big)\,,\\
|z_2\rangle &= \frac{1}{2}\big(\mathrm{e}^{\mathrm{i}\eta_1}|x_1\rangle + \mathrm{e}^{\mathrm{i}\eta_2}|x_2\rangle - \mathrm{e}^{\mathrm{i}\eta_3}|x_3\rangle - \mathrm{e}^{\mathrm{i}\eta_4}|x_4\rangle\big)\,,\\
|z_3\rangle &= \frac{1}{2}\big(\mathrm{e}^{\mathrm{i}\eta_1}|x_1\rangle - \mathrm{e}^{\mathrm{i}\eta_2}|x_2\rangle + \mathrm{e}^{\mathrm{i}\eta_3}|x_3\rangle - \mathrm{e}^{\mathrm{i}\eta_4}|x_4\rangle\big)\,,\\
|z_4\rangle &= \frac{1}{2}\big(\mathrm{e}^{\mathrm{i}\eta_1}|x_1\rangle - \mathrm{e}^{\mathrm{i}\eta_2}|x_2\rangle - \mathrm{e}^{\mathrm{i}\eta_3}|x_3\rangle + \mathrm{e}^{\mathrm{i}\eta_4}|x_4\rangle\big)\,.
\end{aligned} \tag{15.92}$$

[48] This reasoning holds also if some λ_i are equal to zero and the polygon reduces, for example, to a triangle.

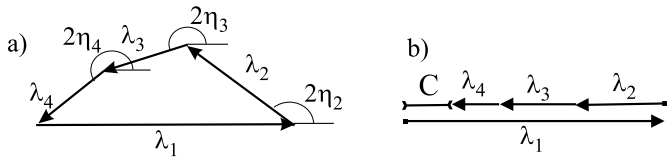


Figure 15.10. Concurrence polygon: (a) quadrangle for a separable state, (b) line for an entangled state with concurrence C.

On the one hand they form a decomposition of the state analysed, $\rho = \sum_i |x_i\rangle\langle x_i| = \sum_i |z_i\rangle\langle z_i|$. On the other hand, due to Eqs. (15.90) and (15.91) $\langle z_i|\tilde{z}_i\rangle = 0$ for $i = 1, ..., 4$, hence each pure state $|z_i\rangle$ of the decomposition is separable and so is ρ.

Consider now a mixed state ρ for which $C > 0$ since λ_1 is so large that the chain cannot be closed (see Figure 15.10(b)). Making use of the pure states $|x_i\rangle$ constructed before we introduce a set of four states

$$|y_1\rangle = |x_1\rangle\ , \quad |y_2\rangle = \mathrm{i}|x_2\rangle\ , \quad |y_3\rangle = \mathrm{i}|x_3\rangle\ , \quad |y_4\rangle = \mathrm{i}|x_4\rangle\ . \tag{15.93}$$

and a symmetric matrix $Y_{ij} = \langle y_i|\tilde{y}_j\rangle$, the relative phases of which are chosen in such a way that relation (15.90) implies

$$\mathrm{Tr}\, Y = \sum_{i=1}^{4} \langle y_i|\tilde{y}_j\rangle = \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4 \ = \ C(\rho)\ . \tag{15.94}$$

The states (15.93) are subnormalized, hence the above expression represents an average of a real quantity, the absolute value of which coincides with the concurrence (15.85). Using Schrödinger's theorem again one may find yet another decomposition, $|z_i\rangle \equiv \sum_i V^*_{ij}|y_j\rangle$, such that every state has *the same* concurrence, $C(|z_i\rangle) = C(\rho)$ for $i = 1, \ldots, 4$. To do so define a symmetric matrix $Z_{ij} = \langle z_i|\tilde{z}_j\rangle$ and observe that $\mathrm{Tr}Z = \mathrm{Tr}(VYV^T)$. This trace does not change if V is real and $V^T = V^{-1}$ follows. Hence one may find such an orthogonal V that all overlaps are equal to concurrence, $Z_{ii} = C(\rho)$, and produce the final decomposition $\rho = \sum_i |z_i\rangle\langle z_i|$.

The above decomposition is optimal for the concurrence of formation (Wootters, 1998),

$$C_F(\rho) \ = \ \min_{\mathcal{E}_\rho} \sum_{i=1}^{M} p_i C(|\phi_i\rangle) \ = \ \min_{V} \sum_{i=1}^{m} |(VYV^T)_{ii}|\ , \tag{15.95}$$

where the minimum over ensembles $\mathcal{E}_\rho$ may be replaced by a minimum over $m \times n$ rectangular matrices V, containing n orthogonal vectors of size $m \leq 4^2$ (Uhlmann, 1998). The function relating E_F and concurrence is convex, hence the decomposition into pure states of equal concurrence is also optimal for entropy of

formation. Thus E_F of any two-qubit state ρ is given by the function (15.82) of concurrence of formation C_F equal to $C(\rho)$, and defined by (15.89).[49]

While the algebraic fact **i)** was used to calculate concurrence, the existence of a general formula for maximal fidelity hinges on property **ii)**. Let us write the state ρ in its Fano form (15.38) and analyse invariants of local unitary transformations $U_1 \otimes U_2$. Due to the relation $SU(2) \approx SO(3)$ this transformation may be interpreted as an independent rotation of both Bloch vectors, $\vec{\tau}^{\,A} \to O_1\vec{\tau}^{\,A}$ and $\vec{\tau}^{\,B} \to O_2\vec{\tau}^{\,B}$. Hence the real correlation matrix $\beta_{ij} = \mathrm{Tr}(\rho\sigma_i \otimes \sigma_j)$ may be brought into diagonal form $K = O_1\beta O_2^T$. The diagonal elements may admit negative values since we have restricted orthogonal matrices to fulfil $\det O_i = +1$. Hence $|K_{ii}| = \kappa_i$, where κ_i stand for singular values of β. Let us order them decreasingly. By construction they are invariant with respect to local unitaries,[50] and govern the maximal fidelity with respect to maximally entangled states (Badziąg et al., 2000),

$$F_m(\rho) = \frac{1}{4}\big[1 + \kappa_1 + \kappa_2 - \mathrm{Sign}[\det(\beta)]\,\kappa_3\big]\,. \tag{15.96}$$

It is instructive to compute explicit formulae for above entanglement measures for several families of two-qubit states. The concurrence of a Werner state (15.42) is equal to the negativity,

$$C\big(\rho_W(x)\big) = \mathcal{N}_T\big(\rho_W(x)\big) = \begin{cases} 0 & \text{if} \quad x \le 1/3 \\ (3x-1)/2 & \text{if} \quad x \ge 1/3 \end{cases}\,, \tag{15.97}$$

its entanglement of formation is given by (15.82), while $F_m = (3x-1)/4$.

Another interesting family of states arises as a convex combination of a Bell state with an orthogonal separable state (Horodecki et al., 2000b)

$$\sigma_H(a) \;\equiv\; a\,|\psi^-\rangle\langle\psi^-| \;+\; (1-a)\,|00\rangle\langle 00|\,. \tag{15.98}$$

Concurrence of such a state is by construction equal to its parameter, $C = a$, while the negativity reads $\mathcal{N}_T = \sqrt{(1-a)^2 + a^2} + a - 1$. The relative entanglement of entropy reads (Vedral and Plenio, 1998) $E_R = (a-2)\ln(1-a/2) + (1-a)\ln(1-a)$. We will use also a mixture of Bell states,

$$\sigma_B(b) \;\equiv\; b\,|\psi^-\rangle\langle\psi^-| + (1-b)\,|\psi^+\rangle\langle\psi^+|\,, \tag{15.99}$$

for which by construction $F_m = \max\{b, 1-b\}$ and $C = \mathcal{N}_T = 2F_m - 1$.

[49] A streamlined proof of this fact was provided in (Audenaert et al., 2001b), while the analogous problem for two rebits was solved in Caves, Fuchs and Rungta (2001a).

[50] Two-qubit density matrix is specified by 15 parameters, the local unitaries are characterized by six variables, so there exist nine functionally independent local invariants. However, two states are locally equivalent if they share additional nine discrete invariants which determine signs of κ_i, τ_i^A and τ_i^B (Makhlin, 2002). A classification of mixed states based on degeneracy and signature of K was worked out in Grassl et al. (1998), Englert and Metwally (2001) and Kuś and Życzkowski (2001).

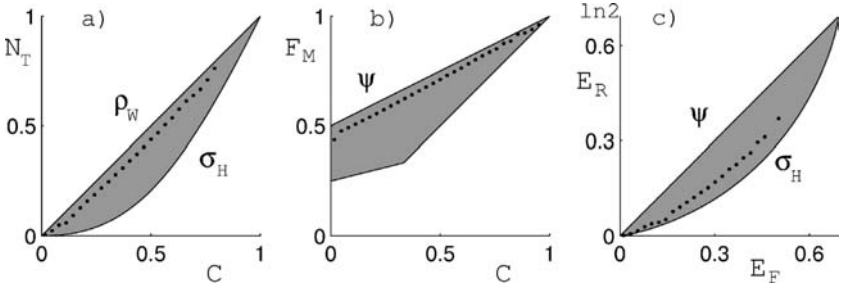


Figure 15.11. Bounds between entanglement measures for two-qubits: (a) negativity versus concurrence (15.100), (b) maximal fidelity versus concurrence (15.101), (c) relative entropy of entanglement versus entanglement of formation (15.103). Labels represent families of extremal states while dots denote averages taken with respect to the HS measure in $\mathcal{M}^{(4)}$.

Entanglement measures are correlated: they vanish for separable states and coincide for maximally entangled states. For two-qubit systems several explicit bounds are known. Concurrence forms an upper bound for negativity (Eisert and Plenio, 1999; Życzkowski, 1999). This statement was proved in Verstraete, Audenaert, Dehaene and DeMoor (2001b), where it was shown that these measures coincide if the eigenvector of ρ^{T_A} corresponding to the negative eigenvalue is maximally entangled. The lower bound

$$C \;\geq\; \mathcal{N}_T \;\geq\; \sqrt{(1-C)^2 + C^2} + C - 1 \tag{15.100}$$

is achieved (Verstraete et al., 2001b) for states (15.98).

Analogous tight bounds between maximal fidelity and concurrence or negativity were established in Verstraete and Verschelde (2002),

$$\frac{1+C}{2} \geq F_m \geq \begin{cases} (1+C)/4 & \text{if} \quad C \leq 1/3 \\ C & \text{if} \quad C \geq 1/3 \end{cases}, \tag{15.101}$$

$$\frac{1+\mathcal{N}_T}{2} \geq F_m \geq \begin{cases} \frac{1}{4} + \frac{1}{8}\left(\mathcal{N}_T + \sqrt{5\mathcal{N}_T^2 + 4\mathcal{N}_T}\right) & \text{if} \quad \mathcal{N}_T \leq \frac{\sqrt{5}-2}{3} \\ \sqrt{2\mathcal{N}_T(\mathcal{N}_T+1)} - \mathcal{N}_T & \text{if} \quad \mathcal{N}_T \geq \frac{\sqrt{5}-2}{3} \end{cases}. \tag{15.102}$$

Upper bound for fidelity is realized for the family (15.99) or for any other state for which $C = \mathcal{N}_T$.

Relative entropy of entanglement is bounded from above by E_F. Numerical investigations suggest (Verstraete et al., 2001b) that the lower bound is achieved

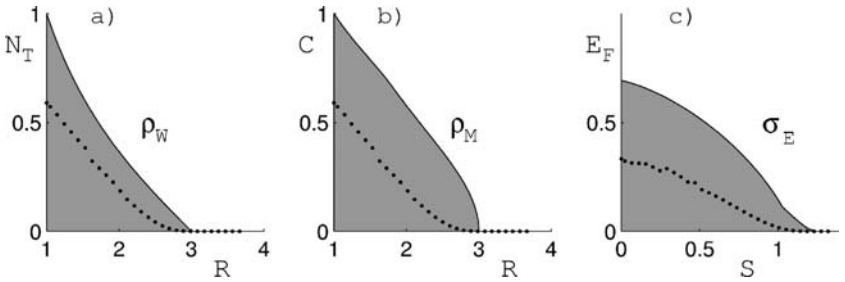


Figure 15.12. Upper bounds for measures of entanglement as a function of mixedness for two-qubits: (a) negativity and (b) concurrence versus participation ratio $R = 1/(\mathrm{Tr}\rho^2)$; (c) entanglement of formation versus von Neumann entropy. Shading shows entire accessible region while dots denote the average taken with respect to the HS measure in $\mathcal{M}^{(4)}$. For pure states it coincides with the average over FS measure.

for the family of (15.98), which implies

$$E_F \geq E_R \geq \Big[(C-2)\ln(1-C/2) + (1-C)\ln(1-C)\Big]. \qquad (15.103)$$

Here $C^2 = 1-(2\mu_1-1)^2$ and $\mu_1 = S^{-1}(E_F)$ denotes the larger of two pre-images of the entropy function (15.82). Similar bounds between relative entropy of entanglement, and concurrence or negativity were studied in Miranowicz and Grudka (2004).

Making use of the analytical formulae for entanglement measures we may try to explore the interior of the 15-dim set of mixed states. In general, the less pure a state is, the less it is entangled: If $R = 1/[\mathrm{Tr}\rho^2] \geq 3$ we enter the separable ball and all entanglement measures vanish. Also movements along an global orbit $\rho \to U\rho\, U^\dagger$ generically changes entanglement. For a given spectrum $\vec{x}$ the largest concurrence C^*, which may be achieved on such an orbit is given by Eq. (15.64) (Ishizaka and Hiroshima, 2000; Verstraete et al., 2001a). Hence the problem of finding maximally entangled mixed states of two qubits does not have a unique solution: it depends on the measure of mixedness and the measure of entanglement used. Both quantities may be characterized, for example, by a Rényi entropy (or its function), and for each choice of the pair of parameters q_1, q_2 one may find an extremal family of mixed states (Wei, Nemoto, Goldbart, Kwiat, Munro and Verstraete, 2003).

Figure 15.12 presents average entanglement plotted as a function of measures of mixedness computed with respect to HS measure. For a fixed purity $\mathrm{Tr}\rho^2$ the Werner states (15.42) produce the maximal negativity $\mathcal{N}_T$. On the other hand, concurrence C becomes maximal for the following states (Munro, James, White

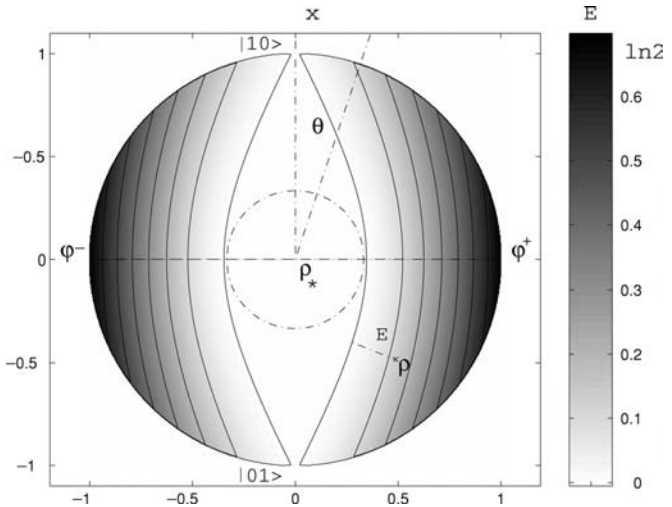


Figure 15.13. Entanglement of formation of generalized Werner states (15.105) in polar coordinates. White set represents separable states.

and Kwiat, 2001)

$$\rho_M(y) \equiv \begin{bmatrix} a & 0 & 0 & y/2 \\ 0 & 1-2a & 0 & 0 \\ 0 & 0 & 0 & 0 \\ y/2 & 0 & 0 & a \end{bmatrix}, \quad \text{where} \quad \begin{cases} a = 1/3 & \text{if} \quad y \le 2/3 \\ a = y/2 & \text{if} \quad y \ge 2/3 \end{cases} \tag{15.104}$$

Here $y \in [0, 1]$ and $C(\rho) = y$ while $\mathrm{Tr}\rho^2 = 1/3 + y^2/2$ in the former case and $\mathrm{Tr}\rho^2 = 1 - 2y(1-y)$ in the latter. A family of states σ_E providing the upper bound of E_F as a function of von Neumann entropy (see the line in Figure 15.12(c)) was found in Wei et al. (2003). Note that HS measure restricted to pure states coincides with Fubini–Study measure. Hence at $S = 0$ the average pure states entropy of entanglement reads $\langle E_1 \rangle_\psi = 1/3$, while for $R = 1$ we obtain $\langle C \rangle_\psi = \langle \mathcal{N}_T \rangle_\psi = 3\pi/16 \approx 0.59$ (see Section 14.6 and Problem 15.9).

To close this section let us show, in Figure 15.13, entanglement of formation for an illustrative class of two-qubit states

$$\rho(x, \vartheta) \equiv x(|\psi_\theta\rangle\langle\psi_\theta|) + (1-x)\rho_* \ \text{ with } \ |\psi_\theta\rangle = \frac{1}{\sqrt{2}}\Big(\sin\frac{\vartheta}{2}|01\rangle + \cos\frac{\vartheta}{2}|10\rangle\Big). \tag{15.105}$$

For $x = 1$ the pure state is separable for $\vartheta = 0, \pi$ and maximally entangled $(*)$ for $\vartheta = \pi/2, 3\pi/2$. The dashed horizontal line represents the Werner states. The set $\mathcal{M}_S$ of separable states contains the maximal ball and touches the set of pure states

in two points. A distance E of ρ from the set $\mathcal{M}_S$ may be interpreted as a measure of entanglement.

We have come to the end of our tour across the space of two qubit mixed states. Since all the properties **i)–x)** break down for higher N, the geometry of quantum entanglement gets correspondingly more complex. Already for the system of two qutrits the bound entangled states appear, while the multipartite problems contain non-equivalent forms of quantum entanglement.

Problems

Problem 15.1 Show that for any bipartite pure state $|\psi\rangle$ the transformed states $|\phi\rangle = (V_1 \otimes \mathbb{1})|\psi\rangle$ and $|\phi'\rangle = (\mathbb{1} \otimes V_2^\dagger)|\psi\rangle$ are equivalent up to the basis selection. Here both unitary matrices are equal,[51] $V_2 = V_1$. Loosely speaking, instead of doing something to the first subsystem, one can undo it in the second one.

Problem 15.2 Prove that the totally symmetric polynomials of the Schmidt vector, $\mu_1 = \sum_{i\neq j} \lambda_i\lambda_j$, $\mu_2 = \sum_{i\neq j\neq k} \lambda_i\lambda_j\lambda_k$ or $\mu_{N-1} = \lambda_1\lambda_2\ldots\lambda_N$ are entanglement monotones (Barnum and Linden, 2001; Sinołęcka et al., 2002). (See also (Gour, 2004) in which entanglement measures based on $(\mu_m)^{1/m}$ are introduced.)

Problem 15.3 Show that the *distillable entanglement*, (the optimal asymptotic efficiency m/n of an entanglement concentration protocol transforming locally n copies of an arbitrary pure state $|\psi\rangle$ into m maximally entangled states $|\phi^+\rangle$ of the $N \times N$ system), is equal to the entanglement entropy $E(|\psi\rangle)$ (Bennett et al., 1996a).

Problem 15.4 Show that separability implies the majorization relation (15.51).

Problem 15.5 Consider the following $N = 4$ mixed states

$$\sigma_1 = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \sigma_2 = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}. \tag{15.106}$$

Show that σ_2 is separable, while σ_1 is entangled (Nielsen and Kempe, 2001), even though both states are globally and locally isospectral (their partial traces have identical spectra).

Problem 15.6 Let us call ρ *locally diagonalizable* if it can be brought to the diagonal form $\rho = U\Lambda U^\dagger$ by a local unitary matrix $U = U_A \otimes U_B$. Are all separable states locally diagonalizable?

[51] Relaxing this constraint one may find such V_1 and V_2 that $|\phi'\rangle$ is exactly equal to $|\phi\rangle$. This symmetry is called environment-assisted invariance (Zurek, 2003), in short *envariance*.

Problem 15.7 Show that if the smallest eigenvalue of a density matrix ρ of size N^2 is larger than $\frac{1}{N}[(N^2-2)/(N^2-1)]$, the state ρ is separable (Pittenger and Rubin, 2002).

Problem 15.8 Show that the norm of the correlation matrix of the Fano form (15.38) satisfies an inequality: $\mathrm{Tr}(\beta\beta^T) = 2\,||\beta||_2^2 \leq [KN(KN-1) - 2K||\vec{\tau}^A||^2 - 2N||\vec{\tau}^B||^2]/4$. Is this bound sufficient to guarantee positivity of ρ?

Problem 15.9 Consider a random pure state of a 2×2 system written in its Schmidt form $|\Psi\rangle = \cos\chi|00\rangle + \sin\chi|11\rangle$ with $\chi \in [0, \pi/4]$. Show that the FS measure on the space $\mathbb{C}\mathbf{P}^3$ induces the probability distributions for the Schmidt angle (Życzkowski and Sommers, 2001)

$$P(\chi) = 3\cos(2\chi)\sin(4\chi) \quad \text{and} \quad P(C) = 3C\sqrt{1-C^2} \tag{15.107}$$

where the *concurrence* is equal $C = \sin(2\chi)$. Find the mean angle and mean concurrence. For which angle χ_m is the volume of the local orbit maximal?

Epilogue

After going through the chapter on entanglement you will have reached the end of our book. As the subtitle suggests, its aim was literally to present an introduction to the subject of quantum entanglement.

We have left untouched several important aspects of quantum entanglement, including multipartite systems, infinite-dimensional systems and continuous variables. Moreover, we believe that some key ideas presented in the book might be extended much further then we have managed to do. For instance the maps–states duality, illustrated in Chapter 11, might be used to find relations between capacities of quantum channels and measures of entanglement of the corresponding states of an extended system.

In the book we have consistently used a geometric approach to highlight similarities and differences between the classical and quantum spaces of states. What is the knowledge gained by studying the book good for? We hope it will contribute to a better understanding of quantum mechanics. We hope also that it will provide a solid foundation for a new, emerging field of science – the theory of quantum information processing. Quantum entanglement plays a decisive role in all branches of the field including quantum cryptography, quantum error correction and quantum computing.

In trying to describe the intricate geometry of the space of quantum states, we have deliberately restricted ourselves to discussing the statics of quantum theory. We have presented the arena, in which quantum information can be processed. We have not attempted to inject any concrete dynamics into our arena, but hope that readers equipped with some knowledge of its properties may introduce into it spectacular action.

In a sense we have characterized all the peculiarities of football fields of various sizes, without even specifying the rules of the game. Having at your disposal a huge flat grassy field, you can play soccer, cricket, American football, rugby or Australian football, according to your mood and wishes.

In a similar way you can play different games in the multi-dimensional arena of quantum states. It stays there right at the centre of the beautiful Platonic world of quantum theory, accessible to all of us.

Its rich structure provides a real challenge especially for young researchers. We wish you a good game in fine company! Good luck!

Appendix 1

Basic notions of differential geometry

This appendix explains things that are explained in every book on differential geometry.[1] It is included to make our book self contained.

A1.1 Differential forms

One-forms were defined in Section 1.4. The *exterior product* of one-forms is defined by

$$\mathrm{d}\theta^i \wedge \mathrm{d}\theta^j \;=\; -\mathrm{d}\theta^j \wedge \mathrm{d}\theta^i \;. \tag{A1.1}$$

The result is called a *two-form*. The exterior product is assumed to be linear over the real numbers so these two-forms can be used as a basis in which to expand any anti-symmetric covariant tensor with two indices. Continuing in the same way we can define three-forms and so on, up to and including n-forms if the dimension of the manifold is n. We can think of the exterior product of two one-forms as an area element spanned by the two one-forms. Then, given an m-dimensional submanifold, we can integrate an m-form over that submanifold. A further interesting definition is the *exterior derivative* of an m-form ω, which is an $(m+1)$-form defined by

$$\mathrm{d}\omega = \partial_{i_1}\omega_{i_2\ldots i_{m+1}}\mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_m} \;. \tag{A1.2}$$

Here $\omega_{i_1\ldots i_m}$ is an anti-symmetric tensor of the appropriate rank. The definition is a generalization of the familiar 'curl' in vector analysis. If $\mathrm{d}\omega = 0$ the form is said to be *closed*. If $\omega = \mathrm{d}\theta$ (where θ is a form of rank one less than that of ω) then ω is *exact*. An exact form is closed because $d^2 = 0$; the converse holds on topologically trivial spaces such as $\mathbb{R}^n$. An example of a closed 2-form is the field strength tensor in classical electrodynamics. An analogue of Stokes' theorem holds; if M is a

[1] Such as Schrödinger (1950) or Murray and Rice (1993).

subspace and ∂M its boundary then

$$\int_M \mathrm{d}\omega \;=\; \int_{\partial M} \omega \,. \tag{A1.3}$$

A1.2 Riemannian curvature

Given a metric tensor g_{ij} we can define the *Levi–Civita connection* by

$$\Gamma_{ij}^{\;\;k} \;=\; \frac{1}{2} g^{km}(\partial_i g_{jm} + \partial_j g_{im} - \partial_m g_{ij}) \,. \tag{A1.4}$$

It is not a tensor. In fact its transformation law contains an inhomogeneous term. Given a scalar field f its gradient $\partial_i f$ is a covariant vector, but given a contravariant (say) vector V^i the expression $\partial_j V^i$ does not transform as a tensor. Instead we use the connection to define its *covariant derivative* as

$$\nabla_j V^i \;=\; \partial_j V^i + \Gamma_{jk}^{\;\;i}\, V^k \,. \tag{A1.5}$$

This does transform like a tensor although the individual terms on the right-hand side do not. The covariant derivative of a covariant vector is then defined so that $\nabla_i(U_j V^j) = \partial_i(U_j V^j)$, which transforms like a vector since $U_j V^j$ is a scalar. An analogous argument determines the covariant derivative of an arbitrary tensor. As an example,

$$\nabla_m T_{ij}^{\;\;\;k} \;=\; \partial_m T_{ij}^{\;\;\;k} + \Gamma_{mn}^{\;\;k}\, T_{ij}^{\;\;\;n} - \Gamma_{mi}^{\;\;n}\, T_{nj}^{\;\;\;k} - \Gamma_{mj}^{\;\;n}\, T_{in}^{\;\;\;k} \,. \tag{A1.6}$$

Both the logic and the pattern should be clear. One can check that

$$\nabla_i g_{jk} \;=\; 0 \,. \tag{A1.7}$$

In fact this is how the *metric compatible affine connection* (A1.4) was defined. Parallel transport of vectors using this connection conserves lengths and scalar products.

With the metric compatible affine connection in hand we can write down a differential equation for a curve $x^i(\sigma)$ whose solution, given suitable initial data $x^i(0)$ and $\dot{x}^i(0)$, is the unique geodesic starting from that point in that direction. This *geodesic equation* is

$$\dot{x}^j \nabla_j x^i \;=\; \ddot{x}^i + \Gamma_{jk}^{\;\;i}\, \dot{x}^j\, \dot{x}^k = 0 \,. \tag{A1.8}$$

(As usual the dot signifies differentiation with respect to σ.)

The *Riemann tensor* is defined by the equation

$$(\nabla_i \nabla_j - \nabla_j \nabla_i) V^k \;=\; -R_{ijl}^{\;\;\;k}\, V^l \,. \tag{A1.9}$$

This four index tensor plays a central role in Riemannian geometry, but we will have to refer elsewhere for its properties (Schrödinger, 1950; Murray and Rice, 1993). Let us just record the slightly frightening explicit expression

$$R_{ijl}{}^{k} = \partial_j \Gamma_{il}{}^{k} - \partial_i \Gamma_{jl}{}^{k} + \Gamma_{jm}{}^{k} \Gamma_{il}{}^{m} - \Gamma_{im}{}^{k} \Gamma_{jl}{}^{m} . \tag{A1.10}$$

Perhaps it becomes slightly less frightening if we lower one index using the metric; using square brackets to denote anti-symmetry in the indices, one can show that

$$R_{ijkl} = R_{[ij][kl]} = R_{[kl][ij]} . \tag{A1.11}$$

Suppose now that we have a two-dimensional plane in the tangent space at some point, spanned by the tangent vectors m^i and n^i. Then we can define the *sectional curvature* associated to that 2-plane ,

$$K = \frac{m^{[i} n^{j]} R_{ijkl} \, m^{[k} n^{l]}}{m^2 \, n^2} . \tag{A1.12}$$

The point is that the Riemann tensor acts like a matrix on the space of 2-planes. Using index contraction we can define the *Ricci tensor*

$$R_{ij} = R_{ikj}{}^{k} . \tag{A1.13}$$

Because of the index symmetries of the Riemann tensor the Ricci tensor turns out to be a symmetric tensor, $R_{ij} = R_{ji}$. The *curvature scalar* is

$$R = g^{ij} R_{ij} . \tag{A1.14}$$

In two-dimensional spaces the Riemann tensor can be reconstructed from the curvature scalar R. (There is only one sectional curvature to worry about.) Moreover the sign of R has a simple interpretation: if $R > 0$ nearby geodesics that start out parallel tend to attract each other, while they diverge if $R < 0$. In three dimensions the Riemann tensor can be reconstructed from the Ricci tensor, while in four dimensions and higher this is no longer possible; the full Riemann tensor is needed in higher dimensions because one can choose many independent two-dimensional tangent planes along which to measure sectional curvatures.

Space is flat if and only if the Riemann tensor vanishes. In flat space parallel transport of vectors between two points is independent of the path, and it is possible to find a coordinate system in which the affine connection vanishes.

A1.3 A key fact about mappings

A key fact about tensors is that they behave well under (reasonable) maps. How they behave depends on whether they have their indices upstairs or downstairs. Suppose we have a map $M \to M'$ between two manifolds M and M'. We assume that the

dimension of the image of M is equal to the dimension of M, but the dimension of M' may be larger, in which case we have an embedding rather than a one-to-one map. Anyway we can describe the map using coordinates as $x^i \to x^{i'} = x^{i'}(x)$. Then we have the following theorem:

Theorem. *A covariant tensor on M' defines a covariant tensor on M. A contravariant tensor on M defines a contravariant tensor on the image of M in M'.*

The proof is simple, given that we know the functions $x^{i'}(x)$:

$$V_i(x) \equiv \frac{\partial x^{i'}}{\partial x^i} V_{i'}\big(x'(x)\big) , \qquad V^{i'}(x') \equiv \frac{\partial x^{i'}}{\partial x^i} V^i\big(x(x')\big) . \tag{A1.15}$$

If the map is not one-to-one the functions $x^i(x^{i'})$ are defined only on the image of M, so the theorem is as general as it can get.

Appendix 2

Basic notions of group theory

This appendix lists a number of formulae that are explained in every book on group theory.[1] Some conventions can be chosen at will, which is why this list is essential.

A2.1 Lie groups and Lie algebras

Lie groups are groups containing a continuously infinite number of elements with the amazing property that they can to a large extent be understood through an analysis of the tangent space at the unit element of the group. This tangent space is known as the Lie algebra of the group. We will deal only with the *classical groups* $SU(N)$, $SO(N)$ and $Sp(N)$ and in fact mostly with the *special unitary groups* $SU(N)$. These are all, in the technical sense, simple and compact groups and have in many respects analogous properties. The unitary group $U(N)$ is not simple, but can be understood in terms of its simple subgroups $U(1)$ and $SU(N)$.

After complexification, the Lie algebra of a compact simple group can be brought to the standard *Cartan form*

$$[H_i, H_j] = 0\,, \qquad [H_i, E_\alpha] = \alpha_i E_\alpha\,, \tag{A2.1}$$

$$[E_\alpha, E_\beta] = N_{\alpha\beta} E_{\alpha+\beta}\,, \qquad [E_\alpha, E_{-\alpha}] = \alpha^i H_i\,, \tag{A2.2}$$

where α_i is a member of the set of positive *root vectors* and $N_{\alpha\beta} = 0$ if $\alpha_i + \beta_i$ is not a root vector. The H_i, $1 \le i \le r$ span the maximal commuting Cartan subalgebra. Their number r is the *rank* of the group, equal to $N - 1$ for $SU(N)$.

The Lie bracket $[A, B]$ is a peculiar kind of product on a vector space. Once a matrix representation is chosen it is also meaningful to consider the usual product AB, and set $[A, B] = AB - BA$.

[1] Such as Gilmore (1974) or Fuchs and Schweigert (2003).

A2.2 SU(2)

The generators of $SU(2)$ in the fundamental representation are precisely the Pauli matrices $\vec{\sigma} = \{\sigma_x, \sigma_y, \sigma_z\}$,

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{A2.3}$$

They form an orthonormal basis in the Lie algebra. There is one unitary representation of $SU(2)$ in every dimension $N = n + 1 = 2j + 1$. Then the generators are

$$J_x = \frac{1}{2}\begin{bmatrix} 0 & \sqrt{n} & 0 & 0 & \cdots \\ \sqrt{n} & 0 & \sqrt{2(n-1)} & 0 & \cdots \\ 0 & \sqrt{2(n-1)} & 0 & \sqrt{3(n-2)} & \cdots \\ 0 & 0 & \sqrt{3(n-2)} & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{A2.4}$$

$$J_y = \frac{\mathrm{i}}{2}\begin{bmatrix} 0 & -\sqrt{n} & 0 & 0 & \cdots \\ \sqrt{n} & 0 & -\sqrt{2(n-1)} & 0 & \cdots \\ 0 & \sqrt{2(n-1)} & 0 & -\sqrt{3(n-2)} & \cdots \\ 0 & 0 & \sqrt{3(n-2)} & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{A2.5}$$

$$J_z = \frac{1}{2}\begin{bmatrix} n & 0 & 0 & \cdots \\ 0 & n-2 & 0 & \cdots \\ 0 & 0 & n-4 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \equiv \begin{bmatrix} j & 0 & 0 & \cdots \\ 0 & j-1 & 0 & \cdots \\ 0 & 0 & j-2 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{A2.6}$$

A crucial choice here is that J_x and J_z are real while J_y is imaginary. To get to the Cartan form, set $H = J_z$ and $E_\pm = J_x \pm \mathrm{i}J_y$.

A2.3 SU(N)

$SU(N)$ is an $(N^2 - 1)$-dimensional group and in the defining representation the Lie algebra consists of $N^2 - 1$ traceless Hermitian N by N matrices, labelled by the index i, $1 \le i \le N^2 - 1$. A complete orthonormal set of generators obey

$$\sigma_i \sigma_j \;=\; \frac{2}{N}\delta_{ij} + d_{ijk}\,\sigma_k + \mathrm{i} f_{ijk}\,\sigma_k \;, \tag{A2.7}$$

where f_{ijk} is totally anti-symmetric in its indices and d_{ijk} is totally symmetric and traceless ($d_{iik} = 0$). For the commutator and the anti-commutator respectively, this

means that

$$[\sigma_i, \sigma_j] = 2\mathrm{i}\, f_{ijk}\, \sigma_k\ , \qquad \{\sigma_i, \sigma_j\} = \frac{4}{N}\delta_{ij} + 2\, d_{ijk}\, \sigma_k\ . \tag{A2.8}$$

The generators are Hermitian matrices and obey

$$\mathrm{Tr}\, \sigma_i\sigma_j = 2\, \delta_{ij}\ , \qquad \mathrm{Tr}\, \sigma_i\sigma_j\sigma_k = 2\, d_{ijk} + 2\mathrm{i}\, f_{ijk} \tag{A2.9}$$

(the first equation here is orthonormality of the generators) as well as the completeness relation

$$(\sigma_i)_A^{\ B}(\sigma_i)_C^{\ D} = 2\delta_A^D\delta_B^C - \frac{2}{N}\delta_A^B\delta_C^D\ . \tag{A2.10}$$

The above normalization is implied by the case $N = 2$, since the Pauli matrices (A2.3) are normalized just like this. In this case $d_{ijk} = 0$, while the group constants form the anti-symmetric tensor, $f_{ijk} = \epsilon_{ijk}$.

The following identities are true for any N:

$$f_{ijm}f_{mkn} + f_{jkm}f_{min} + f_{kim}f_{mjn} = 0 \tag{A2.11}$$

$$d_{ijm}f_{mkn} + d_{jkm}f_{min} + d_{kim}f_{mjn} = 0 \tag{A2.12}$$

$$f_{ijm}f_{mkn} = \frac{2}{N}(\delta_{ik}\delta_{jn} - \delta_{in}\delta_{jk}) + d_{ikm}d_{mjn} - d_{jkm}d_{min} \tag{A2.13}$$

$$f_{imn}f_{jmn} = N\delta_{ij} \qquad d_{imn}d_{jmn} = \frac{N^2-4}{N}\delta_{ij}\ . \tag{A2.14}$$

Scanning this list one realizes that some identities are 'missing'. These identities exist, but they depend on N. For further information consult the literature (Macfarlane, Sudbery and Weisz, 1968; Sudbery, 1990).

A2.4 Homomorphisms between low-dimensional groups

We sometimes makes use of the following isomorphisms between Lie algebras:

$$SU(2) \ \cong\ SO(3)\ , \tag{A2.15}$$

$$SO(4) \ \cong\ SO(3)\times SO(3) \ \cong\ SU(2)\times SU(2)\ , \tag{A2.16}$$

$$SU(4) \ \cong\ SO(6)\ . \tag{A2.17}$$

Globally these are $2 \to 1$ homomorphisms between the corresponding groups; for explanations see (especially) Section 3.7.

Appendix 3

Geometry: do it yourself

In this appendix we provide some additional exercises of a more practical nature.

Exercise 3.1 – Real projective space. Cut out a disc of radius r. Prepare a narrow strip of length πr and glue it into a Möbius strip. The total length of the boundary of a strip is equal to the circumference of the disc so you may try to glue them together.[1] When you are finished, you can contemplate a fine model of a real projective space, $\mathbb{R}\mathbf{P}^2$.

Exercise 3.2 – Hypersphere $\mathbf{S}^3$ may be obtained by identifying points on the surfaces of two identical 3-balls as discussed in Section 3.1. To experience further features of the hypersphere get some playdough and prepare two cylinders of different colours with their length more than three times larger than their diameter. Form two linked tori as shown in Figure A3.2.

Start gluing them together along their boundaries. After this procedure is completed, you will be in position to astonish your colleagues by presenting them a genuine *Heegard decomposition* of a hypersphere.

Exercise 3.3 – Mixed states. Make a ball out of playdough. Glue a string to its surface along the shape of the stitching of a tennis ball (see Figure A3.3). Obtain a convex hull by cutting out the redundant dough with a knife. How much of a ball is taken away?

Convince yourself that the convex hull of a one-dimensional string located at $\mathbf{S}^2$ forms a considerable part of the ball $\mathbf{B}^3$. In a similar way $\mathcal{M}^{(3)}$ – the convex hull of four-dimensional manifold $\mathbb{C}\mathbf{P}^2$ consisting of $N = 3$ pure states placed at $\mathbf{S}^7$ contains a non-negligible part of $\mathbf{B}^8$ (see Section 14.3).

Exercise 3.4 – Entangled pure states. Magnify Figure A3.4 and cut out the net of the cover tetrahedron. It represents the entanglement of formation of the pure

[1] If you happen to work in three-dimensions this simple task gets difficult.

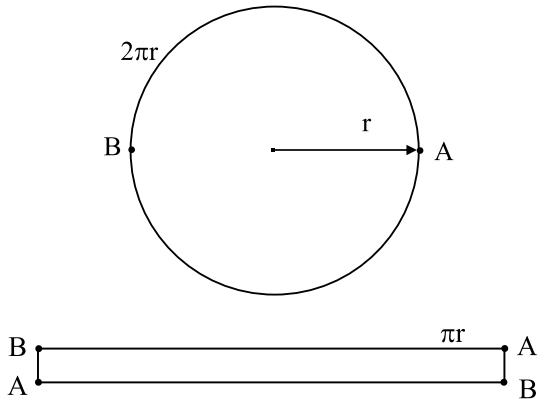


Figure A3.1. A narrow Möbius strip glued with a circle produces $\mathbb{R}\mathbf{P}^2$.

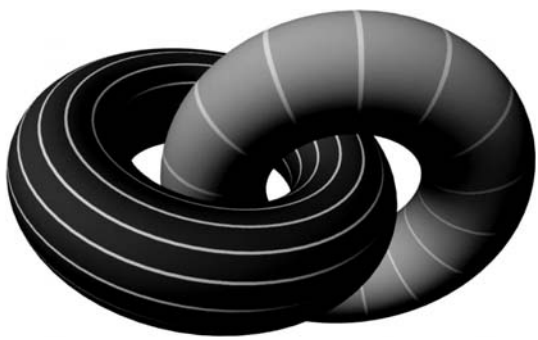


Figure A3.2. Heegard decomposition of a 3-sphere.

Figure A3.3. Imagine a convex hull of the one-dimensional stitching of the tennis ball.

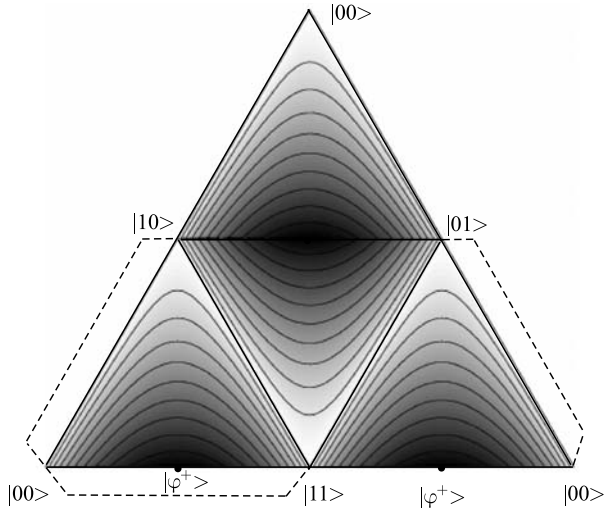


Figure A3.4. Net of the tetrahedron representing entanglement for pure states of two qubits: maximally entangled states plotted in black.

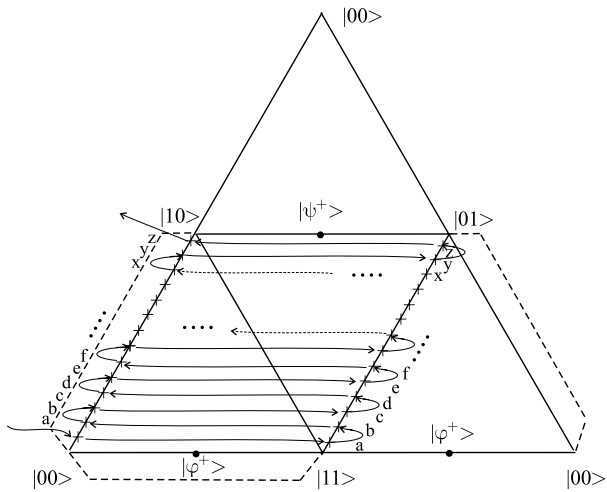


Figure A3.5. Sew with a coloured thread inside a transparent tetrahedron to get the ruled surface consisting of separable pure states of two qubits.

states of two qubits for a cross section of $\mathbb{C}\mathbf{P}^3$ defined in Eqs. (4.70) and (4.71) by setting all phases ν_i in (4.67) to zero.

Glue it together to get the entanglement tetrahedron with four product states in four corners. Enjoy the symmetry of the object and study the contours of the states of equal entanglement.

Exercise 3.5 – Separable pure states. Prepare a net of a regular tetrahedron from transparency according to the blueprint shown in Figure A3.5. Make holes with a needle along two opposite edges as shown in the picture. Thread a needle with a (red) thread and start sewing it through your model. Only after this job is done glue the tetrahedron together.[2] If you pull out the loose thread and get the object sketched in Figure 15.1, you can contemplate how a fragment of the subspace of separable states forms a ruled surface embedded inside the tetrahedron.

[2] Our experience shows that sewing after the tetrahedron is glued together is much more difficult.

Appendix 4

Hints and answers to the exercises

Problem 1.1. A counterexample is easily provided by drawing figures in two dimensions.

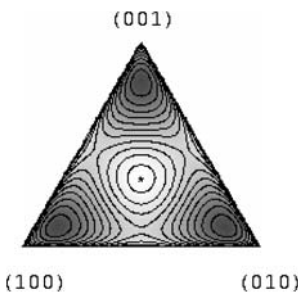
Problem 1.2. One way is to construct the simplex. If we put its centre at the origin the $N = n + 1$ points can be placed at

$$\begin{array}{cccccccc}
(& -r_1, & -r_2, & \cdots, & -r_{n-1}, & -r_n &) \\
(& R_1, & -r_2, & \cdots, & -r_{n-1}, & -r_n &) \\
(& 0, & R_2, & \cdots, & -r_{n-1}, & -r_n &) \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
(& 0, & 0, & \cdots, & 0, & R_n &)
\end{array}$$

This helps.

Problem 2.1.

Here is a plot of the structural entropy. The maximum $S_2 - S_1 \approx 0.223\,66$ is attained for $\vec{p} \approx (0.806, 0.097, 0.097)$ and its two other permutations. They are visible at the contour plot provided as three dark hills.



Problem 2.2. The $N = 3$ case shows the idea. We have $\vec{x} \cdot (\vec{y} - \vec{z}) = x_1(y_1 - z_1) + x_2(y_2 - z_2) + x_3(y_3 - z_3) = (x_1 - x_2)(y_1 - z_1) + (x_2 - x_3)(y_1 + y_2 - z_1 - z_2) + x_3(y_1 + y_2 + y_3 - z_1 - z_2 - z_3) \geq 0$ because of the conditions stated.

Problem 2.3. (a) Let $a, b \in [0, 1]$ and $a' = 1 - a, b' = 1 - b$ and define

$$B = T_1 T_2 = \begin{bmatrix} a & a' & 0 \\ a' & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & b & b' \\ 0 & b' & b \end{bmatrix} = \begin{bmatrix} a & a'b & a'b' \\ a' & ab & ab' \\ 0 & b' & b \end{bmatrix}. \tag{A4.1}$$

B is a bistochastic matrix. It is also orthostochastic since $B_{ij} = (O_{ij})^2$, where

$$O = \begin{bmatrix} \sqrt{a} & \sqrt{a'b} & -\sqrt{a'b'} \\ \sqrt{a'} & -\sqrt{ab} & \sqrt{ab'} \\ 0 & \sqrt{b'} & \sqrt{b} \end{bmatrix} . \tag{A4.2}$$

Problem 2.4. We know that $\vec{x}$ is a (non-unique) convex combination of permutation matrices acting on $\vec{y}$; this defines a bistochastic matrix according to Birkhoff's theorem.

Problem 2.5. One obtains two important cases of the Dirichlet distribution (2.73): the round measure ($s = 1/2$) for real and flat measure ($s = 1$) for complex Gaussian random numbers (Życzkowski and Sommers, 2001). Compare also Problem 7.3.

Problem 2.7. For $q \leq 2$. To see this, study the second derivative in the vicinity of $p = 1$.

Problem 3.1. You will obtain

$$\frac{x^i}{X^I} = \frac{2}{1 + X^0} \quad \Rightarrow \quad X^0 = \frac{4 - r^2}{4 + r^2} \quad \Rightarrow \quad \mathrm{d}s^2 = \left(\frac{4}{4 + r^2}\right)^2 \mathrm{d}x^i \mathrm{d}x^i \, . \tag{A4.3}$$

Problem 3.2. The angles are obtained by intersecting, respectively, the plane and the sphere with two intersecting planes. The angles will be equal if and only if both the plane and the sphere meet the line of intersection of the two planes at the same angle. But this will happen if and only if the line of intersection forms a chord of the great circle.

Problem 3.3. You can try a calculation to see whether the natural map $(x, y) \to (x, 2y)$ between the tori is analytic (it is not). Or you can observe that the tori inherit natural flat metrics from the complex plane. On each torus there will be a pair of special closed geodesics that intersect each other, namely what used to be straight lines along the x- and y-directions on the plane. Their circumferences are equal on one of the tori, and differ by a factor of two on the other. But analytic, hence conformal, transformations do not change the ratio of two lengths, and it follows that no such analytic transformation between the tori can exist.

Problem 3.4. As an intermediate step you must prove

$$\Omega^{il}\partial_l\Omega^{jk} + \Omega^{kl}\partial_l\Omega^{ji} + \Omega^{kl}\partial_l\Omega^{ij} = 0 \, . \tag{A4.4}$$

Problem 3.5. A hyperplane through the origin in embedding space meets the 3-sphere in a 2-sphere given in stereographic coordinates by

$$a_I X^I = 0 \qquad \Rightarrow \qquad 2a_1 x + 2a_2 y + 2a_3 z + 1 - r^2 = 0$$

(where we assumed that the fourth component of the vector equals one). A geodesic is the intersection of two such spheres; choose them to have their centres at $(a, 0, 0)$ and $(b_1, b_2, 0)$ again without loss of generality. If you also demand $r^2 = 1$ (the equator) you get three equations with the solutions $(x, y, z) = (0, 0, \pm 1)$.

Problem 3.6. The key point is that two Hopf circles with the opposite twist meet twice. Only one half of the circumference of a Hopf circle is needed to label the members of the family of circles that twist in the other way. (Draw the torus as a flat square to see this.)

Figure A4.1. Left: the Möbius strip as a vector bundle, with a global section (i.e. an embedding of the circle in the bundle). Right: the principal bundle, with fibres equal to the group $\{\pm\}$.

Problem 3.7. For $\tau = -\phi$ we get

$$X + \mathrm{i}Y = \cos\frac{\theta}{2} \quad \Rightarrow \quad Y = 0 \ \& \ X > 0 \quad \Rightarrow \quad y = 0 \ \& \ x > 0\,. \tag{A4.5}$$

With the exception of one point this maps $\mathbf{S}^2$ onto a half plane. The other two sections provide maps onto a hemisphere and a unit disc, respectively. In all three cases it is geometrically evident that we are selecting one point from each geodesics, except for a single one for which there is no prescription.

Problem 3.8. The group acting on the fibres is the discrete group with two elements, the unit element and the element that turns the fibre upside down. Figure A4.1 tells the rest.

Problem 4.1. Consider the case of three points and put one at the origin, one at $(1, 0)$. The location of the third point can be anywhere. That gives an $\mathbb{R}^2$. This coordinatization of the space of triplets fails if the first two points coincide. That case evidently corresponds to one additional point 'at infinity', so we have a natural one-to-one correspondence between the space of triplets and a plane + the point at infinity, that is $\mathbb{C}\mathbf{P}^1$. Consider the case of four points: if the first two points are distinct we proceed as above; the two remaining points can be coordinatized by $\mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{C}^2$. If the first two points coincide we have only a triplet of points do deal with. This is a $\mathbb{C}\mathbf{P}^1$ according to what was just shown. But $\mathbb{C}^2$ plus a $\mathbb{C}\mathbf{P}^1$ 'at infinity' is a $\mathbb{C}\mathbf{P}^2$. And so on. This is useful in archaeology if we use the Fubini–Study metric to give a measure. Then we can answer questions like 'given $n + 2$ stones, what is the probability that there are k triplets of stones lying (to a given precision) on straight lines?'.

Problem 4.2. A Klein bottle; a bottle without an inside (or outside). It takes a four-dimensional being to make one that does not intersect itself.

Problem 4.3. Integrate the Fubini–Study 2-form Ω over the embedded $\mathbb{C}\mathbf{P}^1$; this gives $\int_{\mathbb{C}\mathbf{P}^1} \Omega =$ the area $= \pi$, since Ω induces the usual Fubini–Study 2-form on $\mathbb{C}\mathbf{P}^1$. But if $\mathbb{C}\mathbf{P}^1$ could be shrunk to a point then this calculation could be done within a single coordinate patch, and there could be no obstruction to the calculation $\int_{\mathbb{C}\mathbf{P}^1} \Omega = \int_{\mathbb{C}\mathbf{P}^1} \mathrm{d}\omega = \int_{\partial(\mathbb{C}\mathbf{P}^1)} \omega = 0$, where we used Stokes' theorem and the fact that $\mathbb{C}\mathbf{P}^1$ has no boundary. This is a contradiction. Alternatively one can stare at the line at infinity in the octant picture of $\mathbb{C}\mathbf{P}^2$ and convince oneself that any attempt to move it will increase its area.

Problem 5.1. The two pure states divide a great circle into two segments. If one of the eigenstates of A lies on the shortest of these segments, the answer is $D_{\mathrm{Bhatt}} = \theta_A + \theta/2$, otherwise it is $D_{\mathrm{Bhatt}} = \theta/2$ (independent of A).

Problem 6.1. From (6.9) the Q-function of a Fock state is $Q_{|n\rangle}(z) = |z|^{2n}\mathrm{e}^{-|z|^2}/n!$ and from (6.47) we must have $\int \mathrm{d}z^2 Q_{|n\rangle} P_{|1\rangle} = \delta_{n1}$. The solution is the moderately singular distribution $P_{|1\rangle} = \mathrm{e}^{|z|^2}\partial_z\partial_{\bar{z}}\delta^{(2)}(z)$.

Problem 6.2. Using spherical polars in phase space (and the integral representation of the gamma function) one finds $S_W(|n\rangle) = 1 + n + \ln n! - n\Psi(n+1)$, where Ψ is the digamma function defined in Eq. (7.52).

Problem 7.1. It is given by $D_{\mathrm{FS}} = \arccos\sqrt{Q_{\max}}$, so it can be obtained by taking the maximum of Q in Eq. (7.23).

Problem 7.2. Wehrl entropy and participation number for pure states of $N = 2$–5 read

N	j	m	$S_W(\|\psi\rangle)$	$R(\|\psi\rangle)$
2	1/2	$\pm 1/2$	$1/2 = 0.5$	$1\frac{1}{2}$
3	1	± 1	$2/3 \approx 0.667$	$1\frac{2}{3}$
3	1	0	$5/3 - \ln 2 \approx 0.974$	$2\frac{1}{2}$
4	3/2	$\pm 3/2$	$3/4 = 0.75$	$1\frac{3}{4}$
4	3/2	$\pm 1/2$	$9/4 - \ln 3 \approx 1.151$	$2\frac{11}{12}$
4	3/2	$\|\psi_{\triangle}\rangle$	$21/8 - \ln 4 \approx 1.239$	$3\frac{2}{11}$
5	2	± 2	$4/5 = 0.8$	$2\frac{1}{4}$
5	2	± 1	$79/30 - \ln 4 \approx 1.247$	$3\frac{3}{20}$
5	2	0	$47/15 - \ln 6 \approx 1.342$	$3\frac{1}{2}$
5	3/2	$\|\psi_{\mathrm{tetr.}}\rangle$	$165/45 - \ln 9 \approx 1.492$	$4\frac{1}{5}$

Problem 7.3. For qubits, the uniform distribution of points at the Bloch sphere according to the measure $\sin\theta \mathrm{d}\theta\mathrm{d}\phi$ leads to the uniform distribution of the component $y_1 = \cos\theta$ in the interval $[-1, 1]$. For larger n write components of a random vector using auxiliary random variables $\xi_k = (\sin\vartheta_k)^{2k}$, distributed uniformly in $[0, 1]$. If $n = 2$ one has $\vec{y} = (\cos^2\vartheta_2, \sin^2\vartheta_2\cos^2\vartheta_1, \sin^2\vartheta_2\sin^2\vartheta_1) = (1 - \xi_2^{1/2}, \xi_2^{1/2}(1-\xi_1), \xi_2^{1/2}\xi_1)$. This vector is distributed uniformly in the triangle $y_1 \in [0, 1]$ and $y_2 \in [0, 1 - y_1]$, while $y_3 = 1 - y_1 - y_2$. The same reasoning performed for arbitrary n allows one to obtain the desired result (Życzkowski, 1999).

Problem 8.2. (a) yes; (b) no. Any permutation is represented by an orthogonal matrix, and multiplication by any unitary matrix does not change the singular values.

Problem 8.3. (a) absolute values of real eigenvalues; (b) equal to unity; (c) absolute values of complex eigenvalues.

Problem 8.4. In fact an even stronger property is true. Directly from the definition of the singular values it follows that $\mathrm{sv}(A) = \mathrm{sv}(UAV)$, for arbitrary unitary U and V. However, the special case $V = U^{-1}$ is often useful in calculations.

Problem 8.5. This is the Cauchy–Schwarz inequality for the scalar product in Hilbert–Schmidt space.

Problem 8.6. We know that $\langle\psi|P|\psi\rangle \geq 0$ for all vectors. Let $|\psi\rangle$ be a basis vector.

Problem 8.7. We can bring an arbitrary vector $\tau_i\sigma_i$ into the Cartan subalgebra, $U\tau_i\sigma_i U^\dagger = \lambda_i H_i$. Generically that is the best we can do, so the number of non-zero elements will equal the dimension of the Cartan subalgebra, that is $N - 1$.

Problem 8.8. Definition (8.30) applied to the Pauli matrices gives

$$O=\begin{bmatrix} c^2\vartheta c(2\phi)-s^2\vartheta c(2\psi) & c^2\vartheta c(2\phi)+s^2\vartheta c(2\psi) & -s(2\vartheta)c(\phi+\psi) \\ s^2\vartheta s(2\psi)-c^2\vartheta s(2\phi) & c^2\vartheta c(2\phi)+s^2\vartheta c(2\psi) & s(2\vartheta)s(\phi+\psi) \\ s(2\vartheta)c(\psi-\phi) & -s(2\vartheta)s(\psi-\phi) & c(2\vartheta) \end{bmatrix}. \tag{A4.6}$$

where $c \equiv \cos$, $s \equiv \sin$. This is the *Cayley parametrization* of the group $SO(3)$ and describes the rotation with respect to the axis $\vec{\Omega} = (\sin\vartheta \sin\psi, \sin\vartheta \cos\psi, \cos\vartheta \sin\phi,)$ by an angle t such that $\cos(t/2) = \cos\vartheta \cos\phi$.

Problem 9.4. The spectrum consists of the MN numbers $\alpha_i\beta_j$, where $i = 1, \ldots, M$ and $j = i, \ldots, N$ (say).

Problem 9.6. Use the Schwarz inequality, $|\mathrm{Tr}(AB)|^2 \leq \mathrm{Tr}(AA^\dagger) \times \mathrm{Tr}(BB^\dagger)$, and replace A by $A\rho^{1/2}$ and B by $B\rho^{1/2}$, respectively (Mehta, 1989).

Problem 10.1. It is enough to consider a pure POVM, so that $E_i = |\phi_i\rangle\langle\phi_i|$. The vectors have components ϕ_i^α, $\alpha = 1, \ldots, N$, $i = 1, \ldots, k$. Let these components be the elements of an $N \times k$ matrix. The completeness relation $\sum_{i=1}^k (E_i)_\beta^\alpha = \delta_\beta^\alpha$ implies that the rows of this matrix are orthonormal. We can always add an additional set of $k - N$ rows to the matrix, so that it becomes unitary. The columns of the new matrix form an orthonormal basis in a k dimensional Hilbert space, and there is an obvious projection of its vectors down to the original Hilbert space.

Problem 10.2. Let $a = \mathrm{diag}(A)$ and $c = \mathrm{diag}(C)$ be diagonals of complex matrices. Show that $ABC^\dagger = (ac^\dagger) \circ B$ and use it with $A = A_i$ and $C = A_i^\dagger$ for all $i = 1, \ldots, k$.

Problem 10.3. To prove positivity, take an arbitrary vector V^i and define $A \equiv V^i A_i$. Then $\bar{V}^i \sigma_{ij} V^j = \mathrm{Tr}\rho AA^\dagger$ is positive because the trace of two positive operators is always positive. For the final part see (10.56).

Problem 10.5. The phase flip channel.

Problem 10.6. The dynamical matrix D_Φ is represented by $D_{\substack{mn \\ \mu\nu}} = \rho_{m\mu}\delta_{n\nu}$. Writing down the elements of the image $\sigma' = \Phi_\rho(\sigma) = D^R\sigma = (\rho \otimes \mathbb{1}_N)^R\sigma$ in the standard basis we obtain the desired result, $\sigma'_{m\mu} = D_{\substack{mn \\ \mu\nu}}\sigma_{n\nu} = \rho_{m\mu}\mathrm{Tr}\sigma = \rho_{m\mu}$.

Problem 11.1. Write both matrices in their eigen representations, $A = \sum_i a_i|\alpha_i\rangle\langle\alpha_i|$ and $B = \sum_i b_i|\beta_i\rangle\langle\beta_i|$. Perform decompositions $A^R = \sum_i a_i\alpha^{(i)} \otimes \bar{\alpha}^{(i)}$ and $B^R = \sum_i b_i\beta^{(i)} \otimes \bar{\beta}^{(i)}$ as in (10.58), multiply them and reshuffle again to establish positivity of $(A^R B^R)^R$. For a different setting see Havel (2003).

Problem 11.2. (a) The spectrum $\vec{d}$ consists of $N(N+1)/2$ elements equal to $+1$ and $N(N-1)/2$ elements equal to -1, so its sum (the trace of D) is equal to N, as required. (b) This canonical form contains one negative term and three positive terms; due to the triple degeneracy the choice of the positive terms is not unique.

Problem 11.3. Using the nonhomogeneous form of (10.36) write down the dynamical matrix D and show that some of its eigenvalues are negative. To study the co-positivity analyse the spectrum of D^{T_a} (equal to the spectrum of D^{T_b}).

Problem 11.4. $\Phi_T = \frac{N-1}{N}\Phi_* + \frac{1}{N}T$

Problem 11.5. The spectrum $\vec{d}$ of D_Ψ reads $(a-3, a, a, b, b, b, c, c, c)$. Hence $cp(\Psi) = \min(a-3, b, c)$ and the map is CP if $a \geq 3$ and $b, c \geq 0$. The spectrum of $D_\Psi^{T_A}$ is threefold degenerated and contains $\{a-1, \lambda_+, \lambda_-\}$, where $\lambda_\pm = (b+c \pm \sqrt{(b-c)^2+4})/2$. Therefore $ccp(\Psi) = \min(a-1, \lambda_-)$ and the map is CcP if $a \geq 1$ and $bc \geq 1$. Note that these results are consistent with (11.9) and (11.10).

Problem 12.1. Open with the observation that

$$\ln(A + xB) = \ln(A + xB + u_0) - \int_0^{u_0} \frac{\mathrm{d}u}{A + xB + u} \,. \tag{A4.7}$$

Use the fact that $A + u$ is invertible for any positive u to rewrite the right-hand side as

$$\ln(A + u_0) + \ln\left(\mathbb{1} + \frac{1}{A+u_0}xB\right) - \int_0^{u_0}\left[\frac{1}{A+u} - \frac{1}{A+u}xB\frac{1}{A+xB+u}\right]\mathrm{d}u \,.$$

Now do the integral over the first term, collect terms, expand the remaining logarithm, and finally let $u_0 \to \infty$.

Problem 12.2. The eigenvalues of $(\mathbb{1} - \rho/z)^{-1}$ are $z/(z-\lambda_i)$. With ρ diagonalized and the contour chosen suitably, the first integral equals the von Neumann entropy and the second is the subentropy,

$$S_Q(\rho) \equiv -\sum_{i=1}^{N}\Big(\prod_{i\neq j}\frac{\lambda_i}{\lambda_i - \lambda_j}\Big)\lambda_i \ln\lambda_i \,. \tag{A4.8}$$

Problem 12.3.

$$\begin{aligned}\sum_k p_k S(\rho_k||\sigma) &= \sum_k p_k(\mathrm{Tr}\rho_k \ln\rho_k - \mathrm{Tr}\rho_k \ln\sigma) \\ &= \sum_k p_k(\mathrm{Tr}\rho_k \ln\rho_k - \mathrm{Tr}\rho_k \ln\sigma + \mathrm{Tr}\rho_k \ln\rho - \mathrm{Tr}\rho_k \ln\rho) \\ &= \sum_k p_k(\mathrm{Tr}\rho_k \ln\rho_k - \mathrm{Tr}\rho_k \ln\rho) + \mathrm{Tr}\rho \ln\rho - \mathrm{Tr}\rho \ln\sigma = \sum_k p_k S(\rho_k||\rho) + S(\rho||\sigma) \,.\end{aligned} \tag{A4.9}$$

Problem 12.4. Write the Husimi function using the Schmidt decomposition $|\Psi\rangle = \sqrt{\lambda_1}|11\rangle + \sqrt{\lambda_2}|22\rangle$. Integration over the Cartesian product of two spheres gives

$$S_W(\Psi) = \frac{\lambda_1^2(1-\ln\lambda_1)}{\lambda_1 - \lambda_2} + \frac{\lambda_2^2(1-\ln\lambda_2)}{\lambda_2 - \lambda_1} \,. \tag{A4.10}$$

Up to an additive constant this result is equal to the Wehrl entropy of one qubit mixed state obtained by partial trace (Mintert and Życzkowski, 2004) or to the subentropy (A4.8) of this state.

Problem 12.5. It is enough to show that $\mathrm{Tr} d_1 d_2 \geq \mathrm{Tr} W d_1 W^\dagger d_2$, where $W = V^\dagger U$ is unitary. We can write this as an inequality for scalar products between vectors: $\vec{d}_1 \cdot \vec{d}_2 \geq (B\vec{d}_1) \cdot \vec{d}_2$ where B is unistochastic. Problem 2.2 shows that this is true.

Problem 12.6. Work in a basis where $A + B$ is diagonal. Note that $\left(\det(A+B)\right)^{1/N} = \prod_i (A_{ii} + B_{ii})^{1/N} \geq \prod_i A_{ii}^{1/N} + \prod_i B_{ii}^{1/N}$. For our purposes the second step is the more interesting: from the Schur–Horn theorem and the Schur concavity of the elementary symmetric functions it follows that $\prod_i A_{ii} \geq \det A$ (and similarly for B).

Problem 12.7. It is sufficient to compute $\mathrm{Tr} L L^\dagger$ using the following representation of the superoperator, $L = \sum_{i=1}^r A_i \otimes A_i^*$.

Problem 13.1. We want to show that $||\mathbf{x}||^2 \geq ||B\mathbf{x}||^2$, so we must show that $\mathbb{1} - B^T B$ is a positive operator. This follows from the Frobenius–Perron theorem.

Problem 13.2. We know that there is a POVM such that $\sqrt{F} = \sum_i \sqrt{p_i q_i}$, with probabilities given in Eq. (13.48). Then

$$\begin{aligned} 2(1 - \sqrt{F}) &= \sum_i (\sqrt{p_i} - \sqrt{q_i})^2 \leq \sum_i |\sqrt{p_i} - \sqrt{q_i}||\sqrt{p_i} + \sqrt{q_i}| \\ &= \sum_i |p_i - q_i| \leq 2 D_{\mathrm{Tr}}(\rho, \sigma)\,, \end{aligned} \tag{A4.11}$$

where Helstrom's theorem was used in the last step.

Problem 13.3. (a) This follows if we set $U = 1$ in the argument that led to the quantum Bhattacharyya coefficient. Equality holds if $[\rho, \sigma] = 0$.
(b) Equality holds if one of the states is pure. The inequality follows from this because of concavity. Incidentally, Uhlmann has proved but not published that $F(\sigma, \rho) \geq \mathrm{Tr}\sigma\rho + \sqrt{2}\sqrt{(\mathrm{Tr}\sigma\rho)^2 - \mathrm{Tr}\sigma\rho\sigma\rho}$, with equality for $N = 2$.

Problem 14.1. No, since $f_{\mathrm{WY}} = (f_{\max} + f_{\mathrm{geom}})/2$, where $f_{\mathrm{geom}} = \sqrt{t}$ is related to the geometric mean, $1/c_{\mathrm{geom}}(x, y) = \sqrt{xy}$.

Problem 14.2.

Table A4.1. *Volumes of orthogonal groups and real flag manifolds*

Manifold	Dimension	$\mathrm{Vol}[X]$, $a = 1/2$	$\mathrm{Vol}'[X]$, $a = 1$
$\mathbb{R}\mathbf{P}^N$	N	$\frac{\pi^{(N+1)/2}}{\Gamma[(N+1)/2]}$	$2^{N/2}\frac{\pi^{(N+1)/2}}{\Gamma[(N+1)/2]}$
$\mathbf{F}_{\mathbb{R}}^{(N)} = \frac{O(N)}{[O(1)]^N}$	$N(N-1)/2$	$\frac{\pi^{N(N+1)/4}}{\Theta_N}$	$2^{N(N-1)/4}\frac{\pi^{N(N+1)/4}}{\Theta_N}$
$O(N)$	$N(N-1)/2$	$2^N\frac{\pi^{N(N+1)/4}}{\Theta_N}$	$2^{N(N+3)/4}\ \frac{\pi^{N(N+1)/4}}{\Theta_N}$
$SO(N)$	$N(N-1)/2$	$2^{N-1}\frac{\pi^{N(N+1)/4}}{\Theta_N}$	$2^{(N(N+3)/4-1}\ \frac{\pi^{N(N+1)/4}}{\Theta_N}$

where $\Theta_N \equiv \prod_{k=1}^N \Gamma(k/2)$.

Problem 14.4. Integrating over respective distributions we obtain $\langle S(\rho)\rangle_{\rm HS} = 1/3$, $\langle S(\rho)\rangle_{\rm B} = 2 - 7\ln 2/6$, $\langle S(\rho)\rangle_o = 2 - \ln 2$ and $\langle S(\rho)\rangle_u = \ln 2/2$.

Problem 14.5. The averages read

$$\langle {\rm Tr}\rho^3\rangle_{N,K} = \frac{(K+N)^2 + KN + 1}{(KN+1)(KN+2)}, \quad \langle {\rm Tr}\rho^4\rangle_{N,K} = \frac{(K+N)[(K+N)^2 + 3KN + 5]}{(KN+1)(KN+2)(KN+3)}, \quad \text{(A4.12)}$$

and are consistent with results of Malacarne et al. (2002).

Problem 14.6. The result

$$P_{\rm HS}(x) = \frac{1}{2\pi}\sqrt{\frac{4}{x} - 1} \quad \text{(A4.13)}$$

is a special case of Eq. (14.59) for $K = N$. In the rescaled variables, $y = \sqrt{x}$, this distribution is equivalent to the quarter-circle law, $P(y) = \sqrt{4 - y^2}/\pi$.

Problem 14.7. Fidelity between pure states is equal to the squared component of $|\phi\rangle$ expanded in a basis containing $|\psi\rangle$. Hence its probability reads $P_N(F) = (N-1)\cdot(1-F)^{N-2}$ (see Section 7.6 on random pure states).

Problem 14.8. Average fidelities, $\langle F\rangle_{\rm HS} = 1/2 + 9\pi^2/512 \approx 0.6735$ and $\langle F\rangle_{\rm B} = 1/2 + 8/(9\pi^2) \approx 0.590$ exceed the average over two random pure states, $\langle F\rangle_{\rm FS} = 1/2$ (Życzkowski and Sommers, 2005).

Problem 15.1. It is enough to observe that $|\phi\rangle = (V_1 \otimes V_1)|\phi'\rangle$. To find envariance one uses the Schmidt decomposition, $|\psi\rangle = \sum_k \sqrt{\lambda_k}|e_k\rangle \otimes |f_k\rangle$ and selects $V_1 = \sum_k {\rm e}^{{\rm i}\alpha_k}|e_k\rangle\langle e_k|$ and $V_2 = \sum_k {\rm e}^{-{\rm i}\alpha_k}|f_k\rangle\langle f_k|$ with arbitrary phases α_k (Zurek, 2003).

Problem 15.3. The Schmidt vector of the state $|\phi_+\rangle^{\otimes m}$ consists of N^m components each equal to N^{-m}. The state consisting of n copies of the initial state $|\psi\rangle$ may be, for large n, approximated by $K = \exp[nE(|\psi\rangle)]$ terms in the Schmidt decomposition described by the vector $\vec{\lambda}$. Choosing $m \approx n[E(|\psi\rangle)]/\ln N$ we see that $\{N^{-m}, \ldots, N^{-m}\} \prec \{\lambda_1, \ldots, \lambda_K\}$. Thus Nielsen's majorization theorem implies that such a conversion may be done (Nielsen, 1999). Asymptotically the reverse transformation is also possible (Bennett et al. 1996a), so for any pure statethe distillable entanglement is just equal to entanglement entropy, $E_D(|\psi\rangle) = E(|\psi\rangle)$.

Problem 15.4. Write a separable state in its eigenbasis, $\rho = \sum_j \lambda_j|\Psi_j\rangle\langle\Psi_j|$ and in its decomposition into pure product states, $\rho = \sum_i p_i|\phi_i^A\rangle\langle\phi_i^A| \otimes |\phi_i^B\rangle\langle\phi_i^B|$. Write also the partial trace $\rho_A = \sum_i p_i|\phi_i^A\rangle\langle\phi_i^A|$ in its eigenbasis, $\rho_A = \sum_k \lambda_k^A|k\rangle\langle k|$. Apply Schrödinger's mixture theorem (8.39) twice, substituting $\sqrt{p_i}|\phi_i^A\rangle = \sum_k V_{ik}\sqrt{\lambda_k^A}|k\rangle$ into $\sqrt{\lambda_j}|\Psi_j\rangle = \sum_i U_{ji}\sqrt{p_i}|\phi_i^A\rangle|\phi_i^B\rangle$, where U and V are unitary. Multiply the result by its adjoint and obtain $\lambda_j = \sum_k B_{jk}\lambda_k^A$ making use of the orthonormality, $\langle k|k'\rangle = \delta_{k,k'}$. Showing that B is bistochastic implies (15.51) due to HLP lemma.

Problem 15.6. Obviously not. A simple dimension counting will do. Consider $N \times N$ problem for which the set of separable states has $N^4 - 1$ dimensions, while the set of localy diagonalizable states forms its $(3N^2 - 3)$-dimensional subset. Note that it contains the set of all product mixed states of dimensionality $2N^2 - 2$.

Problem 15.8. This inequality follows from condition $\mathrm{Tr}\rho^2 \leq 1$. It is not sufficient for positivity, but may be accompanied by additional inequalities involving higher traces $\mathrm{Tr}\rho^k$, with $k = 3, 4, \ldots$ (Kimura, 2003; Schirmer, Zhang and Leahy, 2004).

Problem 15.9. Partial trace induces the HS measure (14.35) with $\lambda_1 = \cos^2\chi$. Changes of variables provide the required distributions, while integrations give the expectation values $\langle\chi\rangle_{\mathbb{C}\mathbf{P}^3} = 1/3$ and $\langle C\rangle_{\mathbb{C}\mathbf{P}^3} = 3\pi/16$. The distribution $P(\chi)$ achieves maximum at $\chi_m = \arccos\left[\sqrt{1/2 + 1/\sqrt{6}}\right]$, while it is most likely to find a two-qubit random pure state with concurrence $C_m = 1/\sqrt{2}$.

References

Abe, S. (1997). A note on the q-deformation-theoretic aspect of the generalized entropy in nonextensive physics, *Phys. Lett.* **A 224**: 326.

Abe, S. and Rajagopal, A. K. (2001). Non-additive conditional entropy and its significance for local realism, *Physica* **A 289**: 157.

Accardi, L. (1976). Non-relativistic quantum mechanic as a noncommutative Markov process, *Adv. Math.* **20**: 329.

Acín, A., Andrianov, A., Costa, L., Jané, E., Latorre, J. I. and Tarrach, R. (2000). Generalized Schmidt decomposition and classification of three-quantum-bit states, *Phys. Rev. Lett.* **85**: 1560.

Acín, A., Andrianov, A., Jané, E. and Tarrach, R. (2001). Three-qubit pure state canonical forms, *J. Phys.* **A 34**: 6725.

Adelman, M., Corbett, J. V. and Hurst, C. A. (1993). The geometry of state space, *Found. Phys.* **23**: 211.

Agarwal, G. (1981). Relation between atomic coherent-state representation state multipoles and generalized phase space distributions, *Phys. Rev.* **A 24**: 2889.

Akhtarshenas, S. J. and Jafarizadeh, M. A. (2003). Robustness of entanglement for Bell decomposable states, *E. Phys. J.* **D 25**: 293.

Akhtarshenas, S. J. and Jafarizadeh, M. A. (2004). Optimal Lewenstein-Sanpera decomposition for some bipartite systems, *J. Phys.* **A 37**: 2965.

Alberti, P. M. (1983). A note on the transition probability over C^*-algebras, *Lett. Math. Phys.* **7**: 25.

Alberti, P. M. and Uhlmann, A. (1981). *Dissipative Motion in State Spaces*, Leipzig: Teubner.

Alberti, P. M. and Uhlmann, A. (1982). *Stochasticity and Partial Order: Doubly Stochastic Maps and Unitary Mixing*, Reidel.

Alberti, P. M. and Uhlmann, A. (2000). On Bures distance and $*$-algebraic transition probability between inner derived positive linear forms over w^*-algebras, *Acta Appl. Math.* **60**: 1.

Albeverio, S., Chen, K. and Fei, S.-M. (2003). Generalized reduction criterion for separability of quantum states, *Phys. Rev.* **A 68**: 062313.

Alfsen, E. M. and Shultz, F. W. (2001). *State Spaces of Operator Algebras*, Boston: Birkhäuser.

Alfsen, E. M. and Shultz, F. W. (2003). *Geometry of State Spaces of Operator Algebras*, Boston: Birkhäuser.

Ali, S. T., Antoine, J.-P. and Gazeau, J.-P. (2000). *Coherent States, Wavelets and Their Generalizations*, Springer-Verlag.
Alicki, R. (1995). Comment on 'Reduced dynamics need not be completely positive', *Phys. Rev. Lett.* **75**: 3020.
Alicki, R. and Fannes, M. (2001). *Quantum Dynamical Systems*, Oxford University Press.
Alicki, R. and Fannes, M. (2004). Continuity of quantum conditional information, *J. Phys.* **A 37**: L55.
Alicki, R. and Lendi, K. (1987). *Quantum Dynamical Semigroups and Applications*, Springer-Verlag.
Alicki, R., Łoziński, A., Pakoński, P. and Życzkowski, K. (2004). Quantum dynamical entropy and decoherence rate, *J. Phys.* **A 37**: 5157.
Alon, N. and Lovasz, L. (2001). Unextendible product bases, *J. Combinat. Theor.* **A 95**: 169.
Amari, S. (1985). *Differential Geometrical Methods in Statistics*, Springer-Verlag.
Amosov, G. G., Holevo, A. S. and Werner, R. F. (n.d.). On some additivity problems in quantum information theory, Preprint math-ph/0003002.
Anandan, J. and Aharonov, Y. (1990). Geometry of quantum evolution, *Phys. Rev. Lett.* **65**: 1697.
Ando, T. (1989). Majorization, doubly stochastic matrices, and comparison of eigenvalues, *Linear Algebra Appl.* **118**: 163.
Ando, T. (1994). Majorizations and inequalities in matrix theory, *Linear Algebra Appl.* **199**: 17.
Ando, T. (2004). Cones and norms in the tensor product of matrix spaces, *Linear Algebra Appl.* **379**: 3.
Aragone, C., Gueri, G., Salamó, S. and Tani, J. L. (1974). Intelligent spin states, *J. Phys.* **A 7**: L149.
Araki, H. (1980). On a characterization of the state space of quantum mechanics, *Commun. Math. Phys.* **75**: 1.
Araki, H. and Lieb, E. (1970). Entropy inequalities, *Commun. Math. Phys.* **18**: 160.
Aravind, P. (1997). The 'twirl', stella octangula and mixed state entanglement, *Phys. Lett.* **A 233**: 7.
Aravind, P. K. (1996). Geometry of the Schmidt decomposition and Hardy's theorem, *Am. J. Phys.* **64**: 1143.
Arnold, V. I. (2000). Symplectic geometry and topology, *J. Math. Phys.* **41**: 3307.
Arrighi, P. and Patricot, C. (2003). A note on the correspondence between qubit quantum operations and special relativity, *J. Phys.* **A 36**: L287.
Arrighi, P. and Patricot, C. (2004). On quantum operations as quantum states, *Ann. Phys. (N.Y.)* **311**: 26.
Arthurs, E. and Kelly, Jr, J. L. (1965). On the simultaneous measurement of a pair of conjugate variables, *Bell Sys. Tech. J.* **44**: 725.
Arveson, W. B. (1969). Sub-algebras of C^*-algebras, *Acta Math.* **123**: 141.
Arvind, Mallesh, K. S. and Mukunda, N. (1997). A generalized Pancharatnam geometric phase formula for three-level quantum systems, *J. Phys.* **A 30**: 2417.
Ashtekar, A. and Magnon, A. (1975). Quantum fields in curved space-times, *Proc. Roy. Soc. A* **346**: 375.
Ashtekar, A. and Schilling, T. A. (1998). Geometrical formulation of quantum mechanics, *in* A. Harvey (ed.), *On Einstein's Path*, Springer, p. 23.
Asoudeh, M., Karimipour, V., Memarzadeh, L. and Rezakhani, A. T. (2004). On a suggestion relating topological and quantum mechanical entanglements, *Phys. Lett.* **A 327**: 380.

Aspect, A., Dalibard, J. and Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analysers, *Phys. Rev. Lett.* **49**: 1804.

Audenaert, K., Eisert, J., Jané, E., Plenio, M. B., Virmani, S. and Moor, B. D. (2001a). Asymptotic relative entropy of entanglement, *Phys. Rev. Lett.* **87**: 217902.

Audenaert, K., Moor, B. D., Vollbrecht, K. G. H. and Werner, R. F. (2002). Asymptotic relative entropy of entanglement for orthogonally invariant states, *Phys. Rev.* **A 66**: 032310.

Audenaert, K., Verstraete, F. and Moor, B. D. (2001b). Variational characterizations of separability and entanglement of formation, *Phys. Rev.* **A 64**: 052304.

Avron, J. E., Sadun, L., Segert, J. and Simon, B. (1989). Chern numbers, quaternions, and Berry's phases in Fermi systems, *Commun. Math. Phys.* **124**: 595.

Ay, N. and Tuschmann, W. (2002). Dually flat manifolds and global information geometry, *Open Sys. Inf. Dyn.* **9**: 195.

Bacry, H. (1974). Orbits of the rotation group on spin states, *J. Math. Phys.* **15**: 1686.

Badziąg, P., Deaur, P., Horodecki, M., Horodecki, P. and Horodecki, R. (2002). Concurrence in arbitrary dimensions, *J. Mod. Optics* **49**: 1289.

Badziąg, P., Horodecki, M., Horodecki, P. and Horodecki, R. (2000). Local environment can enhance fidelity of quantum teleportation, *Phys. Rev.* **A 62**: 012311.

Balian, R., Alhassid, Y. and Reinhardt, H. (1986). Dissipation in many-body systems: A geometric approach based on information theory, *Phys. Rep.* **131**: 1.

Ball, K. M. (1997). An elementary introduction to modern convex geometry, *in* S. Levy (ed.), *Flavors of Geometry*, Cambridge University Press, p. 1.

Bandyopadhyay, S. and Roychowdhury, V. (2002). Supercatalysis, *Phys. Rev.* **A 65**: 042306.

Barbieri, M., Martini, F. D., Nepi, G. D., Mataloni, P., D'Ariano, G. and Macciavello, C. (2003). Detection of entanglement with polarized photons: Experimental realization of an entanglement witness, *Phys. Rev. Lett.* **91**: 227901.

Bargmann, V. (1961). On a Hilbert space of analytic functions and an associated analytic transform, *Commun. Pure Appl. Math.* **14**: 187.

Bargmann, V. (1964). Note on Wigner's theorem on symmetry operations, *J. Math. Phys.* **5**: 862.

Barndorff-Nielsen, O. E. and Gill, R. D. (2000). Fisher information in quantum statistics, *J. Phys.* **A 33**: 4481.

Barnum, H. and Linden, N. (2001). Monotones and invariants for multi-particle quantum states, *J. Phys.* **A 34**: 6787.

Barnum, H., Caves, C. M., Fuchs, C. A., Jozsa, R. and Schumacher, B. (1996). Non-commuting mixed states cannot be broadcast, *Phys. Rev. Lett.* **76**: 2818.

Barnum, H., Knill, E., Ortiz, G. and Viola, L. (2003). Generalizations of entanglement based on coherent states and convex sets, *Phys. Rev.* **A 68**: 032308.

Barros e Sá, N. (2001a). Decomposition of Hilbert space in sets of coherent states, *J. Phys.* **A 34**: 4831.

Barros e Sá, N. (2001b). Uncertainty for spin systems, *J. Math. Phys.* **42**: 981.

Batle, J., Plastino, A. R., Casas, M. and Plastino, A. (2004). Inclusion relations among separability criteria, *J. Phys.* **A 37**: 895.

Beck, C. and Schlögl, F. (1993). *Thermodynamics of Chaotic Systems*, Cambridge University Press.

Belavkin, B. and Staszewski, P. (1982). c^*-algebraic generalization of relative entropy and entropy, *Ann. Inst. H. Poincaré* **A 37**: 51.

Bell, J. S. (1964). On the Einstein–Podolsky–Rosen paradox, *Physics* **1**: 195.

Bell, J. S. (1987). *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press.

Benatti, F. and Narnhofer, H. (2001). Additivity of the entanglement of formation, *Phys. Rev.* **A 63**: 042306.

Benatti, F., Floreanini, R. and Piani, M. (2004). Quantum dynamical semigroups and non-decomposable positive maps, *Phys. Lett.* **A 326**: 187.

Bendat, J. and Sherman, S. (1955). Monotone and convex operator functions, *Trans. Amer. Math. Soc.* **79**: 58.

Bengtsson, I. and Ericsson, Å. (2003). How to mix a density matrix, *Phys. Rev.* **A 67**: 012107.

Bengtsson, I., Brännlund, J. and Życzkowski, K. (2002). CP^n, or entanglement illustrated, *Int. J. Mod. Phys.* **A 17**: 4675.

Bennett, C. H., Bernstein, H. J., Popescu, S. and Schumacher, B. (1996a). Concentrating partial entanglement by local operations, *Phys. Rev.* **A 53**: 2046.

Bennett, C. H., Brassard, G., Crépeau, C., Josza, R., Peres, A. and Wootters, W. K. (1993). Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels, *Phys. Rev. Lett.* **70**: 1895.

Bennett, C. H., DiVincenzo, D. P., Fuchs, C. A., Mor, T., Rains, E., Shor, P. W., Smolin, J. A. and Wootters, W. K. (1999a). Quantum non-locality without entanglement, *Phys. Rev.* **A 59**: 1070.

Bennett, C. H., DiVincenzo, D. P., Mor, T., Shor, P. W., Smolin, J. A. and Terhal, B. M. (1999b). Unextendible product bases and bound entanglement, *Phys. Rev. Lett.* **82**: 5385.

Bennett, C. H., DiVincenzo, D. P., Smolin, J. and Wootters, W. K. (1996b). Mixed-state entanglement and quantum error correction, *Phys. Rev.* **A 54**: 3824.

Bennett, C. H., Popescu, S., Rohrlich, D., Smolin, J. A. and Thapliyal, A. V. (2001). Exact and asymptotic measures of multipartite pure-state entanglement, *Phys. Rev.* **A 63**: 012307.

Berceanu, S. (2001). Coherent states, phases and symplectic areas of geodesic triangles, *in* M. Schliechenmaier, S. T. Ali, A. Strasburger and A. Odzijewicz (eds), *Coherent States, Quantization and Gravity*, Wydawnictwa Uniwersytetu Warszawskiego, p. 129.

Bernevig, B. A. and Chen, H. D. (2003). Geometry of the 3-qubit state, entanglement and division algebras, *J. Phys.* **A 36**: 8325.

Berry, D. W. and Sanders, B. C. (2003). Bounds on general entropy measures, *J. Phys.* **A 36**: 12255.

Berry, M. V. (1984). Quantal phase factors accompanying adiabatic changes, *Proc. Roy. Soc. A* **392**: 45.

Bertrand, J. and Bertrand, P. (1987). A tomographic approach to Wigner's function, *Found. Phys.* **17**: 397.

Bhatia, R. (1997). *Matrix Analysis*, Springer-Verlag.

Bhatia, R. and Rosenthal, P. (1997). How and why to solve the operator equation AX - XB = Y, *Bull. Lond. Math. Soc.* **29**: 1.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.* **35**: 99.

Blanchard, P., Jakóbczyk, L. and Olkiewicz, R. (2001). Measures of entanglement based on decoherence, *J. Phys.* **A 34**: 8501.

Blaschke, W. and Terheggen, H. (1939). Trigonometria Hermitiana, *Rend. Sem. Mat. Univ. Roma, Ser. 4* **3**: 153.

Bloore, F. J. (1976). Geometrical description of the convex sets of states for systems with spin$-1/2$ and spin-1, *J. Phys.* **A 9**: 2059.

Bodmann, B. G. (2004). A lower bound for the Wehrl entropy of quantum spin with sharp high-spin asymptotics, *Commun. Math. Phys.* **250**: 287.
Bogomolny, E., Bohigas, O. and Lebœuf, P. (1992). Distribution of roots of random polynomials, *Phys. Rev. Lett.* **68**: 2726.
Bogomolny, E., Bohigas, O. and Lebœuf, P. (1996). Quantum chaotic dynamics and random polynomials, *J. Stat. Phys.* **85**: 639.
Born, M. and Wolf, E. (1987). *Principles of Optics*, New York: Pergamon.
Boschi, D., Branca, S., De Martini, F., Hardy, L. and Popescu, S. (1998). Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein–Podolsky–Rosen channels, *Phys. Rev. Lett.* **80**: 1121.
Bouda, J. and Bužek, V. (2002). Purification and correlated measurements of bipartite mixed states, *Phys. Rev.* **A 65**: 034304.
Bourennane, M., Eibl, M., Kurtsiefer, C., Weinfurter, H., Guehne, O., Hyllus, P., Bruß, D., Lewenstein, M. and Sanpera, A. (2004). Witnessing multipartite entanglement, *Phys. Rev. Lett.* **92**: 087902.
Bouwmeester, D., Pan, J.-W., Mattle, K., Eibl, M., Weinfurter, H. and Zeilinger, A. (1997). Experimental quantum teleportation, *Nature* **390**: 575.
Boya, L. J., Sudarshan, E. C. G. and Tilma, T. (2003). Volumes of compact manifolds, *Rep. Math. Phys.* **52**: 401.
Braun, D., Kuś, M. and Życzkowski, K. (1997). Time-reversal symmetry and random polynomials, *J. Phys.* **A 30**: L117.
Braunstein, S. L. (1996). Geometry of quantum inference, *Phys. Lett.* **A 219**: 169.
Braunstein, S. L. and Caves, C. M. (1994). Statistical distance and the geometry of quantum states, *Phys. Rev. Lett.* **72**: 3439.
Braunstein, S. L., Caves, C. M., Jozsa, R., Linden, N., Popescu, S. and Schack, R. (1999). Separability of very noisy mixed states and implications for NMR quantum computing, *Phys. Rev. Lett.* **83**: 1054.
Bravyi, S. (2004). Requirements for compatibility between local and multipartite quantum states, *Quant. Inf. Comp.* **4**: 12.
Brehm, U. (1990). The shape invariant of triangles and trigonometry in two-point homogeneous spaces, *Geometriae Dedicata* **33**: 59.
Brennen, G. K. (2003). An observable measure of entanglement for pure states of multi-qubit systems, *Quant. Inf. Comp.* **3**: 619.
Breuer, H.-P. and Petruccione, F. (2002). *The Theory of Open Quantum Systems*, Oxford University Press.
Brody, D. C. and Hughston, L. P. (2001). Geometric quantum mechanics, *J. Geom. Phys.* **38**: 19.
Brun, T. A. and Cohen, O. (2001). Parametrization and distillability of three-qubit entanglement, *Phys. Lett.* **A 281**: 88.
Bruß, D. (2002). Characterizing entanglement, *J. Math. Phys.* **43**: 4237.
Bruß, D. and Peres, A. (2000). Construction of quantum states with bound entanglement, *Phys. Rev.* **A 61**: 30301(R).
Bruß, D., Cirac, J. I., Horodecki, P., Hulpke, F., Kraus, B., Lewenstein, M. and Sanpera, A. (2002). Reflections upon separability and distillability, *J. Mod. Opt.* **49**: 1399.
Bures, D. J. C. (1969). An extension of Kakutani's theorem on infinite product measures to the tensor product of semifinite $W*$- algebras, *Trans. Am. Math. Soc.* **135**: 199.
Busch, P., Lahti, P. and Mittelstaedt, P. (1991). *The Quantum Theory of Measurement*, Springer-Verlag.
Cahill, K. and Glauber, R. (1969). Density operators and quasi-probability distributions, *Phys. Rev.* **177**: 1882.

Campbell, L. L. (1986). An extended Čencov characterization of the information metric, *Proc. Amer. Math. Soc.* **98**: 135.
Carlson, B. and Keller, J. M. (1961). Eigenvalues of density matrices, *Phys. Rev.* **121**: 659.
Carteret, H. A., Higuchi, A. and Sudbery, A. (2000). Multipartite generalization of the Schmidt decomposition, *J. Math. Phys.* **41**: 7932.
Caves, C. M. (n.d.). Measures and volumes for spheres, the probability simplex, projective Hilbert spaces and density operators, http://info.phys.unm.edu/caves/reports.
Caves, C. M., Fuchs, C. A. and Rungta, P. (2001a). Entanglement of formation of an arbitrary state of two rebits, *Found. Phys. Lett.* **14**: 199.
Caves, C. M., Fuchs, C. A. and Schack, R. (2001b). Unknown quantum states: The quantum de Finetti representation, *J. Math. Phys.* **43**: 4537.
Čencov, N. N. (1982). *Statistical Decision Rules and Optimal Inference*, American Mathematical Society.
Cerf, N. J., Adami, C. and Gingrich, R. M. (1999). Reduction criterion for separability, *Phys. Rev.* **A 60**: 898.
Cerf, N. J. and Adami, C. (1999). Quantum extension of conditional probability, *Phys. Rev.* **A 60**: 893.
Chen, K. and Wu, L.-A. (2002). The generalized partial transposition criterion for separability of multipartite quantum states, *Phys. Lett.* **A 306**: 14.
Chen, K. and Wu, L.-A. (2003). A matrix realignment method for recognizing entanglement, *Quant. Inf. Comp.* **3**: 193.
Chen, K. and Wu, L.-A. (2004). Test for entanglement using physically observable witness operators and positive maps, *Phys. Rev.* **A 69**: 022312.
Chen, P.-X., Liang, L.-M., Li, C.-Z. and Huang, M.-Q. (2002). A lower bound on entanglement of formation of $2 \otimes n$ system, *Phys. Lett.* **A 295**: 175.
Chern, S. (1979). *Complex Manifolds Without Potential Theory*, Springer-Verlag.
Cho, S.-J., Kye, S.-H. and Lee, S. G. (1992). Generalized Choi map in 3-dimensional matrix algebra, *Linear Alg. Appl.* **171**: 213.
Choi, M.-D. (1972). Positive linear maps on C^*-algebras, *Can. J. Math.* **3**: 520.
Choi, M.-D. (1975a). Completely positive linear maps on complex matrices, *Linear Alg. Appl.* **10**: 285.
Choi, M.-D. (1975b). Positive semidefinite biquadratic forms, *Linear Alg. Appl.* **12**: 95.
Choi, M.-D. (1980). Some assorted inequalities for positive maps on C^*-algebras, *J. Operator Theory* **4**: 271.
Choi, M.-D. and Lam, T. (1977). Extremal positive semidefinite forms, *Math. Ann.* **231**: 1.
Christandl, M. and Winter, A. (2004). Squashed entanglement: An additive entanglement measure, *J. Math. Phys*. **45**: 829.
Chruściński, D. and JamioŁkowski, A. (2004). *Geometric Phases in Classical and Quantum Mechanics*, Birkhäuser.
Chuang, I. L., Gershenfeld, N., Kubinec, M. G. and Leung, D. W. (1998). Bulk quantum computation with nuclear magnetic resonance: Theory and experiment, *Proc. Royal Soc. London. Ser. A* **454**: 447.
Clauser, J. F. and Shimony, A. (1978). Bell's theorem; experimental tests and implications, *Rep. Prog. Phys.* **41**: 1881.
Clauser, J. F., Horne, M. A., Shimony, A. and Holt, R. A. (1969). Proposed experiment to test local hidden-variable theories, *Phys. Rev. Lett.* **23**: 880.
Clifton, R. and Halvorson, H. (2000). Bipartite mixed states of infinite-dimensional systems are generically non-separable, *Phys. Rev.* **A 61**: 012108.
Clifton, R., Hepburn, B. and Wuthrich, C. (2002). Generic incomparability of infinite-dimensional entangled states, *Phys. Lett.* **A 303**: 121.

Coffman, V., Kundu, J. and Wootters, W. K. (2000). Distributed entanglement, *Phys. Rev.* **A 61**: 052306.
Coleman, A. and Yukalov, V. I. (2000). *Reduced Density Matrices*, Springer.
Coleman, A. J. (1963). The structure of fermion density matrices, *Rev. Mod. Phys.* **35**: 668.
Collins, D. and Popescu, S. (2002). Classical analog of entanglement, *Phys. Rev.* **A 65**: 032321.
Constantinescu, T. and Ramakrishna, V. (2003). Parametrizing quantum states and channels, *Quant. Inf. Proc.* **2**: 221.
Cortese, J. (n.d.). Relative entropy and single qubit Holevo–Schumacher–Westmoreland channel capacity, Preprint quant-ph/0207128.
Cory, D. G., Fahmy, A. F. and Havel, T. F. (1997). Ensemble quantum computing by NMR spectroscopy, *Proc. Natl. Acad. Sci. USA* **94**: 1634.
Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*, Wiley.
Cromwell, P. (1997). *Polyhedra*, Cambridge University Press.
Currie, D. G., Jordan, T. F. and Sudarshan, E. C. G. (1963). Relativistic invariance and Hamiltonian theories of interacting particles, *Rev. Mod. Phys.* **35**: 350.
Dąbrowski, L. and Jadczyk, A. (1989). Quantum statistical holonomy, *J. Phys.* **A 22**: 3167.
Daftuar, S. and Klimesh, M. (2001). Mathematical structure of entanglement catalysis, *Phys. Rev.* **A 64**: 042314.
Davies, E. B. (1976). *Quantum Theory of Open Systems*, London: Academic Press.
Davis, R. I. A., Delbourgo, R. and Jarvis, P. D. (2000). Covariance, correlation and entanglement, *J. Phys.* **A 33**: 1895.
Delbourgo, R. (1977). Minimal uncertainty states for the rotation group and allied groups, *J. Phys.* **A 10**: 1837.
Delbourgo, R. and Fox, J. R. (1977). Maximum weight vectors possess minimal uncertainty, *J. Phys.* **A 10**: L233.
Delbrück, M. and Molière, G. (1936). Statistische Quantenmechanik und Thermodynamik, *Ab. Preuss. Akad. Wiss.* **1**: 1.
DeMarco, B., Ben-Kish, A., Leibfried, D., Meyer, V., Rowe, M., Jelenkovic, B. M., Itano, W. M., Britton, J., Langer, C., Rosenband, T. and Wineland, D. (2002). Experimental demonstration of a controlled-NOT wave-packet gate, *Phys. Rev. Lett.* **89**: 267901.
Deuar, P., Munro, W. J. and Nemoto, K. (2000). Upper bound on the region of separable states near the maximally mixed state, *J. Opt.* **B 2**: 225.
Devetak, I. and Winter, A. (2005). Distillation of secret key and entanglement from quantum states, *Proc. Royal Soc. Lond. A Math.-Phys.* **461**: 207.
Dittmann, J. (1995). On the Riemannian metric on the space of density matrices, *Rep. Math. Phys.* **36**: 309.
Dittmann, J. (1999a). Explicit formulae for the Bures metric, *J. Phys.* **A 32**: 2663.
Dittmann, J. (1999b). The scalar curvature of the Bures metric on the space of density matrices, *J. Geom. Phys.* **31**: 16.
DiVincenzo, D., Fuchs, C. A., Mabuchi, H., Smolin, J. A., Thapliyal, A. and Uhlmann, A. (1999). Entanglement of assistance, *Lecture Notes Comp. Sci.* **1509**: 247.
DiVincenzo, D. P. (1995). Two-bit gates are universal for quantum computation, *Phys. Rev.* **A 50**: 1015.
DiVincenzo, D. P., Mor, T., Shor, P. W., Smolin, J. A. and Terhal, B. M. (2003). Unextendible product bases, uncompletable product bases and bound entanglement, *Comm. Math. Phys.* **238**: 379.
DiVincenzo, D. P., Shor, P. W., Smolin, J. A., Terhal, B. M. and Thapliyal, A. (2000a). Evidence for bound entangled states with negative partial transpose, *Phys. Rev.* **A 61**: 062312.

DiVincenzo, D., Terhal, B. and Thapliyal, A. V. (2000b). Optimal decompositions of barely separable states, *J. Mod. Opt.* **47**: 377.
Doherty, A. C., Parillo, P. A. and Spedalieri, F. M. (2002). Distinguishing separable and entangled states, *Phys. Rev. Lett.* **88**: 187904.
Doherty, A. C., Parillo, P. A. and Spedalieri, F. M. (2004). Complete family of separability criteria, *Phys. Rev.* **A 69**: 022308.
Donald, M. J. (1987). Further results on the relative entropy, *Math. Proc. Cam. Phil. Soc.* **101**: 363.
Donald, M. J. and Horodecki, M. (1999). Continuity of relative entropy of entanglement, *Phys. Lett.* **A 264**: 257.
Donald, M. J., Horodecki, M. and Rudolph, O. (2002). The uniqueness theorem for entanglement measures, *J. Math. Phys.* **43**: 4252.
Dowling, J. P., Agarwal, G. S. and Schleich, W. P. (1994). Wigner function of a general angular momentum state: Applications to a collection of two-level atoms, *Phys. Rev.* **A 49**: 4101.
Dür, W. and Cirac, J. I. (2002). Equivalence classes of non-local operations, *Q. Inform. Comput.* **2**: 240.
Dür, W., Cirac, J. I., Lewenstein, M. and Bruß, D. (2000a). Distillability and partial transposition in bipartite systems, *Phys. Rev.* **A 61**: 062313.
Dür, W., Vidal, G. and Cirac, J. I. (2000b). Three qubits can be entangled in two inequivalent ways, *Phys. Rev.* **A 62**: 062314.
Dyson, F. (1962). The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics, *J. Math. Phys.* **3**: 1199.
Edelman, A. and Kostlan, E. (1995). How many zeros of a random polynomial are real?, *Bull. Amer. Math. Soc.* **32**: 1.
Eggeling, T., Vollbrecht, K., Werner, R. and Wolf, M. (2001). Distillability via protocols respecting the positivity of partial transpose, *Phys. Rev. Lett.* **87**: 257902.
Eggleston, H. G. (1958). *Convexity*, Cambridge University Press.
Einstein, A. (1949). Autobiographical notes, *in* P. A. Schilpp (ed.), *Albert Einstein: Philosopher – Scientist*, Cambridge University Press, p. 1.
Einstein, A., Podolsky, B. and Rosen, N. (1935). Can quantum mechanical description of reality be considered complete?, *Phys. Rev.* **47**: 777.
Eisert, J. (n.d.). Entanglement in quantum information theory, Ph.D. Thesis, University of Potsdam, 2001.
Eisert, J. and Briegel, H. J. (2001). Schmidt measure as a tool for quantifying multiparticle entanglement, *Phys. Rev.* **A 64**: 022306.
Eisert, J. and Plenio, M. B. (1999). A comparison of entanglement measures, *J. Mod. Opt.* **46**: 145.
Eisert, J., Audenaert, K. and Plenio, M. B. (2003). Remarks on entanglement measures and non-local state distinguishability, *J. Phys.* **A 36**: 5605.
Eisert, J., Hyllus, P., Guhne, O. and Curty, M. (2004). Complete hierarchies of efficient approximations to problems in entanglement theory, *Phys. Rev.* **A 70**: 062317.
Ekert, A. and Knight, P. L. (1995). Entangled quantum systems and the Schmidt decomposition, *Am. J. Phys.* **63**: 415.
Emch, G. G. (1972). *Algebraic Methods in Statistical Mechanics and Quantum Field Theory*, Wiley Interscience.
Endres, D. M. and Schindelin, J. E. (2003). A new metric for probability distributions, *IEEE Trans. Inf. Theory* **49**: 1858.
Englert, B.-G. and Metwally, N. (2001). Remarks on 2-q-bit states, *Appl. Phys.* **B 72**: 35.

Eom, M.-H. and Kye, S.-H. (2000). Duality for positive linear maps in matrix algebras, *Math. Scand.* **86**: 130.
Ericsson, A. (2002). Separability and the stella octangula, *Phys. Lett.* **A 295**: 256.
Ericsson, M., Sjöqvist, E., Brännlund, J., Oi, D. K. L. and Pati, A. K. (2003). Generalization of the geometric phase to completely positive maps, *Phys. Rev.* **A 67**: 020101(R).
Evans, D. E. (1984). Quantum dynamical semigroups, *Acta Appl. Math.* **2**: 333.
Everett, H. (1957). 'Relative state' formulation of quantum mechanics, *Rev. Mod. Phys.* **29**: 454.
Ewald, G. (1996). *Combinatorial Convexity and Algebraic Geometry*, New York: Springer.
Faddejew, D. K. (1957). Zum Begriff der Entropie eines endlichen Wahrscheinlichkeitsschemas, *in* H. Grell (ed.), *Arbeiten zur Informationstheorie I*, Deutscher Verlag der Wissenschaften, p. 86.
Fannes, M. (1973). A continuity property of the entropy density for spin lattices, *Commun. Math. Phys.* **31**: 291.
Fano, U. (1983). Pairs of two-level systems, *Rev. Mod. Phys.* **55**: 855.
Fisher, R. A. (1925). Theory of statistical estimation, *Proc. Cambr. Phil. Soc.* **22**: 700.
Fiurášek, J. (2002). Structural physical approximations of unphysical maps and generalized quantum measurements, *Phys. Rev.* **A 66**: 052315.
Foong, S. K. and Kanno, S. (1994). Proof of Page's conjecture on the average entropy of a subsystem, *Phys. Rev. Lett.* **72**: 1148.
Forrester, P. (2006). *Log-gases and Random Matrices*, Princeton University Press, to appear. see www.ms.unimelb.edu.au/∼matpjf.
Fubini, G. (1903). Sulle metriche definite da una forma Hermitiana, *Atti Instituto Veneto* **6**: 501.
Fuchs, C. A. (1996). *Distinguishability and Accessible Information in Quantum Theory*, PhD Thesis, Univ. of New Mexico, Albuquerque.
Fuchs, C. A. and Caves, C. M. (1995). Mathematical techniques for quantum communication theory, *Open Sys. Inf. Dyn* **3**: 345.
Fuchs, C. A. and van de Graaf, J. (1999). Cryptographic distinguishability measures for quantum mechanical states, *IEEE Trans. Inf. Th.* **45**: 1216.
Fuchs, J. and Schweigert, C. (2003). *Symmetries, Lie Algebras and Representations*, Cambridge University Press.
Fujiwara, A. and Algoet, P. (1999). One-to-one parametrization of quantum channels, *Phys. Rev.* **A 59**: 3290.
Gerjuoy, E. (2003). Lower bound on entanglement of formation for the qubit–qudit system, *Phys. Rev.* **A 67**: 052308.
Gibbons, G. W. (1992). Typical states and density matrices, *J. Geom. Phys.* **8**: 147.
Gibbons, G. W. and Pohle, H. (1993). Complex numbers, quantum mechanics and the beginning of time, *Nucl. Phys.* **B 410**: 375.
Gibilisco, P. and Isola, T. (2003). Wigner–Yanase information on quantum state space: The geometric approach, *J. Math. Phys.* **44**: 3752.
Gill, R. (2003). Time, finite statistics, and Bell's fifth position, *in* A. Khrennikov (ed.), *Foundations of Probability and Physics – 2*, Växjö University Press, p. 179.
Gilmore, R. (1974). *Lie Groups, Lie Algebras, and Some of Their Applications*, New York: Wiley.
Gingrich, R. M. (2002). Properties of entanglement monotones for three-qubit pure states, *Phys. Rev.* **A 65**: 052302.
Ginibre, J. (1965). Statistical ensembles of complex quaternion and real matrices, *J. Math. Phys.* **6**: 440.

Gisin, N. (1989). Stochastic quantum dynamics and relativity, *Helv. Phys. Acta* **62**: 363.
Gisin, N. (1991). Bell inequality holds for all non-product states, *Phys. Lett.* **A 154**: 201.
Gitman, D. and Shelepin, A. L. (1993). Coherent states of SU(N) groups, *J. Phys.* **A 26**: 313.
Glauber, R. J. (1963). Coherent and incoherent states of the radiation field, *Phys. Rev.* **131**: 2766.
Gleason, A. M. (1957). Measures on the closed subspaces of a Hilbert space, *J. Math. Mech.* **6**: 885.
Gnutzmann, S. and Kuś, M. (1998). Coherent states and the classical limit on irreducible SU(3) representations, *J. Phys.* **A 31**: 9871.
Gnutzmann, S. and Życzkowski, K. (2001). Rényi-Wehrl entropies as measures of localization in phase space, *J. Phys.* **A 34**: 10123.
Gnutzmann, S., Haake, F. and Kuś, M. (2000). Quantum chaos of $SU(3)$ observables, *J. Phys.* **A 33**: 143.
Golden, S. (1965). Lower bounds for the Helmholtz function, *Phys. Rev.* **B 137**: 1127.
Gorini, V., Kossakowski, A. and Sudarshan, E. C. G. (1976). Completely positive dynamical semigroups of n-level systems, *J. Math. Phys.* **17**: 821.
Gour, G. (2004). Family of concurrence monotones and its applications, *Phys. Rev.* **A 71**: 012318.
Grasseli, M. R. and Streater, R. F. (2001). On the uniqueness of the Chentsov metric in quantum information geometry, *Inf. Dim. Analysis, Quantum Prob.* **4**: 173.
Grassl, M., Rötteler, M. and Beth, T. (1998). Computing local invariants of qubit systems, *Phys. Rev.* **A 58**: 1833.
Greenberger, D. M., Horne, M. and Zeilinger, A. (1989). Going beyond Bell's theorem, *in* M. Kafatos (ed.), *Bell's theorem, quantum theory and conceptions of the Universe*, Dordrecht: Kluwer, p. 69.
Griffiths, J. and Harris, J. (1978). *Principles of Algebraic Geometry*, Wiley.
Gruska, J. (1999). *Quantum Computing*, New York: McGraw-Hill.
Gühne, O. and Lewenstein, M. (2004). Entropic uncertainty relations and entanglement, *Phys. Rev.* **A 70**: 022316.
Gühne, O., Hyllus, P., Bruß, D., Ekert, A., Lewenstein, M., Macchiavello, C. and Sanpera, A. (2002). Detection of entanglement with few local measurements, *Phys. Rev.* **A 66**: 062305.
Gühne, O., Hyllus, P., Bruß, D., Ekert, A., Lewenstein, M., Macchiavello, C. and Sanpera, A. (2003). Experimental detection of entanglement via witness operators and local measurements, *J. Mod. Opt.* **50**: 1079.
Gürsey, F. and Tze, H. C. (1980). Complex and quaternionic analyticity in chiral and gauge theories, I, *Ann. of Phys.* **128**: 29.
Gurvits, L. (2003). Classical deterministic complexity of Edmonds' problem and quantum entanglement, *Proceedings of the 35-th ACM Symposium on Theory of Computing*, New York: ACM Press, p. 10.
Gurvits, L. (2004). Classical complexity and quantum entanglement, *J. Comput. Syst. Sciences* **69**: 448.
Gurvits, L. and Barnum, H. (2002). Largest separable balls around the maximally mixed bipartite quantum state, *Phys. Rev.* **A 66**: 062311.
Gurvits, L. and Barnum, H. (2003). Separable balls around the maximally mixed multipartite quantum states, *Phys. Rev.* **A 68**: 042312.
Gurvits, L. and Barnum, H. (2004). Better bound on the exponent of the radius of the multipartite separable ball, *Phys. Rev*. **A 72**: 032322.

Ha, K.-C. (1998). Atomic positive linear maps in matrix algebra, *Publ. RIMS Kyoto Univ.* **34**: 591.

Ha, K.-C., Kye, S.-H. and Park, Y.-S. (2003). Entangled states with positive partial transposes arising from indecomposable positive linear maps, *Phys. Lett.* **A 313**: 163.

Haake, F. (2001). *Quantum Signatures of Chaos, 2nd. ed.*, Springer.

Hall, M. J. W. (1998). Random quantum correlations and density operator distributions, *Phys. Lett.* **A 242**: 123.

Hammerer, K., Vidal, G. and Cirac, J. I. (2002). Characterization of non-local gates, *Phys. Rev.* **A 66**: 062321.

Hangan, T. and Masala, G. (1994). A geometrical interpretation of the shape invariant geodesic triangles in complex projective spaces, *Geometriae Dedicata* **49**: 129.

Hannay, J. H. (1996). Exact statistics for $2j$ points on a sphere: The Majorana zeros for random spin state, *J. Phys.* **A 29**: L101.

Hannay, J. H. (1998). The Berry phase for spin in the Majorana representation, *J. Phys.* **A 31**: L53.

Hardy, G. H. and Ramanujan, S. S. (1918). Asymptotic formulae in combinatory analysis, *Proc. Lond. Math. Soc.* **17**: 75.

Hardy, G. H., Littlewood, J. E. and Pólya, G. (1929). Some simple inequalities satisfied by convex functions, *Messenger Math.* **58**: 145.

Hardy, L. (1999). Method of areas for manipulating the entanglement properties of one copy of a two-particle pure entangled state, *Phys. Rev.* **A 60**: 1912.

Hardy, L. (n.d.). Quantum theory from five reasonable axioms, Preprint quant-ph/0101012.

Harremoës, P. and Topsøe, F. (2001). Inequalities between entropy and index of coincidence derived from information diagrams, *IEEE Trans. Inform. Theory* **47**: 2944.

Harriman, J. E. (1978). Geometry of density matrices 1, *Phys. Rev.* **A 17**: 1249.

Harris, J. (1992). *Algebraic Geometry. A First Course*, Springer.

Hartley, R. V. L. (1928). Transmission of information, *Bell Sys. Tech. J.* **7**: 535.

Havel, T. F. (2003). Robust procedures for converting among Lindblad, Kraus and matrix representations of quantum dynamical semigroups, *J. Math. Phys.* **44**: 534.

Havel, T. F., Sharf, Y., Viola, L. and Cory, D. G. (2001). Hadamard products of product operators and the design of gradient: Diffusion experiments for simulating decoherence by NMR spectroscopy, *Phys. Lett.* **A 280**: 282.

Havrda, J. and Charvát, F. (1967). Quantification methods of classification processes: Concept of structural α entropy, *Kybernetica* **3**: 30.

Hayden, P., Leung, D. W. and Winter, A. (n.d.a). Aspects of generic entanglement, Preprint quant-ph/0407049.

Hayden, P. M., Horodecki, M. and Terhal, B. M. (2001). The asymptotic entanglement cost of preparing a quantum state, *J. Phys.* **A 34**: 6891.

Hayden, P., Terhal, B. M. and Uhlmann, A. (n.d.b). On the LOCC classification of bipartite density matrices, Preprint quant-ph/0011095.

Hein, M., Eisert, J. and Briegel, H. J. (2004). Multiparty entanglement in graph states, *Phys. Rev.* **A 69**: 062311.

Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory*, London: Academic Press. Mathematics in Science and Engineering vol. 123.

Henderson, L. and Vedral, V. (2000). Information, relative entropy of entanglement and irreversibility, *Phys. Rev. Lett.* **84**: 2263.

Herbert, N. (1982). FLASH—A superluminal communicator based upon a new kind of quantum measurement, *Found. Phys.* **12**: 1171.

Heydari, H. and Björk, G. (2004). Entanglement measure for general pure multipartite quantum states, *J. Phys.* **A 37**: 9251.
Hiai, F. and Petz, D. (1991). The proper formula for relative entropy and its asymptotics in quantum probability, *Commun. Math. Phys.* **143**: 99.
Hiai, F., Ohya, M. and Tsukada, M. (1981). Sufficiency, KMS condition and relative entropy in von Neumann algebras, *Pacific J. Math.* **96**: 99.
Hildebrand, R. (n.d.). PPT from spectra, Preprint quant-ph/0502170.
Hill, S. and Wootters, W. K. (1997). Entanglement of a pair of quantum bits, *Phys. Rev. Lett.* **78**: 5022.
Holevo, A. S. (1973). Information theoretical aspects of quantum measurements, *Prob. Inf. Transmission USSR* **9**: 177.
Holevo, A. S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory*, North-Holland.
Holevo, A. S. (2001). *Statistical Structure of Quantum Theory*, Springer.
Horn, A. (1954). Doubly stochastic matrices and the diagonal of rotation, *Am. J. Math.* **76**: 620.
Horn, R. and Johnson, C. (1985). *Matrix Analysis*, Cambridge University Press.
Horn, R. and Johnson, C. (1991). *Topics in Matrix Analysis*, Cambridge University Press.
Horodecki, K., Horodecki, M., Horodecki, P. and Oppenheim, J. (2005). Secure key from bound entanglement, *Phys. Rev. Lett.* **94**: 160502.
Horodecki, M. (2001). Entanglement measures, *Quant. Inf. Comp.* **1**: 3.
Horodecki, M. and Horodecki, P. (1999). Reduction criterion of separability and limits for a class of distillation protocols, *Phys. Rev.* **A 59**: 4206.
Horodecki, M., Horodecki, P. and Horodecki, R. (1996a). Separability of mixed states: Necessary and sufficient conditions, *Phys. Lett.* **A 223**: 1.
Horodecki, M., Horodecki, P. and Horodecki, R. (1998). Mixed-state entanglement and distillation: Is there a 'bound' entanglement in nature?, *Phys. Rev. Lett.* **80**: 5239.
Horodecki, M., Horodecki, P. and Horodecki, R. (2000a). Limits for entanglement measures, *Phys. Rev. Lett.* **84**: 2014.
Horodecki, M., Horodecki, P. and Horodecki, R. (2000b). Mixed-state entanglement and quantum communication, *in* G. Alber and M. Weiner (eds), *Quantum Information – Basic Concepts and Experiments*, Springer, p. 1.
Horodecki, M., Horodecki, P. and Horodecki, R. (n.d.b). Separability of mixed quantum states: Linear contractions approach, Preprint quant-ph/0206008.
Horodecki, M., Horodecki, P. and Oppenheim, J. (2003a). Reversible transformations from pure to mixed states and the unique measure of information, *Phys. Rev.* **A 67**: 062104.
Horodecki, M., Shor, P. W. and Ruskai, M. B. (2003b). Entanglement breaking channels, *Rev. Math. Phys.* **15**: 621.
Horodecki, P. (1997). Separability criterion and inseparable mixed states with positive partial transposition, *Phys. Lett.* **A 232**: 333.
Horodecki, P. and Ekert, A. (2002). Method for direct detection of quantum entanglement, *Phys. Rev. Lett.* **89**: 127902.
Horodecki, P., Horodecki, M. and Horodecki, R. (1999). Bound entanglement can be activated, *Phys. Rev. Lett.* **82**: 1056.
Horodecki, P., Lewenstein, M., Vidal, G. and Cirac, I. (2000c). Operational criterion and constructive checks for the separability of low rank density matrices, *Phys. Rev.* **A 62**: 032310.
Horodecki, P., Smolin, J. A., Terhal, B. M. and Thapliyal, A. V. (2003c). Rank two bipartite bound entangled states do not exist, *Theor. Comp. Sci.* **292**: 589.

Horodecki, R. and Horodecki, M. (1996). Information-theoretic aspects of quantum inseparability of mixed states, *Phys. Rev.* **A 54**: 1838.
Horodecki, R. and Horodecki, P. (1994). Quantum redundancies and local realism, *Phys. Lett.* **A 194**: 147.
Horodecki, R., Horodecki, P. and Horodecki, M. (1996b). Quantum α-entropy inequalities: Independent condition for local realism?, *Phys. Lett.* **A 210**: 377.
Hua, L. K. (1963). *Harmonic Analysis of Functions of Several Variables in the Classical Domains*, American Mathematical Society. Chinese original 1958, Russian translation, Moskva 1959.
Hübner, M. (1992). Explicit computation of the Bures distance for density matrices, *Phys. Lett.* **A 163**: 239.
Hudson, R. (1974). When is the Wigner quasi-probability density non-negative?, *Rep. Math. Phys.* **6**: 249.
Hughston, L. P. (1995). Geometric aspects of quantum mechanics, *in* S. Huggett (ed.), *Twistor Theory*, S. Marcel Dekker.
Hughston, L. P., Jozsa, R. and Wootters, W. K. (1993). A complete classification of quantum ensembles having a given density matrix, *Phys. Lett.* **A 183**: 14.
Hulpke, F. and Bruß, D. (2005). A two-way algorithm for the entanglement problem, *Phys. Rev.* **A 72**: 012321.
Hunziker, W. (1972). A note on symmetry operations in quantum mechanics, *Helv. Phys. Acta* **45**: 233.
Husimi, K. (1940). Some formal properties of density matrices, *Proc. Phys. Math. Soc. Japan* **22**: 264.
Hyllus, P., Gühne, O., Bruß, D. and Lewenstein, M. (2005). Relations between entanglement witnesses and Bell inequalities, Preprint quant-ph/0504079.
Ingarden, R., Kossakowski, A. and Ohya, M. (1997). *Information Dynamics and Open Systems*, Dordrecht: Kluwer.
Ingarden, R. S. (1981). Information geometry in functional spaces of classical and quantum finite statistical systems, *Int. J. Engng. Sci.* **19**: 1609.
Ishizaka, S. (2003). Analytical formula connecting entangled state and the closest disentangled state, *Phys. Rev.* **A 67**: 060301.
Ishizaka, S. and Hiroshima, T. (2000). Maximally entangled mixed states under non-local unitary operations in two qubits, *Phys. Rev.* **A 62**: 022310.
Ivanović, I. D. (1981). Geometrical description of quantal state determination, *J. Phys.* **A 14**: 3241.
Jakóbczyk, L. and Siennicki, M. (2001). Geometry of Bloch vectors in two-qubit system, *Phys. Lett.* **A 286**: 383.
Jamiołkowski, A. (1972). Linear transformations which preserve trace and positive semi-definiteness of operators, *Rep. Math. Phys.* **3**: 275.
Jamiołkowski, A. (1975). An effective method of investigation of positive maps on the set of positive definite operators, *Rep. Math. Phys.* **5**: 415.
Janik, R. A. and Nowak, M. A. (2003). Wishart and anti-Wishart random matrices, *J. Phys.* **A 36**: 3629.
Jauch, J. M. (1968). *Foundations of Quantum Mechanics*, Addison-Wesley.
Jauch, J. M. and Piron, C. (1967). Generalized localizability, *Helv. Phys. Acta* **45**: 559.
Jaynes, E. T. (2003). *Probability Theory. The Logic of Science*, Cambridge University Press.
Jeffreys, H. (1961). *Theory of Probability*, Oxford: Clarendon Press.
Jenčova, A. (2004). Generalized relative entropies as contrast functionals on density matrices, *Int. J. Theor. Phys.* **43**: 1635.

Jensen, J. G. and Schack, R. (2001). A simple algorithm for local conversion of pure states, *Phys. Rev.* **A 63**: 062303.
Jonathan, D. and Plenio, M. B. (1999a). Entanglement-assisted local manipulation of pure quantum states, *Phys. Rev. Lett.* **83**: 3566.
Jonathan, D. and Plenio, M. B. (1999b). Minimial conditions for local pure-state entanglement manipulation, *Phys. Rev. Lett.* **83**: 1455.
Jones, K. R. W. (1990). Entropy of random quantum states, *J. Phys.* **A 23**: L1247.
Jordan, P., von Neumann, J. and Wigner, E. P. (1934). On an algebraic generalization of the quantum mechanical formalism, *Ann. Math.* **56**: 29.
Jozsa, R. (1994). Fidelity for mixed quantum states, *J. Mod. Opt.* **41**: 2315.
Jozsa, R. and Schumacher, B. (1994). A new proof of the quantum noiseless coding theorem, *J. Mod. Opt.* **41**: 2343.
Jozsa, R., Robb, D. and Wootters, W. K. (1994). Lower bound for accessible information in quantum mechanics, *Phys. Rev.* **A 49**: 668.
Kac, M. (1943). On the average number of real roots of a random algebraic equation, *Bull. Amer. Math. Soc.* **49**: 314.
Kantorovich, L. V. (1942). On translocation of masses, *Doklady Acad. Sci. URSS* **37**: 199.
Kapur, J. (1994). *Measures of Information and Their Applications*, New York: John Wiley & Sons.
Karnas, S. and Lewenstein, M. (2001). Separable approximations of density matrices of composite quantum systems, *J. Phys.* **A 34**: 6919.
Kauffmann, L. H. and Lomonaco Jr., S. J. (2002). Comparing quantum entanglement and topological entanglement, *New. J. Phys.* **4**: 73.1.
Kendon, V. M., Życzkowski, K. and Munro, W. J. (2002). Bounds on entanglement in qudit subsystems, *Phys. Rev.* **A 66**: 062310.
Keyl, M. (2002). Fundamentals of quantum information theory, *Phys. Rep.* **369**: 431.
Khaneja, N., Brockett, R. and Glaser, S. J. (2001). Time optimal control in spin systems, *Phys. Rev.* **A 63**: 0323308.
Khinchin, A. I. (1957). *Mathematical foundations of information theory*, Dover.
Kibble, T. W. B. (1979). Geometrization of quantum mechanics, *Commun. Math. Phys.* **65**: 189.
Kim, H.-J. and Kye, S.-H. (1994). Indecomposable extreme positive linear maps in matrix algebras, *Bull. London. Math. Soc.* **26**: 575.
Kimura, G. (2003). The Bloch vector for n-level systems, *Phys. Lett.* **A 314**: 339.
King, C. and Ruskai, M. B. (2001). Minimal entropy of states emerging from noisy channels, *IEEE Trans. Inf. Th.* **47**: 192.
Klauder, J. R. (1960). The action option and a Feynman quantization of spinor fields in terms of ordinary c-numbers, *Ann. Phys. (N.Y.)* **11**: 60.
Klauder, J. R. and Skagerstam, B.-S. S. (1985). *Coherent States. Applications in Physics and Mathematical Physics*, Singapore: World Scientific.
Klauder, J. R. and Sudarshan, E. C. G. (1968). *Fundamentals of Quantum Optics*, New York: W. A. Benjamin.
Klein, O. (1931). Zur quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre, *Z. f. Physik* **72**: 767.
Klyachko, A. (n.d.). Quantum marginal problem and representations of the symmetric group, Preprint quant-ph/0409113.
Knill, E. and Laflamme, R. (1998). Power of one bit of quantum information, *Phys. Rev. Lett.* **81**: 5672.
Knutson, A. (2000). The symplectic and algebraic geometry of Horn's problem, *Linear Algebra Appl.* **319**: 61.

Kobayashi, S. and Nomizu, K. (1963). *Foundations of Differential Geometry I*, Wiley Interscience.
Korsch, H. J., Müller, C. and Wiescher, H. (1997). On the zeros of the Husimi distribution, *J. Phys.* **A 30**: L677.
Kossakowski, A. (2000). Remarks on positive maps of finite-dimensional simple Jordan algebras, *Rep. Math. Phys.* **46**: 393.
Kossakowski, A. (2003). A class of linear positive maps in matrix algebras, *Open Sys. & Information Dyn.* **10**: 1.
Kraus, B. and Cirac, J. I. (2001). Optimal creation of entanglement using a two-qubit gate, *Phys. Rev.* **A 63**: 062309.
Kraus, K. (1971). General state changes in quantum theory, *Ann. Phys.* **64**: 311.
Kraus, K. (1983). *States, Effects and Operations: Fundamental Notions of Quantum Theory*, Springer-Verlag.
Kubo, F. and Ando, T. (1980). Means of positive linear operators, *Math. Ann.* **246**: 205.
Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Ann. Math. Stat.* **22**: 79.
Kuś, M. and Życzkowski, K. (2001). Geometry of entangled states, *Phys. Rev.* **A 63**: 032307.
Kuś, M., Mostowski, J. and Haake, F. (1988). Universality of eigenvector statistics of kicked tops of different symmetries, *J. Phys.* **A 21**: L1073.
Kye, S.-H. (2003). Facial structures for unital positive linear maps in the two-dimensional matrix algebra, *Linear Algebra Appl.* **362**: 57.
Laflamme, R., Cory, D., Negrevergne, C. and Viola, L. (2002). NMR quantum information processing and entanglement, *Quantum Inform. Compu.* **2**: 166.
Landau, L. J. and Streater, R. F. (1993). On Birkhoff's theorem for doubly stochastic completely positive maps of matrix algebras, *Linear Algebra Appl.* **193**: 107.
Lanford, O. and Robinson, D. (1968). Mean entropy of states in quantum statistical ensembles, *J. Math. Phys.* **9**: 1120.
Lebœuf, P. (1991). Phase space approach to quantum dynamics, *J. Phys.* **A 24**: 4575.
Lebœuf, P. and Shukla, P. (1996). Universal fluctuations of zeros of chaotic wavefunctions, *J. Phys.* **A 29**: 4827.
Lebœuf, P. and Voros, A. (1990). Chaos-revealing multiplicative representation of quantum eigenstates, *J. Phys.* **A 23**: 1765.
Lee, C. T. (1988). Wehrl's entropy of spin states and Lieb's conjecture, *J. Phys.* **A 21**: 3749.
Lee, S., Chi, D. P., Oh, S. D. and Kim, J. (2003). Convex-roof extended negativity as an entanglement measure for bipartite quantum systems, *Phys. Rev.* **A 68**: 062304.
Leonhardt, U. (1997). *Measuring the Quantum State of Light*, Cambridge University Press.
Lesniewski, A. and Ruskai, M. B. (1999). Monotone Riemannian metrics and relative entropy on non-commutative probability spaces, *J. Math. Phys.* **40**: 5702.
Lévay, P. (2004). The geometry of entanglement: Metrics, connections and the geometric phase, *J. Phys.* **A 37**: 1821.
Lewenstein, M. and Sanpera, A. (1998). Separability and entanglement of composite quantum systems, *Phys. Rev. Lett.* **80**: 2261.
Lewenstein, M., Bruß, D., Cirac, J. I., Kraus, B., Kuś, M., Samsonowicz, J., Sanpera, A. and Tarrach, R. (2000a). Separability and distillability in composite quantum systems: A primer, *J. Mod. Opt.* **47**: 2841.
Lewenstein, M., Kraus, B., Cirac, J. I. and Horodecki, P. (2000b). Optimization of entanglement witnesses, *Phys. Rev.* **A 62**: 052310.

Lieb, E. H. (1973). Convex trace functions and the Wigner–Yanase–Dyson conjecture, *Adv. Math.* **11**: 267.
Lieb, E. H. (1975). Some convexity and subadditivity properties of entropy, *Bull. AMS* **81**: 1.
Lieb, E. H. (1978). Proof of an entropy conjecture of Wehrl, *Commun. Math. Phys.* **62**: 35.
Lieb, E. H. and Ruskai, M. B. (1973). Proof of the strong subadditivity of quantum mechanical entropy, *J. Math. Phys.* **14**: 1938.
Lin, J. (1991). Divergence measures based on the Shannon entropy, *IEEE Trans. Inf. Theory* **37**: 145.
Lindblad, G. (1973). Entropy, information and quantum measurement, *Comm. Math. Phys.* **33**: 305.
Lindblad, G. (1974). Expectations and entropy inequalities for finite quantum systems, *Commun. Math. Phys.* **39**: 111.
Lindblad, G. (1975). Completely positive maps and entropy inequalities, *Commun. Math. Phys.* **40**: 147.
Lindblad, G. (1976). On the generators of quantum dynamical semigroups, *Commun. Math. Phys.* **48**: 119.
Lindblad, G. (1991). Quantum entropy and quantum measurements, *Lecture Notes in Physics* **378**: 36. Quantum Aspects of Optical Communication, ed. C. Bendjaballah et al.
Linden, N. and Winter, A. (n.d.). A new inequality for the von Neumann entropy, Preprint quant-ph/0406162.
Lloyd, S. (1995). Almost any quantum logic gate is universal, *Phys. Rev. Lett.* **75**: 346.
Lloyd, S. and Pagels, H. (1988). Complexity as thermodynamic depth, *Ann. Phys. (N.Y.)* **188**: 186.
Lo, H.-K. and Popescu, S. (1998). Concentrating entanglement by local actions: Beyond mean values, *Phys. Rev.* **A 63**: 022301.
Lockhart, R. (2000). Optimal ensemble length of mixed separable states, *J. Math. Phys.* **41**: 6766.
Lockhart, R. B. and Steiner, M. J. (2002). Preserving entanglement under perturbation and sandwiching all separable states, *Phys. Rev.* **A 65**: 022107.
Lockhart, R. B., Steiner, M. J. and Gerlach, K. (2002). Geometry and product states, *Quant. Inf. Comp.* **2**: 333.
Löwner, K. (1934). Über monotone Matrixfunktionen, *Math. Z.* **38**: 177.
Łoziński, A., Buchleitner, A., Życzkowski, K. and Wellens, T. (2003). Entanglement of $2 \times k$ quantum systems, *Europhys. Lett.* **A 62**: 168.
Lubkin, E. (1978). Entropy of an n-system from its correlation with a k-reservoir, *J. Math. Phys.* **19**: 1028.
Lüders, G. (1951). Über die Zustandsänderung durch den Messprozeß, *Ann. Phys. (Leipzig)* **8**: 322.
Luo, S. (2000). A simple proof of Wehrl's conjecture on entropy, *J. Phys.* **A 33**: 3093.
Maassen, H. and Uffink, J. B. M. (1988). Generalized entropic uncertainty relations, *Phys. Rev. Lett.* **60**: 1103.
Macfarlane, A. J., Sudbery, A. and Weisz, P. H. (1968). On Gell–Mann's λ-matrices, d- and f-tensors, octets and parametrization of $SU(3)$, *Commun. Math. Phys.* **11**: 77.
Mahler, G. and Weberruß, V. A. (1995, (2nd. ed.) 1998). *Quantum Networks*, Springer.
Majewski, W. A. (1975). Transformations between quantum states, *Rep. Math. Phys.* **8**: 295.
Majewski, W. A. and Marciniak, M. (2001). On characterization of positive maps, *J. Phys.* **A 34**: 5863.

Majorana, E. (1932). Atomi orientati in campo magnetico variabile, *Nuovo Cimento* **9**: 43.
Makhlin, Y. (2002). Non-local properties of two-qubit gates and mixed states, and the optimization of quantum computations, *Quant. Inf. Proc.* **1**: 243.
Malacarne, L. C., Mandes, R. S. and Lenzi, E. K. (2002). Average entropy of a subsystem from its average Tsallis entropy, *Phys. Rev.* **E 65**: 046131.
Mandel, L. and Wolf, E. (1995). *Optical Coherence and Quantum Optics*, Cambridge University Press.
Marchenko, V. A. and Pastur, L. A. (1967). The distribution of eigenvalues in certain sets of random matrices, *Math. SS*. **72**: 507.
Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*, New York: Academic Press.
McCrimmon, K. (1978). Jordan algebras and their applications, *Bull. AMS.* **84**: 612.
Mehta, M. (1991). *Random Matrices, 2nd. ed.*, New York: Academic Press.
Mehta, M. L. (1989). *Matrix Theory*, Delhi: Hindustan Publishing.
Mermin, N. D. (1998). The Ithaca interpretation of quantum mechanics, *Pramana* **51**: 549.
Meyer, D. A. and Wallach, N. R. (2002). Global entanglement in multiparticle systems, *J. Math. Phys.* **43**: 4273.
Mielnik, B. (1969). Theory of filters, *Commun. Math. Phys.* **15**: 1.
Mielnik, B. (1981). Quantum theory without axioms, *in* C. J. Isham, R. Penrose and D. W. Sciama (eds), *Quantum Gravity II. A Second Oxford Symposium*, Oxford University Press, p. 638.
Mielnik, B. (2001). Nonlinear quantum mechanics: A conflict with the Ptolomean structure?, *Phys. Lett.* **A 289**: 1.
Mintert, F. and Życzkowski, K. (2004). Wehrl entropy, Lieb conjecture and entanglement monotones, *Phys. Rev.* **A 69**: 022317.
Mintert, F., Kuś, M. and Buchleitner, A. (2004). Concurrence of mixed bipartite quantum states in arbitrary dimensions, *Phys. Rev. Lett.* **92**: 167902.
Miranowicz, A. and Grudka, A. (2004). A comparative study of relative entropy of entanglement, concurrence and negativity, *J. Opt. B: Quantum Semiclass. Opt.* **6**: 542.
Mirbach, B. and Korsch, H. J. (1998). A generalized entropy measuring quantum localization, *Ann. Phys. (N.Y.)* **265**: 80.
Miyake, A. (2003). Classification of multipartite entangled states by multidimensional determinants, *Phys. Rev.* **A 67**: 012108.
Molnár, L(2001). Fidelity preserving maps on density operators, *Rep. Math. Phys.* **48**: 299.
Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. and Wineland, D. J. (1995). Demonstration of a fundamental quantum logic gate, *Phys. Rev. Lett.* **75**: 4714.
Morozova, E. A. and Čencov, N. N. (1990). Markov invariant geometry on state manifolds (in Russian), *Itogi Nauki i Tehniki* **36**: 69.
Mosseri, R. and Dandoloff, R. (2001). Geometry of entangled states, Bloch spheres and Hopf fibrations, *J. Phys.* **A 34**: 10243.
Mukunda, N. and Simon, R. (1993). Quantum kinematic approach to the geometric phase. 1. General formalism, *Ann. Phys.* **228**: 205.
Mukunda, N., Arvind, Chaturvedi, S. and Simon, R. (2003). Generalized coherent states and the diagonal representation of operators, *J. Math. Phys.* **44**: 2479.
Munro, W. J., James, D. F. V., White, A. G. and Kwiat, P. G. (2001). Maximizing the entanglement of two mixed qubits, *Phys. Rev.* **A 64**: 030302.
Murray, M. K. and Rice, J. W. (1993). *Differential Geometry and Statistics*, Chapman and Hall.
Nielsen, M. A. (1999). Conditions for a class of entanglement transformations, *Phys. Rev. Lett.* **83**: 436.

Nielsen, M. A. (2000). Continuity bounds for entanglement, *Phys. Rev.* **A 61**: 064301.
Nielsen, M. A. (2002). A simple formula for the average gate fidelity of a quantum dynamical operation, *Phys. Lett.* **303**: 249.
Nielsen, M. A. and Chuang, I. L. (2000). *Quantum Computation and Quantum Information*, Cambridge University Press.
Nielsen, M. A. and Kempe, J. (2001). Separable states are more disordered globally than locally, *Phys. Rev. Lett.* **86**: 5184.
Nielsen, M. A., Dawson, C., Dodd, J., Gilchrist, A., Mortimer, D., Osborne, T., Bremner, M., Harrow, A. and Hines, A. (2003). Quantum dynamics as physical resource, *Phys. Rev.* **A 67**: 052301.
Nogues, G., Rauschenbeutel, A., Osnaghi, S., Bertet, P., Brune, M., Raimond, J., Haroche, S., Lutterbach, L. and Davidovich, L. (2000). Measurement of the negative value for the Wigner function of radiation, *Phys. Rev.* **A 62**: 054101.
Nonnenmacher, S. (1989). Crystal properties of eigenstates for quantum cat maps, *Nonlinearity* **10**: 1569.
O'Connor, K. M. and Wootters, W. K. (2001). Entangled rings, *Phys. Rev.* **A 63**: 052302.
Ohya, M. and Petz, D. (1993). *Quantum Entropy and Its Use*, Springer. Second edition, Springer 2004.
Oi, D. K. L. (n.d.). The geometry of single qubit maps, Preprint quant-ph/0106035.
Orłowski, A. (1993). Classical entropy of quantum states of light, *Phys. Rev.* **A 48**: 727.
Osaka, H. (1991). Indecomposable positive maps in low-dimensional matrix algebra, *Linear Algebra Appl.* **153**: 73.
Oxenrider, C. J. and Hill, R. D. (1985). On the matrix reordering Γ and Ψ, *Linear Algebra Appl.* **69**: 205.
Ozawa, M. (2001). Entanglement measures and the Hilbert–Schmidt distance, *Phys. Lett.* **A 268**: 158.
Page, D. N. (1993). Average entropy of a subsystem, *Phys. Rev. Lett.* **71**: 1291.
Partovi, M. H. (2004). Universal measure of entanglement, *Phys. Rev. Lett.* **92**: 077904.
Pechukas, P. (1994). Reduced dynamics need not to be completely positive, *Phys. Rev. Lett.* **73**: 1060.
Penrose, R. (1960). A spinor approach to general relativity, *Ann. of Phys. (N.Y.)* **10**: 171.
Penrose, R. (2004). *The Road to Reality*, London: Jonathan Cape.
Perelomov, A. M. (1977). Generalized coherent states and some of their applications, *Sov. Phys. Usp.* **20**: 703.
Peres, A. (1995). Higher order Schmidt decomposition, *Phys. Lett.* **A 202**: 16.
Peres, A. (1996). Separability criterion for density matrices, *Phys. Rev. Lett.* **77**: 1413.
Peres, A. and Terno, D. R. (1998). Convex probability domain of generalized quantum measurements, *J. Phys.* **A 31**: L671.
Peres, A. and Wootters, W. K. (1991). Optimal detection of quantum information, *Phys. Rev. Lett.* **66**: 1119.
Petz, D. (1994). Geometry of canonical correlation on the state space of a quantum system, *J. Math. Phys.* **35**: 780.
Petz, D. (1996). Monotone metrics on matrix spaces, *Linear Algebra Appl.* **244**: 81.
Petz, D. (1998). Information-geometry of quantum states, *Quantum Prob. Commun.* **X**: 135.
Petz, D. and Sudár, C. (1996). Geometries of quantum states, *J. Math. Phys.* **37**: 2662.
Picken, R. F. (1990). The Duistermaat–Heckman integration formula on flag manifolds, *J. Math. Phys.* **31**: 616.
Pinsker, M. (1964). *Information and Information Stability of Random Variables and Processes*, San Francisco: Holden-Day.

Pipek, J. and Varga, I. (1992). Universal scheme for the spacial-localization properties of one-particle states in finite d-dimensional systems, *Phys. Rev.* **A 46**: 3148.
Pitowsky, I. (1998). Infinite and finite Gleason's theorems and the logic of indeterminacy, *J. Math. Phys.* **39**: 218.
Pittenger, A. O. (2003). Unextendible product bases and the construction of inseparable states, *Linear Albegra Appl.* **359**: 235.
Pittenger, A. O. and Rubin, M. H. (2000). Separability and Fourier representations of density matrices, *Phys. Rev.* **A 62**: 032313.
Pittenger, A. O. and Rubin, M. H. (2002). Convexity and the separability problem of quantum mechanical density matrices, *Linear Algebra Appl.* **346**: 47.
Pittenger, A. O. and Rubin, M. H. (2003). Geometry of entanglement witnesses and local detection of entanglement, *Phys. Rev.* **A 67**: 012327.
Plesch, M. and Bužek, V. (2003). Entangled graphs: Bipartite entanglement in multiqubit systems, *Phys. Rev.* **A 67**: 012322.
Poon, Y.-T. and Tsing, N.-K. (1987). Inclusion relations between orthostochastic matrices and products of pinching matrices, *Lin. Multilin. Algebra* **21**: 253.
Popescu, S. and Rohrlich, D. (1992). Generic quantum nonlocality, *Phys. Lett.* **A 166**: 293.
Popescu, S. and Rohrlich, D. (1997). Thermodynamics and the measure of entanglement, *Phys. Rev.* **A 56**: R3319.
Poulin, D., Blume–Kohout, R., Laflamme, R. and Olivier, H. (2004). Exponential speedup with a single bit of quantum information, *Phys. Rev. Lett.* **92**: 177906.
Poźniak, M., Życzkowski, K. and Kuś, M. (1998). Composed ensembles of random unitary matrices, *J. Phys.* **A 31**: 1059.
Preskill, J. (n.d.). Lecture notes on quantum computation, 1998, see www.theory.caltech.edu/people/preskill/.
Prosen, T. (1996a). Exact statistics of complex zeros for Gaussian random polynomials with real coefficients, *J. Phys.* **A 29**: 4417.
Prosen, T. (1996b). Parametric statistics of zeros of Husimi representations of quantum chaotic eigenstates and random polynomials, *J. Phys.* **A 29**: 5429.
Rabei, E. M., Arvind, Simon, R. and Mukunda, N. (1990). Bargmann invariants and geometric phases – a generalized connection, *Phys. Rev.* **A 60**: 3397.
Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models*, New York: Wiley.
Rachev, S. T. and Rüschendorf, L. (1998). *Mass Transportation Problems. Vol. 1: Theory*, Springer.
Radcliffe, J. M. (1971). Some properties of coherent spin states, *J. Phys.* **A 4**: 313.
Rains, E. M. (1999a). Bound on distillable entanglement, *Phys. Rev.* **A 60**: 179.
Rains, E. M. (1999b). Rigorous treatment of distillable entanglement, *Phys. Rev.* **A 60**: 173.
Rains, E. M. (2001). A semidefinite program for distillable entanglement, *IEEE Trans. Inform. Theory* **47**: 2921.
Rajagopal, A. K. and Rendell, R. W. (2002). Separability and correlations in composite states based on entropy methods, *Phys. Rev.* **A 66**: 022104.
Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters, *Bull. Calcutta Math. Soc.* **37**: 81.
Reck, M., Zeilinger, A., Bernstein, H. J. and Bertani, P. (1994). Experimental realization of any discrete unitary operator, *Phys. Rev. Lett.* **73**: 58.
Řeháček, J. and Hradil, Z. (2003). Quantification of entanglement by means of convergent iterations, *Phys. Rev. Lett.* **90**: 127904.

Renes, J. M., Blume-Kohout, R., Scott, A. J. and Caves, C. M. (2004). Symmetric informationally complete quantum measurements, *J. Math. Phys.* **45**: 2171.
Rényi, A. (1961). On measures of entropy and information, *Proc. of the Fourth Berkeley Symp. Math. Statist. Prob. 1960, Vol. I*, Berkeley: University of California Press, p. 547.
Robertson, A. G. (1983). Automorphisms of spin factors and the decomposition of positive maps, *Quart. J. Math. Oxford* **34**: 87.
Rossignoli, R. and Canosa, N. (2002). Generalized entropic criterion for separability, *Phys. Rev.* **A 66**: 042306.
Rudolph, O. (2001). A new class of entanglement measures, *J. Math. Phys.* **A 42**: 5306.
Rudolph, O. (2003a). On the cross norm criterion for separability, *J. Phys.* **A 36**: 5825.
Rudolph, O. (2003b). Some properties of the computable cross norm criterion for separability, *Physical Review* **A 67**: 032312.
Rungta, P. and Caves, C. M. (2003). Concurrence-based entanglement measures for isotropic states, *Phys. Rev.* **A 67**: 012307.
Rungta, P., Bužek, V., Caves, C. M., Hillery, M. and Milburn, G. J. (2001). Universal state inversion and concurrence in arbitrary dimensions, *Phys. Rev.* **A 64**: 042315.
Ruskai, M. B. (2002). Inequalities for quantum entropy: A review with conditions for equality, *J. Math. Phys.* **43**: 4358.
Ruskai, M. B., Szarek, S. and Werner, E. (2002). An analysis of completely-positive trace-preserving maps on 2×2 matrices, *Linear Algebra Appl.* **347**: 159.
Saff, E. B. and Kuijlaars, A. B. J. (1997). Distributing many points on a sphere, *Math. Intelligencer* **19**: 5.
Salvemini, T. (1943). Sul calcolo degli indici di concordanza tra due caratteri quantitativi, *Atti della VI Riunione della Soc. Ital. di Statistica*.
Sánchez-Ruiz, J. (1995). Simple proof of Page's conjecture on the average entropy of a subsystem, *Phys. Rev.* **E 52**: 5653.
Sanov, I. N. (1957). On the probability of large deviations of random variables (in Russian), *Mat. Sbornik* **42**: 11.
Sanpera, A., Tarrach, R. and Vidal, G. (1998). Local description of quantum inseparability, *Phys. Rev.* **A 58**: 826.
Schatten, R. (1950). *A Theory of Cross-spaces*, Princeton University Press.
Schirmer, S., Zhang, T. and Leahy, J. V. (2004). Orbits of quantum states and geometry of Bloch vectors for n-level systems, *J. Phys.* **A 37**: 1389.
Schlienz, J. and Mahler, G. (1995). Description of entanglement, *Phys. Rev.* **A 52**: 4396.
Schmidt, E. (1907). Zur Theorie der linearen und nichtlinearen Integralgleichungen, *Math. Ann* **63**: 433.
Schrödinger, E. (1926a). Der stetige Übergang von der Mikro - zur Makromechanik, *Naturwissenschaften* **14**: 664.
Schrödinger, E. (1926b). Die Gesichtsempfindungen, *in* O. Lummer (ed.), *Müller-Pouillets Lehrbuch der Physik, 11th edn, Vol. 2*, Friedr. Vieweg und Sohn, pp. 456–560.
Schrödinger, E. (1935a). Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften* **32**: 446.
Schrödinger, E. (1935b). Discussion of probability relations between separated systems, *Proc. Camb. Phil. Soc.* **31**: 555.
Schrödinger, E. (1936). Probability relations between separated systems, *Proc. Camb. Phil. Soc.* **32**: 446.
Schrödinger, E. (1950). *Space–Time Structure*, Cambridge University Press.
Schumacher, B. (1995). Quantum coding, *Phys. Rev.* **A 51**: 2738.

Schumacher, B. and Nielsen, M. (1996). Quantum data processing and error correction, *Phys. Rev.* **A 54**: 2629.
Schumacher, B. and Westmoreland, M. (n.d.). Relative entropy in quantum information theory, Preprint quant-ph/0004045.
Schupp, P. (1999). On Lieb's conjecture for the Wehrl entropy of Bloch coherent states, *Commun. Math. Phys.* **207**: 481.
Schur, J. (1923). Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie, *Sitzber. Berl. Math. Ges.* **22**: 9.
Schwinger, J. (1965). On angular momentum, *in* L. C. Biedenharn and H. van Dam (eds), *Quantum Theory of Angular Momentum*, Academic Press, p. 229.
Scutaru, H. (n.d.). On Lieb's conjecture, Preprint FT-180 (1979) Bucharest and math-ph/9909024.
Segal, I. E. (1947). Postulates for general quantum mechanics, *Ann. Math.* **48**: 930.
Sen, S. (1996). Average entropy of a quantum subsystem, *Phys. Rev. Lett.* **77**: 1.
Sengupta, A. M. and Mitra, P. P. (1999). Distributions of singular values for some random matrices, *Phys. Rev.* **E 60**: 3389.
Shannon, C. E. (1948). A mathematical theory of communication, *Bell Sys. Tech. J.* **27**: 379, 623.
Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*, Univ. of Illinois Press.
Shapere, A. and Wilczek, F. (1989). *Geometric Phases in Physics*, World Scientific.
Shor, P. W. (n.d.). Equivalence of additivity questions in quantum information theory, Preprint quant-ph/0305035.
Shor, P. W., Smolin, J. A. and Terhal, B. M. (2001). Non-additivity of bipartite distillable entanglement follows from a conjecture on bound entangled Werner states, *Phys. Rev. Lett.* **86**: 2681.
Sinołęcka, M., Życzkowski, K. and Kuś, M. (2002). Manifolds of equal entanglement for composite quantum systems, *Acta Phys. Pol.* **B 33**: 2081.
Slater, P. B. (1999a). Hall normalization constants for the Bures volumes of the n-state quantum systems, *J. Phys.* **A 32**: 8231.
Slater, P. B. (1999b). A priori probabilities of separable quantum states, *J. Phys.* **A 32**: 5261.
Slater, P. B. (2005). Silver mean conjectures for 15-d volumes and 14-d hyperareas of the separable two-qubit systems, *J. Geom. Phys.* **53**: 74.
Słomczyński, W. (2002). Subadditivity of entropy for stochastic matrices, *Open Sys. & Information Dyn.* **9**: 201.
Słomczyński, W. and Życzkowski, K. (1998). Mean dynamical entropy of quantum maps on the sphere diverges in the semi-classical limit, *Phys. Rev. Lett.* **80**: 1880.
Smithey, D. T., Beck, M., Raymer, M. and Faridani, A. (1993). Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum, *Phys. Rev. Lett.* **70**: 1244.
Sommers, H.-J. and Życzkowski, K. (2003). Bures volume of the set of mixed quantum states, *J. Phys.* **A 36**: 10083.
Sommers, H.-J. and Życzkowski, K. (2004). Statistical properties of random density matrices, *J. Phys.* **A 37**: 8457.
Spanier, J. and Oldham, K. B. (1987). *An Atlas of Functions*, Washington: Hemisphere Publishing Corporation.
Steiner, M. (2003). Generalized robustness of entanglement, *Phys. Rev.* **A 67**: 054305.
Stinespring, W. F. (1955). Positive functions on c^* algebras, *Proc. Am. Math. Soc.* **6**: 211.

Størmer, E. (1963). Positive linear maps of operator algebras, *Acta Math.* **110**: 233.
Størmer, E. (1982). Decomposable positive maps on c^*-algebras, *Proc. Amer. Math. Soc.* **86**: 402.
Streater, R. F. (1995). *Statistical Dynamics*, London: Imperial College Press.
Study, E. (1905). Kürzeste Wege in komplexen Gebiet, *Math. Annalen* **60**: 321.
Sudarshan, E. C. G. (1963). Equivalence of semiclassical and quantum mechanical descriptions of light beams, *Phys. Rev. Lett.* **10**: 277.
Sudarshan, E. C. G. and Shaji, A. (2003). Structure and parametrization of stochastic maps of density matrices, *J. Phys.* **A 36**: 5073.
Sudarshan, E. C. G., Mathews, P. M. and Rau, J. (1961). Stochastic dynamics of quantum-mechanical systems, *Phys. Rev.* **121**: 920.
Sudbery, A. (1990). Computer-friendly d-tensor identities for $SU(n)$, *J. Phys.* **A 23**: L705.
Sudbery, A. (2001). On local invariants of pure three-qubit states, *J. Phys.* **A 34**: 643.
Sugita, A. (2002). Proof of the generalized Lieb–Wehrl conjecture for integer indices more than one, *J. Phys.* **A 35**: L621.
Sugita, A. (2003). Moments of generalized Husimi distribution and complexity of many-body quantum states, *J. Phys.* **A 36**: 9081.
Sylvester, J. (1884). Sur l'équation en matrices $px = xq$, *C. R. Acad. Sci. Paris* **99**: 67, 115.
Szarek, S. (2005). The volume of separable states is super-doubly-exponentially small, *Phys. Rev.* **A 72**: 032304.
Takesaki, T. and Tomiyama, J. (1983). On the geometry of positive maps in matrix algebras, *Math. Z.* **184**: 101.
Tanahashi, K. and Tomiyama, J. (1988). Indecomposable positive maps in matrix algebra, *Canad. Math. Bull.* **31**: 308.
Tang, W. (1986). On positive linear maps between matrix algebra, *Linear Algebra Appl.* **79**: 33.
Terhal, B., Chuang, I., DiVincenzo, D., Grassl, M. and Smolin, J. (1999). Simulating quantum operations with mixed environments, *Phys. Rev.* **A 60**: 88.
Terhal, B., Horodecki, M., Leung, D. W. and DiVincenzo, D. (2002). The entanglement of purification, *J. Math. Phys.* **43**: 4286.
Terhal, B. M. (2000a). Bell inequalities and the separability criterion, *Phys. Lett.* **A 271**: 319.
Terhal, B. M. (2000b). A family of indecomposable positive linear maps based on entangled quantum states, *Linear Algebra Appl.* **323**: 61.
Terhal, B. M. (2002). Detecting quantum entanglement, *Theor. Comput. Sci.* **287**: 313.
Terhal, B. M. and Vollbrecht, K. G. H. (2000). Entanglement of formation for isotropic states, *Phys. Rev. Lett.* **85**: 2625.
Thiele, E. and Stone, J. (1984). A measure of quantum chaos, *J. Chem. Phys.* **80**: 5187.
Topsøe, F. (2001). Bounds for entropy and divergence for distributions over a two element set, *J. Ineq. P. Appl. Math.* **2**: 25.
Tribus, M. and McIrvine, E. C. (1971). Energy and information, *Scient. Amer.* **224**: 178.
Tsallis, C. (2002). Entropic non-extensivity: a possible measure of complexity, *Chaos, Solitons, Fract.* **13**: 371.
Tsallis, C., Lloyd, S. and Baranger, M. (2001). Peres criterion for separability through non-extensive entropy, *Phys. Rev.* **A 63**: 042104.
Tucci, R. R. (n.d.a). All moments of the uniform ensemble of quantum density matrices, Preprint quant-ph/0206193.
Tucci, R. R. (n.d.b). Entanglement of formation and conditional information transmission, Preprint quant-ph/0010041.

Tucci, R. R. (n.d.c). Relaxation method for calculating quantum entanglement, Preprint quant-ph/0101123.
Uhlmann, A. (1971). Endlich dimensionale Dichtematrizen, *Wiss. Z. Karl-Marx-Univ.* **20**: 633.
Uhlmann, A. (1977). Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory, *Commun. Math. Phys.* **54**: 21.
Uhlmann, A. (1992). The metric of Bures and the geometric phase, *in* R. Gielerak (ed.), *Quantum Groups and Related Topics*, Dordrecht: Kluwer, p. 267.
Uhlmann, A. (1993). Density operators as an arena for differential geometry, *Rep. Math. Phys.* **33**: 253.
Uhlmann, A. (1995). Geometric phases and related structures, *Rep. Math. Phys.* **36**: 461.
Uhlmann, A. (1996). Spheres and hemispheres as quantum state spaces, *J. Geom. Phys.* **18**: 76.
Uhlmann, A. (1998). Entropy and optimal decompositions of states relative to a maximal commutative subalgebra, *Open Sys. & Information Dyn.* **5**: 209.
Uhlmann, A. (2000). Fidelity and concurrence of conjugated states, *Rep. Rev.* **A 62**: 032307.
Uhlmann, A. (2003). Concurrence and foliations induced by some 1-qubit channels, *Int. J. Theor. Phys.* **42**: 983.
Umegaki, H. (1962). Conditional expectation in an operator algebra IV: entropy and information, *Ködai Math. Sem. Rep.* **14**: 59.
van Dam, W. and Hayden, P. (n.d.). Rényi-entropic bounds on quantum communication, Preprint quant-ph/0204093.
Varadarajan, V. (1985). *Geometry of Quantum Theory*, Springer.
Vedral, V. (2002). The role of relative entropy in quantum information theory, *Rev. Mod. Phys.* **74**: 197.
Vedral, V. and Kashefi, E. (2002). Uniqueness of the entanglement measure for bipartite pure states and thermodynamics, *Phys. Rev. Lett.* **89**: 037903.
Vedral, V. and Plenio, M. B. (1998). Entanglement measures and purification procedures, *Phys. Rev.* **A 57**: 1619.
Vedral, V., Plenio, M. B., Rippin, M. A. and Knight, P. L. (1997). Quantifying entanglement, *Phys. Rev. Lett.* **78**: 2275.
Verstraete, F. and Verschelde, H. (2002). Fidelity of mixed states of two qubits, *Phys. Rev.* **A 66**: 022307.
Verstraete, F., Audenaert, K. and DeMoor, B. (2001a). Maximally entangled mixed states of two qubits, *Phys. Rev.* **A 64**: 012316.
Verstraete, F., Audenaert, K., Dehaene, J. and DeMoor, B. (2001b). A comparison of the entanglement measures negativity and concurrence, *J. Phys.* **A 34**: 10327.
Verstraete, F., Dahaene, J. and DeMoor, B. (2002a). On the geometry of entangled states, *J. Mod. Opt.* **49**: 1277.
Verstraete, F., Dahaene, J., DeMoor, B. and Verschelde, H. (2002b). Four qubits can be entangled in nine different ways, *Phys. Rev.* **A 65**: 052112.
Vidal, G. (1999). Entanglement of pure states for a single copy, *Phys. Rev. Lett.* **83**: 1046.
Vidal, G. (2000). Entanglement monotones, *J. Mod. Opt.* **47**: 355.
Vidal, G. and Tarrach, R. (1999). Robustness of entanglement, *Phys. Rev.* **A 59**: 141.
Vidal, G. and Werner, R. F. (2002). A computable measure of entanglement, *Phys. Rev.* **A 65**: 032314.
Vidal, G., Dür, W. and Cirac, J. I. (2002). Entanglement cost of bipartite mixed states, *Phys. Rev. Lett.* **89**: 027901.

Vidal, G., Jonathan, D. and Nielsen, M. A. (2000). Approximate transformations and robust manipulation of bipartite pure state entanglement, *Phys. Rev.* **A 62**: 012304.
Virmani, S. and Plenio, M. (2000). Ordering states with entanglement measures, *Phys. Lett.* **A 268**: 31.
Vollbrecht, K. G. H. and Werner, R. F. (2000). Why two qubits are special, *J. Math. Phys.* **41**: 6772.
Vollbrecht, K. G. H. and Werner, R. F. (2001). Entanglement measures under symmetry, *Phys. Rev.* **A 64**: 062307.
Vollbrecht, K. G. H. and Wolf, W. W. (2002). Conditional entropies and their relation to entanglement criteria, *J. Math. Phys.* **43**: 4299.
von Mises, R. (1957). *Probability, Statistics and Truth*, New York: Dover.
von Neumann, J. (1927). Thermodynamik quantummechanischer Gesamtheiten, *Gött. Nach.* **1**: 273.
von Neumann, J. (1955). *Mathematical Foundations of Quantum Mechanics*, Princeton University Press.
Wehrl, A. (1978). General properties of entropy, *Rev. Mod. Phys.* **50**: 221.
Wehrl, A. (1979). On the relation between classical and quantum mechanical entropies, *Rep. Math. Phys.* **16**: 353.
Wei, T.-C. and Goldbart, P. M. (2003). Geometric measure of entanglement and applications to bipartite and multipartite quantum states, *Phys. Rev.* **A 68**: 042307.
Wei, T. C., Nemoto, K., Goldbart, P. M., Kwiat, P. G., Munro, W. J. and Verstraete, F. (2003). Maximal entanglement versus entropy for mixed quantum states, *Phys. Rev.* **A 67**: 022110.
Wellens, T. and Kuś, M. (2001). Separable approximation for mixed states of composite quantum systems, *Phys. Rev.* **A 64**: 052302.
Werner, R. F. (1989). Quantum states with Einstein–Podolski–Rosen correlations admitting a hidden-variable model, *Phys. Rev.* **A 40**: 4277.
Wigner, E. P. (1932). On the quantum correction for thermodynamic equilibrium, *Phys. Rev.* **40**: 749.
Wigner, E. P. (1959). *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*, Academic Press.
Williamson, S. J. and Cummins, H. Z. (1983). *Light and Color in Nature and Art*, Wiley.
Witte, C. and Trucks, M. (1999). A new entanglement measure induced by the Hilbert–Schmidt norm, *Phys. Lett.* **A 257**: 14.
Wódkiewicz, K. (2001). Stochastic decoherence of qubits, *Optics Express* **8**: 145.
Woerdeman, H. J. (2004). The separability problem and normal completions, *Linear Algebra Appl.* **376**: 85.
Wong, A. and Christensen, N. (2001). Potential multiparticle entanglement measure, *Phys. Rev.* **A 63**: 044301.
Wong, Y.-C. (1967). Differential geometry of Grassmann manifolds, *Proc. Nat. Acad. Sci. USA* **47**: 589.
Wootters, W. K. (1981). Statistical distance and Hilbert space, *Phys. Rev.* **D 23**: 357.
Wootters, W. K. (1987). A Wigner function formulation of finite-state quantum mechanics, *Ann. Phys. (N.Y.)* **176**: 1.
Wootters, W. K. (1990). Random quantum states, *Found. Phys.* **20**: 1365.
Wootters, W. K. (1998). Entanglement of formation of an arbitrary state of two qubits, *Phys. Rev. Lett.* **80**: 2245.
Wootters, W. K. (2001). Entanglement of formation and concurrence, *Quant. Inf. Comp.* **1**: 27.

Wootters, W. K. and Fields, B. F. (1989). Optimal state determination by mutually unbiased measurements, *Ann. Phys. (N.Y.)* **191**: 363.
Woronowicz, S. L. (1976a). Non-extendible positive maps, *Commun. Math. Phys.* **51**: 243.
Woronowicz, S. L. (1976b). Positive maps of low-dimensional matrix algebra, *Rep. Math. Phys.* **10**: 165.
Wu, N. and Coppins, R. (1981). *Linear Programming and Extensions*, New York: McGraw-Hill.
Yopp, D. A. and Hill, R. D. (2000). On completely copositive and decomposable linear transformations, *Linear Algebra Appl.* **312**: 1.
Yu, Y.-K. and Zhang, Y.-C. (2002). On the anti-Wishart distribution, *Physica* **A 312**: 1.
Zalka, C. and Rieffel, E. (2002). Quantum operations that cannot be implemented using a small mixed environment, *J. Math. Phys.* **43**: 4376.
Zanardi, P. and Lidar, D. A. (2004). Purity and state fidelity of quantum channels, *Phys. Rev.* **A 70**: 012315.
Zhang, W.-M., Feng, D. H. and Gilmore, R. (1990). Coherent states: Theory and some applications, *Rev. Mod. Phys.* **62**: 867.
Zurek, W. H. (2003). Environment-assisted invariance, entanglement, and probabilities in quantum physics, *Phys. Rev. Lett.* **90**: 120404.
Życzkowski, K. (1990). Indicators of quantum chaos based on eigenvector statistics, *J. Phys.* **A 23**: 4427.
Życzkowski, K. (1999). Volume of the set of separable states II, *Phys. Rev.* **A 60**: 3496.
Życzkowski, K. (2001). Localization of eigenstates & mean Wehrl entropy, *Physica* **E 9**: 583.
Życzkowski, K. (2003). Rényi extrapolation for Shannon entropy, *Open Sys. & Information Dyn.* **10**: 297.
Życzkowski, K. and Słomczyński, W. (1998). The Monge distance between quantum states, *J. Phys.* **A 31**: 9095.
Życzkowski, K. and Słomczyński, W. (2001). Monge metric on the sphere and geometry of quantum states, *J. Phys.* **A 34**: 6689.
Życzkowski, K. and Sommers, H.-J. (2000). Truncations of random unitary matrices, *J. Phys.* **A 33**: 2045.
Życzkowski, K. and Sommers, H.-J. (2001). Induced measures in the space of mixed quantum states, *J. Phys.* **A 34**: 7111.
Życzkowski, K. and Sommers, H.-J. (2003). Hilbert–Schmidt volume of the set of mixed quantum states, *J. Phys.* **A 36**: 10115.
Życzkowski, K. and Sommers, H.-J. (2005). Average fidelity between random quantum states, *Phys. Rev.* **A 71**: 032313.
Życzkowski, K., Horodecki, P., Sanpera, A. and Lewenstein, M. (1998). Volume of the set of separable states, *Phys. Rev.* **A 58**: 883.

Index