Ingemar Bengtsson

GEOMETRY OF QUANTUM MECHANICS

Winter 1997-1998

These are the notes from my "course" in the geometry of quantum mechanics. The idea was to introduce the mathematics in its own right, but not to introduce anything that is not directly relevant to the subject. It gets a bit sketchy towards the end, and I missed some of the really interesting things (I blame time limitations). The contents became

- Mostly about the 2-sphere
- Mostly about the 3-sphere
- Complex projective spaces
- Real quantum mechanics
- Colours and statistical geometry
- The space of density matrices
- Purified density matrices

I gave the course again in the spring of 2000. Moreover Karol Życzkowski read the notes. As a result the first three chapters were rewritten.

THE 2-SPHERE: A COMPLEX MANIFOLD

Spheres

We start with what appears to be a rather long detour; eventually it will lead us into the heart of quantum mechanics, and more especially to the geometry of its pure states. In the case of two dimensional Hilbert spaces the space of pure states is the sphere $\mathbf{S}^2$. In higher dimensions the pure states do not form spheres, but they do share a special property of $\mathbf{S}^2$, namely that their geometry is Kähler. So Kähler geometry is what this chapter is really about.

The opening move is the introduction of a vector space having $n+1$ dimensions. It is equipped with a *scalar product*

$$X \cdot Y = X^I \delta_{IJ} Y^J \ . \tag{1}$$

In flat spaces we will use the Kronecker delta to raise and lower indices without comment, so $X_I = X^I$. The *sphere* $\mathbf{S}^n$ is defined as an n dimensional surface in the *embedding* vector space, given by the equation

$$X \cdot X = 1 \ . \tag{2}$$

The scalar product induces a metric on the sphere, namely the usual "*round*" metric. We will think of spheres as Riemannian manifolds equipped with this natural metric.

We want to introduce an intrinsic coordinate system on the sphere. There are many ways to do this. Begin by decomposing the vectors according to

$$X^I = (X^0, X^i) \ ; \qquad 1 \leq i \leq n. \tag{3}$$

One obvious choice of intrinsic coordinates is now given by perpendicular projection to the equatorial plane (known as orthographical projection among mapmakers):

$$X^i = x^i \qquad X^0 = \sqrt{1-r^2} \ , \qquad r^2 \equiv \sum_{i=1}^{d} x^i x^i < 1 \ . \tag{4}$$

This *coordinate patch* covers the region where $X^0 > 1$. The metric when expressed in these coordinates is

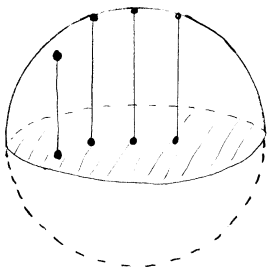

Figure 1: Orthographical projection onto the equatorial plane

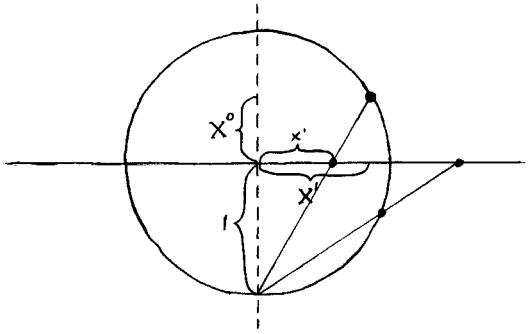

Figure 2: Stereographic projection from the South Pole

$$ds^2 = dX^0 dX^0 + \sum_{i=1}^{n} dX^i dX^i = (\delta_{ij} + \frac{x_i x_j}{1 - r^2}) dx^i dx^j \ . \tag{5}$$

(Somewhat illogically, it is to be understood that $x_i = x^i$ in this formula.) We need several coordinate patches of this kind to cover the entire sphere.

An alternative choice of intrinsic coordinates—rather more useful for our purposes—is given by *stereographic projection* from the South Pole to the equatorial plane, so that

$$\frac{x^i}{X^i} = \frac{1}{1 + X^0} \quad \Leftrightarrow \quad X^i = \frac{2x^i}{1 + r^2} \qquad X^0 = \frac{1 - r^2}{1 + r^2} \ . \tag{6}$$

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:

$$ds^2 = \Omega^2 \delta_{ij} dx^i dx^j = \frac{4}{(1+r^2)^2} \delta_{ij} dx^i dx^j \ . \tag{7}$$

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.

Both coordinate systems have their special advantages. The topology of coordinate space is $\mathbf{R}^n$. When stereographic coordinates are used the sphere has one further point which is not covered by these coordinates, so the topology of $\mathbf{S}^n$ is the topology of $\mathbf{R}^n$ with one extra point attached "at infinity". The conformal factor ensures that the round metric is smooth at the added point, so that the "infinity" of flat space lies at finite distance on the sphere. An great advantage of stereographic coordinates is that all angles come out correctly if we draw a picture in coordinate space, although distances are badly distorted when we move away from the origin. We say that the map between the sphere and the coordinate space is a *conformal map*, that is it preserves angles.

Our coordinate systems can also be used to describe the *quotient space*

$$\mathbf{RP}^n = \mathbf{S}^n / \mathbf{Z}^2 \ . \tag{8}$$

The notation on the right hand side means the following: $\mathbf{Z}^2$ denotes the discrete isometry group that transforms a given point on the sphere to its antipodal point. As a group it is isomorphic to the multiplicative group with two elements $1, -1$, and the specific way in which it acts on $\mathbf{S}^n$ has to be understood. When we take the quotient of the sphere with this group we mean that we consider the space whose points are the *orbits* of the group, an orbit being a set of points that can be transformed into each other by the group. The quotient space is therefore the sphere with antipodal points identified. On the left hand side of the equation we have the *real projective space* $\mathbf{RP}^n$. By definition this is the space whose points are lines through the origin in $\mathbf{R}^{n+1}$. Since any such line cuts the sphere in a pair of antipodal points, it is clear that the equation that we wrote down is correct.

The orthographical projection covers all of $\mathbf{RP}^n$ in a one-to-one fashion, except for the points coming from the equator of the sphere. Topologically the equator is an $\mathbf{S}^{n-1}$ that becomes an $\mathbf{RP}^{n-1}$ when antipodal points are

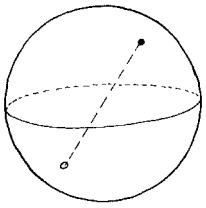

Figure 3: Real projective space is the sphere with antipodal points identified

identified. Hence real projective n-space has the topology of $\mathbf{R}^n$ (the topology of a hemisphere) with a real projective $n-1$ space attached "at infinity". In mathematics real projective space is famous for many reasons, one of them being that it provides a model for Riemann's elliptic geometry in which the parallel axiom of Euclid is changed so that every pair of lines meet in a unique point, while the remaining axioms remain in force. We will see rather more of projective space later on, and most especially complex projective space.

Back to the sphere: It has $(n+1)n/2$ Killing vectors

$$J_{IJ} = X_I \partial_J - X_J \partial_I \ . \tag{9}$$

A *Killing vector* is a tangent vector that points in the direction of an infinitesimal *isometry* (a transformation that preserves distances). These are precisely those Killing vectors of the flat embedding space that leave the sphere intact. If we exponentiate an infinitesimal transformation we get a finite isometry and the Killing vectors become the tangent vectors of the *Killing fields* (or Killing flow lines); on the 2-sphere these are always circles at constant distance from a pair of antipodal *fixed points* where the flow vanishes but in higher dimensions the situation gets more interesting, as we will see later.

A *geodesic* is a curve having the property that it is the shortest curve connecting any pair of nearby points on the curve. On the sphere the geodesics are conveniently described by means of the embedding coordinates. They can be obtained as the solutions of the Euler-Lagrange equations coming from the constrained Lagrangian

$$L = \frac{1}{2}\dot{X} \cdot \dot{X} + \Lambda(X \cdot X - 1) \ . \tag{10}$$

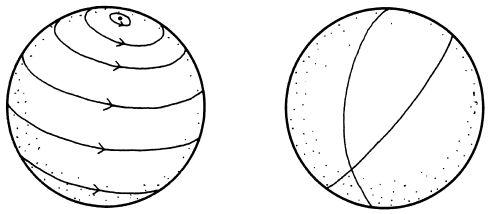

Figure 4: Killing flows and geodesics on the 2-sphere

Here the overdot denotes differentiation with respect to the *affine parameter* along the curve. It is not hard to find a useful set of constants of the motion. It includes

$$X \cdot X = 1 \qquad\qquad \dot{X} \cdot \dot{X} = 1 \ , \tag{11}$$

where the last equality follows from a convenient normalization of the affine parameter. Another set of constants of the motion is

$$k^{IJ} = X^I \dot{X}^J - X^J \dot{X}^I \ . \tag{12}$$

With these in hand we deduce that

$$k_{IJ} X^J = -\dot{X}_I \quad \& \quad k_{IJ} \dot{X}^J = X_I \quad \Rightarrow \quad \ddot{X}^I = -X^I \ . \tag{13}$$

Hence the general solution for a geodesic takes the form

$$X^I(\tau) = k^I \cos\tau + l^I \sin\tau \ , \qquad k \cdot k = l \cdot l = 1 \ , \ k \cdot l = 0 \ . \tag{14}$$

Topologically all the geodesics are circles, namely great circles on the sphere. A great circle is given by the intersection of the sphere with a plane through the origin in the embedding space. The vectors k and l span such a plane.

Let us now choose two points along the curve characterized by two choices of the affine parameter τ, and compute

$$X(\tau_1) \cdot X(\tau_2) = \cos(\tau_1 - \tau_2) \ . \tag{15}$$

But 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

$$\cos d = X(\tau_1) \cdot X(\tau_2) \ , \tag{16}$$

where d is the *geodesic distance* between the two points.

Even dimensional spheres

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

$$x^i = (x^a, x^{m+a}) \ , \tag{17}$$

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

$$z^a = x^a + ix^{m+a} \qquad \bar{z}^{\bar{a}} = x^a - ix^{m+a} \ . \tag{18}$$

So the new coordinates come in pairs connected by complex conjugation,

$$(z^a)^* = \bar{z}^{\bar{a}} \ . \tag{19}$$

This equation ensures that the original coordinates take real values only. Clearly, if the latter serve as coordinates, then so do the complex numbers z^a and $\bar{z}^{\bar{a}}$. In fact because of equation (19) the $\bar{z}^{\bar{a}}$ are not needed—the m complex numbers z^a will serve as coordinates. If we choose stereographic coordinates for definiteness we find that the round metric becomes

$$ds^2 \equiv g_{ab}dz^a dz^b + 2g_{a\bar{b}}dz^a d\bar{z}^{\bar{b}} + g_{\bar{a}\bar{b}}d\bar{z}^{\bar{a}}d\bar{z}^{\bar{b}} = \frac{4}{(1+r^2)^2}\delta_{a\bar{b}}dz^a d\bar{z}^{\bar{b}} \ . \tag{20}$$

Note that we are not complexifying the manifold or anything like that. 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. (Equivalently we could use both z^a and $\bar{z}^{\bar{a}}$ as independent coordinates, and forget about eq. (19)). What we actually did may seem like a cheap trick in comparison, and indeed in general it is just that. For the 2-sphere it is anything but cheap, as we will see.

The way to see whether the introduction of complex coordinates is more than a trick is to 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:

$$x'^a = \frac{X^a}{1 - X^0} \qquad x'^{n+a} = \frac{-X^{m+a}}{1 - X^0} \ . \tag{21}$$

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

$$z'^a = x'^a + i x'^{m+a} = \frac{X^a - i X^{m+a}}{1 - X^0} = \frac{2(x^a - i x^{m+a})}{1 + r^2 - 1 + r^2} = \frac{\bar{z}^a}{r^2} \ . \tag{22}$$

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

$$z'(z) = \frac{1}{z} \ . \tag{23}$$

This is an important observation: We have found that the transition functions between the two patches that cover the 2-sphere are analytic functions of the complex coordinates. In higher dimensions this simple manœuvre fails.

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

$$g_{z\bar{z}} = \frac{2}{(1 + |z|^2)^2} = \frac{2}{1 + |z|^2} \left(1 - \frac{|z|^2}{1 + |z|^2} \right) = 2 \partial_z \partial_{\bar{z}} \ln (1 + |z|^2) \ . \tag{24}$$

In some sense there exists a "potential" for the metric of the 2-sphere. Spaces for which this is true are called Kähler manifolds, and it will turn out that—unlike higher dimensional spheres—the pure states of quantum mechanics always have this property.

Complex manifolds

We pause 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 analytic.

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

$$z' = z'(z, \bar{z}) \qquad\qquad \bar{z}' = \bar{z}'(z, \bar{z}) \ . \tag{25}$$

The manifold is complex iff it can be covered by coordinate patches such that

$$z' = z'(z) \qquad\qquad \bar{z}' = \bar{z}'(\bar{z}) \ . \tag{26}$$

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

$$\bar{z}^{\bar{a}} \equiv (z^a)^* \ . \tag{27}$$

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-analytic, but the idea is to restrict oneself to analytic 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 $\mathbf{C}^n = \mathbf{R}^{2n}$ 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. To underline how rare complex manifolds really are we may try to define a compact submanifold of $\mathbf{C}^n$ as the locus of a set of analytic functions

$$f(z^1, ..., z^n) = 0 \ . \tag{28}$$

It turns out that the only connected compact complex manifold that can be obtained in this way is the zero dimensional point—although non-compact complex submanifolds are easy to find.

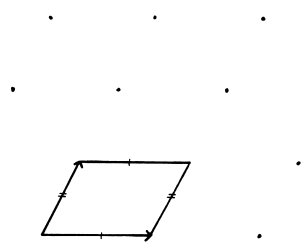

Figure 5: A flat torus is a parallelogram with sides identified; it is also defined by a pair of vectors, or by the lattice of points that can be reached by translations with these two vectors

To get an example of a manifold that can be turned into a complex manifold in several inequivalent ways, consider the *torus* $\mathbf{T}^2 = \mathbf{C}/\Gamma$. 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 $\mathbf{C}$. We can describe the group by the lattice of points in the complex plane that can be reached from the origin by translations belonging to Γ. Topologically it does not matter how we choose our lattice, but it is a fact that in general there is no analytic coordinate transformation connecting two tori constructed from two different lattices. In fact there is a two dimensional "moduli space" of inequivalent complex manifolds with the topology of a torus.

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

$$V = V^a \partial_a + \bar{V}^{\bar{a}} \partial_{\bar{a}} \ , \tag{29}$$

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

$$\mathbf{V^C} = \mathbf{V}^{(1,0)} \oplus \mathbf{V}^{(0,1)} \ , \tag{30}$$

where

$$V^a \partial_a \in \mathbf{V}^{(1,0)} \qquad \bar{V}^{\bar{a}} \partial_{\bar{a}} \in \mathbf{V}^{(0,1)} \ . \tag{31}$$

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.

Hermitian and Kähler manifolds

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

$$ds^2 = 2 g_{a\bar{b}} dz^a d\bar{z}^{\bar{b}} \ . \tag{32}$$

The metric is a symmetric tensor, hence

$$g_{\bar{a}b} = g_{b\bar{a}} \ . \tag{33}$$

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

$$(g_{a\bar{b}})^* = g_{b\bar{a}} \ . \tag{34}$$

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

$$\begin{pmatrix} 0 & g_{a\bar{b}} \\ g_{\bar{a}b} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & g_{1\bar{1}} & g_{1\bar{2}} \\ 0 & 0 & g_{2\bar{1}} & g_{2\bar{2}} \\ g_{\bar{1}1} & g_{\bar{1}2} & 0 & 0 \\ g_{\bar{2}1} & g_{\bar{2}2} & 0 & 0 \end{pmatrix} . \tag{35}$$

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, only "one block" of it.

The metric will preserve its Hermitian 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 $\mathbf{C}^n$ equipped with its flat metric.

Given an Hermitian metric we can construct a differential form

$$J = 2ig_{a\bar{b}}dz^a \wedge d\bar{z}^{\bar{b}} . \tag{36}$$

This trick—to use an n by 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 two-form. The manifold is *Kähler*—and J is said to be a *Kähler form*—if J is closed, that is to say if

$$dJ = 2ig_{a\bar{b},c}dz^c \wedge dz^a \wedge d\bar{z}^{\bar{b}} + 2ig_{a\bar{b},\bar{c}}d\bar{z}^{\bar{c}}dz^a \wedge d\bar{z}^{\bar{b}} = 0 , \tag{37}$$

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

$$g_{a\bar{b},\bar{c}} = g_{a\bar{c},\bar{b}} \qquad g_{a\bar{b},c} = g_{c\bar{b},a} . \tag{38}$$

It is not hard to see that 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

$$g_{a\bar{b}} = \partial_a \partial_{\bar{b}} K . \tag{39}$$

This is clearly a highly non-trivial property.

We have seen that $\mathbf{S}^2$ is a Kähler manifold. This happened because any two-form on a two dimensional manifold is closed by default (there are no

three-forms), so that every Hermitian two dimensional manifold has to be Kähler.

It is interesting—and not very difficult to show—that a complex manifold is Kähler iff its metric *osculates* the flat metric to order two everywhere. This means that one can always find (using analytic coordinate transformations only) a coordinate system such that the metric takes the form

$$ds^2 = 2(\delta_{a\bar{b}} + [2])dz^a dz^{\bar{b}} \ , \tag{40}$$

where [2] stands for an expression that is at least quadratic in the coordinates. This is clearly the case for the 2-sphere.

Parallel transport and curvature

When we do differential geometry we will want to compare vectors belonging to tangent spaces at different points. Hence we need a notion of *parallel transport* of a vector along a curve that connects two different points. In Riemannian geometry the standard way to define parallel transport is known as the *Levi-Civita connection*. In formulæ, we construct the *Christoffel symbols*

$$\Gamma_{ij}{}^{k} = \frac{1}{2} g^{km}(g_{mi,j} + g_{mj,i} - g_{ij,m}) \tag{41}$$

and—given a curve with tangent vector t^i—write down the ordinary differential equation

$$t^j D_j V^i \equiv t^j(\partial_j V^i + \Gamma_{jk}{}^{i}) = \dot{V}^i + t^j \Gamma_{jk}{}^{i} V^k = 0 \ . \tag{42}$$

This will determine the vector V^i all along the curve. In pictures, if the metric comes from an embedding in a flat space the rule is simply that the vector is projected down on the manifold as we go along the curve. A glance at the 2-sphere shows that parallel transport depends on the path; if the curve is closed the transported vector differs from what it was originally. This happens because the 2-sphere is curved. Technically its Riemann tensor $R_{ijl}{}^{k}$ is non-zero;

$$(D_i D_j - D_j D_i)V^k \equiv R_{ijl}{}^{k} V^l \neq 0. \tag{43}$$

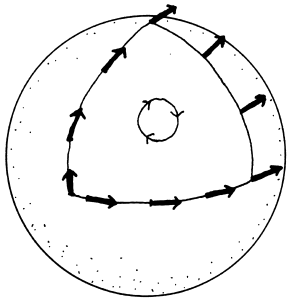

Figure 6: Parallel transport on the sphere

The Levi-Civita connection is constructed in such a way that the length of 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 turns out to be "yes" without any further ado. On a Kähler manifold the only non-vanishing components of the Christoffel symbols are

$$\Gamma_{ab}{}^{c} = g^{c\bar{d}} g_{\bar{d}a,b} \qquad \Gamma_{\bar{a}\bar{b}}{}^{\bar{c}} = g^{\bar{a}d} g_{d\bar{a},\bar{b}} \ . \tag{44}$$

Now take a holomorphic tangent vector, that is a vector of type $(1, 0)$. If the vector is

$$V = V^a \partial_a \tag{45}$$

then the equation for parallel transport becomes

$$\dot{V}^a + t^b \Gamma_{bc}{}^{a} V^c = 0 \ . \tag{46}$$

The point is that the components $\bar{V}$ do not appear in this equation since certain components of the Christoffel symbols are zero. This means that 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. (Parenthetically we remark that if the manifold is Hermitian but not Kähler then dJ is a non-vanishing

three-form which can be used to define a connection which preserves length but which differs from the Levi-Civita connection in that it adds an extra rotation to the transported vector. This is known as a connection with torsion, and it can be used to define a notion of parallel transport that preserves the split of the complexified tangent space also in this more general case. The Kähler case is simpler and also more interesting for our purposes, so we do not go into this here.)

All formulæ become simple on a Kähler manifold. The only non-vanishing components of the Riemann tensor are those that can be obtained by means of index permutations from

$$R_{a\bar{b}c\bar{d}} = g_{\bar{d}d}\Gamma_{ac}{}^{d}{}_{,\bar{b}} \ . \tag{47}$$

The Ricci tensor is

$$R_{a\bar{b}} = -\Gamma_{ca}{}^{c}{}_{,\bar{b}} \ . \tag{48}$$

It is easy to show that these formulæ are implied by the special properties of the metric. Finally, a useful concept is that 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 dz^a and $d\bar{z}^{\bar{a}}$ (and here we use the old fashioned notation according to which dz^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

$$R(z,dz) = \frac{R_{a\bar{b}c\bar{d}}dz^a d\bar{z}^{\bar{b}} dz^c d\bar{z}^{\bar{d}}}{(ds^2)^2} \ . \tag{49}$$

Holomorphic sectional curvature is clearly analogous to ordinary scalar curvature on real manifolds and, unsurprisingly, 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.

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 two-form as well. This is to say that a manifold is *symplectic* iff there exist two tensor fields

$$\Omega_{ij} \qquad \Omega^{ij} \tag{50}$$

(not related by raising indices with a metric—indeed no metric is assumed) such that

$$\Omega_{ij} = -\Omega_{ji} \qquad \Omega^{ik}\Omega_{kj} = \delta^i_j \ ; \tag{51}$$

moreover the *symplectic 2-form* Ω is required to be closed,

$$d\Omega = 0 \quad \Leftrightarrow \quad \Omega_{[ij,k]} = 0 \ . \tag{52}$$

These requirements are actually a bit non-trivial: It may well be that a (compact) manifold does not admit a symplectic structure, although it essentially always admits a metric.

In classical mechanics the 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. The point there is that the symplectic form can be used to associate a vector field with any function $H(x)$ on the manifold through the equation

$$V^i_H = \Omega^{ij}\partial_j H \ . \tag{53}$$

This is known as a *Hamiltonian vector field.* The *Poisson bracket* of two arbitrary functions F and G is then defined by

$$\{F, G\} = \partial_i F \Omega^{ij} \partial_j G \ . \tag{54}$$

It is clearly bilinear and anti-symmetric and it obeys the Jacobi identity precisely because the symplectic form is closed.

On a Kähler manifold the Kähler form is always there to provide the symplectic form. It is non-degenerate because the metric is, and it is closed by definition. 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 it is easy to see that

$$\Omega_{ij} = J_{ij} = \sqrt{g}\epsilon_{ij} \ , \tag{55}$$

where an ϵ-tensor with two indices is available because we are in two dimensions. In higher dimensions it is clear that the manifold has to be rather special if a closed two-form is to arise in a natural manner. Thus when $n > 2$ the spheres $\mathbf{S}^n$ fail to be Kähler manifolds (indeed they are neither complex nor symplectic), but we will find other examples of Kähler manifolds as we proceed. This means that we will have occasion to return to various aspects of complex manifolds that were glossed over in this chapter.

Exercises:

- 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.
- 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.
- 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.
- Verify eq. (48) from scratch—that is, start by verifying eq. (44).

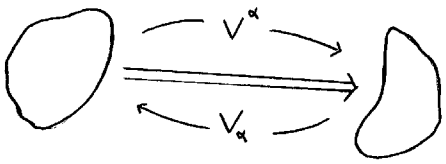

Figure 7: Mapping of vectors

THE 3-SPHERE: A FIBRE BUNDLE

The 3-sphere

All the odd dimensional spheres are fibre bundles. Indeed the theory of fibre bundles was created in response to the treatment that the 3-sphere was given in 1931 by Hopf and (independently) by Dirac. The part of the general theory that is relevant to quantum mechanics is roughly the same as that which comes into play on the odd dimensional spheres, and as always we will develop the mathematics only to the extent that it is directly relevant to quantum mechanics. Almost everything that we will say applies to all odd dimensional spheres, but we will concentrate on the 3-sphere for the simple reason that this is something that can be visualized. The 3-sphere is of course defined as the hypersurface

$$X^2 + Y^2 + Z^2 + U^2 = 1 \tag{56}$$

embedded in a flat four dimensional space with (X, Y, Z, U) as its Cartesian coordinates, and its intrinsic metric is defined by the embedding. It may be worthwhile to recall why, since later on in this chapter we will have to remember which way tensors go under maps between two manifolds: If there is a map from $\mathbf{M}$ to $\mathbf{N}$ then a contravariant tensor on $\mathbf{M}$ is "*pushed forward*" to a contravariant tensor on $\mathbf{N}$, while a covariant tensor on $\mathbf{N}$ is "*pulled back*" to a covariant tensor on $\mathbf{M}$. Our *embedding* is really a map from $\mathbf{S}^3$ to $\mathbf{R}^4$, and the flat metric on $\mathbf{R}^4$ is pulled back to $\mathbf{S}^3$ by this map.

The best way to visualize the 3-sphere is to use a stereographic projection from the South Pole ($U = -1$); then $\mathbf{S}^3$ is conformally mapped to ordinary flat space 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. This training consists in looking at geodesics and Killing vectors in the picture.

A general formula for the geodesics was given in the first chapter. It is easy to prove that they appear in the 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 take a Killing vector field. A straightforward calculation shows that

$$J_{XY} = X\partial_Y - Y\partial_X = x\partial_y - y\partial_x \ , \tag{57}$$

where x, y (and z) are the stereographic coordinates in our picture. This is as simple as it can be. There is a 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

$$J_{ZU} = Z\partial_U - U\partial_Z = 0 \quad \Leftrightarrow \quad Z = U = 0 \ . \tag{58}$$

This is a geodesic (as it must be), namely a great circle on the equator. By analogy with J_{XY} the flowlines 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

$$(\rho - a)^2 + z^2 = a^2 - 1 > 0 \ ; \qquad \rho^2 \equiv x^2 + y^2 \tag{59}$$

invariant. 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".

Now the striking thing about the 3-sphere—and about all odd dimensional spheres—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 $\mathbf{S}^2$—Killing field or no Killing field—has to have a fixed point somewhere. An example of a Killing field without fixed points on $\mathbf{S}^3$ is clearly

$$\xi = J_{XY} + J_{ZU} \ ; \qquad ||\xi||^2 = X^2 + Y^2 + Z^2 + U^2 = 1 \ . \tag{60}$$

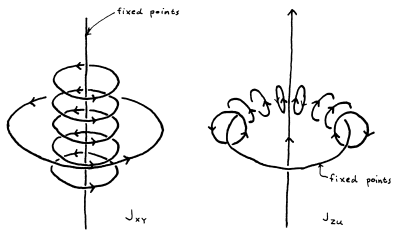

Figure 8: Flow lines and fixed points of J_{XY} and J_{ZU}

Given our pictures of Killing vector fields it is clear that this combination must have flowlines 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 is the key to understanding the 3-sphere as a fibre bundle.

A complex embedding space

We will now prove that the flow lines of the Killing vector field ξ are geodesics as well, and we will do so in a somewhat roundabout way that has the advantage of bringing complex manifolds back in. The point is simply that the embedding space $\mathbf{R}^4$ is also the complex vector space $\mathbf{C}^2$. Therefore we can introduce the complex embedding coordinates

$$\begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix} = \begin{pmatrix} X + iY \\ Z + iU \end{pmatrix} \ . \tag{61}$$

The generalization to n complex dimensions is immediate. We often use P, Q, R ... to denote vectors in complex vector spaces. The scalar product in $\mathbf{R}^{2n}$ now becomes a *Hermitian form* on $\mathbf{C}^n$, namely

$$P \cdot \bar{Q} = \delta_{\alpha\bar{\alpha}} P^\alpha \bar{Q}^{\bar{\alpha}} = P^\alpha \bar{Q}_\alpha \ . \tag{62}$$

Here we made the obvious move of defining

$$(Z^\alpha)^* = \bar{Z}^{\bar{\alpha}} \equiv \bar{Z}_\alpha \ , \tag{63}$$

so that we get rid of the barred indices.

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

$$Z \cdot \bar{Z} = 1 \ . \tag{64}$$

In complex embedding coordinates a general geodesic takes the form

$$Z^\alpha(\tau) = e^{i\tau} m^\alpha + e^{-i\tau} n^\alpha \ , \quad m \cdot \bar{m} + n \cdot \bar{n} = 1 \ , \ m \cdot \bar{n} = 0 \ , \tag{65}$$

where again the affine parameter has been normalized so that the geodesic distance between two points

$$P^\alpha = Z^\alpha(\tau_1) \qquad Q^\alpha = Z^\alpha(\tau_2) \tag{66}$$

is given by

$$d = |\tau_2 - \tau_1| \ . \tag{67}$$

A short calculation reveals that

$$\cos d = \frac{1}{2}(P \cdot \bar{Q} + Q \cdot \bar{P}) \ . \tag{68}$$

This is a useful formula to have.

The everywhere non-vanishing Killing vector field ξ on $\mathbf{S}^3$ is

$$\xi = J_{XY} + J_{ZU} = i(Z^\alpha \partial_\alpha - \bar{Z}^{\bar{\alpha}} \partial_{\bar{\alpha}}) \ . \tag{69}$$

The latter expression can be used on $\mathbf{S}^{2n+1}$ for any n. Now consider the family of geodesics given by

$$Z^\alpha = e^{i\tau} m^\alpha \ . \tag{70}$$

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

$$\dot{Z}^\alpha = i Z^\alpha \tag{71}$$

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

$$\partial_\tau = \dot{Z}^\alpha \partial_\alpha + \dot{\bar{Z}}^{\bar{\alpha}} \partial_{\bar{\alpha}} = i(Z^\alpha \partial_\alpha - \bar{Z}^{\bar{\alpha}} \partial_{\bar{\alpha}}) \ . \tag{72}$$

But this is precisely the everywhere non-vanishing Killing vector field that we found before. So we have found that $\mathbf{S}^{2n+1}$ can be foliated with a family

Figure 9: Diverging geodesics in flat space, converging in curved space, and Clifford parallells

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.

Clifford parallels

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 the question naturally arises whether 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. Actually a more stringent definition requires a notion of parallel transport that always takes tangent vectors of the Clifford parallels into each other; we will return to discuss this point.

As a matter of fact 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

$$P^\alpha = e^{i\tau} P_0^\alpha \qquad Q^\alpha = e^{i(\tau+\tau_0)} Q_0^\alpha \ . \tag{73}$$

We will soon exercise our right to choose the constant τ_0. The scalar product of the constant vectors will be some complex number

$$P_0 \cdot \bar{Q}_0 = re^{i\phi} \ . \tag{74}$$

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

$$\cos d = \frac{1}{2}(P \cdot \bar{Q} + Q \cdot \bar{P}) = r\cos(\phi - \tau_0) \ . \tag{75}$$

The point is that this is independent of the affine parameter τ, 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 τ_0 so that $\cos d_0 = r$, i.e. so that the distance attains its minimum value. A concise way to express d_0 is by means of the equation

$$\cos^2 d_0 = r^2 = P_0 \cdot \bar{Q}_0 Q_0 \cdot \bar{P}_0 = P \cdot \bar{Q} Q \cdot \bar{P} \ . \tag{76}$$

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

The final touch of this discussion concerns the question whether there is a notion of parallel transport that takes the tangent vectors of two putative Clifford parallels into each other. It turns out that on the 3-sphere there is a tensor with three indices that can be added to the Levi-Civita connection $\Gamma_{ij}{}^k$ so that the Riemann tensor of the new connection vanishes—in this special sense the 3-sphere is flat. The new connection is said to be a connection with *torsion*; it gives the extra "twist" to parallel transport that we need here. That this works has to do with the fact that $\mathbf{S}^3$ is a *group manifold*—namely the space whose points are the elements of the Lie group $SU(2)$. It does not work for odd dimensional spheres in general, but then it does not matter much for our present purposes.

The Hopf fibration of the 3-sphere

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

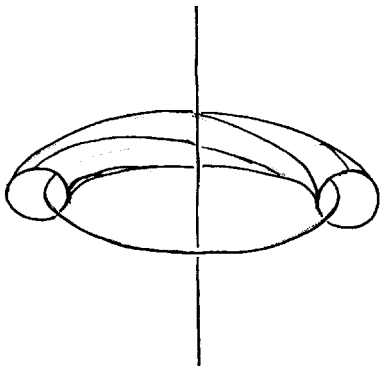

Figure 10: The Hopf fibration of the 3-sphere

get 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. (59). Each torus is itself foliated by a one parameter family of circles and it is interesting to observe that every pair of circles in the congruence are linked with each other (their linking number is one). The whole construction is known as the *Hopf fibration* of the 3-sphere. By the way, the circles are called Villarceau circles. Another remark by the way is that the metric induced on the tori by the metric on the 3-sphere is flat (as you can easily check from eq. (78) 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.

Now the interesting question is: How many circles are there altogether, or more precisely what is the space whose points consists of the Clifford parallels? A little thinking gives the answer directly. On each torus there is a one parameter set of Clifford parallels, 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 Clifford parallels 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 geodesic in the congruence in

some smooth manner. Equivalently, we want to choose the zero point of the coordinate τ 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. We can almost do it though. For instance, we could 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; it maps all of the 2-sphere except one point onto an open unit disk 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 Clifford parallels 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 $\mathbf{S}^3$ of precisely the type that we discussed above. They are defined by

$$\begin{pmatrix} Z^1 \\ Z^2 \end{pmatrix} = \begin{pmatrix} X + iY \\ Z + iU \end{pmatrix} = \begin{pmatrix} e^{\frac{i}{2}(\tau+\phi)} \cos\frac{\theta}{2} \\ e^{\frac{i}{2}(\tau-\phi)} \sin\frac{\theta}{2} \end{pmatrix} . \tag{77}$$

The coordinate τ goes along the Clifford parallels (although the normalization differs with a factor one half from the one that we have used), and a little calculation verifies that the coordinate θ labels the tori in eq. (59), with $\cos(\theta/2) = 1/a$. The intrinsic metric on the 3-sphere becomes

$$ds^2 = |dZ^1|^2 + |dZ^2|^2 = \frac{1}{4}(d\tau^2 + d\theta^2 + d\phi^2 + 2\cos\theta d\tau d\phi) . \tag{78}$$

Since the coordinates θ and ϕ labels the 2-sphere's worth of geodesics on the congruence we can try to map this $\mathbf{S}^2$ into $\mathbf{S}^3$ through an equation of the form

$$\tau = \tau(\theta, \phi). \tag{79}$$

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. (76) 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

$$\cos^2 d_0 = \frac{1}{2}\left(1 + \cos\theta_1 \cos\theta_2 + \cos(\phi_1 - \phi_2)\sin\theta_1 \sin\theta_2\right) \ . \tag{80}$$

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

$$\cos^2 d_0 \approx 1 - d_0^2 \equiv 1 - ds^2 \ . \tag{81}$$

A short calculation then verifies that that the metric is

$$ds^2 = \frac{1}{4}(d\theta^2 + \sin^2\theta d\phi^2) \ . \tag{82}$$

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.

Fibre bundles

Let us now bring in the general theory of *fibre bundles*. 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: By definition 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 Us 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.

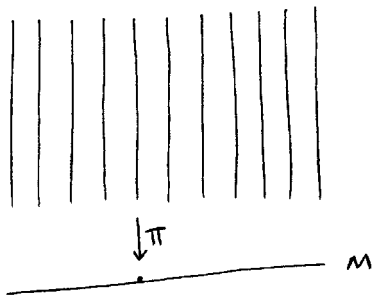

Figure 11: A fibre bundle—the space of fibres is M

A *Cartesian product* $M \times F$ (as in $\mathbf{R}^2 \equiv \mathbf{R} \times \mathbf{R}$) 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. To keep things simple in this chapter we will from now until further notice limit ourselves to the case of one dimensional fibres only.

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 coordinates on the base manifold and τ a coordinate along the fibres. 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

$$x'^a = x'^a(x) \qquad\qquad \tau' = \tau'(x, \tau) \; . \tag{83}$$

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,

$$\tau = \tau(x) \; . \tag{84}$$

This defines an embedding of the base manifold into the bundle. For a non-trivial bundle such as the 3-sphere we have seen that 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

$$\tau = \tau(x) \quad \rightarrow \quad \tau' = \tau'(x) \; . \tag{85}$$

Such a transformation along the fibres is known as a *local gauge transformation*, for reasons that will become more evident later on.

One can show that 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!

Parallel transport on the 3-sphere

It is essential to realize that there is no canonical way in which the fibres in a fibre bundle can be identified with each other. Let us try to see why this is so, using the 3-sphere as our example. The natural way to proceed is clearly to start from a point on a given fibre and then to draw a curve to some other fibre that we wish to identify point by point with the first one. Let σ be a parameter along the curve. Since the 3-sphere carries a natural metric it is natural to prescribe that the tangent vector

$$\partial_\sigma = \dot{\tau}\partial_\tau + \dot{\theta}\partial_\theta + \dot{\phi}\partial_\phi \equiv t^i \partial_i \tag{86}$$

of the connecting curve shall be orthogonal to the tangent vector $\partial_\tau \equiv \xi^i \partial_i$ of the fibre. Hence we require

$$g_{ij}\xi^i t^j = 0 \; . \tag{87}$$

Using eq. (78) for the metric we find that this equation becomes

$$g_{\tau\tau}\dot{\tau} + g_{\tau\theta}\dot{\theta} + g_{\tau\phi}\dot{\phi} = \dot{\tau} + \cos\theta\dot{\phi} = 0 \; . \tag{88}$$

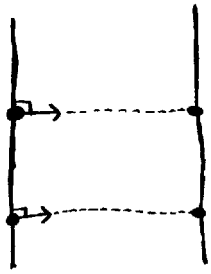

Figure 12: Trying to identify a pair of fibres

It tells us how τ changes with θ and ϕ along the curve. The operation is often referred to as *parallel transport*, but it is a somewhat different parallel transport than the one encountered in the previous chapter—then we were transporting tangent vectors, now we are transporting the position along the fibres.

A priori parallel transport can be defined in many ways. Our definition relied on the existence of a preferred metric on the 3-sphere. On a general fibre bundle this option will not be there, since there may be no natural metric. In any case we have in fact not solved the problem of identifying the fibres, not even for the 3-sphere. Let us look at the situation in this way: Choose any curve on $\mathbf{S}^2$, and also choose an arbitrary point on a particular fibre. What we have shown is that the curve can be "lifted" in a unique way to a curve in $\mathbf{S}^3$ that projects to the specified curve and that goes through the specified point on the given fibre. When the curve passes another fibre it will pass through some unique point there. Can this result be used to identify points on different fibres? The answer is no, because the identification will depend on the curve selected in the base manifold. To see this, consider parallel transport around a closed curve ∂S (assumed to bound a region S) on the 2-sphere. We find that

$$\tau_{final} - \tau_{initial} = -\int_{\partial S} \cos\theta d\phi = \int_S \sin\theta d\theta d\phi \ , \tag{89}$$

where we used Stokes' theorem in the last step. The conclusion is that when we come back to the original fibre the coordinate τ has changed by an amount that is proportional to the area enclosed by the curve in the base manifold. Hence the operation of parallel transport depends on the path, a situation that is often described by saying that there is a non-trivial *holonomy*. This

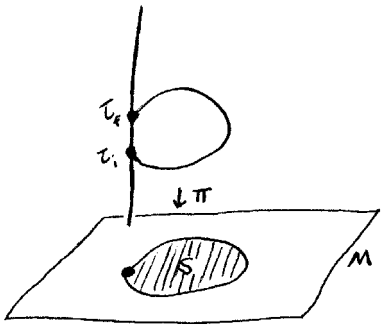

Figure 13: Holonomy

is the origin of the geometric phase that occurs in cyclic quantum evolution, as we will see.

Just so that there can be no misunderstanding, note that the curves may well lie in a single patch of the form $U \times F$. Hence the path dependence of parallel transport is to be expected also in fibre bundles with trivial topology.

Connections on principal fibre bundles

The operation of parallel transport on a fibre bundle deserves to be formalized. The assumption that the fibres be one dimensional simplifies the discussion quite a bit—that is in fact why we make it. Anyway, what we need is a decomposition of the tangent space into "*vertical*" and "*horizontal*" directions, that is directions that are respectively tangential and "perpendicular" to the fibres. Such a structure is called a *connection* by mathematicians. To a physicist a connection is the same thing as the vector potential in electrodynamics, and it may not be altogether evident that the meaning is the same. To make things a little more difficult we will not assume that we have any natural metric on the bundle space, so that there is no definition of "perpendicular" to rely on.

We want to decompose tangent space so that

$$\mathbf{T}_p = \mathbf{V}_p \oplus \mathbf{H}_p \ . \tag{90}$$

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 $\mathbf{H}_p$. It is clear what a vector belonging to the former should like in our adapted coordinate system: Since it should point along the fibres it must be proportional to

$$\partial_\tau \in \mathbf{V}_p \ . \tag{91}$$

For the 3-sphere we declared that any vector that is orthogonal to a vertical vector is horizontal, but if we do not have a natural metric available it is not so clear what a horizontal vector should look like.

The trick is to choose a special one-form ω (that is a covariant vector ω_i) and to declare that a horizontal vector W^i is any vector that obeys

$$W^i \in \mathbf{H}_p \quad \Leftrightarrow \quad \omega_i W^i = 0 \ . \tag{92}$$

Such a statement does not require any metric. But how do we choose ω? The obvious guess $d\tau$ has a problem since we are permitting coordinate transformations under which

$$d\tau = \frac{\partial \tau}{\partial \tau'} d\tau' + \frac{\partial \tau}{\partial x'^a} dx'^a \ . \tag{93}$$

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

$$\omega = d\tau + U = d\tau + U_a dx^a \ . \tag{94}$$

In this way a connection, that is a decomposition of the tangent space, is equivalent to the specification of a special one-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. We can lift this curve to a curve in the bundle by insisting that the one-form induced on the lifted curve by the connection one-form ω vanishes. Thus

$$\omega = d\tau + U = \left(\frac{d\tau}{d\sigma} + U_a \frac{dx^a}{d\sigma} \right) d\sigma = 0 \ . \tag{95}$$

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

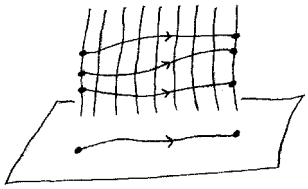

Figure 14: A connection, and the horizontal lift of a curve from the base manifold

associate a unique family of curves $(\tau(\sigma), x^a(\sigma))$ in the bundle to any curve in the base manifold. Through any point in the fibre there passes a unique curve. It is called the *horizontal lift* of the curve in the base manifold.

The condition that parallel transport be independent of the path, or in other words that the horizontal lift of a closed curve is itself closed, is just

$$\int_{\partial S} U = \int_S dU = 0 \ , \tag{96}$$

where S is any surface enclosed by the curve ∂S in the base manifold. The 2-form

$$\Omega = d\omega = dU = \frac{1}{2}(\partial_a U_b - \partial_b U_a) dx^a \wedge dx^b \tag{97}$$

is known as the *curvature 2-form*. This formula is valid only for the case of one-dimensional fibres; a similar but non-linear formula exists in the general case but we will not use it much.

Back to the 3-sphere

Let us see how these things work out for the 3-sphere. The connection that we used was defined by

$$\omega = d\tau + \cos\theta d\phi \ . \tag{98}$$

This is clearly an example of the general construction; it is a preferred example since it was suggested to us by the natural metric on $\mathbf{S}^3$. It is a perfectly regular one-form on the bundle.

We will now attempt to take a section of $\mathbf{S}^3$, that is to say that we wish to view the base manifold $\mathbf{S}^2$ as an embedded submanifold of the 3-sphere so that ω induces a one-form on the 2-sphere. Physicists normally refer to this induced one-form as the connection. We will work in some coordinate patch, say using the Euler angles. Then a local section is defined by the equation

$$\tau = \phi \ . \tag{99}$$

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

$$\omega^+ = (1 + \cos\theta) d\phi \ . \tag{100}$$

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

$$\omega^+ = \frac{i}{1 + |z|^2} \left(-\frac{dz}{z} + \frac{d\bar{z}}{\bar{z}} \right) \ . \tag{101}$$

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

$$\omega^+ = \frac{i}{1 + |z'|^2} (\bar{z}' dz' - z' d\bar{z}') \ , \tag{102}$$

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 $\tau = -\phi$ it will be well-defined everywhere except at the South Pole; explicitly

$$\omega^- = (\cos\theta - 1) d\phi \ . \tag{103}$$

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

$$\omega^+ - \omega^- = 2d\phi \ . \tag{104}$$

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

$$\Omega_{\theta\phi} = \sin\theta \tag{105}$$

is an everywhere regular 2-form on the base manifold—it is in fact equal to a constant factor times the area element of $\mathbf{S}^2$.

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 explains what Dirac was doing in 1931, while Hopf was fibering the 3-sphere. More to the point as far as we are concerned the next two chapters are going to reveal that the Hopf fibration of the odd dimensional spheres is a key ingredient in the geometry of the pure states in quantum mechanics.

Generalizations

Possibly the reader has had enough of fibre bundles for the time being; nevertheless some generalizations must be mentioned. First of all the assumption that the fibres be one dimensional (and the structure group G Abelian) was made for convenience only. The general case is important in physics, and it does occur in quantum mechanics (for instance in the treatment of the geometric phase for degenerate Hamiltonians). 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. The main difficulty is that it is hard to get explicit about the holonomies—Stokes' theorem is no longer helpful because the simple line integrals along closed curves that we used above must now be "path ordered" due to the fact that the connections at different points along the curve do not commute with each other.

We get a rich supply of new fibre bundles if we first look at the 3-sphere, and more especially at those properties that are special to the 3-sphere. Indeed $\mathbf{S}^3$ and $\mathbf{S}^1$ are the only spheres that are also group manifolds. Given a group its group manifold is the space whose points are the elements of the group; the group is a Lie group if whenever $g_3 = g_1 g_2$ then the coordinates of the group element g_3 is "nice" function of the coordinates of g_1 and g_2. A standard example of a Lie group is $SU(2)$, the group of unitary two-by-two matrices with unit determinant. Such matrices can be parametrized by two complex numbers Z^1 and Z^2 obeying one condition, that is to say that any $SU(2)$ matrix can be written in the form

$$g = \begin{pmatrix} Z^1 & iZ^2 \\ i\bar{Z}^2 & \bar{Z}^1 \end{pmatrix} , \qquad |Z^1|^2 + |Z^2|^2 = 1 . \tag{106}$$

We see that indeed there is a one-to-one correspondence between the elements of $SU(2)$ and points on the 3-sphere. Moreover the group manifold of $SU(2)$ is a round 3-sphere in a natural way. This is not the place to present group theory in full detail, but let us see how it works. The basic point is that the group acts on itself; indeed a given group element h acts on the group manifold in three ways, through the *left action* $g \to hg$, the *right action* $g \to gh^{-1}$ and the *adjoint action* $g \to hgh^{-1}$. These three transformations preserve distances as defined by the round metric, and in this sense the latter is natural. To see this, let us choose a one parameter subgroup of $SU(2)$ that consists of all matrices of the form

$$h(\tau) = \begin{pmatrix} e^{-i\tau} & 0 \\ 0 & e^{i\tau} \end{pmatrix} , \qquad 0 \leq \tau < 2\pi . \tag{107}$$

This subgroup considered by itself is the group $U(1)$, whose group manifold is a circle (parametrized by the real number τ). Next let the group act on itself by right action, say. Letting τ vary this yields a family of curves through the group manifold, namely

$$g(\tau) = g_0 h^{-1}(\tau) \quad \Leftrightarrow \quad Z^\alpha = e^{i\tau} Z_0^\alpha . \tag{108}$$

But this family of curves is by now very familiar to us: It is the congruence of circles that constitute the Hopf fibration. Choosing other $U(1)$ subgroups will just have the effect of rotating this congruence, while using left instead of right action will produce Hopf fibrations of the opposite "twist". Clearly

it is natural to define a metric on $SU(2)$ through the requirement that left and right action produce Killing vector fields on the group manifold—then $SU(2)$ becomes a round 3-sphere.

The way to generalize this way of looking at the 3-sphere is this: Choose any compact Lie group G and any subgroup H. Then form the *coset space* G/H by identifying all points in the group manifold of G that can be reached from each other through right action with H. The result is a principal fibre bundle with bundle space G whose base manifold is the space of *orbits* of H, namely G/H. This construction will occur again and again. To see how common it is we show that every sphere is a coset space: First recall that a sphere $\mathbf{S}^n$ is by definition the space of unit vectors in the vector space $\mathbf{R}^{n+1}$. Now nothing prevents us from regarding a given unit vector as the first vector in an orthonormal *frame* (that is basis) in $\mathbf{R}^{n+1}$. Of course there will be many frames corresponding to each unit vector. But what is the space of frames? We can go from any frame to any other frame by means of a rotation and conversely every rotation will transform the frames. This means that the space of frames can be identified with the group manifold of the rotation group $SO(n+1)$. Given a frame whose first member is some particular unit vector, all those frames that can be reached from this one through a rotation that leaves the first member fixed correspond to the same unit vector. But that subgroup of $SO(n+1)$ that leaves a given unit vector invariant is $SO(n)$. Pulling this argument together we find that

$$\mathbf{S}^n = SO(n+1)/SO(n) \ . \tag{109}$$

Further study reveals that the round metric on $\mathbf{S}^n$ arises in a natural manner from the natural metric on $SO(n+1)$.

In a different direction 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. We may 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 is the Möbius strip. A very important example is the tangent bundle over a space $\mathbf{M}$. Then the fibre at a point p is the tangent space $\mathbf{T}_p$ at this point, which is indeed a vector space. The group of general linear transformations acts on the fibres

of the tangent bundle. A section of the tangent bundle is something quite undramatic: One element of the fibre at each point in the base space, that is to say a vector field. If you think about it, you see that parallel transport of vectors as defined in the previous chapter is an operation that takes place in the tangent bundle.

Exercises:

- Verify that geodesics on $\mathbf{S}^3$ meet the equator in two antipodal points.
- 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.
- 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.
- We found that $\mathbf{S}^2 = SU(2)/U(1) = SO(3)/SO(2)$. Now $SO(2) = U(1)$ and $SU(2) = SO(3)/Z^2$. Is there a contradiction?

COMPLEX PROJECTIVE SPACES

A definition of $\mathbf{CP}^n$

We have seen in detail that the Hopf fibration of the 3-sphere gives the ordinary sphere. The same construction can be done for spheres of arbitrary odd dimension $2n+1$. It turns out that the base manifold will always be a complex manifold of real dimension $2n$ known as *complex projective n-space*, $\mathbf{CP}^n$. The most effective presentation of $\mathbf{CP}^n$ starts directly in the flat vector space $\mathbf{C}^{n+1}$. In quantum mechanics this vector space is known as Hilbert space, and we have indeed arrived at the geometry of pure states in quantum mechanics but we do not make that connection just yet, so this chapter will stay on the level of pure mathematics. Complex projective n-space is the space of one dimensional subspaces in the vector space, that is to say that its points are in one-to-one correspondence with the equivalence classes

$$Z^\alpha \sim \lambda Z^\alpha \ , \qquad \lambda \neq 0 \ . \tag{110}$$

The coordinates Z^α on $\mathbf{C}^{n+1}$ are known as *homogeneous coordinates* on $\mathbf{CP}^n$; any algebraic expression which is homogeneous in these coordinates is a well defined expression on the projective space.

To see that this definition is equivalent to the Hopf fibration of the odd dimensional spheres we observe that we can choose the absolute value of λ so that

$$Z \cdot \bar{Z} = 1 \ . \tag{111}$$

This is $\mathbf{S}^{2n+1}$, and we still have to identify all points on the sphere that are related by

$$Z^\alpha \sim e^{i\phi} Z^\alpha \ . \tag{112}$$

This is precisely the Hopf fibration, and it follows that

$$\mathbf{CP}^n = \mathbf{S}^{2n+1}/\mathbf{S}^1 \ . \tag{113}$$

We will discuss the fibre bundle aspects of this construction in more detail later. But first we wish to discuss the reason why the base manifold is called a projective space. What is projective geometry?

Real projective space

To understand projective geometry it is helpful to begin with *real projective space* $\mathbf{RP}^n$, which is defined as the space of one dimensional subspaces in a real vector space of dimension $n+1$. The advantage gained by looking at the real case is that we can really do this by looking, at least if we keep n small. Then we turn to formulæ when we discuss complex projective space—and many of our observations in the real case will be found to hold true in the complex case as well. We gave a brief discussion of the topology and geometry of $\mathbf{RP}^n$ in chapter 2, where we regarded it as an n-sphere with antipodal points identified. For the time being we wish to forget about the metric and see what kind of geometry that can be formulated without it.

First we visualize the vector space $\mathbf{R}^3$. The projective plane is the space of *rays* or lines through the origin. How can we coordinatize these? One possibility is to use homogeneous coordinates: To any point in $\mathbf{R}^3$ (given by a triple of numbers) we can assign the ray that passes through this point. The ray does not determine these triples of numbers uniquely since if we multiply all three numbers with a common factor we get the same ray. Hence a point of $\mathbf{RP}^2$ is given by an equivalence class

$$(X^0, X^1, X^2) \sim k(X^0, X^1, X^2) \; ; \qquad k \neq 0 \; . \tag{114}$$

If we want to use true coordinates one possibility is to choose a plane in $\mathbf{R}^3$ such as $X^0 = 1$. We think of this plane as an affine plane, which means that it is like a vector space but 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 called 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 the vector space $\mathbf{R}^2$ and the space of rays lying in it is the projective line $\mathbf{RP}^1$. Hence the affine coordinates display the projective plane as an infinite plane to which we must add an extra projective line "at infinity".

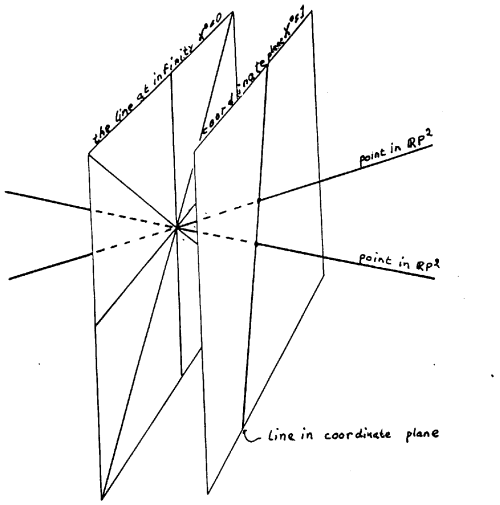

Figure 15: Affine coordinates for (almost all of) the projective plane

Projective geometry originated in Renaissance art, when painters discovered that parallel lines on the ground (a flat plane) must intersect on a "*vanishing line*" that exists on the canvas but not on the ground. This happens because both the ground and the canvas can be regarded as different affine coordinate systems for the same projective plane; the points of this projective 2-space are the rays in a vector space having the eye of the painter as its origin. A line, either on the canvas or on the ground, then arises as the intersection of the canvas or the ground with a two dimensional subspace of the vector space. The vanishing line is given by a plane that is parallel to the ground.

There is nothing special about the line "at infinity"; the name is just an artifact of a special choice of affine coordinates. Every projective line is the space of rays in some two dimensional subspace of $\mathbf{R}^3$ and it has the topology of a circle—to its image in the affine plane we must add one point "at infinity". There are obvious advantages in adding the line at infinity to the affine plane since we no longer have to worry about exceptional pairs of parallel lines in formulating theorems about *incidence*. Thus it is true that two points on the plane (affine or projective) determine a unique line, but on the projective plane it is also true that every pair of lines determines a

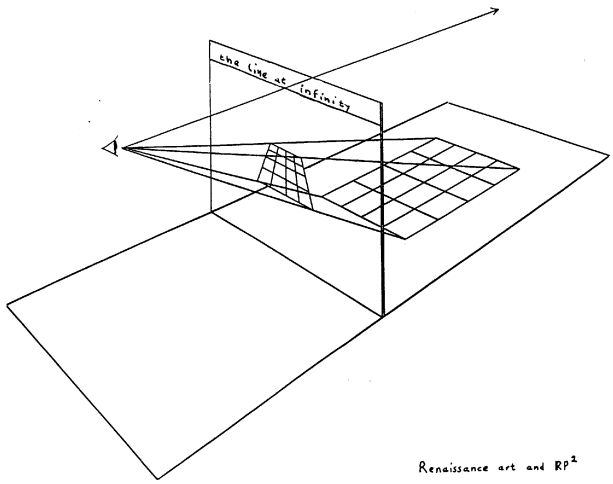

Figure 16: Renaissance art and $\mathbf{RP}^2$

unique point of intersection (since a pair of two dimensional subspaces in $\mathbf{R}^3$ intersect in a unique ray). Projective geometry arose as the study of such incidence properties between points and lines. It is interesting to observe that the space of lines in $\mathbf{RP}^2$ is another $\mathbf{RP}^2$ since one can set up a one-to-one correspondence between rays and planes through the origin in $\mathbf{R}^3$. In fact there is a complete *duality* between the space of points and the space of lines in the projective plane—and this is where the word "dual" first entered mathematics. The natural way to select a specific one-to-one correspondence between $\mathbf{RP}^2$ and the dual $\mathbf{RP}^2$ is to introduce a metric in the vector space, and to declare that a ray there 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 on the sphere as the natural metric to use on $\mathbf{RP}^2$, along the lines of chapter 2. 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 $\mathbf{R}^3$. In one affine coordinate system the points of intersection of the cone with the affine plane will appear as a circle, in

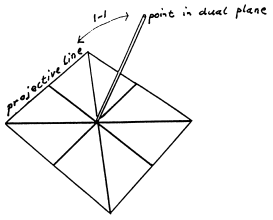

Figure 17: Polarity: A metric for the projective plane

another as a hyperbola. Hence circles, ellipses and hyperbolæ are projectively equivalent. In homogeneous coordinates such a conic section appears as

$$(X^1)^2 + (X^2)^2 = (X^0)^2 \ . \tag{115}$$

This equation defines a submanifold of $\mathbf{RP}^2$ 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 $\mathbf{RP}^2$: If we draw our space as a disk with antipodal points on the boundary identified—this would be the picture of the topology implied by the discussion in the first chapter—then we see that the hyperbola becomes a topological circle because we are adding two points "at infinity". Its interior is an ordinary disk while its exterior is a *Möbius strip* because of the identifications at the boundary—if you glue a disk to the boundary of a Möbius strip you obtain a projective plane. By the way $\mathbf{RP}^2$, unlike $\mathbf{RP}^1$, 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. In general $\mathbf{RP}^{2n+1}$ is orientable while $\mathbf{RP}^{2n}$ is not.

If you think a little bit further about the $\mathbf{RP}^2$ 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 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

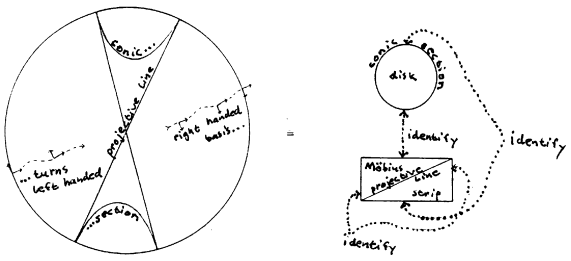

Figure 18: The topology of the real projective plane

section.

Complex projective space

Projective geometry can be defined starting from vector spaces defined over arbitrary fields. Real projective space $\mathbf{RP}^n$ and *quaternionic projective space* $\mathbf{HP}^n$ do have roles to play in quantum mechanics, but it is the complex projective space $\mathbf{CP}^n$ that is all important. Moreover complex projective space is in some respects easier to study than is its real cousin, mainly because the complex numbers form a *closed field* (in the sense that every polynomial equation has a root). Therefore we now switch to the complex case, although much of the discussion can be taken over to the real case in an obvious way.

By definition $\mathbf{CP}^n$ is the space of rays in the complex vector space $\mathbf{C}^{n+1}$, or equivalently the space of equivalence classes of $n+1$ complex numbers under

$$(Z^0, Z^1, \ \ldots\ , Z^n) \sim \lambda(Z^0, Z^1, \ \ldots\ , Z^n)\ ; \qquad \lambda \neq 0 \quad \lambda \in \mathbf{C}\ . \tag{116}$$

We can cover $\mathbf{CP}^n$ with coordinate patches like

$$Z^\alpha = (1, z^1, \ \ldots\ , z^n)\ , \qquad Z^\alpha = (z'^n, 1, z'^1, \ \ldots\ , z'^{n-1}) \tag{117}$$

and so on. This leads to two important observations. The first is that one coordinate patch covers all of $\mathbf{CP}^n$ except that subset which has (say)

$$Z^\alpha = (0, Z^1, \ \ldots \ , Z^n) \ . \tag{118}$$

But the set of all such rays is a $\mathbf{CP}^{n-1}$. Hence we conclude that topologically $\mathbf{CP}^n$ is like $\mathbf{C}^n$ with a $\mathbf{CP}^{n-1}$ attached "at infinity". The second observation is that in a region where the coordinate systems overlap they are related by (say)

$$z'^a = \frac{Z^{a+1}}{Z^1} = \frac{Z^0}{Z^1}\frac{Z^{a+1}}{Z^0} = \frac{z^{a+1}}{z^1} \ , \tag{119}$$

which is clearly an analytic function on the overlap region. The conclusion this time is that $\mathbf{CP}^n$ is a complex manifold.

Of major importance are the *linear subspaces* of $\mathbf{CP}^n$. They are defined as the images of the subspaces of $\mathbf{C}^{n+1}$ 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 $\mathbf{CP}^n$ defined by the equation

$$k_\alpha Z^\alpha = 0 \tag{120}$$

for some fixed k_α. This definition is unaffected by a change of scale for the homogeneous coordinates, and also by a change of scale in the numbers k_α. Hence the $n+1$ complex numbers k_α themselves form a $\mathbf{CP}^n$; 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 $m \leq n$ independent linear equations on the Z^αs we obtain a linear subspace of complex dimension $n-m$ which is itself a $\mathbf{CP}^{n-m}$. Geometrically this is the intersection of m hyperplanes. The space whose points consist of such $n-m$ dimensional subspaces of $\mathbf{CP}^n$ 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 $\mathbf{CP}^n$. 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 $\mathbf{CP}^0$ is just a point. The complex projective line is a $\mathbf{CP}^1$—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 $\mathbf{C}^{n+1}$ 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 a 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.

All statements about linear subspaces—such as when two linear subspaces intersect—are invariant under general linear transformations of $\mathbf{C}^{n+1}$. This is the group $GL(n+1, \mathbf{C})$, that is to say the group of all complex $(n+1)\times(n+1)$ matrices. It is only a subgroup that acts *effectively* on $\mathbf{CP}^n$. (A transformation is said to act effectively on $\mathbf{CP}^n$ if it moves at least one point—but some elements of $GL(n+1, \mathbf{C})$ do not since they only change the homogeneous coordinates with an irrelevant factor). The *projective group* is $GL(n+1, \mathbf{C})$ with the irrelevant transformations divided out. According to Klein's conception of geometry projective geometry is completely characterized by the projective group; its subgroups include the group of affine transformations that preserves the $\mathbf{CP}^{n-1}$ at infinity and that 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 $(1, 0, \ ... \ , 0)$, $(0, 1, \ ... \ , 0)$, ... ; $(1, 1, \ ... \ , 1)$ by means of a projective transformation. For $\mathbf{CP}^1$ 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.

The stellar representation and the spinor notation

The *stellar representation* is a delightful way of visualizing $\mathbf{CP}^n$ in real terms that works for any n. Majorana invented it when he observed that vectors in $\mathbf{C}^{n+1}$ are in one-to-one correspondence with a set of polynomials in one complex variable z, of the form

$$w(z) = Z^0 z^n + Z^1 z^{n-1} + ... + Z^n \ . \tag{121}$$

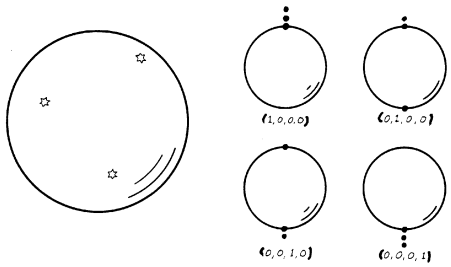

Figure 19: Some points in $\mathbf{CP}^3$

Therefore the set of rays will be in one-to-one correspondence with unordered sets of n complex numbers, namely with the complex roots of the equation

$$Z^0 z^n + Z^1 z^{n-1} + ... + z^{n+1} = 0 = Z^0(z - z_1)(z - z_2)...(z - z_n) \ . \qquad (122)$$

If the multiplicity of a root is larger than one, some of the points in the set coincide. If $Z^0 = 0$ then infinity counts as a root (of multiplicity higher than one if also $Z^1 = 0$). Finally, by means of a stereographic projection the points of $\mathbf{CP}^n$ can be represented as unordered sets of n points on an ordinary 2-sphere. Evidently it follows that $\mathbf{CP}^1 = \mathbf{S}^2$, but $\mathbf{CP}^2 = \mathbf{S}^2 \times \mathbf{S}^2/\Gamma$ where Γ is a discrete group, and so on in higher dimensions.

There is a piece of notation that is conveniently introduced at this juncture. Let us denote the homogeneous coordinates on $\mathbf{CP}^1$ by

$$\zeta^A = (u, v) \ . \qquad (123)$$

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

$$z = \frac{v}{u} \ : \quad \zeta^A \sim (1, z) \ . \qquad (124)$$

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

$$\zeta_A \equiv \zeta^B \epsilon_{BA} \quad \Leftrightarrow \quad \zeta^A = \epsilon^{AB} \zeta_B \ . \qquad (125)$$

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}. Just stick to the convention just made and all will be well. Note that $\zeta_A \zeta^A = 0$.

We can go on to consider totally symmetric *multispinors*

$$\Psi^{AB} = \Psi^{(AB)} , \qquad \Psi^{ABC} = \Psi^{(ABC)} \tag{126}$$

and so on. (Recall that 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 $\mathbf{CP}^n$. To see that this works, consider the equation

$$\Psi_{AB \ldots M} \zeta^A \zeta^B \ldots \zeta^M = 0 . \tag{127}$$

If we now choose the scale so that $\zeta^A = (1, z)$ then the above equation turns into an n-th degree polynomial in z, having n complex roots. We can use this fact together with eq. (122) 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. (122) we find that

$$\Psi_{AB \ldots M} \zeta^A \zeta^B \ldots \zeta^M = 0 = (\alpha_0 + \alpha_1 z)(\beta_0 + \beta_1 z) \ldots (\mu_0 + \mu_1 z) . \tag{128}$$

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:

$$\begin{aligned} \Psi_{AB \ldots M} \zeta^A \zeta^B \ldots \zeta^M &= \alpha_A \zeta^A \beta_B \zeta^B \ldots \mu_M \zeta^M \\ \Rightarrow & \\ \Psi_{AB \ldots M} &= \alpha_{(A} \beta_B \ldots \mu_{M)} . \end{aligned} \tag{129}$$

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

$\mathbf{CP}^1$) in $\mathbf{CP}^2$ for definiteness. A general point in $\mathbf{CP}^2$ is described by a pair of unordered points on the 2-sphere, or equivalently as a spinor

$$\Psi^{AB} = \alpha^{(A}\beta^{B)} \ . \tag{130}$$

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.

Complex curves, quadrics and the Segre embedding

The equation that defines a subspace of $\mathbf{CP}^n$ does not have to be linear; any homogeneous polynomial equation in the 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 polynomial. The locus of a number of homogeneous equations

$$w_1(Z) = w_2(Z) = \ \ldots \ = w_m(Z) = 0 \tag{131}$$

is also a subspace known as an *algebraic* (or projective) *variety.* If we remember that the only compact complex submanifold of $\mathbf{C}^n$ is a point, it is interesting to learn that all non-singular algebraic varieties are complex submanifolds of $\mathbf{CP}^n$ and conversely that every compact complex submanifold is the locus of a set of homogeneous equations (a statement known as *Chow's theorem*).

There are two kinds of submanifolds that are immediate interest in quantum mechanics. One of them is the *complex curve*; by definition this is a map of $\mathbf{CP}^1$ into $\mathbf{CP}^n$. In real terms it is a 2-surface in a $2n$ dimensional space which does not sound much like a curve, but once it is accepted that $\mathbf{CP}^1$ deserves the name of line it will be admitted that the name curve is reasonable too. 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 $\mathbf{CP}^1$ we clearly get a map into $\mathbf{CP}^2$ if we set

$$(u, v) \quad \rightarrow \quad (u^2, uv, v^2) \ . \tag{132}$$

This is a well defined map because the expression is homogeneous in the u and v. Evidently the resulting complex curve in $\mathbf{CP}^2$ obeys the equation

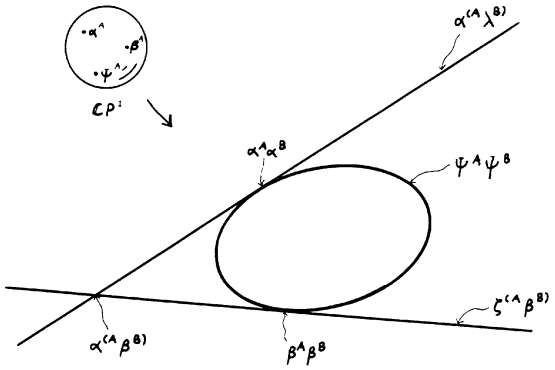

Figure 20: A conic section in $\mathbf{CP}^2$ and a pair of tangent lines

$$Z^0 Z^2 = Z^1 Z^1 \ . \tag{133}$$

Hence it is a quadric hypersurface as well and it is not too difficult to show that 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 spinor notation we get a conic section in $\mathbf{CP}^2$ as the set of points for which the principal spinors coincide. That is to say that

$$\Psi^{AB} = \Psi^A \Psi^B \tag{134}$$

for some spinor Ψ^A. Through any point on the quadric (for which $\Psi^A = \alpha^A$ say) there goes a complex projective line

$$\Psi^{AB} = \zeta^{(A} \alpha^{B)} \tag{135}$$

(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.

In higher dimensions we encounter complex curves that are increasingly difficult to grasp. (For dimensional reasons they cannot be quadric surfaces.) The next simplest case is the *twisted cubic* curve in $\mathbf{CP}^3$, defined by

$$(u, v) \quad \rightarrow \quad (u^3, u^2 v, u v^2, v^3) \ . \tag{136}$$

We leave it aside for the moment though.

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*

$$\mathbf{C}^{n+1} \otimes \mathbf{C}^{m+1} = \mathbf{C}^{(n+1)(m+1)} \ . \tag{137}$$

Then there should be an embedded submanifold

$$\mathbf{CP}^n \times \mathbf{CP}^m \in \mathbf{CP}^{(n+1)(m+1)-1} \ . \tag{138}$$

Indeed this is true. In terms of homogeneous coordinates the submanifold can be parametrized as

$$Z^\alpha = Z^{\mu\mu'} = P^\mu Q^{\mu'}, \tag{139}$$

in a fairly obvious notation—the P^μ and $Q^{\mu'}$ are homogeneous coordinates on $\mathbf{CP}^n$ and $\mathbf{CP}^m$ respectively. (Please do not confuse this notation with the spinor notation—this is something completely different!) 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

$$Z^{\mu\mu'} Z^{\nu\nu'} - Z^{\mu\nu'} Z^{\nu\mu'} = 0 \ . \tag{140}$$

In quantum mechanics the Segre submanifold reappears as the set of separable states of a composite system.

Let us consider the simplest case

$$\mathbf{CP}^1 \times \mathbf{CP}^1 \in \mathbf{CP}^3 \ , \tag{141}$$

when only one equation is needed. Write

$$(Z^0, Z^1, Z^2, Z^3) = (Z^{00'}, Z^{01'}, Z^{10'}, Z^{11'}) \ . \tag{142}$$

Then the submanifold $\mathbf{CP}^1 \times \mathbf{CP}^1$ is obtained as the quadric surface

$$Z^0 Z^3 - Z^1 Z^2 = 0 \ . \tag{143}$$

In general the non-degenerate quadrics in $\mathbf{CP}^3$ are in one-to-one correspondence to all possible embeddings of $\mathbf{CP}^1 \times \mathbf{CP}^1$. 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 $\mathbf{CP}^m$ that is characterized by that particular $Q^{\mu'}$—will be given by the equations

$$\frac{Z^{00'}}{Z^{01'}} = \frac{P^0 Q^{0'}}{P^0 Q^{1'}} = \frac{Q^{0'}}{Q^{1'}} = \frac{P^1 Q^{0'}}{P^1 Q^{1'}} = \frac{Z^{10'}}{Z^{11'}} \ . \tag{144}$$

Since $Q^{\mu'}$ is fixed this implies that there is a complex number λ such that

$$Z^0 = \lambda Z^1 \qquad Z^2 = \lambda Z^3 \ . \tag{145}$$

This is a pair of linear equations and they define a projective line in $\mathbf{CP}^3$. 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

$$\mathbf{RP}^1 \times \mathbf{RP}^1 \in \mathbf{RP}^3 \ . \tag{146}$$

This time we choose to diagonalize the quadric; define

$$Z^0 = X + U \quad Z^1 = X - U \quad Z^2 = V + Y \quad Z^3 = V - Y \ . \tag{147}$$

We then obtain

$$Z^0 Z^3 - Z^1 Z^2 = X^2 + Y^2 - U^2 - V^2 = 0 \ . \tag{148}$$

Now let us choose affine coordinates by dividing through with V. The quadric becomes

$$x^2 + y^2 - u^2 = 1 \ . \tag{149}$$

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).

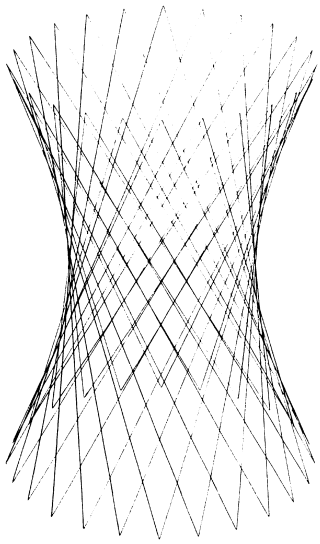

Figure 21: The real Segre embedding: A hyperboloid ruled by straight lines

The Fubini-Study metric

So far we have been thinking of $\mathbf{CP}^n$ as a projective space—and of $\mathbf{CP}^1$ as a projective line. However, $\mathbf{CP}^1$ is naturally a round 2-sphere as well, and it gets this additional structure from the Hopf fibration of the 3-sphere. Since we can obtain $\mathbf{CP}^n$ from the Hopf fibration of a higher dimensional sphere there is a natural metric available for all values of n. This metric is ultimately inherited from the Hermitian form

$$Z \cdot \bar{Z} \equiv Z^\alpha \bar{Z}_\alpha \tag{150}$$

on $\mathbf{C}^{n+1}$. Note that this is the first time that we have brought complex conjugation into the game. Now the Hermitian form provides a polarity, that is to say a specific map from the points of $\mathbf{CP}^n$ to the dual space of hyperplanes. We simply map the point P^α to the hyperplane of points Q^α that obey the equation

$$\bar{P} \cdot Q = 0 \ . \tag{151}$$

As you may recall a polarity on the projective plane gives rise to the natural

metric there, and we are now about to see how this construction generalizes.

With the Hermitian form in hand we can define the distance between two points carrying the homogeneous coordinates P^α and Q^α as the shortest distance between the pair of Clifford parallels that are defined by the equivalence classes

$$P^\alpha \sim e^{i\phi} P^\alpha \tag{152}$$

(and similarly for Q^α) on the odd dimensional sphere $Z \cdot \bar{Z} = 1$. To every point in $\mathbf{CP}^n$, alternatively to each equivalence class $P^\alpha \sim \lambda P^\alpha$, there corresponds precisely one such Clifford parallel, which is why this definition works.

We worked out the required distance in the previous chapter. It only remains to write the result in a manifestly projective invariant form, since we wish to regard all of $\mathbf{C}^{n+1}$ (with the origin deleted) as the bundle space. So we define the *projective cross ratio*

$$\kappa = \frac{P \cdot \bar{Q} Q \cdot \bar{P}}{P \cdot \bar{P} Q \cdot \bar{Q}} \ . \tag{153}$$

We use d to denote the distance between P and Q. Then our definition of distance is

$$\cos^2 d = \kappa \ . \tag{154}$$

This metric is known to mathematicians as the *Fubini-Study metric*.

To obtain the line element in a local form we assume that

$$Q^\alpha = P^\alpha + dP^\alpha \tag{155}$$

and expand to second order in the vector dP^α (and to second order in d). The result is

$$ds^2 = \frac{dP \cdot d\bar{P} P \cdot \bar{P} - dP \cdot \bar{P} P \cdot d\bar{P}}{P \cdot \bar{P} \ P \cdot \bar{P}} = 2 \frac{P^{[\alpha} dP^{\beta]} \bar{P}_{[\alpha} d\bar{P}_{\beta]}}{P \cdot \bar{P} P \cdot \bar{P}} \ . \tag{156}$$

(The square brackets denote antisymmetrization of indices.) The second form of the line element is useful because it makes it manifest that the Fubini-Study metric really is a metric on the space of Clifford parallels—i.e. on $\mathbf{CP}^n$— even though it is written in homogeneous coordinates. The point is that it shows that the line element is invariant under

$$dP^\alpha \to dP^\alpha + zP^\alpha \ , \tag{157}$$

where $z = x + iy$ is some complex number. But the vector xP^α is orthogonal to the sphere, while iyP^α points along the Clifford parallels. The metric has to be unaffected by a change of dP^α in these directions if it is to be a metric on the space of Clifford parallels.

For $n = 1$ the Fubini-Study metric is in fact just the standard round metric on $\mathbf{S}^2$ (with radius one half), but we have some work to do before we are familiar with the higher dimensional case.

A picture of $\mathbf{CP}^2$

We would now like to draw a picture of $\mathbf{CP}^2$ as a real four dimensional manifold, in such a way that the picture generalizes to all n in a straightforward way. A rather brutal way to do so is to introduce real positive numbers n^I and a set of n phases ϕ_i and then scale the homogeneous coordinates so that

$$Z^\alpha \sim (n^0, e^{i\nu_1} n^1, \ \ldots \ , e^{i\nu_n} n^n) \ , \qquad\qquad n^I n^I = 1 \ . \tag{158}$$

We can then look at $\mathbf{CP}^n$ as consisting of n-tori (coordinatized by the phases) lying over the positive hyperoctant of an n-sphere. This picture has its limitations but it is rather amusing to draw it for the low dimensional cases. In particular, let us choose $n = 2$. In these coordinates the Fubini-Study metric becomes

$$ds^2 = dn_0^2 + dn_1^2 + dn_2^2 + n_1^2(1-n_1^2)d\nu_1^2 + n_2^2(1-n_2^2)d\nu_2^2 - 2n_1^2 n_2^2 d\nu_1 d\nu_2 \ . \tag{159}$$

So our picture is metrically accurate as well: We can think of $\mathbf{CP}^2$ as flat tori lying over the positive octant of a round 2-sphere. The metric on the torus at (n^0, n^1, n^2) is

$$g_{ij} = \begin{pmatrix} n_1^2(1-n_1^2) & -n_1^2 n_2^2 \\ -n_1^2 n_2^2 & n_2^2(1-n_2^2) \end{pmatrix} \ . \tag{160}$$

Although the tori are flat their shape depends on their position on the octant; at the edges of the octant they degenerate to circles. Indeed the edges of the

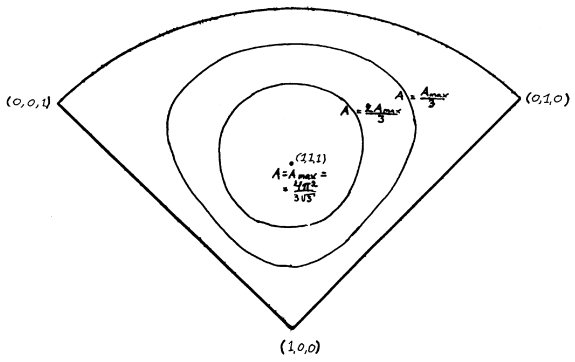

Figure 22: $\mathbf{CP}^2$ from the point of view of $(1,0,0)$: There is a torus over each point of the octant except at the boundaries, where there are circles, and at the corners, which are just points. The level curves show the areas of the tori (zero at the boundaries). The dashed lines are Berger spheres; for comparison a round 3-sphere is drawn as a sequence of rectangular tori. We also give instructions for how to make a similar picture of $\mathbf{RP}^2$. [You can see by inspection that the figure actually shown is incomplete!]

octant represent $\mathbf{CP}^1$s—evidently we can think of a 2-sphere as "circles over a line segment" with the circles degenerating to points at the ends of the segment, i.e. at the corners of the octant.

If we use stereographic coordinates x, y rather than the embedding coordinates n^I it is easy to draw the picture. This is then $\mathbf{CP}^2$ "from the point of view of $(1,0,0)$", the point that occupies the bottom corner. Of course $\mathbf{CP}^n$ is really a homogeneous space and no point is intrinsically special. The corners of the octant look special in the picture only because of the way we introduced our coordinate system.

Over each point in the interior there is a flat torus, and we can decorate our picture by indicating how the shape of the torus changes as its position on the octant is changing. First of all the area of the torus is, using eq. (160),

$$A = \int_0^{2\pi} \int_0^{2\pi} d\nu_1 d\nu_2 \sqrt{g} = 4\pi^2 n_0 n_1 n_2 \ . \tag{161}$$

So we add some level curves for the area to our picture, and observe that the maximal area occurs at $(1,1,1)$. Finally we picture our tori as rhomboids

with edges along the coordinate directions, compute the angle between these edges, and draw some of them at appropriate points of the octants.

A 3-sphere is topologically like tori over a line segment, with the tori degenerating to circles at the ends (for the round 3-sphere, this is just a description of the Hopf fibration). It is easy to identify such topological 3-spheres on the octant. Of special interest are the 3-spheres defined as the set of points lying at constant distance from a given point in $\mathbf{CP}^2$. When this distance is so small that the curvature of the embedding space is unimportant they will look like round 3-spheres, but as their radii grow they will become "squashed". To see how, let us use stereographic coordinates on the octant, i.e.

$$n_0 = \frac{1-r^2}{1+r^2} \qquad n_1 = \frac{2x}{1+r^2} \qquad n_2 = \frac{2y}{1+r^2} \ . \tag{162}$$

This is how our picture was drawn. Then we change coordinates to α and θ, where

$$x = \tan\frac{\alpha}{2}\cos\frac{\theta}{2} \qquad y = \tan\frac{\alpha}{2}\sin\frac{\theta}{2} \ , \qquad 0 \le \alpha \le \frac{\pi}{2} \ , \quad 0 \le \theta \le \pi \ . \tag{163}$$

The special way of coordinatizing the radial direction (with α) is useful because α now measures the proper distance from the origin to the points for which the coordinate takes the value α, as is evident when—in a few simple steps—we compute the Fubini-Study metric in these coordinates. It is

$$ds^2 = d\alpha^2 + \sin^2\alpha\, dl^2 \ , \tag{164}$$

where the metric dl^2 on the squashed sphere at distance α from the origin is

$$\begin{aligned} dl^2 = {} & \frac{1}{4}d\theta^2 + \cos^2\frac{\theta}{2}(1-\sin^2\alpha\cos^2\frac{\theta}{2})d\nu_1^2 + \\ & + \sin^2\frac{\theta}{2}(1-\sin^2\alpha\sin^2\frac{\theta}{2})d\nu_2^2 - 2\sin^2\alpha\cos^2\frac{\theta}{2}\sin^2\frac{\theta}{2}d\nu_1 d\nu_2 \ . \end{aligned} \tag{165}$$

Finally we set

$$\tau = \nu_1 + \nu_2 \qquad \phi = \nu_1 - \nu_2 \tag{166}$$

and obtain

$$dl^2 = \frac{1}{4}(d\theta^2 + d\tau^2 + d\phi^2 + 2\cos\theta d\tau d\phi - \sin^2\alpha(d\tau + \cos\theta d\phi)^2) \ . \qquad (167)$$

In the limit of vanishing α, that is to say when the radius of our sphere is very small, this is just the round metric on the 3-sphere expressed in Euler angles. When $\alpha = \pi/2$ on the other hand $d\tau$ disappears from the metric and the expression collapses to the metric on the 2-sphere, that is to the "$\mathbf{CP}^1$ at infinity" in our picture. Hence our squashed spheres at constant distance from a point in $\mathbf{CP}^2$ interpolate between $\mathbf{S}^3$ and $\mathbf{CP}^1$.

What is actually happening is that we are squashing our spheres along the fibres in the Hopf fibration. It is easy to check that a vector pointing along the fibres, such as $\xi = \partial_\tau$, shrinks as α grows. On the other hand a vector orthogonal to ξ has a length that is independent of α and therefore suffers no squashing. A sphere that is squashed in this particular way is known as a *Berger sphere*.

Since the squashed 3-spheres interpolate between the origin and a complex projective line—topologically a 2-sphere—one may be led to suspect that there may be no family of topological 2-spheres that serve the same end, or otherwise expressed that a complex projective line cannot be deformed to a point within $\mathbf{CP}^n$, in analogy with the case of the real projective lines in $\mathbf{RP}^n$. Indeed this is true, and we will soon find a snappy proof for this.

Volume of $\mathbf{CP}^n$

The coordinate system that we used for our picture is also useful for computing the measure $\sqrt{g}$, and hence the volume, of $\mathbf{CP}^n$. Whatever the dimension we will find that we can think of $\mathbf{CP}^n$ as a set of flat real n-tori lying over the points of a hyperoctant of a round n-sphere. The volume element induced on each torus is

$$\begin{aligned} & n_0 n_1 \ ... \ n_n d\nu_1 \ ... \ d\nu_n = \\ & = \cos\phi \sin\phi \cos\theta_1 \sin^2\theta_1 \ .... \ \cos\theta_{n-1} \sin^n\theta_{n-1} d\nu_1 \ ... \ d\nu_n \ , \end{aligned} \qquad (168)$$

where we used spherical polar coordinates on the sphere in the second step. Using the volume element on the sphere that we found in chapter 1 then leads to the Fubini-Study volume element on the $\mathbf{CP}^n$:

$$\sqrt{g} = \cos\phi \sin\phi \cos\theta_1 \sin^3\theta_1 \cdot \ \ldots \cdot \cos\theta_{n-1} \sin^{2n-1}\theta_{n-1} \ . \tag{169}$$

Computing the total volume now leads to an embarrassingly simple integral, with the answer

$$\mathrm{vol}(\mathbf{CP}^n) = \frac{\pi^n}{n!} \ . \tag{170}$$

Because we fixed the linear scale of the space by the requirement that a closed geodesic has circumference π this goes to zero when n goes to infinity. Asymptotically the volume of $\mathbf{CP}^n$ 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 $\mathbf{CP}^n$ is always larger than that of the sphere (except when $n = 1$, when they coincide).

Differential geometry of $\mathbf{CP}^n$

Let us now return to the case of a complex projective space of arbitrary dimension and see what we can say about the Fubini-Study geometry in general. A key fact is that $\mathbf{CP}^n$ is Kähler manifold. To see this we use the affine coordinates

$$Z^\alpha = (1, z^a) \ . \tag{171}$$

We have already used these coordinates to show that $\mathbf{CP}^n$ is a complex manifold. When expressed in terms of them the Fubini-Study metric becomes

$$ds^2 = \frac{1}{1+|z|^2}\left(dz^a d\bar{z}_a - \frac{\bar{z}_a dz^a d\bar{z}_b z^b}{1+|z|^2}\right) \ , \qquad |z|^2 \equiv z^a \bar{z}_a \ . \tag{172}$$

(We are using $\delta_{a\bar{b}}$ to change $\bar{z}^{\bar{a}}$ into $\bar{z}_a$.) This metric is Hermitian and it is also a Kähler metric; indeed

$$2g_{a\bar{b}} = \frac{1}{1+|z|^2}\left(\delta_{a\bar{b}} - \frac{\bar{z}_a z_{\bar{b}}}{1+|z|^2}\right) = \partial_a \partial_{\bar{b}} \ln\left(1+|z|^2\right) . \tag{173}$$

Other quantities that we are interested in follow quickly:

$$g^{a\bar{b}} = 2(1+|z|^2)(\delta^{a\bar{b}} + z^a \bar{z}^{\bar{b}}) \tag{174}$$

$$\Gamma_{bc}{}^{a} = -\frac{1}{1+|z|^2}(\delta^a_b \bar{z}_c + \delta^a_c \bar{z}_b) \tag{175}$$

$$R_{a\bar{b}c\bar{d}} = -2(g_{a\bar{b}} g_{c\bar{d}} + g_{a\bar{d}} g_{c\bar{b}}) \tag{176}$$

$$R_{a\bar{b}} = 2(n+1) g_{a\bar{b}} . \tag{177}$$

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 $\mathbf{CP}^n$ or else it has the topology of an open unit ball in $\mathbf{C}^n$, depending on the sign of the curvature. (If it vanishes the space is $\mathbf{C}^n$.) 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.

Geodesics

We now bring in some one real dimensional curves in $\mathbf{CP}^n$ (generically called "*threads*" in the older literature). First we have the geodesics. Actually it is enough to study one geodesic, determined as usual once a tangent vector at some specified point is given. This is because of the fact that the isometry group—that we not yet introduced, but we will do so in a moment—can be

used to transform any two points and any two tangent vectors at a point into each other, and hence all geodesics can be transformed into each other by isometries. Indeed the geodesics are all closed circles! The geodesics also have equal lengths. Since

$$\cos^2 d = \kappa \leq 1 \tag{178}$$

the circumference of a closed geodesic is always equal to π, and no two points can be situated at a distance larger than $\pi/2$ from each other. Manifolds whose geodesics enjoy these properties are called C_π *manifolds*, and are quite exceptional. Apart from $\mathbf{CP}^n$, the spheres are obvious examples of C_π manifolds.

To see the geodesic between two arbitrary points in $\mathbf{CP}^n$ we first observe that there is a unique complex projective line between the points. This complex projective line is really a 2-sphere, and the points are points on this 2-sphere. Now it happens that the complex projective line is a *totally geodesic submanifold*, as is the case if and only if a curve on the submanifold which is a geodesic with respect to the intrinsic geometry on the submanifold is a geodesic with respect to the embedding space too. Hence our sought for geodesic between two points in $\mathbf{CP}^n$ is an arc of a great circle on the unique 2-sphere containing the two points.

It only remains to point out that the property of being totally geodesic is non-trivial. In flat space a plane is totally geodesic, but a cylinder is not (even though it is intrinsically flat). Technically a submanifold is totally geodesic if and only if its *extrinsic curvature* vanishes. In $\mathbf{CP}^n$ all the linear subspaces are totally geodesic. The simplest way to prove this is to rely on the fact that all geodesics in $\mathbf{CP}^n$ are equivalent; then all that we have to do is to compute one geodesic in any convenient coordinate system and check that it is contained within a $\mathbf{CP}^1$ and also that it is a great circle there.

Isometries

Now we turn to isometries—and to another important class of threads, namely the Killing flow lines. The Fubini-Study metric is manifestly invariant under *unitary* transformations of the underlying vector space. However, not all of these transformations act effectively on $\mathbf{CP}^n$ (i.e., there are elements of the

unitary group $U(n+1)$ that do not move any points in $\mathbf{CP}^n$, because they only change the phase of the homogeneous coordinates). Indeed we can use the freedom of rescaling the homogeneous coordinates to remove an arbitrary factor from the unitary transformation. It follows that the isometry group must belong to the subgroup $SU(n+1)$. More precisely it is the subgroup

$$SU(n+1)/Z^{n+1} \ . \tag{179}$$

This is a subgroup of the group of all projective transformations. The discrete group Z^{n+1} is given by multiplication with the $n+1$ roots of unity, and we must take it out because multiplication with an overall phase is irrelevant. To see how this goes, consider $n = 1$. Then we know that the isometry group is the group of isometries of the round sphere, which is $SO(3)$. Indeed $SU(2)/Z^2 = SO(3)$, so it works. The reader may object that the group of isometries of the sphere includes reflections and that the true isometry group is $O(3)$. This is correct; there are further discrete isometries present on $\mathbf{CP}^n$. To be precise about it we rely on *Wigner's theorem* (a result that may be familiar from quantum mechanics) to tell us that all isometries of $\mathbf{CP}^n$ arise from unitary or *anti-unitary* transformations of $\mathbf{C}^{n+1}$. Only the former can be reached by exponentiating an infinitesimal isometry, and therefore we do not discuss the latter just yet.

We obtain an infinitesimal isometry by choosing an Hermitian matrix—a generator of $U(n+1)$—and writing

$$i\dot{Z}^\alpha = H^\alpha{}_\beta Z^\beta \ . \tag{180}$$

This equation—also known as the *Schrödinger equation*—determines the Killing flow on $\mathbf{CP}^n$. A part of it represents "gauge"; we are not interested in movement orthogonal to the sphere or along the Clifford parallells, since such movement does not affect the pure state. Therefore we write the projective equation

$$iZ^{[\alpha}\dot{Z}^{\beta]} = Z^{[\alpha}H^{\beta]}{}_\gamma Z^\gamma \ . \tag{181}$$

This equation contains all the information concerning the Killing flow in directions orthogonal to Z, and hence all the information concerning the Killing flow on $\mathbf{CP}^n$. We will call it the *projective Schrödinger equation.*

There is an interesting point to be made here about the velocity of the Killing flow. Using the Fubini-Study metric together with the projective Schrödinger equation we can write

$$\left(\frac{ds}{dt}\right)^2 = 2\frac{Z^{[\alpha}d\dot{Z}^{\beta]}\bar{Z}_{[\alpha}d\dot{\bar{Z}}_{\beta]}}{\bar{Z}\cdot Z\bar{Z}\cdot Z} = \langle H^2\rangle - < H >^2 \ , \tag{182}$$

where a minor calculation precedes the last step and

$$\langle H\rangle \equiv \frac{\bar{Z}_\alpha H^\alpha{}_\beta Z^\beta}{\bar{Z}\cdot Z} \tag{183}$$

and so on. The final result is nice:

$$\frac{ds}{dt} = \sqrt{\langle H^2\rangle - \langle H\rangle^2} \ . \tag{184}$$

The fixed points of the flow occur when the right hand side vanishes, that is to say precisely when

$$H^\alpha{}_\beta Z^\beta = EZ^\alpha \ . \tag{185}$$

If we keep in mind that the eigenvectors of an Hermitian operator are orthogonal we find for the geodesic distance d between the fixed points that

$$\cos^2 d = 0 \quad \Rightarrow \quad d = \pi/2 \ . \tag{186}$$

This holds for any pair of fixed points. Altogether the picture is a nice generalization of the situation for $\mathbf{CP}^1$, where the Killing fields are circles of constant latitude around a pair of antipodal points.

Let us use our naive picture of $\mathbf{CP}^2$ to get a feeling for how the Killing fields look in general. Since any Hermitian matrix can be diagonalized, one can always choose bases so that a given Hermitian three by three matrix takes the form

$$H^\alpha{}_\beta = \begin{pmatrix} E_0 & 0 & 0 \\ 0 & E_1 & 0 \\ 0 & 0 & E_2 \end{pmatrix} \ . \tag{187}$$

The fixed points of the flow—determined by the eigenvectors—therefore occur at the corners of our octant picture (if the matrix is not diagonal in

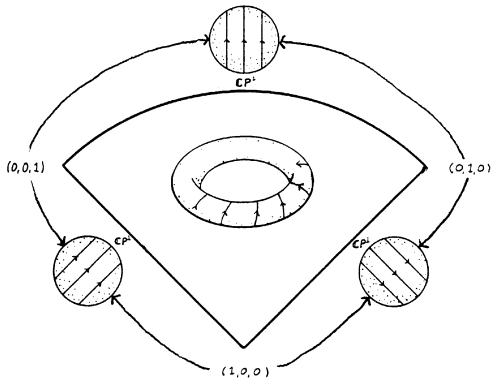

Figure 23: Killing vector flows on $\mathbf{CP}^2$

this basis, we can adapt the picture so that it suits the matrix). If we now exponentiate eq. (180) we find that the isometry becomes

$$\begin{pmatrix} n^0 \\ n^1 e^{i\nu_1} \\ n^2 e^{i\nu_2} \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} e^{-iE_0 t} & 0 & 0 \\ 0 & e^{-iE_1 t} & 0 \\ 0 & 0 & e^{-iE_2 t} \end{pmatrix} \begin{pmatrix} n^0 \\ n^1 e^{i\nu_1} \\ n^2 e^{i\nu_2} \end{pmatrix} \tag{188}$$

where t is a parameter along the flow lines. Taking out an overall phase, we find that this implies that

$$n^I \rightarrow n^I \qquad \nu_1 \rightarrow \nu_1 + (E_0 - E_1)t \quad \nu_2 \rightarrow \nu_2 + (E_0 - E_2)t \ . \tag{189}$$

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. The orbits do not close in general.

Fibre bundle aspects

Let us now return to the fact that $\mathbf{CP}^n$ can be regarded as the base manifold of a bundle whose total space is $\mathbf{S}^{2n+1}$ or even $\mathbf{C}^{n+1}$ with the origin deleted. The latter bundle is known as the *tautological bundle* for fairly obvoius reasons. In both cases there is a preferred connection available that allows us to lift curves in the base manifold to the bundle space in a unique manner—in fact we relied on this already when we wrote down the Schrödinger equation (180), since this is here supposed to be the lift of a Killing vector field to the tautological bundle.

In the case of the 3-sphere the preferred connection was (in Euler angles)

$$\omega = d\tau + \cos\theta d\phi \ . \tag{190}$$

In general a connection picks out a horizontal subspace of the tangent space at every point (while the vertical subspace points along the fibres). Given a connection we can lift any curve in the base manifold to the bundle because a 1-form will be induced on the curve by the connection when we embed the curve in the bundle space; requiring this 1-form to vanish leads to an ordinary differential equation that determines the curve. The connection chosen is preferred in the sense that the lifted curves will be orthogonal to the fibres with respect to the natural metric on the bundle space—in this case the round metric on $\mathbf{S}^3$. Now we rewrite the preferred connection using the embedding coordinates; a straightforward calculation verifies that

$$\omega = -2idZ \cdot \bar{Z} = i(Z \cdot d\bar{Z} - dZ \cdot \bar{Z}) \tag{191}$$

(and the second step follows because $Z \cdot \bar{Z} = 1$). This equation in fact defines a preferred connection on any odd dimensional sphere—the equation $\omega = 0$ expresses the requirement that dZ^α be "Hermitian orthogonal" to Z^α, so that the lifted curve will be perpendicular both to the normal vectors of the sphere and to the Clifford parallells. If we change the homogeneous coordinates with an overall phase the connection transforms as

$$Z'^\alpha = e^{i\lambda} Z^\alpha \quad \Rightarrow \quad \omega' = -2idZ' \cdot \bar{Z}' = \omega + 2d\lambda \ . \tag{192}$$

This is a gauge transformation of the Abelian connection.

Let us see that this is indeed the connection that was implicit in our description of the Killing vector fields. We wrote the Schrödinger equation as

$$idZ^\alpha = H^\alpha{}_\beta Z^\beta dt \ . \tag{193}$$

Evidently

$$\omega = (Z^\alpha \bar{H}_\alpha{}^\beta \bar{Z}_\beta - \bar{Z}_\alpha H^\alpha{}_\beta Z^\beta)dt = 0 \tag{194}$$

along the resulting curve in the bundle, so that this is really the lift of the curve defined in the base manifold by the projective Schrödinger equation.

We would like to have a connection on the tautological bundle, which means that we would like to relax the condition $Z \cdot \bar{Z} = 1$. This poses no problem. The result is

$$\omega = i\frac{Z \cdot d\bar{Z} - dZ \cdot \bar{Z}}{Z \cdot \bar{Z}} \ . \tag{195}$$

This is the preferred connection on the tautological bundle over $\mathbf{CP}^n$.

When we compute the curvature 2-form we will see how economical the whole structure is. A short calculation yields

$$\Omega = d\omega = 2i\frac{Z \cdot \bar{Z} dZ \cdot \wedge d\bar{Z} - dZ \cdot \bar{Z} \wedge Z \cdot d\bar{Z}}{Z \cdot \bar{Z} Z \cdot \bar{Z}} \ . \tag{196}$$

If we compare this expression to the Fubini-Study metric and recall the definition of the Kähler form J we see that

$$\Omega = J \ . \tag{197}$$

The structure could not be tighter. We also see an answer to the question that suggested itself when we showed that the 2-sphere is the only sphere that has the Kähler property: How can a two-form, and a closed two-form to boot, ever arise in a natural manner beyond two dimensions? The answer provided by the complex projective spaces is that it may arise as the curvature tensor of an Abelian connection. Just as it does in electrodynamics.

Finally we can use this machinery to prove that a complex projective line cannot be shrunk to a point within $\mathbf{CP}^n$, even though it is a 2-sphere. In outline, the idea is that the curvature 2-form on the bundle gives rise to a well defined 2-form on the base manifold, while the connection in general does not—if we attempt to embed the base manifold in the bundle space by taking a section of the bundle there are global obstructions that mean that

the connection cannot be globally defined on the base manifold. Now we can take a complex projective line, think of it as a sphere, and compute

$$\int_{\mathbf{S}^2} \Omega > 0 \ . \tag{198}$$

(A modest effort is required to actually do this.) But now suppose that we could shrink our 2-sphere to a point. If so it must be contained in a single coordinate patch. Then there can be no global obstruction to the calculation

$$\int_{\mathbf{S}^2} \Omega = \int_{\mathbf{S}^2} d\omega = \int_{\partial \mathbf{S}^2} = 0 \ , \tag{199}$$

where $\partial \mathbf{S}^2$ is the boundary of $\mathbf{S}^2$, which is zero because there is no boundary. This is a contradiction showing that global obstructions must be present, and therefore one cannot shrink the 2-sphere in this way.

Generalizations

The story is over for the moment, although we will return to $\mathbf{CP}^n$ later and provide physical content and further perspective. Certain important generalizations can be mentioned though; before doing this let us make the convention that $N = n + 1$ because we will pass back and forth between projective space (dimension n) and its tautological bundle (dimension N), or—which is the same thing—between the space of pure states and the Hilbert space.

The construction of projective space can be generalized so that we consider the space of two dimensional subspaces in a vector space as a space in its own right. In general the space of M dimensional subspaces in an N dimensional linear space is known as the *Grassmannian manifold* $G_{M,N}$. We can also think of this space as the space of $m = M - 1$ dimensional linear subspaces in projective m-space. Then we call it $\mathbf{G}_{m,n}$. In this notation then complex projective space is the special case

$$\mathbf{CP}^n = G_{1,n+1}(\mathbf{C}) = \mathbf{G}_{0,n}(\mathbf{C}) = \mathbf{G}_{n-1,n}(\mathbf{C}) \ . \tag{200}$$

Here the notation also tells us which field ($\mathbf{R}$ or $\mathbf{C}$) we are using. As an exercise, let us compute the dimension of the real Grassmannian consisting of the space of lines in real projective 3-space. A little thought will convince you that we need 4 parameters to describe a line in 3-space: Two angles

to describe the direction of the shortest path connecting the origin to the line, one parameter to describe the length of this path, and one to tell the direction in which the line is pointing. Hence the dimension of $\mathbf{G}_{1,3}(\mathbf{R})$ is four—and the complex dimension of $\mathbf{G}_{1,3}(\mathbf{C})$ is also four as we will see in a moment.

All Grassmannian have a natural description as coset spaces. To see this, let us first describe $\mathbf{CP}^n$ in a new way. Consider the space of all orthonormal frames in $\mathbf{C}^{n+1}$. Since any such frame can be reached from a given one by means of a unitary transformation it is clear that the manifold of orthonormal frames is the same as the group manifold of $U(n+1)$. Now the sphere $\mathbf{S}^{2n+1}$ can be regarded as the space of unit vectors in $\mathbf{C}^{n+1}$. If we try to regard such a unit vector as the first vector in some orthonormal frame we must take into account that if we fix the unit vector the frame can still be changed by means of unitary transformations belonging to the subgroup $U(n)$. Hence we conclude that

$$\mathbf{S}^{2n+1} = \frac{U(n+1)}{U(n)} \qquad \Rightarrow \qquad \mathbf{CP}^n = \frac{U(n+1)}{U(n) \otimes U(1)} \ , \tag{201}$$

where we made use of the Hopf fibration in the second step.

Now for the obvious generalization. The space of sets of M orthonormal vectors in an N dimensional linear space is known as a *Stiefel manifold* $\mathcal{S}$. It is clearly the same as the homogeneous space

$$\mathcal{S} = \frac{U(N)}{U(N-M)} \ . \tag{202}$$

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. This is easily corrected for, and we conclude that

$$G_{M,N}(\mathbf{C}) = \frac{U(N)}{U(N-M) \otimes U(M)} \ . \tag{203}$$

From this equation we can easily compute the dimension of $G_{M,N}(\mathbf{C})$; since $U(N)$ has N^2 real dimensions the complex dimension of the Grassmannian is

$$\frac{1}{2}(N^2 - M^2 - (N-M)^2) = M(N-M) \ . \tag{204}$$

In particular the complex dimension of $G_{2,4}(\mathbf{C}) = \mathbf{G}_{1,3}(\mathbf{C})$ is equal to four.

We also learn that the Grassmannian is a coset space, and at the same time it is a fibre bundle with multidimensional fibres and the non-Abelian structure group $U(N-M) \otimes U(M)$. There is a preferred connection in this bundle that transforms like a non-Abelian Yang-Mills field. The construction appears in quantum mechanics when we consider subspaces spanned by degenerate eigenvectors of the Hamiltonian.

Exercises:

- Consider $n+2$ ordered points on the plane, not all of which coincide. Consider two such sets eqivalent if they can be transformed into each other by translations, rotations, and scalings. Show that the topology of the resulting set is that of $\mathbf{CP}^n$. What does $\mathbf{CP}^n$ have to do with archeology?

- If you manage to glue together a Möbius strip and a hemisphere you get $\mathbf{RP}^2$. What will you obtain if you glue two Möbius strips together?

- Derive the line element of the Fubini-Study metric from the expression for the projective cross ratio and the definition of geodesic distance.

- It is stated above that it is easy to check that, on the Berger sphere, distances along the Hopf fibres are squashed while distances orthogonal to the fibres are unaffected by the squashing. Do this.

- Carry through the argument needed to prove that a linear subspace is a totally geodesic submanifold.

- Carry through the argument needed to prove that a complex projective line cannot be shrunk to a point within $\mathbf{CP}^n$.

REAL QUANTUM MECHANICS

Quantum mechanical notation

In quantum mechanics, the space of pure states is precisely $\mathbf{CP}^n$. In some sense this is so also when Hilbert space is infinite dimensional (in which case n is infinite), but as always we will restrict ourselves to the finite dimensional case for which the geometry of state space can be explored by experiments. To see that the vector space $\mathbf{C}^{n+1}$ is the familiar Hilbert space in disguise, we introduce a basis in the latter and use Dirac notation to write

$$|\psi >= Z^\alpha |\psi_\alpha > \quad . \tag{205}$$

We will use Dirac notation now and then in these notes; the index notation is nicely provocative and superior for geometry, but the Dirac notation deals nicely with the algebraic aspects. A brief reminder might therefore be helpful.

In quantum mechanics it is conventional to normalize all the state vectors to unit length, using the Hilbert space norm

$$< \psi|\psi >= Z \cdot \bar{Z} \ . \tag{206}$$

Moreover an overall phase in the state vector has no physical significance;

$$|\psi > \sim e^{i\phi}|\psi > \quad . \tag{207}$$

Hence the space of physically inequivalent states is precisely complex projective n-space, as advertized.

Let us see how some definitions familiar from quantum mechanics appear in our language. The Schrödinger equation is

$$i\partial_t|\psi >= H|\psi > \quad \Leftrightarrow \quad i\dot{Z}^\alpha = H^\alpha{}_\beta Z^\beta \ . \tag{208}$$

It would be more difficult to write the projective Schrödinger equation in Dirac notation. Also, because of our geometrical perspective we now have an interpretation of the time evolution as the lift of a Killing vector field on $\mathbf{CP}^n$. Our discussion of the velocity of the evolution gives us a precise statement of the time-energy uncertainty relation in quantum mechanics,

$$ds = \sqrt{< H^2 > - < H >^2}dt = \Delta H dt \ . \tag{209}$$

The non-vanishing curvature of the connection that lifts the projective Schrödinger equation to the tautological bundle appears in quantum mechanics as the "geometrical phase" known to physicists as the Berry phase. The Fubini-Study metric itself provides a notion of distance between pure states in quantum mechanics. In Dirac notation, the definition is

$$\cos^2 d = \kappa = \frac{| < P|Q > |^2}{||P||^2||Q||^2} \ . \tag{210}$$

Evidently this distance is some kind of measure of the distinguishability of the two states; if the distance is large they can easily be distinguished through the result of suitable measurements. We will make this precise later. The projective cross ratio κ is known in quantum communication theory as the fidelity function.

We can go on in this way to provide geometrical interpretations of all manners of familiar concepts in quantum mechanics. However, the main task of this chapter is somewhat different: To take away the complex numbers, and to see how the formalism appears without them. This will bring the symplectic form, rather than the metric, to the fore.

Hilbert space from a real point of view

Why do we use complex projective space to describe the space of pure states in quantum mechanics? This is clearly a grand question. To cut it down to size we will first assume that the space of states is a vector space, and try to see where the complex numbers come from. While doing this it is interesting to keep classical mechanics in mind. Hamiltonian mechanics hinges on the idea that observables are functions of the state, and that it is possible to associate a vector field with any such function by means of a symplectic form. Both classical and quantum mechanics can be axiomatized starting from an algebra of observables. According to this general scheme, the observables play two roles. On the one hand, they form an algebra. If we represent them as linear transformations $\mathbf{T}$ on a vector space they obey

$$T_1^I{}_J T_2^J{}_K = T_3^I{}_K \tag{211}$$

and so on. On the other hand the states are to be regarded as functionals on the algebra, taking values among the real numbers. We can define such a map from an observable **O** to the reals by

$$\mathbf{O} \rightarrow < \mathbf{O} > = X^I O_{IJ} X^J \ . \tag{212}$$

Here the observables are naturally symmetric matrices. The problem is that it is not *a priori* clear how the two notions of "observable" can be related to each other; even the index structures on the **O**s and the **T**s are different. One way to solve this problem is to postulate the existence of a non-degenerate symplectic form

$$\Omega_{IJ} = -\Omega_{JI} \ . \tag{213}$$

(This requires that the dimension of our vector space is even, so we are edging towards complex vector spaces.) Our index convention for the inverse is

$$\Omega^{IK} \Omega_{KJ} = \delta^I_J \ . \tag{214}$$

We can now use the symplectic form to set up a one-to-one correspondence between the **O**s and the **T**s:

$$O_{IJ} \equiv \Omega_{IK} T^K_J \ . \tag{215}$$

The **O** defined in this way will indeed be a symmetric matrix provided that we restrict ourselves to transformations **T** that leave the symplectic form invariant, so that

$$0 = \delta \Omega_{IJ} = \Omega_{IK} T^K{}_J + \Omega_{KJ} T^K{}_I \ . \tag{216}$$

This is the way that classical mechanics enables the observables to play their two roles.

Quantum mechanics differs from classical mechanics in that there is a distinguished metric tensor on the state space. Could we use this metric rather than a symplectic form to set up the correspondence between the **O**s and the **T**s? The answer is no, since we would like to restrict ourselves to observables that leave the metric invariant. Then

$$0 = \delta g_{IJ} = g_{IK} T^K{}_J + g_{KJ} T^K{}_I \quad \Leftrightarrow \quad T_{IJ} = -T_{JI} \ . \tag{217}$$

It will not do to set $O_{IJ} = T_{IJ}$, since the former are symmetric matrices. The general scheme therefore demands a symplectic form as well as a metric.

At this stage the group that is generated by the observables is the intersection of the orthogonal and the symplectic group. To recover the structure of conventional quantum mechanics an additional decision will have to be made. There will be some $\mathbf{T}$ that corresponds to g_{IJ}. Call this operator $J^I{}_J$:

$$g_{IJ} = \Omega_{IK} J^K{}_J \quad \Leftrightarrow \quad J^I{}_J = \Omega^{IK} g_{KJ} \ . \tag{218}$$

By taking successive powers of J we can now generate any number of distinguished tensors, something that we may find unpalatable. We decide to shortcircuit this process by the demand that

$$J^I{}_K J^K{}_J = -\delta^I_J \ . \tag{219}$$

This is in fact a compatibility condition imposed on the pair g and Ω. Note that the operator J behaves like the imaginary unit i, indeed it is the imaginary unit written in real notation. It is also known as a complex structure on our real vector space. This is not quite our previous definition of complex structure, which amounted to a direct sum decomposition of the complexified vector space such that the summands were related by complex conjugation. But the two definitions are equivalent. The point is that the eigenvalues of J will be $\pm i$. The two eigenvalue spaces are defined by

$$(1 \pm iJ) X_\pm = 0 \ . \tag{220}$$

Together they will provide a direct sum decomposition of the complexified vector space with the required properties.

Only two more steps are needed to arrive at quantum mechanics. First we use the metric to normalize the states, and then we insist that an observable is allowed only if it commutes with the complex structure. A minor calculation shows that this requirement is

$$[T, J] = 0 \quad \Leftrightarrow \quad O_{IJ} = O_{KL} J^K{}_I J^L{}_J \ . \tag{221}$$

Matrices that obey this condition are called Hermitian. We are now able to construct a Hermitian form with the linearity properties

$$< X, JY >= i < X, Y >= - < JX, Y > \ . \tag{222}$$

The definition is

$$< X, Y > \equiv Xg(1 - iJ)Y = X^I g_{IJ} Y^J + iX^I \Omega_{IJ} Y^J \ . \tag{223}$$

In this way we have been led to precisely the conventional Hilbert space of quantum mechanics.

Note that for all Hermitian observables (including J) we have that

$$< X, TY >= - < TX, Y > \ . \tag{224}$$

Thus "anti-Hermitian" might have been a better designation as far as the **T**s are concerned, but clearly if you have the one you have the other. The group generated by the Hermitian observables is now the unitary group.

To see what is going on it is helpful to choose an explicit matrix representation in our $2m$ dimensional vector space. We choose

$$g_{IJ} = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{1} \end{pmatrix} \qquad \Omega_{IJ} = \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix} \ . \tag{225}$$

We deduce

$$J^I{}_J = \begin{pmatrix} 0 & -\mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix} \qquad O_{IJ} = \begin{pmatrix} m & a \\ -a & m \end{pmatrix} \ . \tag{226}$$

In the latter equation **O** is a general observable, m is a symmetric matrix and a is anti-symmetric. One symmetric and one anti-symmetric real $n \times n$ matrix is of course precisely what is needed to construct a complex matrix that is Hermitian in the usual sense. At this point it is not difficult to introduce complex coordinates such that J acts through multiplication by i, while the operators **T** act like i times a matrix that is Hermitian in the ordinary sense.

Finally the formalism is in place and we are in a position to associate a vector field with any observable. Let us choose the Hamiltonian as the observable that we consider. Then the Schrödinger equation written in real notation takes the form

$$\dot{X}^I = -\Omega^{IJ} H_{JK} X^K \ . \tag{227}$$

Here H is any observable **O**—and the symplectic form was introduced precisely for the purpose of enabling us to write this equation. Because of the various requirements that we have introduced the flow that it defines is a Killing vector field of the metric g_{IJ}.

We can polish the Schrödinger equation a bit further. Instead of thinking of the states as functions of H we can think of H as a function of the states. (This is duality.) Then we have the state space function

$$< H > \equiv X^I H_{IJ} X^J \ . \tag{228}$$

Therefore the Schrödinger equation associates a vector field with the particular class of state space functions that we are considering, namely

$$\dot{X}^I = -\frac{1}{2}\Omega^{IJ}\partial_J < H > \ . \tag{229}$$

(The factor $-1/2$ is unfortunate but does not matter.) This is Hamilton's equations—from the present point of view quantum mechanics appears to be a special case of classical mechanics. It is special because we do not allow arbitrary state space functions to serve as Hamiltonians, but only those that can be written as expectation values of an Hermitian matrix.

It is special in another sense too (although the two senses turn out to be related): Classical mechanics allows arbitrary symplectic manifolds as state spaces, while quantum mechanics requires a vector space, or more precisely a projective space. The discussion given here could be generalized to arbitrary Kähler manifolds, since they possess both metrics and symplectic forms connected via a complex structure. What, if anything, is wrong with that?

Kähler manifolds revisited

General Kähler spaces can also be described from a real point of view. In our previous description the complex structure was built in. In real terms what we require is that there exists a complex structure on the tangent spaces, which means that there must exist a tensor field

$$J^i{}_j \ ; \qquad J^i{}_k J^k{}_j = -1 \ . \tag{230}$$

A non-trivial compatibility condition has to be satisfied to ensure that the commutator of two vector fields of type $(1,0)$ is again a vector field of this type. We state without proof that the required condition is

$$J^k_{\ m} \nabla_{[i} J^m_{\ j]} - J^m_{\ [i} \nabla_{|m|} J^k_{\ j]} = 0 \ , \tag{231}$$

where any symmetric connection can be used to define the covariant derivative. In addition to being a complex manifold all Kähler manifolds are simultaneously Riemannian and symplectic manifolds (there exist a non-degenerate metric and a non-degenerate closed 2-form), and there is a compatibility condition that relates the complex structure, the metric and the symplectic form. We skip the details, which are similar to the flat case.

The fact that all Kähler spaces are symplectic manifolds means that they can serve as state spaces in classical mechanics. The symplectic structure enables us to associate a vector field to any function on the space of states. What is special about quantum mechanics is that the formalism is much more restrictive—one does not choose to consider any function, but only those special functions which can be written as "expectation values" of a Hermitian matrix. Because of the compatibility condition between the symplectic form and the metric g_{IJ} another (and more interesting) way to say this is that one considers only those Hamiltonian vector fields that are also Killing vector fields of the metric g_{IJ}. Quantum mechanics has a metric worth preserving. In effect this means that only a very small subset of the functions is being used. Clearly this scheme can be made to work at all only if there is a reasonably large number of Killing fields to work with. In this way we are led to try $\mathbf{CP}^n$ as the space of pure states—this is the unique compact simply connected Kähler manifold that has constant holomorphic sectional curvature, which means that it possesses a maximal number of Killing vectors.

There is another reason why $\mathbf{CP}^n$ appears to be preferred over general Kähler spaces, namely the existence of the Segre embedding. The fact that this embedding is always available makes it possible to treat composite systems in just the manner 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. In some sense it is also the large dimension of the quantum me-

chanical state space that enables general symplectic manifolds to arise in the classical limit—we must recall that the six dimensional symplectic manifold used to describe a free classical particle is to emerge from an infinite dimensional quantum state space. Quantum mechanics may be rigid, but it is big.

The discussion up to this point is meant to provide a partial answer to why $\mathbf{CP}^n$ is a natural candidate for a space of pure states. It is not meant to rule out alternatives to the standard formalism.

Time and the complex numbers

Let us return to our vector space for a moment, and ask ourselves how the complex structure is to be selected. Assume that the dimension of the vector space is $2m$. A complex structure will be an orthogonal transformation that squares to minus one, and there are many such. Just choose m orthogonal 2-planes and let the complex structure be represented by quarter turns in any direction in all of these planes. In terms of matrices, I am thinking of an $SO(2m)$ matrix written on the block diagonal form

$$\begin{pmatrix} \pm\Omega & & & \\ & \pm\Omega & & \\ & & \pm\Omega & \\ & & & \ddots \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{232}$$

How should we choose the signs ?

This is where time enters. There is one operator that has a fundamental asymmetry: The Hamiltonian H, all of whose eigenvalues are positive. In our vector space H will be associated with an $SO(2m)$ transformation that can be represented as a rotation in m orthogonal planes. Because its (skew) eigenvalues are all positive the rotation always takes place in the positive direction. Since we assume that the complex structure commutes with the Hamiltonian it now follows that a unique complex structure is singled out by the requirement that the quarter turns occur in the planes defined by H and in the direction of the Hamiltonian flow. It also follows that if we reverse the direction of time then the complex structure changes sign—time reversal will be represented in quantum mechanics by an anti-unitary transformation.

Wigner's theorem states that every isometry of $\mathbf{CP}^n$ can be effected by a transformation of $\mathbf{C}^{n+1}$ 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. It is illuminating to adopt the real point of view on this. 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. But this will be so if the transformation is effected by a matrix which obeys either

$$UgU^T = g \qquad\qquad U\Omega U^T = \Omega \tag{233}$$

(the unitary case) or

$$\Theta g\Theta^T = g \qquad\qquad \Theta\Omega\Theta^T = -\Omega \tag{234}$$

(the anti-unitary case). Hence anti-unitary transformations are anti-canonical. This still leaves two options, though:

$$\Theta^2 = \pm 1 \ . \tag{235}$$

The choice of sign depends on the system. Wigner found that if it is required that time reversal changes the sign of the angular momentum operator (conserved or not) then the plus sign must be employed for systems with integer spin and the minus sign for systems with half-integer spin.

Since the result hinges on precisely how the angular momentum operator is defined the discussion cannot be made fully coherent without going into this definition in detail. What follows below is not quite self-contained.

For bosons the conclusion is that Θ squares to unity. We can then restrict our attention to states that are invariant under time reversal, and this leads us to the real state space $\mathbf{RP}^n$. For fermions no such states exist. Also, if H and Θ commute then all the eigenstates of H must be doubly degenerate. This much follows from the fact that Θ is anti-unitary; if its only effect on some state vector was to multiply it with a phase we could show that

$$\Theta^2|E>= \Theta(e^{i\phi}|E>) = e^{-i\phi}\Theta|E>= |E> \tag{236}$$

—and this is not true for fermions. The degeneracy of the energy eigenstates that saves the situation is known as the Kramers degeneracy. There is no

real projective space of time reversal invariant states, but instead there is a quaternionic structure present. Let us sketch how this structure emerges. We know that

$$\Theta J + J\Theta = 0 \ . \tag{237}$$

We define

$$\mathbf{i} = J \qquad \mathbf{j} = \Theta \qquad \mathbf{k} = J\Theta \ . \tag{238}$$

It follows that

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1 \tag{239}$$

$$\mathbf{ij} = \mathbf{k} \qquad \mathbf{jk} = \mathbf{i} \qquad \mathbf{ki} = \mathbf{j} \ . \tag{240}$$

This is the algebra of quaternions. A general quaternion can be written in terms of four real numbers as

$$\mathbf{q} = a_0 + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \ . \tag{241}$$

The quaternions form a non-commutative algebra with distinguished properties; the set of quaternions is denoted by $\mathbf{H}$ (for Hamilton, who invented them).

We now impose a superselection rule, that is to say that we will agree that only observables that commute with the operator Θ will be measured. Since all observables commute with J it follows that under this restriction we can not distinguish between states in our original Hilbert space that differ by a multiplicative factor that is a quaternion. In this situation the distinguishable states of a fermionic system form the quaternionic projective space $\mathbf{HP}^k$, where $4k + 4 = 2n + 2$ if the full state space is $\mathbf{CP}^n$. In brief

$$[T, \mathbf{i}] = 0 \ \Rightarrow \ \mathbf{CP}^n \qquad [T, \mathbf{q}] = 0 \ \Rightarrow \ \mathbf{HP}^n \ . \tag{242}$$

In some ways quaternionic projective geometry is very similar to complex projective geometry, if due care is taken with the non-commutativity. In particular there is a quaternionic Hopf fibration of the sphere $\mathbf{S}^{4k+3}$. Occasionally it has been suggested that quaternionic quantum mechanics could

offer an alternative to the standard formalism. On the other hand we see that both $\mathbf{RP}^n$ and $\mathbf{HP}^k$ have a role to play within the latter.

Exercises:

- Put complex numbers in, and verify that the real Schrödinger equation is the usual Schrödinger equation.

Figure 24: S is convex, S' is not

COLOURS AND STATISTICAL GEOMETRY

Convex sets

In order to appreciate the geometry of quantum mechanics we need to understand statistical geometry; both the metrical geometry on the space of probability distributions and the notion of convex sets that is of fundamental importance in this context. We begin with the latter. The idea is that a convex set is a set such that one can form "mixtures" of any pair of points in the set; geometrically we insist that the straight line between two points in the set also belongs to the set. It will therefore be necessary to have a definition of "straight lines" available. To ensure this the convex set is always regarded as a subset of an affine space, that is to say of a vector space without any distinguished point of origin. Turning to formulæ, let the real number μ belong to the interval $[0,1]$. Then the set S is convex if

$$\mathbf{x} \in S \;\&\; \mathbf{y} \in S \quad \Rightarrow \quad \mu\mathbf{x} + (1-\mu)\mathbf{y} \in S \; . \tag{243}$$

The affine structure gives a notion of distance along the straight line, but we do not need a metric since convexity properties are preserved under general affine transformations. On the other hand it does no harm to introduce a flat metric on the affine space, turning it into Euclidean space $\mathbf{E}^n$. We will frequently do so.

Some terminology: 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; that is to say that the p-simplex

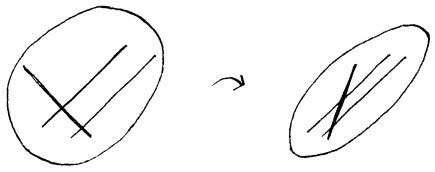

Figure 25: Affine transformation of a convex set

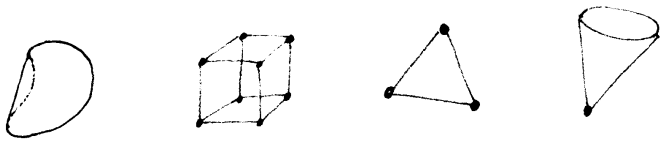

Figure 26: The convex hull of S', a convex polytope, a 3-simplex, and a convex cone

consists of all points of the form

$$\mathbf{x} = \mu_0 \mathbf{x}_0 + \mu_1 \mathbf{x}_1 + ... + \mu_p \mathbf{x}_p \ , \quad \mu_0 + \mu_1 + ... + \mu_p = 1 \ . \tag{244}$$

(All the μs are positive; they are sometimes referred to as barycentric coordinates.) The dimension of a convex set is the largest number p such that the set contains an p-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). Then the boundary of the convex set always has the topology of a sphere, although it may look very squashed from the point of view of the metric. A compact convex set is called a convex body. A convex cone is a set consisting of all the points of the form

$$\mathbf{x} = \mu_0 \mathbf{x}_0 + \mu_1 \mathbf{x}_1 + ... + \mu_p \mathbf{x}_p \ , \quad \mu_0 \geq 0 \ , \ \mu_1 \geq 0 \ , \ ..., \ \mu_p \geq 1 \ . \tag{245}$$

The cone has an apex at the origin.

A theorem due to Carathéodory deserves to be mentioned. It says that the convex hull of a subset M of $\mathbf{R}^n$ is the union of all the convex hulls of subsets of M containing at most $n+1$ points. It should also be noted that

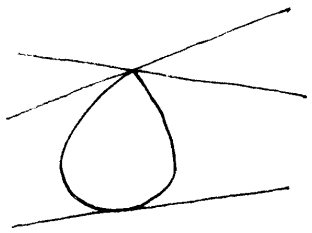

Figure 27: Supporting hyperplanes

there is a dual description of convex sets in terms of supporting hyperplanes. By definition a hyperplane bounds a set S if it partitions the affine space into two half-spaces and if the set S is contained in one of those. A supporting hyperplane of S is a hyperplane that intersects the set and which is such that the entire set lies in one of the closed half spaces formed by the hyperplane. Hence a supporting hyperplane just touches the boundary of S. One can prove that there is a supporting hyperplane passing through every point of the boundary of a convex body, and that every supporting hyperplane of a convex cone passes through its apex. A convex function is a function defined on a convex set that obeys

$$f(\mu \mathbf{x}_1 + (1-\mu)\mathbf{x}_2) \leq \mu f(\mathbf{x}_1) + (1-\mu) f(\mathbf{x}_2) \ . \tag{246}$$

There is a rich mathematical theory concerning convex sets and convex functions.

Colour theory

Convex sets occur in various physical contexts. A special piece of terminology that is used in physical applications: A point in a convex set which can not be expressed as a convex combination of two other points is called a pure state. The pure states belong to the boundary of the convex set, but they may form a proper subset of the boundary (as in the case of a simplex). A point in the set which is not pure is called a mixture, or a mixed state.

An instructive example of a convex set that has nothing to do with quantum mechanics arises in colour theory, and more particularly in the psy-

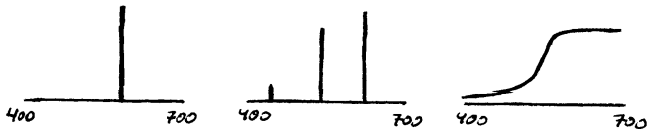

Figure 28: Spectral distributions giving rise to "yellow"

chophysical 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.) We know that light can be characterized by its spectral distribution, which is some positive function V 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 $V(\lambda)$ can be defined as a convex combination

$$V(\lambda) = \int d\lambda' V(\lambda')\delta(\lambda - \lambda') \ , \quad V(\lambda') \geq 0 \ . \tag{247}$$

The pure states are the delta functions.

It is crucial that the space of colours is not the same as the space of spectral distributions, since it is a known fact 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 $[V(\lambda)]$.

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". (The sensitivity curves of the detectors are allowed to overlap; in fact they do, in the eye.) Then it is clear that our three detectors will give us only two real numbers to parametrize the space of colours. It is also clear—and absolutely crucial—that if we add light belonging to the equivalence classes $[V_1(\lambda)]$ and $[V_2(\lambda)]$ together then the response of the detectors will not depend on which representatives of the equivalence classes that we choose. Hence the equation

$$[V(\lambda)] = [V_1(\lambda)] + [V_2(\lambda)] \tag{248}$$

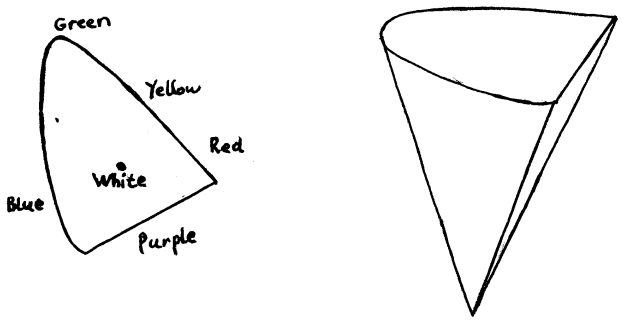

Figure 29: The chromaticity diagram and the full colour space

is well defined. It says that if we add light of two different colours together then a definite colour will result; according to this theory colour space will inherit the property of being a convex set from the space of spectral distributions. The pure states will be those equivalence classes that contain the pure spectral distributions. However, the dimension of colour space will be determined by the number of detectors, and not by the nature of the pure states.

Experimentally it is known that colour space is a three dimensional convex cone (to the extent that we can ignore the fact that very intense light will cause the eyes to boil rather than make them see a colour). If we normalize the total luminosity in some way we obtain a two dimensional convex body which is known as the chromaticity diagram. The pure states form the curved part of the boundary; the straight edge is made of purple. An interesting feature is that a general state can be expressed as a mixture of pure state in many ways.

Naturally one might worry that colour space may differ from person to person. As a matter of fact this is not so—the perception of colour is remarkably universal for human beings, except for colour blind persons that have one or more detectors missing and therefore see a colour space of dimension two or less. To establish the chromaticity diagram experimentally one chooses three reference colours consisting of (say) red, green and blue light. One then adjusts the mixture of these colours until the observer is unable to distinguish the resulting mixture from a given colour C, whose position in the chromaticity diagram in then determined by the equation

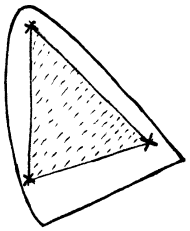

Figure 30: Mixing red, green and blue

$$C = xR + yG + zB \ . \tag{249}$$

In this way we can get only colours inside the triangle spanned by the reference colours. Note that the experiment is a null measurement.

If it is the case that C cannot be reproduced as a mixture of the reference colours then one shines a certain amount of red (say) on the sample to be measured. If the two samples become indistinguishable then C is determined by the equation

$$C + xR = yG + xB \ . \tag{250}$$

If not, repeat with R replaced by G or B.

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 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.

Historically the colour metric is of some interest in that it was perhaps the first example of an application of Riemannian geometry in a context that has nothing to do with Space. This gave it a special epistemological significance. If there is a lesson from colour theory to quantum mechanics, it

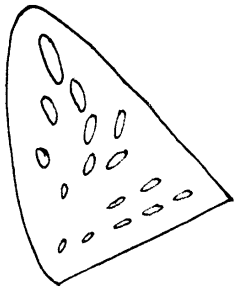

Figure 31: MacAdam ellipses

is probably that the structure of the convex set is determined by the nature of the detectors at least as much as by the nature of the pure states. Moreover the existence of a natural metric that has nothing to do with the convex structure is a theme that will recur.

Statistical geometry

Convex sets figure prominently in probability theory. Thus consider an experiment that can yield $n+1$ possible outcomes. The probability distribution for the outcome is then given by $n+1$ real numbers p_i such that

$$p_i \geq 0 \ , \qquad \sum_{i=0}^{n} p_i = 1 \ . \tag{251}$$

We see that the space of all possible probability distributions forms an n-simplex. This is in many ways the simplest kind of convex set. The pure states are those for which the outcome is certain. They are sitting at the edges of the simplex and hence they form a zero dimensional subset of its boundary. A major simplifying feature is 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 sets that arise in quantum mechanics.

For the purpose of studying the convexity properties of the statistical simplex a flat metric is helpful (and available). However, just as in the case of colour space there is another metric that arises naturally. Indeed the

geometry of the n-sphere is almost manifest here. Define

$$\xi_i \equiv \sqrt{p_i} \qquad \Rightarrow \qquad \sum_{i=0}^{n} \xi_i^2 = 1 \ . \tag{252}$$

We see from this that the space of all probability distributions lies in a sphere embedded in a flat space, and therefore it carries the natural metric of the sphere, namely

$$ds^2 = \sum_{i=0}^{n} d\xi_i d\xi_i = \frac{1}{4} \sum_{i=0}^{n} \frac{dp_i dp_i}{p_i} \ . \tag{253}$$

In statistics this is known as the Bhattacharrya or Fisher-Rao metric. It enables us to define the geodesic distance between two arbitrary probability distributions; if we consider the case where there are only two possible outcomes then the geodesic distance between two probability distributions $(1-p, p)$ and $(1-p', p')$ is

$$\cos d = \sqrt{(1-p)(1-p')} + \sqrt{pp'} \ . \tag{254}$$

This is just our formula for geodesic distance on the 2-sphere. The generalization to higher dimensions should be obvious. Note that the largest possible distance is $\pi/2$.

The only question is whether the number d has (as it were) any statistical significance. It has. The idea behind it was introduced by Fisher in the 20's. If we attempt to establish a probability distribution by means of experiments, the best we can do is to perform a large number N of independent samplings on an ensemble of objects for which the probability distribution is supposed to hold. In this way we will obtain $n+1$ frequencies f_i that will approximate the probabilities p_i. The probability to obtain a given frequency f_j for the jth outcome can be calculated from the ps by means of the multinomial distribution, and we can then ask whether the observed frequencies lie within (say) one standard deviation σ from the predicted probability distribution. For simplicity, let us consider the case where there are only two possible outcomes. If the probability to obtain the first outcome is p and if we perform N samplings then the probability to obtain the first outcome m times—that is to say to obtain the frequency m/N—is

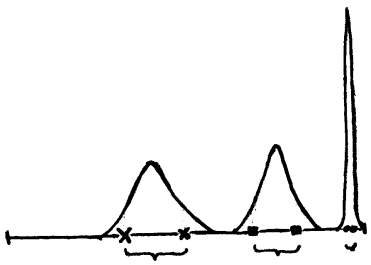

Figure 32: Statistical distance: Some equally spaced points

$$P(m/N) = \left(\begin{array}{c} N \\ m \end{array} \right) p^m (1-p)^{N-m} \ . \tag{255}$$

The expectation value of the frequency is p. The standard deviation is

$$\sigma = \sqrt{\frac{p(1-p)}{N}} \ . \tag{256}$$

Now let us say that two probability distributions are distinguishable in N samplings if they are separated by at least one standard deviation. More generally, we can define a measure of the distinguishability of two arbitrary probability distributions by counting the number of distinguishable distributions that lie in between, and dividing by $\sqrt{N}$ to ensure that the limit exists. In this way we are led to define the statistical distance as

$$d \equiv \frac{1}{2} \lim_{N\to\infty} \frac{1}{\sqrt{N}} \int_p^{p'} \frac{dp}{\sigma} = \int_p^{p'} \frac{dp}{2\sqrt{p(1-p)}} \ . \tag{257}$$

The overall factor is chosen for convenience. When the integral is computed we learn that the statistical distance is precisely equal to the geodesic distance according to the Bhattacharrya metric, as given above. Of course it is even easier to check that

$$ds^2 = \frac{dp^2}{4p(1-p)} = \frac{1}{4} \left(\frac{dp^2}{p} + \frac{d(1-p)^2}{1-p} \right) \ . \tag{258}$$

The basic reason why the statistical metric is curved is that the statistical fluctuations are much smaller when we are close to the pure states, i.e. to

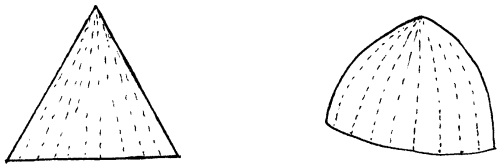

Figure 33: The convex and the statistical geometry of a simplex

the edges of the simplex. A somewhat different derivation (valid for any number of outcomes) notes that for large N the multinomial distribution can be approximated by a normal distribution, so that the probability to obtain the frequencies f_i will be proportional to

$$P(f_j) \propto \exp\left(-\frac{N}{2}\frac{(f_j - p_j)^2}{p_j}\right) . \tag{259}$$

It follows that we will be able to distinguish between two nearby probability distributions $\{p_i\}$ and $\{p_i'\}$ precisely when

$$\frac{(\Delta p)^2}{p_i} = \frac{(p_i - p_i')^2}{p_i} \tag{260}$$

is sufficiently large. This again leads to the Bhattacharrya metric. It should not escape observation that the idea that leads to the notion of statistical distance is the same idea as the one that led us to introduce the colour metric by means of the MacAdam ellipses.

From this point of view a statistical simplex should not be thought of as a flat triangle (say), but rather as the positive octant of a sphere.

Statistical inference

A standard problem in statistic is how to suggest a probability distribution in a rational way, given only the results of a finite set of experiments. We might say that in the absence of any information the uniform probability distribution is the best bet, and then the problem becomes how to upgrade this suggestion when the results of the experiments are coming in. This problem occurs in statistical mechanics, and the method used there is to maximize the Shannon entropy

$$H(p) = -\sum_{i=1}^{n} p_i \ln p_i \tag{261}$$

subject to the constraints

$$\sum_{i=1}^{n} p_i = 1 \qquad < a > \equiv \sum_{i=1}^{n} a_i p_i = \text{constant} \ . \tag{262}$$

The second constraint is to be thought of as the result of some experiment. (The Shannon entropy has some quite unique features which we do not go into now.) This is an optimization problem that is easily treated with the Lagrange multiplier method, and the result is

$$p_i = \frac{e^{-ta_i}}{\sum_{j=1}^{n} e^{-ta_j}} \ . \tag{263}$$

The value of the Lagrange multipler t is to be determined by satisfying the constraint on $< a >$. An efficient numerical method to do so is to integrate the differential equation

$$\frac{dp_i}{dt} = -p_i(a_i - < a >_t) \ . \tag{264}$$

This equation is easily seen to follow from the form of the solution; $< a >_t$ is the expectation value at the "time" t and the integration is to stop when it reaches the prescribed value. (This may happen at a negative value of t but this poses no problem—certainly not since our statistical simplex is finite dimensional so that all sums exist.)

The Bhattacharrya metric enters here in a curious way. First, if we think properly of $\dot{p}^i$ as a contravariant vector and of $\Delta a_i \equiv a_i - < a >$ as a covariant vector we see that the equation is

$$\dot{p}^i = -\frac{1}{4} g^{ij} \Delta a_j \ . \tag{265}$$

Second, if we compute the velocity of the "evolution" in t as defined by the Bhattacharya distance we find that

$$4\left(\frac{ds}{dt}\right)^2 = \sum_{i=1}^{n} \frac{\dot{p}_i \dot{p}_i}{p_i} = < a^2 >_t - < a >_t^2 \ . \tag{266}$$

This is analogous to the velocity of quantum mechanical evolution as driven by an operator H.

Statistical distance and the Fubini-Study metric

We are now in a position to give a proper interpretation of the Fubini-Study metric in quantum mechanics. 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 perform, or in mathematical terms to choose some Hermitian operator A 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.

Suppose that the two states are $|P>$ and $|Q>$. Choose an operator A, and observe that it has $n+1$ orthogonal eigenstates $|i>$ in terms of which we can expand

$$|P>=\sum_{i=0}^{n} c_i|i> \qquad |Q>=\sum_{i=0}^{n} d_i|i> \quad . \tag{267}$$

For convenience, normalize all the states to unity. The probability to obtain a given outcome of the measurement is given by the standard interpretation of quantum mechanics, and the statistical distance between the given states, given the operator, can be computed from the square roots of the probabilities:

$$\cos d_{\mathbf{o}} = \sum_{i=0}^{n} |c_i||d_i| = \sum_{i=0}^{n} | < P|i > || < Q|i > | \; . \tag{268}$$

According to the definition of distance between quantum mechanical states we should now choose the operator A in such a way that $d_{\mathbf{o}}$ becomes as small as possible, that is to say that the right hand side should be as close to one as it can get. This will be so if either $|P>$ or $|Q>$ is an eigenstate of $\mathbf{O}$, in which case the expression collapses to

$$\cos d = | < P|Q > | \ . \tag{269}$$

But this is precisely the geodetic distance between the two states as computed by means of the Fubini-Study metric. We have therefore established that the Fubini-Study metric measures the distinguishability of quantum states in the sense of statistical distance.

Exercises:

- A set of equally spaced points on the interval $[0, 1]$, in the sense of statistical distance, can be obtained by drawing a semi-circle having the interval as its diameter, placing equally spaced points on the semi-circle, and then projecting down on the diameter by means of the equatorial projection. Verify this. Also try to generalize the construction to one dimension higher.

THE SPACE OF DENSITY MATRICES

A survey of the landscape

So far we have taken for granted that the space of all states in quantum mechanics is the same as complex projective space. If we ask the somewhat slippery question "is this really so?", then we find that there are basically two schools of opinion on this: According to the first complex projective space is just a small subset of the space of all density matrices, and it is the latter which is the true space of states. Hence a general state is given by a positive operator ρ of unit trace, and the result of measuring an observable A is the real number

$$< A >= \mathrm{Tr}\rho A \ . \tag{270}$$

This equation can be viewed as a map from the algebra of observables to the real numbers if you are so inclined. There is a one-to-one correspondence between complex projective space and a very special kind of density matrix, namely projection operators onto rays in Hilbert space. The other school does not deny the importance of density matrices, but regards the statement that a state is described by a density matrix as a shorthand description of the true state of affairs, which is that the system under study is really a part of a composite system to be described by a much larger Hilbert space. The density matrices can be obtained by taking partial traces over the unknown part of the entire system. The viewpoints of both schools are consistent with the mathematics of quantum mechanics. Our immediate purpose is to survey the landscape that allows them to be so; in particular we will sketch how the space of states as defined by the first school is (at least morally speaking) the base space of a fibre bundle that can serve as the true space of states according to the second school. In the remainder of this chapter we will study this base space in its own right, especially in its capacity of a convex set, while the next chapter is devoted to the fibre bundle aspects that arise from the viewpoint of the second school. Either way the properties of the space of density matrices are important, regardless of which of the schools—if any—that is right in a fundamental sense.

So, to begin with we assume that we have a Hilbert space $\mathcal{H}$. There is

then always a dual Hilbert space $\mathcal{H}'$ defined as the space of maps from $\mathcal{H}$ to the complex numbers; in the finite dimensional case these two spaces are isomorphic. Thus

$$\mathcal{H} = \mathcal{H}' \qquad \dim\mathcal{H} = 2n \ . \tag{271}$$

(We give the real dimension of the various spaces as they are introduced—note that we now assume that the complex dimension of our Hilbert space is n, in disagreement with our earlier convention.) 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

$$< A, B >= \mathrm{Tr}A^\dagger B \ . \tag{272}$$

We call it the Hilbert-Schmidt space $\mathcal{HS}$. (In the infinite dimensional case the trace may diverge; operators for which $\mathrm{Tr}A^\dagger A$ is finite are called Hilbert-Schmidt operators.) It is not hard to see that

$$\mathcal{HS} = \mathcal{H} \otimes \mathcal{H}' = \mathcal{H} \otimes \mathcal{H} \qquad \dim\mathcal{HS} = 2n^2 \ . \tag{273}$$

The point is that any operator can be written as a linear combination of outer products of vectors and dual vectors, i.e.

$$A^\alpha{}_\beta = \lambda P^\alpha \bar{Q}_\beta + \ \ldots \ . \tag{274}$$

The number of terms that occur on the right hand side is at most n^2.

Now there is a natural projection from the Hilbert-Schmidt space to the space $\mathcal{P}$ of positive operators on $\mathcal{H}$:

$$\mathcal{HS} \to \mathcal{P} \ : \qquad A \to AA^\dagger \qquad \dim\mathcal{P} = n^2 \ . \tag{275}$$

We recall that by definition a positive operator is an operator such that the inequality

$$< \psi|A|\psi > \geq 0 \tag{276}$$

holds for all the states $|\psi >$ in the Hilbert space. It can be shown that a positive operator is always Hermitian. Evidently there is a further projection down to the space of positive operators of unit trace, that is to say to the space $\mathcal{S}$ of density matrices

$$\mathcal{HS} \to \mathcal{S} \ : \qquad A \to \frac{AA^\dagger}{\mathrm{Tr}AA^\dagger} \qquad \dim\mathcal{S} = n^2 - 1 \ . \tag{277}$$

Later on we will show that these projections are actually projections in a principal fibre bundle (or at least almost so—some little restriction has to be enforced to make this statement quite true). For the moment we just collect the three defining properties of a density matrix:

$$\rho^\dagger = \rho \qquad \rho \geq 0 \qquad \mathrm{Tr}\rho = 1 \ . \tag{278}$$

(The middle equation being shorthand for the statement that all the eigenvalues are positive or zero.) Then we consider the subspace of $\mathcal{S}$ defined by

$$\mathbf{CP}^{n-1} \in \mathcal{S} \ : \qquad \rho^2 = \rho \qquad \dim\mathbf{CP}^{n-1} = 2(n-1) \ . \tag{279}$$

The reason why the embedded subspace that is defined by the equation can be identified with $\mathbf{CP}^{n-1}$ is that an operator that squares to itself must have eigenvalues equal to either zero or one; since its trace is required to be unity the eigenvalue one has to occur exactly once, in other words this is an operator of rank one that projects down to a one dimensional subspace of $\mathcal{H}$. Such a subspace can be regarded as a point in $\mathbf{CP}^{n-1}$.

There is another obvious point worth making, which is that our complex projective space lies in the boundary of $\mathcal{S}$. This is obvious because if one eigenvalue is zero there will be an operator with one eigenvalue negative in its neighbourhood. Note also that a similar set of states can be found in the boundary of the space $\mathcal{P}$. It will be given by the equation

$$\rho^2 = \rho\mathrm{Tr}\rho \ . \tag{280}$$

A matrix obeying this condition has rank one. In both cases these states form only a part of the boundary; the entire boundary is made up of all operators of less than maximal rank.

To summarize, we have defined the spaces $\mathcal{H}$, $\mathcal{HS}$, $\mathcal{P}$, $\mathcal{S}$ and $\mathbf{CP}$ and we have shown how they are related to each other.

Positive operators

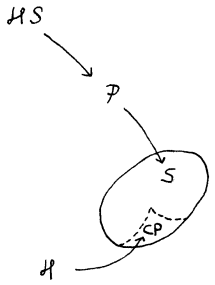

Figure 34: Our spaces—the maps are bundle projections and everybody acts on $\mathcal{H}$

Some brief reminders about positive operators may prove useful. A basic property is that every positive operator P is Hermitian,

$$P^\dagger = P \ . \tag{281}$$

Another is that it always admits a unique positive square root $\sqrt{P}$,

$$(\sqrt{P})^2 = P \ . \tag{282}$$

If A is any operator (not necessarily Hermitian) then the operator

$$|A| \equiv \sqrt{AA^\dagger} \tag{283}$$

is positive. Furthermore, if A is invertible then it can be uniquely decomposed into polar form, which means that

$$A = |A|U \tag{284}$$

where U is a unitary operator. The polar decomposition is clearly analogous to the representation $z = re^{i\phi}$ of a complex number.

A central fact about the space $\mathcal{P}$ of positive operators is that it is a convex set. Clearly the operator

$$P = \mu P_1 + (1-\mu)P_2 \ , \qquad 0 \leq \mu \leq 1 \tag{285}$$

is positive whenever P_1 and P_2 are. In fact it is a convex cone over the space of density matrices with its apex at the operator 0.

The space of density matrices

The space of density operators is a convex set in itself. Its dimension D grows rather quickly with the dimension n of the Hilbert space; for $n = 2$ we get $D = 3$, if $n = 3$ then $D = 8$, and if $n = 4$ then $D = 15$. We must learn to live with these numbers.

Let us now focus on the convexity properties of $\mathcal{S}$. The boundary of a compact convex set always has the topology of a sphere, but metrically this is usually not the case. It is clear that the boundary of $\mathcal{S}$, as a subset in the vector space of Hermitean matrices of trace one, lies in the hypersurface defined by the equation

$$\det \rho = 0 \ . \tag{286}$$

This hypersurface also separates a region with (say) one negative eigenvalue from a region with two negative eigenvalues, so that it is clear that our boundary is a subset of the hypersurface only. Upon reflection we see a kind of "simplex" structure emerging; there will be "corners" where the rank $r = 1$, "edges" where $r = 2$, "faces" where $r = 3$ and so on. (The quotation marks must be kept—the set of "corners" is a connected space, so the "simplex" is not a simplex! Nevertheless I will use this terminology on occasion.) It is not hard to show that the dimension of that part of the boundary where the rank is r is

$$D = r(2n - r) - 1 \ . \tag{287}$$

The "corners" therefore have dimension $D = 2(n - 1)$; the set of "corners" is of course the set of projectors of rank one and hence complex projective space. Now any density matrix can be diagonalized and written as

$$\rho = \lambda_1 Z_1 \bar{Z}_1 + \lambda_2 Z_2 \bar{Z}_2 + \ ... \ + \lambda_n Z_n \bar{Z}_n \ , \quad \lambda_1 + \lambda_2 + \ ... + \lambda_n = 1 \ , \tag{288}$$

where all the λs are positive and the vectors have unit norm. This equation shows that an arbitrary density matrix is a convex combination of rank one projectors, or in other words that $\mathbf{CP}^{n-1}$ is the set of pure states of the convex set $\mathcal{S}$.

We would like to go further and introduce a geometry on the space of density matrices that induces the Fubini-Study metric on the pure states.

This is not really needed as far as the convexity properties are concerned; as a convex set the space of density matrices is just a subset of an affine space. Nevertheless it may be convenient to introduce a flat metric, and we proceed to do so. The Killing metric on the Lie algebra of traceless hermitian matrices—i.e. $su(n)$—will serve our purposes. Thus we define the distance between two arbitrary density matrices ρ_1 and ρ_2 as

$$D^2(1,2) = \frac{1}{2}\mathrm{Tr}(\rho_1 - \rho_2)^2 \ . \tag{289}$$

We can also define the tangent vector $d\rho$ and write the line element

$$ds^2 = \frac{1}{2}\mathrm{Tr}d\rho d\rho \ . \tag{290}$$

Of course there are many other flat metrics available on the space of density matrices, related to the present one by affine transformations. Why should we select this particular one? The answer is that we insist that it should induce the Fubini-Study metric on the pure states, and indeed this is what this metric does. We proceed to show this. Set

$$\rho_1{}^\alpha{}_\beta = \frac{P^\alpha \bar{P}_\beta}{P \cdot \bar{P}} \qquad \rho_2{}^\alpha{}_\beta = \frac{Q^\alpha \bar{Q}_\beta}{Q \cdot \bar{Q}} \ . \tag{291}$$

This yields

$$D^2(1,2) = 1 - \kappa = 1 - \cos^2 d \ , \tag{292}$$

where κ is the projective cross ratio and d is the Fubini-Study distance, i.e. the distance between the points along a curve lying in the pure states. If the points are very close we can expand the square of the cosine to second order and obtain

$$D = d + \text{higher order terms} \ . \tag{293}$$

This proves our point.

The simplest case occurs if the dimension of the Hilbert space is $n = 2$. Then the pure states lie on a sphere ($\mathbf{CP}^1$), and a general density matrix has the form

$$\rho = \begin{pmatrix} \frac{1}{2} + z & x - iy \\ x + iy & \frac{1}{2} - z \end{pmatrix} \ . \tag{294}$$

In this case

$$\rho^2 = \rho \quad \Leftrightarrow \quad \det \rho = 0 \quad \Leftrightarrow \quad x^2 + y^2 + z^2 = \frac{1}{4} \ . \tag{295}$$

The space $\mathcal{S}$ of density matrices form the interior of a sphere, and this sphere is precisely the space $\mathbf{CP}^1$ of pure states. With the metric we adopted, the entire ball is embedded in flat space. As a convex set the ball is of course strikingly different from the simplices that arise in statistics; like in colour theory it is no longer possible to express a given state as a convex combination of pure states in any unique way. Unlike both the pure states now make up the entire boundary, but we already know that this only happens to be true for $n = 2$. For $n > 2$ things are not so simple.

Time evolution

There is an analogue of Wigner's theorem that applies to density matrices; its conditions focus on the convex structure rather than any metric properties. We require a map from $\mathcal{S}$ to $\mathcal{S}$ which is one-to-one and onto, and which preserves the convex structure in the sense that

$$\mu\rho_1 + (1-\mu)\rho_2 \to \mu\rho_1' + (1-\mu)\rho_2' \ . \tag{296}$$

Then Kadison's theorem asserts that such a map must take the form

$$\rho' = U\rho U^{-1} \ , \tag{297}$$

where the operator U is either unitary or anti-unitary. If we take it that the conditions are desirable ones we conclude that time evolution must take the form

$$\dot{\rho} = i[\rho, H] \ , \tag{298}$$

where H is an Hermitian operator. This is a Killing flow with respect to the flat metric since the maps allowed by Kadison's theorem preserve distance.

What are the fixed points of the time evolution? Evidently the set of fixed points consists of all density matrices that commute with H. If we want to be more explicit about it, we can be so by choosing a basis such that

$$U(t) = \mathrm{diag}(e^{iE_1 t}, e^{iE_2 t}, ..., e^{iE_n t}) \ . \tag{299}$$

Then we find for the matrix elements that

$$\rho_{\underline{\beta}}^{\underline{\alpha}}(t) = \rho_{\underline{\beta}}^{\underline{\alpha}}(0) e^{i(E_{\underline{\alpha}} - E_{\underline{\beta}})t} \ . \tag{300}$$

If all the eigenvalues of H are non-degenerate then the set of fixed points consists precisely of all density matrices that are diagonal in the chosen basis; if degeneracies occur the set of fixed points is larger.

Since time evolution preserves distance all the one parameter families $\rho(t)$ form circles at constant distance from any given fixed point. This is hard to visualize except for the case $n = 2$, where time evolution simply means that the ball is rotated (rigidly) around some arbitrary axis of fixed points determined by H.

Exploring the space

It is interesting—or at least amusing—to explore the space of density matrices in a naive manner. The problem that we have on our hands is that the dimension of $\mathcal{S}$ is high, so that the space is hard to visualize. Nevertheless, we proceed. As a point in $\mathcal{S}$ to start from we choose the matrix of ignorance,

$$\rho_0 = \frac{1}{n}\mathbf{1} \ . \tag{301}$$

The distance between ρ_0 and an arbitrary pure state ρ_ψ is given by

$$D^2 = \frac{1}{2}\mathrm{Tr}(\frac{1}{n} - \rho_\psi)^2 = \frac{1}{2}\mathrm{Tr}(\frac{1}{n^2} + \rho_\psi^2 - \frac{2}{n}\rho_\psi) = \frac{n-1}{2n} \ . \tag{302}$$

For $n = 2$ we have confirmed that the pure states form a sphere of radius one half. For $n > 2$ this cannot be true since the topology of $\mathbf{CP}^{n-1}$ is not spherical in general, but we have shown—in effect—that $\mathbf{CP}^{n-1}$ can be embedded in $\mathbf{S}^{n^2-2}$.

Somewhat more generally, we can study a maximally commuting set of density matrices. We first choose a basis of $\mathcal{H}$ and then consider all the density matrix that in this basis take the form

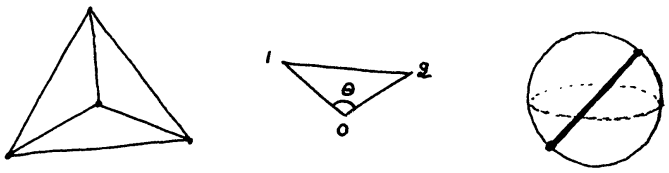

Figure 35: A statistical simplex for $n = 4$; and how the statistical simplex sits in $\mathcal{S}$ for $n = 2$

$$\rho = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n) \ . \tag{303}$$

The distance between a pair of density matrices like this is easy to compute. In effect this subset of density matrices forms a true statistical simplex equipped with its flat metric. Its corners, edges and faces (and so on) consist of density matrices of rank less than n. There are n corners of orthogonal pure states, all of which lie at the distance

$$D = \sqrt{\frac{n-1}{2n}} \tag{304}$$

from the matrix of ignorance and at unit distance from each other. (We are taking a short cut through $\mathcal{S}$—the Fubini-Study distance within the space of pure states is always $\pi/2$!) The angle at ρ_0 in the isosceles triangle formed by ρ_0 and a pair of pure states ρ_1 and ρ_2 follows from simple trigonometry:

$$\sin\frac{\theta}{2} = \frac{D(1,2)}{2D(0,1)} = \sqrt{\frac{n}{2(n-1)}} \ . \tag{305}$$

Note that

$$n = 2 \quad \Rightarrow \quad \theta = \pi \qquad\qquad n \to \infty \quad \Rightarrow \quad \theta \to \pi/2 \ . \tag{306}$$

This gives us a reasonably accurate view of the statistical simplices formed by commuting density matrices. There is nothing peculiarly quantum mechanical about this—we have just described a simplex in a flat $n-1$ dimensional space.

Of course any density matrix can be written on diagonal form if we choose the basis after we choose ρ. Hence for any given ρ we can assume that

$$\rho = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_n) \ , \qquad \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \ . \tag{307}$$

We would like to find the nearest pure state. It is easy to compute the distance to an arbitrary pure state; if we set $Z \cdot \bar{Z} = 1$ for simplicity we get

$$D^2 = \frac{1}{2}(1 + \lambda_1^2 + \ ... \ + \lambda_n^2 - 2\lambda_1 |Z^1|^2 - \ ... \ - \lambda_n |Z^n|^2) \ . \tag{308}$$

If the largest eigenvalue is non-degenerate one finds that the closest pure state is

$$Z^\alpha = (1, 0, ..., 0) \ ; \quad D^2_{min} = \frac{1}{2}((\lambda_0 - 1)^2 + \lambda^2 + \ ... \ + \lambda_n^2) \ . \tag{309}$$

If the two largest eigenvalues are degenerate then there is a $\mathbf{CP}^1$'s worth of pure states

$$Z^\alpha = (Z^0, Z^1, 0, ..., 0) \tag{310}$$

that minimize the distance, and so on. This is as one might have expected.

A picture gallery

Let us try to draw some pictures. This requires us to take three dimensional sections through $\mathcal{S}$. Let us choose $n = 4$ for definiteness, in which case $\mathcal{S}$ has fifteen dimensions. (We choose $n = 4$ rather than 3 because 3 is a prime number and then no entanglement is possible.) Since the dimension is atrociously high any picture will be of limited value, but it can be fun to draw it anyway. To get some interesting structure in it we note that there are two kinds of interesting circles present: The geodesics of $\mathbf{CP}^3$ and the flow lines of some time evolution. Let us begin by choosing a three dimensional section that includes a geodesic circle. All geodesics in $\mathbf{CP}^n$ are equivalent, so we can choose any great circle on any $\mathbf{CP}^1$ to start out with. We take

$$Z^\alpha = (\cos\phi, \sin\phi, 0, 0) \ . \tag{311}$$

We draw it as a circle of radius one half that surrounds the point

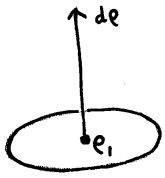

Figure 36: Beginning to draw a picture (twelve dimensions suppressed)

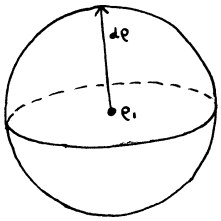

Figure 37: Aiming for a point on the nearby $\mathbf{CP}^1$

$$\rho_1 = \mathrm{diag}(\frac{1}{2}, \frac{1}{2}, 0, 0) \ . \tag{312}$$

The entire disk lies in the "edge" of the "simplex" since all of its points have rank two. We have one dimension left to play with, and the natural strategy seems to be to draw a line through ρ_1 that is orthogonal to the radius of the circle. We continue along this line until we hit the boundary of $\mathcal{S}$, and then we complete the picture by going out from this line in planes that are parallell with the plane in which the circle lies, again until we hit the boundary. The problem is that the line going out from ρ_1 is not unique; all we know is that it lies in a thirteen dimensional hyperplane going through $\mathcal{S}$. Thus we have to select some interesting point to aim for.

Here are some suggestions: A pure state at minimal distance from ρ_1. A pure state at maximal distance from ρ_1. The matrix of ignorance. Combinations or superpositions of those. The first suggestion is easily carried out—our picture becomes a sphere of radius of radius one half surrounding ρ_1. This is the near by $\mathbf{CP}^1$ consisting of the states

$$Z^\alpha = (Z^0, Z^1, 0, 0) \ . \tag{313}$$

The entire ball lies in the "edge" and its surface in a "corner" (i.e. the surface

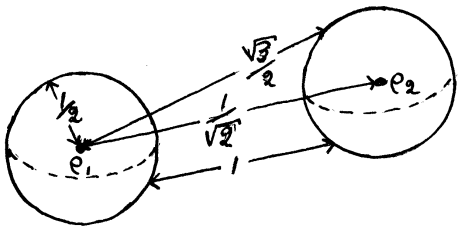

Figure 38: A picture that cannot be drawn in three dimensions

consists of pure states). If we choose to go out in one of the twelve remaining orthogonal directions instead the situation rapidly becomes confusing. There is a far away $\mathbf{CP}^1$ formed by the pure states

$$Z^\alpha = (0, 0, Z^2, Z^3) \ . \tag{314}$$

They form a sphere of radius $\sqrt{3}/2$ surrounding ρ_1. They also form a sphere of radius one half surrounding the point

$$\rho_2 = \mathrm{diag}(0, 0, \frac{1}{2}, \frac{1}{2}) \ . \tag{315}$$

The situation is symmetrical, so the near by $\mathbf{CP}^1$ forms a sphere of radius $\sqrt{3}/2$ surrounding ρ_2. The distance between ρ_1 and ρ_2 is $1/\sqrt{2}$ while the distance from any point on the near by $\mathbf{CP}^1$ to the far away $\mathbf{CP}^1$ is 1. These statements manage to be consistent because we are in a fifteen dimensional space, but if we insist on having our original geodesic circle in our three dimensional picture then we can have at most either one point on the far away $\mathbf{CP}^1$ or the point ρ_2 in the same picture.

Here is a sketch of how the calculation proceeds. Form a unit vector pointing from ρ_1 to ρ_2 (say); it is

$$d\rho_z = \sqrt{2}(\rho_2 - \rho_1) = \frac{1}{\sqrt{2}}\mathrm{diag}(-1, -1, 1, 1) \ \Rightarrow \ ||d\rho_z||^2 = \frac{1}{2}\mathrm{Tr}d\rho_z^2 = 1 \ . \tag{316}$$

(Using old fashioned notation for tangent vectors.) Similarly, form a one parameter family $d\rho_r(\phi)$ of unit vectors pointing from ρ_1 to the geodesic circle, and check that

$$d\rho_z \cdot d\rho_r = \frac{1}{2}\mathrm{Tr} d\rho_z d\rho_r = 0 \ . \tag{317}$$

Now we can draw a line from ρ_1 straight up from the plane where the circle lies. We meet the points

$$\rho(z) \equiv \rho_1 + z d\rho_z = \mathrm{diag}(\frac{1}{2} - \frac{z}{\sqrt{2}}, \frac{1}{2} - \frac{z}{\sqrt{2}}, \frac{z}{\sqrt{2}}, \frac{z}{\sqrt{2}}) \ . \tag{318}$$

We hit the boundary (where the rank is less than maximal) at ρ_2, having encountered the matrix of ignorance along the way. To be precise

$$\rho_0 = \rho(\frac{1}{2\sqrt{2}}) \qquad\qquad \rho_2 = \rho(\frac{1}{\sqrt{2}}) \ . \tag{319}$$

Then we go out orthogonally from the line that we have constructed to get the points

$$\rho(r,z) \equiv \rho_1 + z d\rho_z + r d\rho_r \ . \tag{320}$$

We hit the boundary when

$$\det \rho(r,z) = 0 \ . \tag{321}$$

Solving this equation for $z = z(r)$ gives us an equation for the boundary, and we are ready to draw the picture.

Actually the picture that we obtain in this way is very simple: It is a solid cone with the disk surrounded by the geodesic circle as its base. We repeat the construction for the points

$$\rho_3 = \mathrm{diag}(0, 0, \frac{3}{4}, \frac{1}{4}) \qquad\qquad \rho_4 = \mathrm{diag}(0,0,1,0) \ . \tag{322}$$

The point ρ_4 lies on the far away $\mathbf{CP}^1$ and ρ_3 on the line connecting ρ_4 and ρ_2. Again we get solid cones of varying heights.

Let us now turn to the circles that are flowlines of the time evolution. The time evolution is given by a unitary operator $U(t)$, and we choose a basis in which this operator is diagonal (as discussed above). The fixed points of the flow forms a three dimensional statistical simplex made up of the density matrices that are diagonal in the chosen basis. To draw a picture of the flow

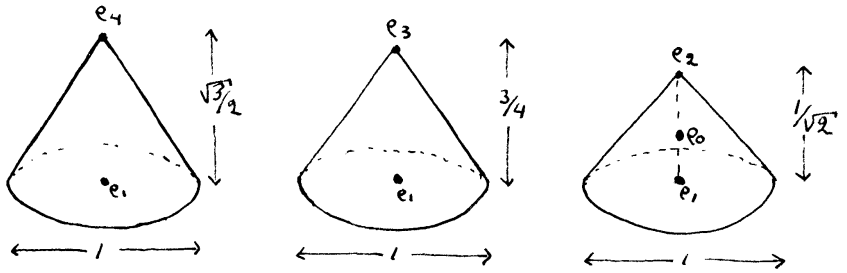

Figure 39: Cones of density matrices, all based on the same geodesic circle

we have to suppress two of the dimensions of the fixed point set. Let us choose the line

$$\rho(0,z,0) = z\rho_2 + (1-z)\rho_1 \tag{323}$$

that connects the two pure fixed points

$$\rho_1 = \mathrm{diag}(1,0,0,0) \quad \rho_2 = \mathrm{diag}(0,1,0,0)\ . \tag{324}$$

The idea is now to use r, z and t as coordinates in our picture, where

$$\rho(r,z,t) = \rho_1 + z d\rho_z + rU(t)d\rho_r U^\dagger(t)\ ; \tag{325}$$

here $d\rho_z$ is a unit vector pointing along the line of fixed points and $d\rho_r$ is a unit vector orthogonal to $d\rho_z$. The boundary of the set of density matrices in our picture will then be given by the equation

$$\det \rho(r,z,t) = 0\ . \tag{326}$$

The picture that we will obtain depends on the choice of $d\rho_r$. A choice made more or less at random is

$$d\rho_r = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\ . \tag{327}$$

Working this out we find that the equation for the boundary in our picture becomes

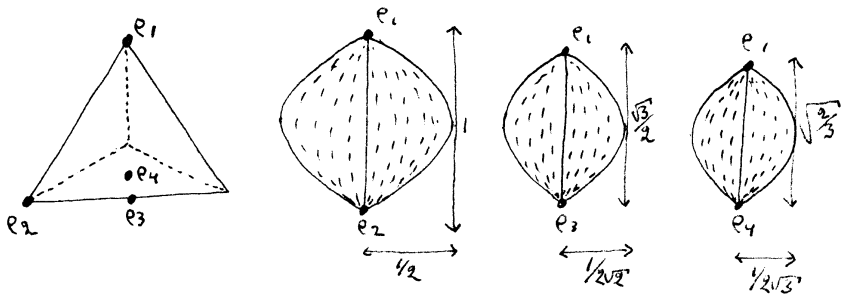

Figure 40: Density matrices surrounding lines of fixed points

$$\det \rho = 0 \quad \Leftrightarrow \quad z(1-z) = r^2 \ . \tag{328}$$

So our picture is a kind of cigar of length 1 and width 1/2, with pure states at the ends. The time evolution takes place along circles around the vertical axis. We repeat the construction for lines of fixed points starting at ρ_1 and ending either at an edge of the simplex, say at

$$\rho_3 = \mathrm{diag}(0, \frac{1}{2}, \frac{1}{2}, 0) \ , \tag{329}$$

or in the middle of a face at

$$\rho_4 = \mathrm{diag}(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \ . \tag{330}$$

We use the same $d\rho_r$ as before. In this way we obtain a set of solid bodies of slightly varied form. They should be compared with the ball that we obtained when the dimension of the Hilbert space was two.

I admit that it is difficult to see what these pictures are good for. But they are the best that I can do.

Exercise:

- Prove that the polar decomposition of an invertible operator is unique.

- Draw a few more pictures containing circles that are flow lines of the time evolution, using different choices of fixed lines and orthogonal directions.

PURIFIED DENSITY MATRICES

The Schmidt decomposition

We will now take the other point of view on density matrices. It will be assumed that the state of the world is described by a point in complex projective space. Nevertheless density matrices do arise, because in practice the world splits in two parts; a part 1 that we study and another part 2 that we may refer to as the environment. In quantum mechanics these two parts are entangled, and if we do not take this entanglement into account it appears as if part 1 is in a state described by a density matrix acting on a small Hilbert space $\mathcal{H}_1$. Thus, suppose that the true state is described by a vector in the Hilbert space

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \ . \tag{331}$$

The system can alternatively be described by a density matrix ρ of rank one acting on $\mathcal{H}$. However, if we can perform measurements only on the first system then the relevant density matrix is

$$\rho_1 = \mathrm{Tr}_2 \rho \ , \tag{332}$$

where the trace is a partial trace over the second system. This is so because as long as we perform measurements only on the first system we will always deal with operators of the form

$$A = A_1 \otimes \mathbf{1}_2 \ . \tag{333}$$

Their expectation values will then always take the form

$$< A >= \mathrm{Tr} \rho A = \mathrm{Tr}_1 \rho_1 A_1 \ . \tag{334}$$

It follows that the reduced density matrix ρ_1 contains all the relevant information about the state. In general ρ_1 will not describe a pure state of the first subsystem, and neither will

$$\rho_2 = \mathrm{Tr}_1 \rho \tag{335}$$

be a pure state of the second. The procedure is known as reduction of the state vector, and according to the second school of thought on quantum states this is how density matrices always arise.

The opposite to reduction of the state vector is purification of the density matrix. We want to find a pure state in some larger product Hilbert space whose reduction gives us the original density matrix back. This problem does not have a unique solution and in fact (as we will see) it can be looked on as a bundle projection.

There is a simple algebraic fact that clarifies the situation. To find it we first observe that whatever the state in $\mathcal{H}$ and whatever orthonormal basis $|e_i>$ we choose in $\mathcal{H}_1$ we can always write the state in $\mathcal{H}$ as

$$|\Psi>=\sum_i |e_i>|\phi_i> \tag{336}$$

—where there is no implication that the states $|\phi_i>$ have any special relation to each other. Taking a partial trace we find

$$\rho_1 = \sum_i \sum_j <\phi_j|\phi_i> \ |e_i><e_j| \ . \tag{337}$$

Now comes the trick. We can always perform unitary transformations to a new basis $|\hat{e}_i>$ in $\mathcal{H}_1$ so that ρ_1 takes the form

$$\rho_1 = \sum_i \lambda_i |\hat{e}_i><\hat{e}_i| \ . \tag{338}$$

We can go back and use this basis from the start. So we write

$$|\Psi>=\sum_i \sqrt{\lambda_i}|\hat{e}_i>|\hat{f}_i> \ \ . \tag{339}$$

The argument then tells us that the $|\hat{f}_i>$s are orthonormal vectors in $\mathcal{H}_2$. They can be used as part of a basis for $\mathcal{H}_2$; they form a complete basis if the rank of ρ_1 equals the dimension of $\mathcal{H}_2$. In conclusion, we have shown that by an appropriate choice of basis vectors in $\mathcal{H}_1$ and $\mathcal{H}_2$ an arbitrary state in the product Hilbert space can be written as a single sum of the particular form known as Schmidt's polar form. This is the clarifying fact that we were after.

Immediately we learn something interesting, namely that

$$\rho_2 = \mathrm{Tr}_1 |\Psi><\Psi| = \sum_i \lambda_i |\hat{f}_i><\hat{f}_i| \ . \tag{340}$$

Thus the density matrices that describe the subsystems are completely isomorhic to each other. In particular the entropies of the two subsystems will be the same because in quantum mechanics the entropy is defined as a function of the density matrix,

$$S = -\mathrm{Tr}\rho \ln \rho \ . \tag{341}$$

The entropy vanishes if and only if the state is pure and the similarity to Shannon's entropy is evident. But this is by the way.

A warning is appropriate as well: There is no similar strong result available for Hilbert spaces that are direct products of more than two factor spaces. This is kind of evident because if there are N factor spaces, all of dimension n, then the number of parameters describing a general state grows like n^N, while the number of unitary transformations one can use to choose basis vectors within the factors grows like $N \times n^2$. For $N > 2$, things get messy and no analogue of the Schmidt decomposition exists in general.

The Hilbert-Schmidt bundle

The clarifying fact that the Schmidt decomposition exists makes itself felt everywhere. We see directly that any density matrix ρ acting on a Hilbert space $\mathcal{H}$ can be purified in the Hilbert-Schmidt space $\mathcal{H} \otimes \mathcal{H}$. Any attempt to use a smaller Hilbert space will fail in general, and 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 $\mathcal{HS}$ provides a canonical arena for the purification of density matrices.

The vectors of $\mathcal{HS}$ can be represented as arbitrary operators A acting on $\mathcal{H}$, and there is a projection down to $\mathcal{S}$ defined by

$$P \ : \quad A \to \rho = \frac{AA^\dagger}{\mathrm{Tr}AA^\dagger} \ . \tag{342}$$

Moreover the unitary group acts on the fibres as

$$A \to A' = AU \ . \tag{343}$$

Is this a fibre bundle? Not quite, because the fibres are not isomorphic. If the rank of A is less than n the set of A's that can be reached in this way is smaller than it would have been if the rank had been maximal. However, if we restrict ourselves to the space $\underline{\mathcal{HS}}$ of invertible operators on $\mathcal{H}$ and the space $\underline{\mathcal{S}}$ of density matrices of maximal rank then the projection

$$P \; : \qquad \underline{\mathcal{HS}} \to \underline{\mathcal{S}} \tag{344}$$

does have isomorphic fibers. Moreover all points on a given fibre can be reached from any special point on the same fibre by acting from the right with a unitary matrix U, together with a scaling to adjust the trace of the operator. (This is clear from the uniqueness of the polar decomposition of invertible operators.) Hence we do have a principal fibre bundle with bundle space $\underline{\mathcal{HS}}$, base space $\underline{\mathcal{S}}$ and structure group $U(n)$ times rescalings. From a topological point of view this is a trivial bundle since it admits a global section; a prescription for assigning a unique point in $\underline{\mathcal{HS}}$ to any point in $\underline{\mathcal{S}}$ is

$$\rho \to \sqrt{\rho} \; . \tag{345}$$

The map is well defined because a positive operator admits a unique positive square root, it is a section because the projection P takes us back to ρ, and it is global because it works everywhere. What is interesting about the Hilbert-Schmidt bundle is its geometry.

The fact that we had to restrict our base space a bit should not by too serious. Clearly the set $\underline{\mathcal{S}}$ of density matrices of maximal rank is dense in the set of all density matrices. It is therefore legitimate to hope that whatever structure on $\underline{\mathcal{S}}$ that we are able to extract from the bundle can be extended to all of $\mathcal{S}$ (although in the event this hope is not fully borne out).

Let us try to find a preferred connection on our bundle. There should be a preferred connection here, since—just like the 3-sphere—the bundle space has a preferred metric defined by the Hermitian form on $\mathcal{HS}$. This metric is

$$X \cdot Y = \frac{1}{2}(< X, Y > + < Y, X >) = \frac{1}{2}\mathrm{Tr}(X^\dagger Y + Y^\dagger X) \; , \tag{346}$$

where X and Y are tangent vectors represented as matrices. To simplify the discussion, let us choose the space of positive operators (of maximal

rank) as our base space. The extra rescaling needed to get down to $\underline{\mathcal{S}}$ is not very interesting. Since the structure group acts on the fibres through right multiplication with a unitary matrix U the vertical tangent vectors at A will be given by

$$dA^{ver} = Au \ , \tag{347}$$

where as usual I am using the old fashioned notation for tangent vectors of which I am inordinately fond and u is an anti-Hermitian matrix such that

$$U = e^u \ . \tag{348}$$

(Recall that a vertical tangent vector points along the fibre.) A connection is defined as soon as we define a set of additional horizontal tangent vectors at every point such that together with the vertical ones they form a basis for the tangent space. Our preferred metric on the bundle space provides a preferred way to define the horizontal subspace of tangent vectors; a tangent vector X at the point A will be horizontal provided that

$$X \cdot dA^{ver} = \frac{1}{2}(X^\dagger Au + u^\dagger A^\dagger X) = 0 \ . \tag{349}$$

Since u is anti-Hermitian this translates to

$$X^\dagger A - A^\dagger X = 0 \ . \tag{350}$$

We could now proceed to define a connection form and use it to lift curves $\rho(\sigma)$ of density matrices in the base space to the bundle. The lifted curve will be unique once it is required to start out from some specified point along the fibre over $\rho(0)$. However, since we have arrived at the sketchy part of these notes we confine ourselves to the remark that this can be done, and to a remark on consistency: If we compute the holonomy of a closed curve in $\mathbf{CP}^{n-1}$ by first deforming the curve into the space of $\underline{\mathcal{S}}$ of density matrices of maximal rank, then computing the holonomy there using the preferred connection on the Hilbert-Schmidt bundle, and finally taking the limit that brings us back to the boundary, we do recover the Berry phase for a closed curve in $\mathbf{CP}^{n-1}$.

The Bures distance

With the connection in hand we have a natural metric on the space of density matrices—we can compute the length of any curve $\rho(\sigma)$ by lifting it into the bundle, computing the length of the lifted curve by means of the natural metric on $\mathcal{HS}$, and defining the length of the original curve to be precisely that. This idea should be familiar from our treatment of the Hopf fibration, where it gave the Fubini-Study metric on $\mathbf{CP}^{n-1}$. A quick way to arrive at the latter result is as follows: Choose a pair of points on the odd dimensional sphere, represented by normalized vectors P and Q. Then compute

$$\min(P-Q)\cdot(\bar{P}-\bar{Q}) \tag{351}$$

—where we move the pair of points up and down along their respective fibres until the minimum is attained. This does not directly give the geodesic distance between points in $\mathbf{CP}^{n-1}$, but if we set $Q = P + dP$ it gives the squared length of the shortest possible tangent vector that connects two neighbouring fibres. This tangent vector has to be orthogonal to the fibre, and hence the result is precisely the line element of the Fubini-Study metric.

We can now repeat this procedure to calculate a line element on the space of density matrices. We choose two points in the Hilbert-Schmidt space that project to a given pair of density matrices; they can be written by means of a polar decomposition as

$$A = \sqrt{\rho_1}V \qquad B = \sqrt{\rho_2}V' \ , \tag{352}$$

where V and V' are unitary operators that move us around the fibres. Then we recall the definition of the norm in Hilbert-Schmidt space, and seek the minimum of

$$||A-B||^2 = \mathrm{Tr}(AA^\dagger + BB^\dagger - AB^\dagger - BA^\dagger) = 2 - \mathrm{Tr}(BA^\dagger + AB^\dagger) \ . \tag{353}$$

Equivalently, we seek the maximum of

$$\mathrm{Tr}(BA^\dagger + \mathrm{h.c.}) = \mathrm{Tr}(\sqrt{\rho_1}\sqrt{\rho_2}V'V^\dagger + \mathrm{h.c}) \ . \tag{354}$$

Making a polar decomposition of the operator $\sqrt{\rho_1}\sqrt{\rho_2}$ and collecting all the unitaries into one operator U this means that we seek the maximum of

$$\mathrm{Tr}|\sqrt{\rho_1}\sqrt{\rho_2}|U + \mathrm{h.c.} \ . \tag{355}$$

In the eigenbasis of the Hermitian operator $|\sqrt{\rho_1}\sqrt{\rho_2}|$ it is easy to see—since the moduli of the diagonal elements of a unitary matrix never exceed unity—that the maximum is attained for $U = 1$. The result of our calculation is known as the Bures distance;

$$D_B^2 \equiv \min||A - B||^2 = 2 - 2\mathrm{Tr}|\sqrt{\rho_1}\sqrt{\rho_2}| \ . \tag{356}$$

This can be rewritten as

$$D_B^2 = 2 - 2\mathrm{Tr}\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}} \ . \tag{357}$$

The drawback of this expression is evident: 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.

Nevertheless it is a geometrically natural construction. If we set $\rho_2 = \rho_1 + d\rho$ we obtain a line element that defines the Bures-Uhlmann metric on the space of density matrices. To a certain extent it is possible to explore the resulting Riemannian geometry, but since time is short I confine myself to some scattered remarks.

In order to be at all interesting the Bures metric had better extend smoothly to that part of the boundary that describes pure states (density matrices of rank one), and moreover it should induce the Fubini-Study metric there. Fortunately it is easy to check that this is so. Let us write

$$\rho_1 = |\psi><\psi| \ . \tag{358}$$

The nice thing about such a density matrix is that it squares to itself and therefore equals its own square root. A very short calculation now shows that

$$D_B^2 = 2 - 2\mathrm{Tr}\sqrt{\rho_1\rho_2\rho_1} = 2 - 2\sqrt{<\psi|\rho_2|\psi>} \ . \tag{359}$$

If ρ_2 is a pure state as well we can introduce the projective cross ratio κ of these two states and write the result as

$$D_B^2 = 2 - 2\sqrt{\kappa} \ . \tag{360}$$

This can be compared with the distance between the two pure states as defined by the flat metric that we used earlier,

$$D^2 = 1 - \kappa \ , \tag{361}$$

as well as with the Fubini-Study distance d along a geodesic that is constrained to lie in the space of pure states,

$$\cos^2 d = \kappa \ . \tag{362}$$

By means of a Taylor expansion to second order we see that for points ρ_1 and ρ_2 that are close enough all these expressions coincide, and this proves that the necessary amount of agreement between these three metrics is indeed present.

The simplest case

Unfortunately the Bures distance between an arbitrary pair of points is considerably hard to come to grips with. The only easy case is $n = 2$, where the density matrices are two-by-two matrices. It helps that any two-by-two matrix obeys the characteristic equation

$$M^2 - M\mathrm{Tr}M + \det M = 0 \ . \tag{363}$$

It follows that

$$(\mathrm{Tr}M)^2 = \mathrm{Tr}M^2 + 2\det M \ . \tag{364}$$

We now set

$$M = \sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}} \ , \tag{365}$$

so that

$$\mathrm{Tr}M^2 = \mathrm{Tr}\rho_1\rho_2 \qquad\qquad \det M = \sqrt{\det\rho_1 \det\rho_2} \ . \tag{366}$$

We can now express the Bures distance in terms of easily computed matrix invariants:

$$D_B^2 = 2 - 2\sqrt{\mathrm{Tr}\rho_1\rho_2 + 2\sqrt{\det\rho_1 \det\rho_2}} \ . \tag{367}$$

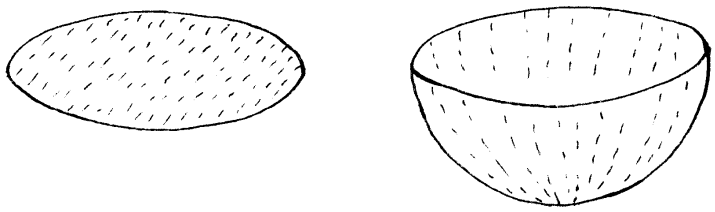

Figure 41: The flat and curved geometries of $\mathcal{S}$; one dimension is suppressed

It is not so easy to perform this trick for $N > 2$.

We can now introduce a coordinate system on the space of density matrices, say

$$\rho = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} , \qquad x^2 + y^2 + z^2 \leq 1 . \tag{368}$$

When we set $\rho_1 = \rho$, $\rho_2 = \rho + d\rho$ in the expression for the Bures distance and expand to second order we find that the Bures metric on two-by-two matrices becomes

$$ds^2 = \frac{1}{4}\left(dx^2 + dy^2 + dz^2 + \frac{(x\,dx + y\,dy + z\,dz)^2}{1 - x^2 - y^2 - z^2} \right) . \tag{369}$$

This is precisely one fourth times the round metric on the 3-sphere, with the intrinsic coordinates given by projection onto the equatorial plane. The pure states make up the equator of the 3-sphere, which is an ordinary two dimensional sphere and hence isometric to $\mathbf{CP}^1$, as we already knew. The resulting picture should be compared to that arising from our previous distance D^2, which regards the interior of the sphere as being flat.

There are rather few explicit results available when $n > 2$, but we do know that the simplest case is quite misleading in some respects. In particular, the Bures metric is not in general a constant curvature metric. Worse than that, the Riemann curvature tensor becomes singular as we approach density matrices of less than maximal rank.

The statistical geometry of density matrices

The reason why the Bures metric is important is that the notion of distance that it defines turns out to be equal to the statistical distance between density matrices. Let us show how this comes about in a special case, namely that of commuting density matrices. We know already that such density matrices define a statistical simplex in the standard sense, and we wish to verify that the Bures metric when restricted to this simplex agrees with the Bhattacharrya metric. Set

$$\rho_1 = \mathrm{diag}(p_1, p_2, ..., p_n) \qquad \rho_2 = \mathrm{diag}(p_1 + dp_1, ..., p_n + dp_n) \ . \tag{370}$$

Insertion in the expression for the Bures distance gives the line element

$$ds_B^2 = 2 - 2\sum_{i=1}^{n} \sqrt{p_i^2 \left(1 + \frac{dp_i}{p_i}\right)} \ . \tag{371}$$

Naturally we have

$$\sum_{i=1}^{n} p_i = 1 \qquad \sum_{i=1}^{n} dp_i = 0 \ . \tag{372}$$

Expanding the square roots and performing the sum therefore gives

$$ds_B^2 = \frac{1}{4}\sum_{i=1}^{n} \frac{dp_i dp_i}{p_i} \tag{373}$$

—precisely the statistical metric on the simplex. And this is the end of the introduction to the geometry of quantum mechanics.