Spheres, Hyperspheres and Quaternions

Lloyd Connellan

May 25, 2014

Submitted as a report for the Final Year Project for the MMath degree at the University of Surrey

Supervisor: Prof. Tom Bridges

Abstract

We principally examine the sphere and the 3-sphere, and the link between them in the form of the Hopf map. We prove the result that points on the 3-sphere correspond to circles on the sphere, and from this we are able to construct the Hopf fibration, $\mathbf{S}^1 \hookrightarrow \mathbf{S}^3 \xrightarrow{\pi} \mathbf{S}^2$, which we establish to be both a fibre bundle, and further a principal bundle. We also generalise the Hopf fibration to other dimensional cases such as $\mathbf{S}^1 \hookrightarrow \mathbf{S}^{2n+1} \xrightarrow{\pi} \mathbb{CP}^n$.

We extensively cover holonomy on the sphere and the generalisations of this. We firstly demonstrate in the context of classical differential geometry, that vectors that are parallel transported on the sphere are rotated due to the curvature of the sphere upon return to their original position. We then generalise this to the Hopf bundle as a principal bundle, using the Ehresmann connection, which is a deconstruction of the tangent space of the 3-sphere into the horizontal and vertical subspaces. We then redefine the notion of parallel transport in terms of the horizontal space of the 3-sphere. In the last couple of sections, we look at how this extends to Hermitian matrices, namely the effect of holonomy on the eigenvectors of a matrix, and an application of this in quantum mechanics in the form of the Berry phase is discussed.

We further explore other properties of n-spheres such as the isomorphism between $\mathbb{R}^n$ and $\mathbf{S}^n$ using stereographic projection, and how we can use this to project the Hopf fibration on to $\mathbb{R}^3$, which we show gives a series of nested tori. We will also examine geodesics on $\mathbf{S}^n$, namely showing that geodesics lie on great circles. Throughout this project, the link of various concepts to quaternions is discussed, in particular that unit quaternions are isomorphic to the 3-sphere.

Contents

1 Introduction **4**

2 Quaternions **5**

3 The Sphere and the Hypersphere **8**

4 Holonomy **8**

5 Geodesics **14**

6 Stereographic Projection **16**

7 The Hopf Map **18**

7.1 Properties of the Hopf Map 20
7.2 The Hopf Fibration 22
7.3 Generalisations of the Hopf Fibration 24

8 Stereographic Projection of the Hopf Fibration **25**

9 Ehresmann Connection **27**

10 Principal Bundles **31**

10.1 Principal Connections 33
10.2 Holonomy on a Principal Bundle 36

11 Hermitian Matrices **40**

11.1 Diabolic Points 41
11.2 Symmetric Matrices 44
11.3 Relation to 3-sphere 47

12 Berry Phase **48**

12.1 Properties of Berry Phase 50

A Appendix **52**

A.1 Proof that the Hopf Map is a Submersion 52

A.2 Proof that $T_p\mathbf{S}^3 = \text{span}\{J_1\mathbf{n}, J_2\mathbf{n}, J_3\mathbf{n}\}$ 53
A.3 Form of a Hermitian Matrix with a Double Eigenvalue 54

1 Introduction

In this project I will be investigating primarily the properties of the 3-sphere, as well as the ordinary sphere, and the many interesting geometrical properties they can be shown to have. The 3-sphere is an intriguing mathematical object to study, since it is very easily defined (as a locus of points around the origin), and is one of the simplest examples of a 3-dimensional manifold, and yet it does not even exist in the world we live in, $\mathbb{R}^3$. Throughout the course of this project, I will be introducing the tools necessary for understanding the geometry of the 3-sphere, including the Hopf mapping, stereographic projection and connections, and how they can be derived.

One of the first things we look at is holonomy, which is essentially the idea that the curvature of the sphere (or 3-sphere) causes vectors that are *parallel* transported (i.e. moved preserving the direction of the vector) around the sphere, will nonetheless end up pointing a different direction upon return to their original point. An example of this is the *rolling ball* phenomenon described by Hanson in [2, p. 123-132], which can be observed by placing one's hand on a ball, and moving it parallel to the floor in small "rubbing" circles. What you will find is that the ball will end up being rotated despite not rotating one's hand at all, which is essentially an example of holonomy at work.

Another key concept we will cover is the Hopf fibration, which uses the Hopf map (a map from the 3-sphere to the sphere), to describe the 3-sphere in terms of the sphere and circles (which we will call fibres). We will go on to show that this is an example of a fibre bundle, a structure that we can examine by using stereographic projection (a way of mapping an n-sphere to $\mathbb{R}^n$). In later sections we will go even further and show that with additional structure it is an example of a principal bundle, and we will see that with the help of connections, we can revisit holonomy on the 3-sphere as a principal bundle.

Finally we will see how the idea of holonomy applies to parameter dependent Hermitian matrices, which it can be shown have a relation to the 3-sphere and to quaternions. We will look at how the eigenvectors of the matrix undergo a *phase shift* when the parameters are moved along a closed curve, and what happens when we restrict the case to symmetric matrices. We then look at an application of this in quantum mechanics with the Berry phase, which essentially looks at what happens when the time-evolution of the eigenvectors is described by the Schrödinger equation.

2 Quaternions

We first introduce the Quaternions, denoted by $\mathbb{H}$, a natural extension to the complex numbers $\mathbb{C}$ that correspond to $\mathbb{R}^4$ as opposed to $\mathbb{R}^2$. Whereas complex numbers are written in the form $z = x + yi$ or as a vector (x, y), quaternions are written in the form $q = a + bi + cj + dk$ or as a 4-vector (a, b, c, d).

As we know, complex numbers can be multiplied together by observing the property $i^2 = -1$, e.g.

$$\begin{aligned}(x_1 + y_1 i)(x_2 + y_2 i) &= x_1 x_2 + y_1 y_2 i^2 + x_1 y_2 i + x_2 y_1 i \\ &= (x_1 x_2 - y_1 y_2) + (x_1 y_2 + x_2 y_1) i\end{aligned}$$

Similarly, quaternions can be multiplied together by simply defining the multiplication rules between i, j and k. We have

$$i^2 = j^2 = k^2 = -1$$

$$ij = k,\ jk = i,\ ki = j$$

This definition of multiplication, whilst closed, is non-commutative and for instance

$$ji = -k = -ij$$

More explicitly, we can write out the product of two quaternions in terms of their components. A useful way to do this is to split a quaternion into its first entry (the real part) and its (i, j, k) part as $q = (a, \mathbf{u})$ where $\mathbf{u} = (b, c, d)$. Then with another quaternion $p = (w, \mathbf{v})$, with $\mathbf{v} = (x, y, z)$, we have the following relation.

$$q * p = (aw - \mathbf{u} \cdot \mathbf{v}, a\mathbf{v} + w\mathbf{u} + \mathbf{u} \times \mathbf{v}) \tag{1}$$

Another standard operation on quaternions is the dot product. As for any vector, the dot product of two quaternions is the summation of each pair of entries multiplied together. i.e. for the quaternions defined above, we have

$$\begin{aligned} q \cdot p &= aw + bx + cy + dz \\ &= aw + \mathbf{u} \cdot \mathbf{v}. \end{aligned}$$

This allows us to define the length or norm $\|q\|$ of a quaternion. Just as for complex numbers, this is found by taking the dot product of q with itself, i.e.

$$\|q\|^2 = q \cdot q = a^2 + b^2 + c^2 + d^2$$

We will pretty much exclusively be dealing with unit quaternions in this report, i.e. quaternions q such that $\|q\| = 1$.

In $\mathbb{C}$, we define the complex conjugate of a complex number $z = x + yi$ to be $\overline{z} = x - yi$. Similarly we can take the conjugate of a quaternion in the following way.

$$\overline{q} = a - bi - cj - dk \tag{2}$$

This is defined to be such so that

$$q * \overline{q} = (q \cdot q, \mathbf{0})$$

It can be shown that complex numbers can be written in terms of matrices. For example an arbitrary complex number $z = x + yi$ is equivalent to:

$$z = x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + y \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Similarly, quaternions can be represented in terms of matrices. For a quaternion q = a + bi + cj + dk, we have

$$\begin{aligned} q &= a \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + b \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\ &+ c \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

or more concisely

$$q = aI + bJ_1 + cJ_2 + dJ_3 \tag{3}$$

It can even be shown that these matrices possess the same qualities as the quaternion basis components i, j and k. For instance, we have that

$$J_1^2 = J_2^2 = J_3^2 = -I,$$

and that

$$J_1J_2 = J_3,\ \ J_2J_3 = J_1,\ \ J_3J_1 = J_2.$$

A more succinct way to write this relation, which we may make use of later on, is the following. For $i, j = 1, 2, 3$, we have

$$J_iJ_j = \sum_{k=1}^{3} \epsilon_{ijk} J_k$$

where ϵ_{ijk} is the Levi-Civita symbol (an antisymmetric scalar).

3 The Sphere and the Hypersphere

This report will deal primarily with the unit sphere $\mathbf{S}^2$ and the unit 3-sphere $\mathbf{S}^3$, which are defined as follows.

$$\begin{aligned}\mathbf{S}^2 &= \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1\} \\ \mathbf{S}^3 &= \{(a,b,c,d) \in \mathbb{R}^4 : a^2+b^2+c^2+d^2 = 1\}\end{aligned}$$

Note the usage of a, b, c and d to refer to the entries of an element $\in \mathbf{S}^3$, the same notation as that used in section 2 to refer to quaternions. The reason behind this is that the space of unit quaternions is essentially identical to $\mathbf{S}^3$. Indeed, as stated in [2, p. 80], if a quaternion given by $q = (a,b,c,d)$ obeys the constraint $q \cdot q = 1$, then the locus of these points is the 3-sphere $\mathbf{S}^3$.

In general, although we will primarily deal with the 3-sphere, a hypersphere is any n-dimensional sphere where n is greater than 2. The equation for an n-sphere is simply:

$$\mathbf{S}^n = \{\mathbf{x} \in \mathbb{R}^{n+1} : \|\mathbf{x}\| = 1\}$$

There also exists another way of defining an n-sphere (when n is an odd number) using complex numbers that we will want to use later on. Namely we have:

$$\mathbf{S}^{2n+1} = \{\mathbf{z} \in \mathbb{C}^{n+1} : \|\mathbf{z}\| = 1\}$$

We can see that this is consistent with the first definition, since each entry in an $(n+1)$-vector $\mathbf{z} \in \mathbb{C}^{n+1}$, i.e. an element of the form $x+iy$ where $x, y \in \mathbb{R}$, corresponds to two entries x and y in $\mathbf{x} \in \mathbb{R}^{2n+2}$.

4 Holonomy

In this section, we will introduce the idea of holonomy that we will go into more detail in in later sections, using the sphere as a basis to understand this concept. Since we will be working in $\mathbb{R}^3$, results that we can obtain from classical differential geometry, i.e. the study of 2D surfaces in $\mathbb{R}^3$, will prove to be useful. For this reason I'll start this section by defining the tools we want to work with, namely coordinate charts, tangent spaces and Christoffel symbols.

Let M be a manifold embedded in $\mathbb{R}^3$, and let U be a subset of $\mathbb{R}^2$. Then we define a chart of a surface in M to be $\mathbf{X}(u,v) : U \to M$ where $(u,v) \in U$. Furthermore the *tangent space* $T_p(M)$ at each point $p \in M$ is given by:

$$T_p(M) = \text{span}\{\mathbf{X}_u, \mathbf{X}_v\} \tag{4}$$

Where $\mathbf{X}_u$ and $\mathbf{X}_v$ are the derivatives of $\mathbf{X}$ with respect to u and v. We can also define the normal $\mathbf{n}$ to the surface to be:

$$\mathbf{n} = \frac{\mathbf{X}_u \times \mathbf{X}_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|}$$

This definition of $\mathbf{n}$ allows us to form a moving frame along the surface, $\text{span}\{\mathbf{X}_u, \mathbf{X}_v, \mathbf{n}\}$, that is a basis for $\mathbb{R}^3$ at each point $\in M$. This is important to know, since it means that any vector $\in \mathbb{R}^3$ can be written as a linear combination of the basis. In particular, we can use this fact to write out an expression for the second derivatives of $\mathbf{X}$ in terms of the moving frame. Since it will be useful for us to do so, rather than using arbitrary letters for the coefficients, we will use *Christoffel symbols* which take the form Γ^i_{jk}.

$$\mathbf{X}_{uu} = \Gamma^1_{11}\mathbf{X}_u + \Gamma^2_{11}\mathbf{X}_v + L\mathbf{n} \tag{5}$$
$$\mathbf{X}_{uv} = \Gamma^1_{12}\mathbf{X}_u + \Gamma^2_{12}\mathbf{X}_v + M\mathbf{n} \tag{6}$$
$$\mathbf{X}_{vv} = \Gamma^1_{22}\mathbf{X}_u + \Gamma^2_{22}\mathbf{X}_v + N\mathbf{n} \tag{7}$$

As we can see, the upper index of the Christoffel symbol corresponds to which first derivative it is a coefficient of, and the lower indices correspond to the partial derivatives of the second derivative. Note that $\mathbf{n}$ doesn't use Christoffel symbols since we ideally want to exclude these terms, as we will see when we introduce the covariant derivative later. Also note the property that $\Gamma^i_{jk} = \Gamma^i_{kj}$, since smooth partial differentiation is not dependent on the order.

As the name of this section implies, we now want to deal with *holonomy* on the sphere, which is essentially the idea that a vector moved around the sphere in a parallel fashion will have its direction altered due to the curvature of the sphere. As illustrated in figure 1 below, a vector can be transported from a point A, around a closed curve without rotating it, and yet when it returns to A, it points in a different direction.

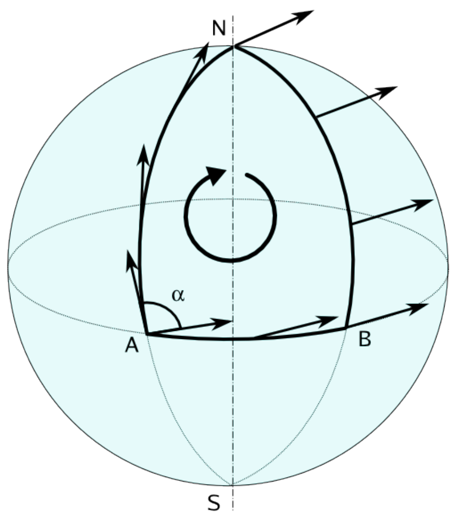

Figure 1: Example of holonomy on the sphere

To study this phenomenon we will first need to define the *covariant derivative* of a tangent vector field $\mathbf{w} \in T_p(M)$ in the direction of a tangent vector $\mathbf{v} \in T_p(M)$ on a surface. Note that $\mathbf{w}$ and $\mathbf{v}$ are of the form.

$$\begin{aligned} \mathbf{w}(u,v) &= w_1(u,v)\mathbf{X}_u + w_2(u,v)\mathbf{X}_v \\ \mathbf{v}(u,v) &= v_1(u,v)\mathbf{X}_u + v_2(u,v)\mathbf{X}_v \end{aligned}$$

Since they belong to the tangent space defined in (4). We define the covariant derivative $\nabla_{\mathbf{v}}\mathbf{w} : T_p(M) \to T_p(M)$ as follows.

$$\nabla_{\mathbf{v}}\mathbf{w} = \mathbf{w}_{\mathbf{v}} - \langle \mathbf{w}_{\mathbf{v}}, \mathbf{n} \rangle \mathbf{n}$$

Where $\langle , \rangle$ is the inner product in our space. The expression $\mathbf{w_v}$, which is actually taking the derivative of a tangent vector field with respect to a tangent vector, is defined as the following.

$$\begin{aligned}\mathbf{w_v} &= (\frac{\partial w_1}{\partial u}v_1 + \frac{\partial w_1}{\partial v}v_2)X_u + (\frac{\partial w_2}{\partial u}v_1 + \frac{\partial w_2}{\partial v}v_2)X_v \\ &\quad + w_1(X_{uu}v_1 + X_{uv}v_2) + w_2(X_{vu}v_1 + X_{vv}v_2)\end{aligned}$$

We will see that this expression can easily be simplified using Christoffel symbols.

The way to understand this definition is that we want to take away any derivatives in the direction $\mathbf{n}$ when we define the covariant derivative. For instance let us examine the expression we had for the derivative with respect to u of $\mathbf{X}_u$ in (5).

$$\mathbf{X}_{uu} = \Gamma^1_{11}\mathbf{X}_u + \Gamma^2_{11}\mathbf{X}_v + L\mathbf{n}$$

Supposing we want to exclude the $\mathbf{n}$ part of this expression, we would need to take away $L\mathbf{n}$. L can be computed explicitly using the inner product $\langle , \rangle$ in the following way.

$$\begin{aligned}\langle \mathbf{X}_{uu}, \mathbf{n} \rangle &= \Gamma^1_{11}\langle \mathbf{X}_u, \mathbf{n} \rangle + \Gamma^2_{11}\langle \mathbf{X}_v, \mathbf{n} \rangle + L\langle \mathbf{n}, \mathbf{n} \rangle \\ \langle \mathbf{X}_{uu}, \mathbf{n} \rangle &= L\end{aligned}$$

Where we have used the fact that $\mathbf{n} \cdot \mathbf{X}_u = \mathbf{n} \cdot \mathbf{X}_v = 0$. Thus we can see that in this case the formula for the covariant derivative where we take away $\langle (\mathbf{X}_u)_u, \mathbf{n} \rangle \mathbf{n}$ is appropriate.

After having defined the covariant derivative, I now want to define the notion of a vector being parallel along a curve. Let $\gamma(t) = \mathbf{X}(u(t), v(t))$ be a curve in M, and let $\mathbf{w}(t) = w_1(t)\mathbf{X}_u + w_2(t)\mathbf{X}_v$ be a tangent vector field. Then $\mathbf{w}$ is *parallel* along γ if:

$$\nabla_{\dot{\gamma}}\mathbf{w} = 0$$

The idea of this definition, is that we can *parallel transport* a vector along a curve, by requiring that the vector remains parallel to the curve at all times (i.e. obeys the above equation). Furthermore we can also define that γ is a *geodesic* if:

$$\nabla_{\dot{\gamma}}\dot{\gamma} = 0$$

Geodesics can be thought of as the shortest path between two points on a surface. In the context of a sphere, we will see that this means the curve must lie on a circle. For now we will just use the first definition and try to understand it. Directly applying the formula, we get for an arbitrary $\mathbf{w}$:

$$\begin{aligned} \nabla_{\dot{\gamma}}\mathbf{w} &= \frac{d}{dt}(w_1\mathbf{X}_u + w_2\mathbf{X}_v) - \langle \dot{\mathbf{w}}, \mathbf{n}\rangle\mathbf{n} \\ &= \dot{w}_1\mathbf{X}_u + \dot{w}_2\mathbf{X}_v + w_1(\mathbf{X}_{uu}\dot{u} + \mathbf{X}_{uv}\dot{v}) + w_2(\mathbf{X}_{vu}\dot{u} + \mathbf{X}_{vv}\dot{v}) - \langle \dot{\mathbf{w}}, \mathbf{n}\rangle\mathbf{n} \end{aligned}$$

Now we will use equations (5), (6) and (7) to simplify the second derivatives in this expression, thus obtaining the following expression in terms of $\mathbf{X}_u$ and $\mathbf{X}_v$ ($\mathbf{n}$ terms will disappear by definition).

$$\begin{aligned} \nabla_{\dot{\gamma}}\mathbf{w} &= (\dot{w}_1 + \Gamma^1_{11}w_1\dot{u} + \Gamma^1_{12}(w_1\dot{v} + w_2\dot{u}) + \Gamma^1_{22}w_2\dot{v})\mathbf{X}_u \\ &\quad +(\dot{w}_2 + \Gamma^2_{11}w_1\dot{u} + \Gamma^2_{12}(w_1\dot{v} + w_2\dot{u}) + \Gamma^2_{22}w_2\dot{v})\mathbf{X}_v \end{aligned} \tag{8}$$

Since $\mathbf{X}_u$ and $\mathbf{X}_v$ are linearly independent, this expression vanishes iff its coefficients are zero. This gives us a pair of simultaneous equations to solve if we want $\mathbf{w}$ to be parallel.

Example: Suppose $\gamma(t)$ is a curve of constant latitude $v(t) = v_0$ on the sphere and $u(t) = t$. i.e.

$$\gamma(t) = \mathbf{X}(u(t), v(t)) = \begin{pmatrix} r\cos t \sin v_0 \\ r\sin t \sin v_0 \\ r\cos v_0 \end{pmatrix}, \quad 0 \le u \le 2\pi,\ 0 < v < \pi$$

Note that v cannot take values of 0 or π since the chart would cease to be regular. To compute the Christoffel symbols of this chart, we need to compute the first and second derivatives with respect to u and v of $\mathbf{X}$. We get:

$$\mathbf{X}_u = \begin{pmatrix} -r\sin u \sin v \\ r\cos u \sin v \\ 0 \end{pmatrix}, \mathbf{X}_v = \begin{pmatrix} r\cos u \cos v \\ r\sin u \cos v \\ -r\sin v \end{pmatrix},$$

$$\mathbf{X}_{uu} = \begin{pmatrix} -r\cos u \sin v \\ -r\sin u \sin v \\ 0 \end{pmatrix}, \mathbf{X}_{uv} = \begin{pmatrix} -r\sin u \cos v \\ r\cos u \cos v \\ 0 \end{pmatrix}, \mathbf{X}_{vv} = \begin{pmatrix} -r\cos u \sin v \\ -r\sin u \sin v \\ -r\cos v \end{pmatrix}$$

Then we use the same trick we used earlier in this section with the inner product, using the linear independence of $\{\mathbf{X}_u, \mathbf{X}_v, \mathbf{n}\}$. Namely we have:

$$\Gamma^1_{11} = \frac{\mathbf{X}_{uu}\cdot\mathbf{X}_u}{\mathbf{X}_u\cdot\mathbf{X}_u} = 0, \quad \Gamma^1_{12} = \frac{\mathbf{X}_{uv}\cdot\mathbf{X}_u}{\mathbf{X}_u\cdot\mathbf{X}_u} = \cot v, \quad \Gamma^1_{22} = \frac{\mathbf{X}_{vv}\cdot\mathbf{X}_u}{\mathbf{X}_u\cdot\mathbf{X}_u} = 0$$

$$\Gamma^2_{11} = \frac{\mathbf{X}_{uu}\cdot\mathbf{X}_v}{\mathbf{X}_v\cdot\mathbf{X}_v} = -\sin v\cos v, \quad \Gamma^2_{12} = \frac{\mathbf{X}_{uv}\cdot\mathbf{X}_v}{\mathbf{X}_v\cdot\mathbf{X}_v} = 0, \quad \Gamma^2_{22} = \frac{\mathbf{X}_{vv}\cdot\mathbf{X}_v}{\mathbf{X}_v\cdot\mathbf{X}_v} = 0$$

With these Christoffel symbols, we can simply put them in (8) to obtain our two simultaneous equations (and we use that $v = v_0$, $u = t$, $\dot{v} = \dot{v_0} = 0, \dot{u} = \dot{t} = 1$):

$$\begin{aligned} \dot{w_1} + w_2 \cot v_0 &= 0 \\ \dot{w_2} - w_1 \sin v_0 \cos v_0 &= 0 \end{aligned}$$

By differentiating the first equation and substituting in the second, we can get an ODE in terms of w_1 (which we can then easily solve).

$$\begin{aligned} 0 &= \ddot{w_1} + \cos^2 v_0 w_1 \\ \Rightarrow w_1 &= w_{11}\cos\omega t + w_{12}\sin\omega t \\ w_2 &= w_{21}\cos\omega t + w_{22}\sin\omega t \end{aligned}$$

Where $\omega = |\cos v_0|$, and w_{11}, w_{12}, w_{21} and w_{22} are arbitrary constants. In conclusion we see the result of holonomy on a sphere, namely that for any $v_0 \neq \frac{\pi}{2}$ (the equator of the sphere), $\mathbf{w}(0) \neq \mathbf{w}(2\pi)$, and so any tangent vector $\mathbf{w}$ traveling around the sphere on γ will not be equal upon return to its original position.

5 Geodesics

In section 4 we briefly mentioned the condition a curve must have to be a geodesic in relation to the covariant derivative. In this section we will be examining this concept further, and since it is easily generalisable, we will be able to look at geodesics on $\mathbf{S}^n$ (recall from chapter 3, $\mathbf{S}^n = \{\mathbf{x} \in \mathbb{R}^{n+1} : \|\mathbf{x}\| = 1\}$). Since it is easier to do so, we will use an alternate definition for a geodesic in terms of the regular derivative as follows. Let $\gamma(t)$ be a curve on a surface, let $\mathbf{n}(t)$ be a unit vector normal to the surface, and let $\lambda(t) \in \mathbb{R}$. Then γ is a geodesic if:

$$\ddot{\gamma} = \lambda \mathbf{n}$$

We can simplify this equation in the case of $\mathbf{S}^n$ by noting that any vector $\in \mathbf{S}^n$ is normal to $\mathbf{S}^n$ at that point since it points directly out from the origin (and is additionally of unit length). Since $\gamma \in \mathbf{S}^n$, we can thus substitute in γ for $\mathbf{n}$ to get:

$$\ddot{\gamma} = \lambda \gamma \tag{9}$$

In order to solve this equation, we use the property of γ being unit length (since it belongs to $\mathbf{S}^n$). In other words, $\langle \gamma, \gamma \rangle = 1$. By differentiating this expression (twice), we get:

$$\begin{aligned} \langle \dot{\gamma}, \gamma \rangle &= 0 & (10) \\ \Rightarrow \langle \dot{\gamma}, \dot{\gamma} \rangle + \langle \ddot{\gamma}, \gamma \rangle &= 0 & (11) \end{aligned}$$

From equation (9), we can take the inner product with γ on the right to get:

$$\langle \ddot{\gamma}, \gamma \rangle = \lambda \langle \gamma, \gamma \rangle = \lambda$$

Then we use equation (11) to get:

$$\lambda = -\langle \dot{\gamma}, \dot{\gamma} \rangle$$

It is possible to prove that $\langle \dot{\gamma}, \dot{\gamma} \rangle$ is in fact a constant (not automatic) by differentiating it:

$$\frac{d}{dt}\langle \dot{\gamma}, \dot{\gamma} \rangle = 2\langle \ddot{\gamma}, \dot{\gamma} \rangle = 2\lambda \langle \gamma, \dot{\gamma} \rangle = 0$$

Where we have used (9) followed by (10). After deducing that $\langle \dot{\gamma}, \dot{\gamma} \rangle$ is a constant, say α^2, we can easily solve equation (9):

$$\gamma(t) = \mathbf{a}\cos(\alpha t) + \mathbf{b}\sin(\alpha t)$$

Where $\mathbf{a}$ and $\mathbf{b}$ are constant vectors $\in \mathbb{R}^n$ constrained by $\mathbf{a} \cdot \mathbf{a} = \mathbf{b} \cdot \mathbf{b} = 1$, and $\mathbf{a} \cdot \mathbf{b} = 0$ due to $\gamma \in \mathbf{S}^n$. This leads to the result that a geodesic on any n-sphere lies on a great circle.

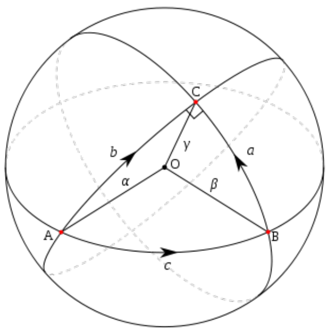

Figure 2: Geodesics on a sphere

As can be seen in figure 2 above, geodesics also represent the shortest distance between two points on the sphere, e.g. the shortest distance between A and B in the diagram is the great circle arc c.

6 Stereographic Projection

Having looked at some results on the ordinary sphere, we now want to turn our attention to $\mathbf{S}^3$, however the problem immediately presents itself that we cannot *see* an object embedded in $\mathbb{R}^4$. If we want to gauge an idea of what $\mathbf{S}^3$ looks like in $\mathbb{R}^3$, i.e. the real world, then stereographic projection is one way to go about doing this, as it is the method of mapping an n-sphere $\mathbf{S}^n$ to $\mathbb{R}^n$.

To explain how stereographic projection works, we first examine the case of the 2-D counterpart which maps $\mathbf{S}^2$ (the sphere) to $\mathbb{R}^2$.

The general idea starts by taking a fixed point on the sphere (for this example we'll take the north pole). Then for each point we want to apply the map to, we draw a line through it and the north pole, and take this line's intersection with the x-y plane (which is essentially $\mathbb{R}^2$). Simply using the equation for a straight line, we can compute that this map $f_1 : \mathbf{S}^2/(0,0,1) \to \mathbb{R}^2$ is given precisely by the following equation.

$$f_1(x,y,z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$$

A diagram of how stereographic projection works on the sphere is shown in figure 3 below.

It is necessary to note that the stereographic projection must exclude the fixed point you project from (in this case $(0,0,1)$) and so you would need two different maps to map the entirety of $\mathbf{S}^2$ to $\mathbb{R}^2$. One simple way to add another stereographic map is to simply choose a different point on the sphere as a reference point. For instance we could cover $\mathbf{S}^2$ entirely by also mapping from the south pole $(0,0,-1)$. The function $f_2 : \mathbf{S}^2/(0,0,-1) \to \mathbb{R}^2$ would be as follows.

$$f_2(x,y,z) = \left(\frac{x}{1+z}, \frac{y}{1+z}\right)$$

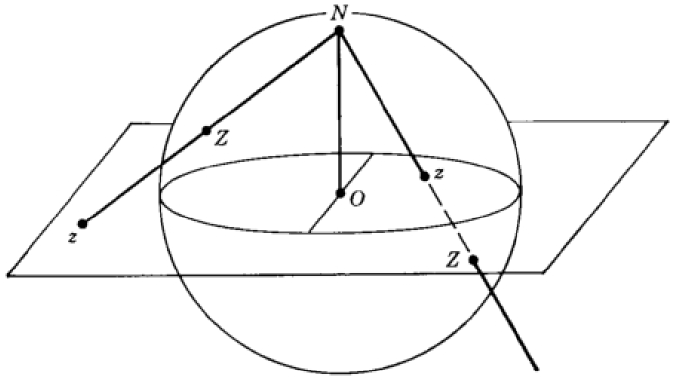

Figure 3: Diagram of stereographic projection on $\mathbf{S}^2$

As it turns out, on the intersection of the two domains of f_1 and f_2, i.e. $\mathbf{S}^2/\{(0,0,1)(0,0,-1)\}$, f_1 and f_2 can be written as smooth functions of each other. For instance

$$f_1(x,y,z) = \left(\frac{x}{1-z}, \frac{y}{1-z}\right) = \left(\frac{x}{1+z}, \frac{y}{1+z}\right) \cdot \frac{1+z}{1-z} = f_2(x,y,z) \cdot \frac{1+z}{1-z}$$

This is essentially the basis behind why $\mathbf{S}^2$ is a smooth manifold. We will use this fact in a later section when we mention fibre bundles. It is also possible to form the inverse map $f_1^{-1} : \mathbb{R}^2 \to \mathbf{S}^2/(0,0,1)$ with the following equation.

$$f_1^{-1}(x,y) = \left(\frac{2x}{x^2+y^2+1}, \frac{2y}{x^2+y^2+1}, \frac{x^2+y^2-1}{x^2+y^2+1}\right)$$

The reason we can't include the north pole in this inverse map, is that if we set $x = y = 0$, we would get $(0,0,-1)$ instead. The same would happen with the south pole if we took the inverse of f_2.

The stereographic projection for $\mathbb{R}^3$ is just an extension of this case. It is harder to visualise how it works since $\mathbf{S}^3$ only exists in $\mathbb{R}^4$, but the equations are pretty much the same. The map $f : \mathbf{S}^3/(1,0,0,0) \to \mathbb{R}^3$ is given by:

$$f(w,x,y,z) = \left(\frac{x}{1-w}, \frac{y}{1-w}, \frac{z}{1-w}\right)$$

As before, this can also be done using the opposite pole as well. Together with both maps, we then end up with the same result that all of $\mathbf{S}^3$ can be mapped to $\mathbb{R}^2$ using this stereographic projection, and from there it can be shown that $\mathbf{S}^3$ is also a smooth manifold. It is also possible to form the inverse map $f^{-1} : \mathbb{R}^3 \to \mathbf{S}^3$ with the following equation.

$$f^{-1}(x, y, z) = \frac{1}{1 + x^2 + y^2 + z^2}(x^2 + y^2 + z^2 - 1, 2x, 2y, 2z) \tag{12}$$

Finally, the general form for stereographic projection of the n-sphere to $\mathbb{R}^{n-1}$, with $\mathbf{x} \in \mathbf{S}^n/(1, 0, 0, ..., 0)$, is the following

$$f(\mathbf{x}) = \frac{\mathbf{x}}{1 - x}$$

As you would expect, from this we can determine that any $\mathbf{S}^n$ is in fact a smooth manifold.

7 The Hopf Map

Another idea we can apply to $\mathbf{S}^3$ is relate it to $\mathbf{S}^2$, which is of course embedded in $\mathbb{R}^3$. Namely, it is possible to form a mapping $\pi : \mathbf{S}^3 \to \mathbf{S}^2$, which we call the Hopf map. It is defined as follows, in [1, p. 87].

$$\pi(a, b, c, d) = (a^2 + b^2 - c^2 - d^2\ ,\ 2(bc + ad)\ ,\ 2(bd - ac)) \tag{13}$$

It is actually possible to derive this equation using quaternions. To show this, we start by taking a unit quaternion $q = (a, b, c, d)$ such that $q \in \mathbf{S}^3$. Then we take another quaternion with first entry equal to zero and the rest denoted by a vector $\mathbf{w}$ such that $W = (0, \mathbf{w}) \in \mathbf{S}^3$ where $\mathbf{w} \in \mathbf{S}^2$. We then examine the following expression which uses quaternion multiplication.

$$\mathbf{w} \mapsto q * W * \overline{q}$$

Where we have defined $\overline{q}$ in (2). I now wish to prove that this map is equivalent to $\mathbf{w} \mapsto (0, Q\mathbf{w})$ where $Q \in SO(3)$ is a rotation matrix, and that in fact it is possible to extract the Hopf map from this rotation matrix. I will show this by doing the calculations step by step below. For the purpose of convenience, let us follow the notation used in (1) and denote the last three entries of q (i.e. (b, c, d)) by $\mathbf{u}$. Then we have:

$$\begin{aligned}
W * \overline{q} &= (0, \mathbf{w}) * (a, -\mathbf{u}) \\
&= (0 + \mathbf{u} \cdot \mathbf{w}, a\mathbf{w} + \mathbf{u} \times \mathbf{w}) \\
q * W * \overline{q} &= (a, \mathbf{u}) * (\mathbf{u} \cdot \mathbf{w}, a\mathbf{w} + \mathbf{u} \times \mathbf{w}) \\
&= (a\mathbf{u} \cdot \mathbf{v} - a\mathbf{u} \cdot \mathbf{v}, a^2\mathbf{w} + a\mathbf{u} \times \mathbf{w} + (\mathbf{u} \cdot \mathbf{w})\mathbf{u} + a\mathbf{u} \times \mathbf{w} + \mathbf{u} \times (\mathbf{u} \times \mathbf{w})) \\
&= (0, a^2\mathbf{w} + 2a\mathbf{u} \times \mathbf{w} + (\mathbf{u} \cdot \mathbf{w})\mathbf{u} + \mathbf{u} \times (\mathbf{u} \times \mathbf{w}))
\end{aligned}$$

We will use the following identity.

$$\mathbf{u} \times (\mathbf{u} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{u} - (\mathbf{u} \cdot \mathbf{u})\mathbf{w}$$

To obtain:

$$q * W * \overline{q} = (0, a^2\mathbf{w} + 2a\mathbf{u} \times \mathbf{w} + 2(\mathbf{u} \cdot \mathbf{w})\mathbf{u} - (\mathbf{u} \cdot \mathbf{u})\mathbf{w})$$

As mentioned above, I want to show that this is equivalent to $(0, Q\mathbf{w})$. We have already shown that the first entry is equal to zero, so now it remains to show that the second part is equivalent to a rotation matrix acting on $\mathbf{w}$. To show this we will utilise the following alternate ways of writing the terms.

$$(\mathbf{u} \cdot \mathbf{w})\mathbf{u} = \mathbf{u}\mathbf{u}^T\mathbf{w}$$

$$\mathbf{u} \times \mathbf{w} = \begin{pmatrix} cw_3 - dw_2 \\ dw_1 - bw_3 \\ bw_2 - cw_1 \end{pmatrix} = \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \hat{u}\mathbf{w}$$

Additionally, recall that we required $q \in \mathbf{S}^3$, so we have:

$$|q|^2 = a^2 + \mathbf{u} \cdot \mathbf{u} = 1$$

Thus we are able to rewrite the above expression as follows.

$$\begin{aligned} q * W * \overline{q} &= (0, (2a^2 - 1)\mathbf{w} + 2a\hat{u}\mathbf{w} + 2\mathbf{u}\mathbf{u}^T\mathbf{w}) \\ &= (0, Q\mathbf{w}) \end{aligned}$$

Where $Q = (2a^2 - 1)I + 2a\hat{u} + 2\mathbf{u}\mathbf{u}^T$. Explicitly we can write out this matrix Q in terms of the components of q.

$$Q = \begin{pmatrix} a^2 + b^2 - c^2 - d^2 & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & a^2 - b^2 + c^2 - d^2 & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & a^2 - b^2 - c^2 + d^2 \end{pmatrix}$$

It is then verifiable that $Q^T Q = I$ and $detQ = +1$, and so $Q \in SO(3)$ as we set out to prove. Furthermore, we can see that the first column of this matrix is equivalent to the formula for the Hopf map that was stated in (13).

As Q is an orthogonal matrix, it has the property that all three columns and all three rows have norm equal to one, which means that they belong to $\mathbf{S}^2$, and so this matrix essentially gives us six different maps from $\mathbf{S}^3 \to \mathbf{S}^2$.

In fact there is a correspondence between the columns and the rows to each other that can be seen through the use of a rotation on $q = (a, b, c, d)$. For instance:

$$\hat{q} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

This map would send the second column to a permutation of the first. This can be quite easily seen from the inherent symmetry of the matrix.

7.1 Properties of the Hopf Map

Now that we have a relation between $\mathbf{S}^3$ and $\mathbf{S}^2$, we want to examine the properties of this linking. Since the Hopf map is surjective, i.e. its image is the whole of $\mathbf{S}^2$, we can form the inverse map $\pi^{-1} : \mathbf{S}^2 \to \mathbf{S}^3$. From (13), taking an arbitrary point $\mathbf{w} = (w_1, w_2, w_3) \in \mathbf{S}^2$, we have:

$$\begin{aligned} a^2 + b^2 - c^2 - d^2 &= w_1 && (14) \\ 2(bc + ad) &= w_2 && (15) \\ 2(bd - ac) &= w_3 && (16) \\ a^2 + b^2 + c^2 + d^2 &= 1 && (17) \end{aligned}$$

Where the last equation is a consequence of $q \in \mathbf{S}^3$. We can solve these equations for a, b, c and d to obtain an expression for the inverse map.

$$(17) - (14) : c^2 + d^2 = \frac{1 - w_1}{2}$$

We can put (15) and (16) into matrix form to obtain:

$$\begin{pmatrix} d & c \\ -c & d \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{2} \begin{pmatrix} w_2 \\ w_3 \end{pmatrix}$$

The determinant of this matrix is $c^2 + d^2$ which we calculated in the line above. If $w_1 = 1$ then the determinant is zero and we have the special case of the north pole of $\mathbf{S}^2$, in which $c = d = 0$ and $a^2 + b^2 = 1$. In that case the inverse Hopf map defines a circle in $\mathbf{S}^3$. Otherwise we can invert the matrix as follows.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{2} \cdot \left(\frac{2}{1 - w_1} \right) \begin{pmatrix} d & -c \\ c & d \end{pmatrix} \begin{pmatrix} w_2 \\ w_3 \end{pmatrix}$$

$$\Rightarrow (1 - w_1) \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} d & -c \\ c & d \end{pmatrix} \begin{pmatrix} w_2 \\ w_3 \end{pmatrix}$$

This gives us two linear equations for a, b, c and d in terms of w_1, w_2 and w_3 which we rewrite below.

$$(1 - w_1)a + w_3 c - w_2 d = 0$$
$$(1 - w_1)b - w_2 c - w_3 d = 0$$

Note that these equations are of the form $\mathbf{x} \cdot q = 0$ and $\mathbf{y} \cdot q = 0$ where $\mathbf{x}$ and $\mathbf{y}$ are the following.

$$\mathbf{x} = \begin{pmatrix} 1 - w_1 \\ 0 \\ w_3 \\ -w_2 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 0 \\ 1 - w_1 \\ -w_2 \\ -w_3 \end{pmatrix}$$

These vectors are clearly linearly independent as their first and second entries respectively are non-zero by our assumption. This means that together, $\mathbf{x} \cdot q = 0$ and $\mathbf{y} \cdot q = 0$ form the equation for a 2-dimensional subspace of $\mathbb{R}^4$ i.e. a plane.

To see how this plane sits in $\mathbf{S}^3$, consider an alternate yet equivalent form for a plane in $\mathbb{R}^4$, $q = \alpha\mathbf{u} + \beta\mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are linearly independent vectors lying inside the plane and α, $\beta \in \mathbb{R}$. In other words, since $\mathbf{x}$ and $\mathbf{y}$ are normal vectors to the plane, we are choosing orthogonal vectors $\mathbf{u}$ and $\mathbf{v}$ to these. i.e. $\mathbf{x} \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} = \mathbf{x} \cdot \mathbf{v} = \mathbf{y} \cdot \mathbf{v} = 0$.

Additionally we can choose these vectors to be orthonormal to each other, i.e. $\mathbf{u} \cdot \mathbf{u} = \mathbf{v} \cdot \mathbf{v} = 1$ and $\mathbf{u} \cdot \mathbf{v} = 0$.

Now once again we enforce the condition that $q \in \mathbf{S}^3$ to see how this plane intersects $\mathbf{S}^3$. We have:

$$q \cdot q = \alpha^2\mathbf{u} \cdot \mathbf{u} + \alpha\beta\mathbf{u} \cdot \mathbf{v} + \beta^2\mathbf{v} \cdot \mathbf{v} = \alpha^2 + \beta^2 = 1$$

Just as we saw in the first case above, this is the equation for a circle. Recall that this inverse Hopf map acts on an arbitrary point $\in \mathbf{S}^2$. Thus we discover the interesting property that the inverse Hopf map sends any single point on $\mathbf{S}^2$ to a great circle on $\mathbf{S}^3$, i.e. a plane through the origin intersecting the 3-sphere.

7.2 The Hopf Fibration

The significance of the fact that the Hopf map sends an arbitrary point in $\mathbf{S}^2$ to a circle in $\mathbf{S}^3$, is that it allows us to create a correspondence between

$\mathbf{S}^1$, $\mathbf{S}^2$ and $\mathbf{S}^3$. We call this structure the *Hopf fibration*, and it is an example of a fibre bundle, which we will define momentarily. To explain this concept loosely, it allows us to describe the 3-sphere in terms of the sphere and circles. That is to say locally, we have the property that $\mathbf{S}^3 = \mathbf{S}^2 \times \mathbf{S}^1$, because each point in $\mathbf{S}^3$ is the same as a point on $\mathbf{S}^2$ and the whole of $\mathbf{S}^1$. It should be noted that globally, we do not have the same structure[1], and in fact this will always be true for any non-trivial fibre bundle.

To solidify this idea, we define that a *fibre bundle* is a structure (P, M, G, π), where P, M and G are topological spaces, and $\pi : P \to M$ is a surjective map (called the *projection*), which locally satisfies the above quality. We call P the *total space*, M the *base space*, and G the *fibre space*. A way to succinctly write a fibre bundle is the following

$$G \hookrightarrow P \xrightarrow{\pi} M$$

We can see that the Hopf fibration does indeed satisfy these qualities, where the fibre space $\mathbf{S}^1$ is embedded in the total space $\mathbf{S}^3$, and $\pi : \mathbf{S}^3 \to \mathbf{S}^2$ is a surjective map from $\mathbf{S}^3$ on to the base space $\mathbf{S}^2$. It can be shown that all of the components are topological spaces (in fact as shown in section 6, any $\mathbf{S}^n$ is a smooth manifold).

In terms of the fibre bundle notation, we have the following structure.

$$\mathbf{S}^1 \hookrightarrow \mathbf{S}^3 \xrightarrow{\pi} \mathbf{S}^2$$

In later chapters we will come back to this idea in the context of principal bundles, and show that the Hopf fibration has even further structure.

[1]It can be shown that $\mathbf{S}^3$ is not homeomorphic to $\mathbf{S}^2 \times \mathbf{S}^1$, by pointing out for instance that

$$\pi_1(\mathbf{S}^3) = 0, \qquad \pi_1(\mathbf{S}^2 \times \mathbf{S}^1) = \mathbb{Z}$$

where $\pi_1(\cdot)$ is the first fundamental group. Thus they do not have the same structure and are not homeomorphic.

7.3 Generalisations of the Hopf Fibration

After noticing the fibre bundle structure between $\mathbf{S}^3$, $\mathbf{S}^2$ and $\mathbf{S}^1$ that we get as a result of the Hopf mapping, a natural question would be whether or not we can generalise this result to other n-spheres. As it turns out there are indeed other forms of the Hopf fibration, which I will list below (also see [2, p. 386-390] for further details).

$$\mathbf{S}^0 \hookrightarrow \mathbf{S}^1 \xrightarrow{\pi} \mathbf{S}^1 \tag{18}$$

$$\mathbf{S}^3 \hookrightarrow \mathbf{S}^7 \xrightarrow{\pi} \mathbf{S}^4 \tag{19}$$

$$\mathbf{S}^7 \hookrightarrow \mathbf{S}^{15} \xrightarrow{\pi} \mathbf{S}^8 \tag{20}$$

Here, (18) is actually an example of the Möbius strip, i.e. a surface with only one side. To see this, first of all we identify that in this case $\mathbf{S}^1$ is the full space, and is mapped by π to another copy of $\mathbf{S}^1$ which is the base space. If we let $(x_1, x_2) \in \mathbf{S}^1$ (i.e. $(x_1)^2 + (x_2)^2 = 1$) then the mapping π is given by.

$$\pi(x_1, x_2) = ((x_1)^2 - (x_2)^2, 2x_1x_2) = (v_1, v_2)$$

This mapped vector (v_1, v_2) does indeed belong to $\mathbf{S}^1$ since $(v_1)^2 + (v_2)^2 = ((x_1)^2 + (x_2)^2)^2 = 1$. The important thing to note, is that just as we noticed in the $\mathbf{S}^3$ case, the inverse map sends a single point in $\mathbf{S}^1$ to multiple (in this case two) points in the total space $\mathbf{S}^1$. Since $\mathbf{S}^0$ is simply composed of two points in $\mathbb{R}$, this means we once again retrieve the property of the total space being a product of the base space and the fibre space, i.e. locally: $\mathbf{S}^1 = \mathbf{S}^1 \times \mathbf{S}^0$.

Visually we can think of the total space being a Möbius strip, where the mapping π sends two points on either side of the strip to each individual point on the circle.

As for the bundles (19) and (20), we can think of these as being quaternionic and octonionic extensions of the original example (we can express the $\mathbf{S}^3$ example in terms of complex numbers by using the construction $\mathbf{S}^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}$, see section 10).

Finally, there exists a further generalisation of the above of the form.

$$\mathbf{S}^1 \hookrightarrow \mathbf{S}^{2n+1} \xrightarrow{\pi} \mathbb{CP}^n$$

Where $\mathbb{CP}^n$ is the *complex projective space*. This generalisation is in fact identical to our original result when $n = 1$, since $\mathbb{CP}^1 \cong \mathbf{S}^2$.

8 Stereographic Projection of the Hopf Fibration

We now know from section 7 that via the Hopf map, a point on $\mathbf{S}^2$ corresponds to a circle on $\mathbf{S}^3$, but we do not know what this looks like since $\mathbf{S}^3$ is embedded in $\mathbb{R}^4$. However we now have a way of mapping $\mathbf{S}^3$ to $\mathbb{R}^3$ using the stereographic projection, which will allow us to form a picture of this. Recall the inverse stereographic map $f^{-1} : \mathbb{R}^3 \to \mathbf{S}^3$ we mentioned in equation (12). This allows us to write an arbitrary point $\mathbf{X} = (X_1, X_2, X_3, X_4) \in \mathbf{S}^3$ in terms of $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$ as follows.

$$X_1 = \frac{2x}{x^2 + y^2 + z^2 + 1}, \quad X_2 = \frac{2y}{x^2 + y^2 + z^2 + 1} \tag{21}$$

$$X_3 = \frac{2z}{x^2 + y^2 + z^2 + 1}, \quad X_4 = \frac{x^2 + y^2 + z^2 - 1}{x^2 + y^2 + z^2 + 1} \tag{22}$$

We want to vary $X \in \mathbf{S}^3$ along a circle and see what the picture looks like in $\mathbb{R}^3$. To do this we use the complex number form of $\mathbf{S}^3 = \{\mathbf{z} = (z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}$, which is an equivalent way of writing $\mathbf{S}^3$ mentioned at the end of chapter 3. z_1 and z_2 are complex numbers, so using polar coordinates, they can be written as $z_1 = r_1 e^{i\phi_1}$ and $z_2 = r_2 e^{i\phi_2}$, where $r_1, r_2 > 0$, and $0 \leq \phi_1, \phi_2 < 2\pi$. Then we notice that by varying ϕ_1 or ϕ_2 through 0 to 2π, we describe a circle in $\mathbf{S}^3$, which is what we want to happen. Due to the condition that $|z_1|^2 + |z_2|^2 = 1$, we have that $r_1^2 + r_2^2 = 1$, which leads to the following further parameterisation of $\mathbf{z}$.

$$z_1 = \cos(\theta/2) e^{i\phi_1}$$

$$z_2 = \sin(\theta/2)e^{i\phi_2} \quad 0 \leq \theta \leq \pi$$

The choice of range of θ means that r_1 and r_2 are always between 0 and 1. Now we want to find an expression for $(x, y, z) \in \mathbb{R}^3$ in terms of z_1 and z_2. We see that $|z_1|^2 = X_1^2 + X_2^2 = \cos^2(\theta/2)$, so from the equations in (21), we get the following

$$\frac{4(x^2 + y^2)}{(x^2 + y^2 + z^2 + 1)^2} = |z_1|^2 = \cos^2(\theta/2)$$

It should be noted that by taking the modulus of z_1, we are now considering all points lying on a circle in $\mathbb{C}$. We can rewrite this equation in the following form, provided we assume $\theta \neq 0, \pi$.

$$4R^2(x^2 + y^2) = (x^2 + y^2 + z^2 + R^2 - r^2)^2$$

Where $R = 1/\cos(\theta/2)$, and $r = \tan(\theta/2)$ for a given θ (can be seen by using the equation $1 + \tan^2(\theta) = 1/\cos^2(\theta)$). This is the quartic form of the equation for a torus about the z-axis, specifically where R is the distance from the centre of the tube to the centre of the torus (the z-axis), and r is the radius of the tube. Thus we obtain the result that a circle in $\mathbf{S}^3$ corresponds to a torus in $\mathbb{R}^3$.

Since θ can vary between 0 and π, this means we get a different torus for each value of θ, since R and r are dependent on θ. We can see that close to $\theta = 0$, the tori have radii approaching 0 (since $r = \tan(\theta/2)$), and distance from the origin approaching 1 from above (since $R = 1/\cos(\theta/2)$). In fact at $\theta = 0$ we just get the equation for a circle around the z-axis with radius 1. As θ increases, both R and r increase, and will tend to infinity as θ approaches π. These tori are nested within each other, i.e. the larger ones contain the smaller ones inside them, and concentric, since they are all centred around the z-axis.

A visualisation of how this allows us to map the Hopf fibration to $\mathbb{R}^3$ is shown in figure 4 below.

We have this picture since the fibres in $\mathbf{S}^3$ are copies of $\mathbf{S}^1$, i.e. circles,

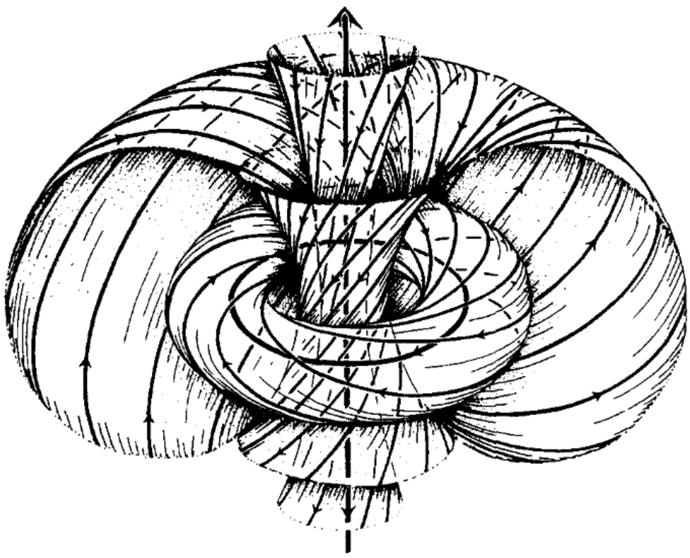

Figure 4: Visualisation of stereographic projection of the Hopf fibration

and we have just shown that in $\mathbb{R}^3$ these circles are tori. We can see that the larger tori converge to the z-axis, since the distance from the inner ring to the z-axis is given by $R - r = (1 - \sin(\theta/2))/\cos(\theta/2)$, which tends to zero as θ approaches π (can be seen by using l'Hôpital's rule).

9 Ehresmann Connection

Recall that we obtained a correspondence between single points $\in \mathbf{S}^2$ and $\mathbf{S}^3$ using the Hopf mapping in section 7. I now want to explore how we can *lift* curves on $\mathbf{S}^2$ to $\mathbf{S}^3$. We can do this using the Ehresmann connection.

Generally when we talk about this connection, it can apply to any fibre bundle (P, M, G, π) provided $\pi : M \to N$ is a submersion, but we want to specifically look at the Hopf fibration for $\mathbf{S}^3$, with $\pi : \mathbf{S}^3 \to \mathbf{S}^2$,[2] which we looked at in section (13).

Since we want to look at curves being mapped by the Hopf map, we let $\xi(t)$ be a curve $\in \mathbf{S}^3$, and $\gamma(t) = \pi(\xi(t))$ be the corresponding curve $\in \mathbf{S}^2$.

[2]Proof that the Hopf map is a submersion is detailed in the appendix A.1.

Then note that the Jacobian of π is a map between the tangent spaces of $\mathbf{S}^2$ and $\mathbf{S}^3$, that it to say $d_p\pi : T_p\mathbf{S}^3 \to T_{\pi(p)}\mathbf{S}^2$. Thus using our notation, we have that $\dot{\gamma}(t) = d_p\pi(\dot{\xi}(t))$ where $\dot{\xi} \in T_p\mathbf{S}^3$ and $\dot{\gamma} \in T_{\pi(p)}\mathbf{S}^2$.

The Ehresmann connection is a way of decomposing the tangent space $T_p\mathbf{S}^3$ into two complementary subspaces, the *vertical* and *horizontal* spaces, using π. Firstly, we define $V_pP = \mathrm{Ker}(d_p\pi)$. Clearly V_pP is a subspace of $T_p\mathbf{S}^3$ since the kernel of any map is a subspace of its domain. Then we define H_pP to be the complement to V_pP in $T_p\mathbf{S}^3$. This automatically ensures the following.

$$T_p\mathbf{S}^3 = V_pP \oplus H_pP$$

Thus we have split the tangent space into two spaces, with one, V_pP, being tangent to the fibre $\mathbf{S}^1$, and the other, H_pP, being transverse to the fibre. Any tangent vector in $T_p\mathbf{S}^3$ is thus the sum of a vertical component and a horizontal component. The purpose of defining this connection, as we will later see in section 10.2, is that the tangent vectors of the *horizontal lift* of a curve always remain in the horizontal space. Therefore we will be able to specify how a curve being lifted from $\mathbf{S}^2$ to $\mathbf{S}^3$ can move.

An equivalent condition to the one above (i.e. where we require that H_pP is the complement to V_pP), is to require that $d_p\pi(H_pP) = T_{\pi(p)}\mathbf{S}^2$. To intuitively understand why this is the case, consider that ordinarily, $d_p\pi(T_p\mathbf{S}^3) = T_{\pi(p)}\mathbf{S}^2$, since $d_p\pi$ is surjective as we required earlier. Since we define $V_pP = \mathrm{Ker}(d_p\pi)$, we know that $d_p\pi(V_pP) = 0$, thus if H_pP is the remaining part of $T_p\mathbf{S}^3$ by virtue of being the complement to V_pP, it must map to the entirety of $T_{\pi(p)}\mathbf{S}^2$.

However we can also show this explicitly if we derive expressions for V_pP and H_pP, and apply $d_p\pi$ to the result. Firstly we find $d_p\pi$ (the Jacobian of

π) as follows (using the notation from equations (14), (15) and (16))

$$\begin{pmatrix} dw_1 \\ dw_2 \\ dw_3 \end{pmatrix} = 2 \begin{pmatrix} a & b & -c & -d \\ d & c & b & a \\ -c & d & -a & b \end{pmatrix} \begin{pmatrix} da \\ db \\ dc \\ dd \end{pmatrix} \tag{23}$$

Where $d_p\pi$ is the 3 x 4 matrix. Firstly we want to compute V_pP, i.e. the kernel (or null space) of the matrix. Through linear algebra, or just by inspection and noticing that the kernel must be 1 dimensional, we find:

$$V_pP = \mathrm{Ker}(d_p\pi) = \mathrm{span}\left\{ \begin{pmatrix} b \\ -a \\ -d \\ c \end{pmatrix} \right\}$$

Now in order to find H_pP, we have to take the complement of V_pP in $T_p\mathbf{S}^3$ as we defined earlier. To do this, I will utilise the result (proved in appendix A.2) that $T_p\mathbf{S}^3 = \mathrm{span}\{J_1\mathbf{n}, J_2\mathbf{n}, J_3\mathbf{n}\}$ where we defined the J_i in equation (3), and $\mathbf{n}$ is a normal vector to $\mathbf{S}^3$. We can easily see that $V_pP = \mathrm{span}\{J_1\mathbf{n}\}$, and so $H_pP = \mathrm{span}\{J_2\mathbf{n}, J_3\mathbf{n}\}$ by consequence, which turns out to be the following.

$$H_pP = \mathrm{span}\left\{ \begin{pmatrix} c \\ d \\ -a \\ -b \end{pmatrix}, \begin{pmatrix} d \\ -c \\ b \\ -a \end{pmatrix} \right\}$$

Recall we want to show that $d_p\pi(H_pP) = T_{\pi(p)}\mathbf{S}^2$. Directly putting in what we just calculated for $d_p\pi$ and H_pP, we get:

$$\begin{pmatrix} a & b & -c & -d \\ d & c & b & a \\ -c & d & -a & b \end{pmatrix} \begin{pmatrix} c & d \\ d & -c \\ -a & b \\ -b & -a \end{pmatrix} = \begin{pmatrix} 2(ac+bd) & 2(ad-bc) \\ 2(cd-ab) & -a^2+b^2-c^2+d^2 \\ a^2-b^2-c^2+d^2 & 2(-cd-ab) \end{pmatrix}$$

So we want to show that these columns (let them be ξ_1 and ξ_2) span $T_{\pi(p)}\mathbf{S}^2$. To do this we have to prove that:

1. The two columns are linearly independent, i.e. $\xi_1 \times \xi_2 \neq 0, \forall\, q \in \mathbf{S}^3$.
2. $\mathbf{w} \cdot \xi_1 = \mathbf{w} \cdot \xi_2 = 0$ where the entries of $\mathbf{w}$ were defined in (14), (15) and (16).

To show the first, we simply compute this cross product and find that it is:

$$\begin{pmatrix} (a^2+b^2)^2 - (c^2+d^2)^2 \\ 2(ad+bc) \\ 2(bd-ac) \end{pmatrix}$$

For this vector to be zero, all three entries must be zero, which gives us three simultaneous equations:

$$\begin{aligned} a^2+b^2 &= c^2+d^2 \\ ad &= -bc \\ ac &= bd \end{aligned}$$

It is fairly easy to see that these equations have no solution for $q \in \mathbf{S}^3$, and so ξ_1 and ξ_2 must be linearly independent.

The second requirement is a straightforward calculation, albeit lengthy. We have

$$\begin{aligned} \mathbf{w} \cdot \xi_1 &= \begin{pmatrix} a^2+b^2-c^2-d^2 \\ 2(bc+ad) \\ 2(bd-ac) \end{pmatrix} \cdot \begin{pmatrix} 2(ac+bd) \\ 2(cd-ab) \\ a^2-b^2-c^2+d^2 \end{pmatrix} \\ &= 2(a^3c+ab^2c-ac^3-acd^3+a^2bd+b^3d-bc^2d-bd^3)+4(acd^2 \\ &\quad -ab^2c-a^2bd+a^2bd)+2(a^2bd-b^3d-bc^2d-bc^d+bd^3-a^3c \\ &\quad +ab^2c+ac^3-acd^2) \\ &= 0 \end{aligned}$$

Similarly for ξ_2, we have $\mathbf{w} \cdot \xi_2 = 0$. Thus we have indeed shown that $d_p\pi(H_pP) = T_{\pi(p)}\mathbf{S}^2$.

10 Principal Bundles

As we mentioned at the end of section 7, the Hopf fibration (or Hopf bundle) is an example of a fibre bundle. We now want to extend this definition and show that it has the structure of a principal bundle as well, and that we can study it with regards to this.

For ease of notation a bit later on, we can rewrite $\mathbf{S}^3$ and the Hopf map in terms of complex numbers i.e. let $\mathbf{S}^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}$. Then the Hopf map can be written in the following way.

$$\pi(z_1, z_2) = (|z_1|^2 - |z_2|^2 \ , \ -(z_1z_2 + \overline{z_1z_2}) \ , \ (\overline{z_1z_2} - z_1z_2)i)$$

To show that this is equivalent to equation (13), let $z_1 = a + bi$ and $z_2 = c + di$. Then we can see that:

$$\begin{aligned}
|z_1|^2 - |z_2|^2 &= |a + bi|^2 - |c + di|^2 \\
&= a^2 + b^2 - c^2 - d^2 \\
(\overline{z_1z_2} - z_1z_2)i &= ((a - bi)(c - di) - (a + bi)(c + di))i \\
&= (ac - bci - adi - bd - ac - bci - adi + bd)i \\
&= -2(bc + ad)i^2 \\
&= 2(bc + ad) \\
-(z_1z_2 + \overline{z_1z_2}) &= -(a + bi)(c + di) - (a - bi)(c - di) \\
&= -ac + bci + adi + bd - ac - bci - adi + bd \\
&= 2(bd - ac)
\end{aligned}$$

So the two equations are indeed equivalent.

Now, a principal bundle is a form of fibre bundle, with additional structure attached. Before, we merely required the fibre G to be a topological space, but we now require it to be a group, and that this group has a group action on the total space P. Specifically we want to define this group action to be a diffeomorphism.

We define that a *principal bundle* is a structure (P, M, G, π), just as for a fibre bundle, with the following properties.

1. P is locally equal to the product space $M \times G$ (where M is the base space and G is a group).

2. It is equipped with a projection (i.e. a mapping) $\pi : P \to M$.

3. For each $g \in G$, there exists a diffeomorphism $R_g : P \to P$ such that $R_g(p) = pg, \forall\, p \in P$.

Under this new notation for a principal bundle, we note that the Hopf fibration does satisfy such qualities, if we let the total space P be $\mathbf{S}^3$, the base space M be $\mathbf{S}^2$, the group G be $\mathbf{S}^1$ and the projection π be the Hopf mapping, and when we define a diffeomorphism as follows. Let $p \in \mathbf{S}^3$, and $g = e^{i\theta} \in \mathbf{S}^1$, where $\theta \in \mathbb{R}$. Then the diffeomorphism is defined to be

$$R_g(p) = (z_1, z_2)e^{i\theta}$$

As required, this diffeomorphism maps any $p \in \mathbf{S}^3$ to another point on $\mathbf{S}^3$ since it preserves the norm of p. We can easily see this since

$$\|R_g(z_1, z_2)\|^2 = \|(z_1, z_2)e^{i\theta}\|^2 = (|z_1|^2 + |z_2|^2)^2 |e^{i\theta}|^2 = 1$$

Note that G acts freely on P in this way, which is to say that for any $\theta \neq 2n\pi$ (where $n \in \mathbb{N}$), p will be moved from its original position (specifically along a circle). Thus we can say that we can think of $\mathbf{S}^1$ acting on $\mathbf{S}^3$ by sending points on $\mathbf{S}^3$ along a particular fibre.

Below in figure 5 from [6, p. 8] is a visualisation of how a principal bundle works. Note the slight difference in notation in that Φ_g is used instead of R_g.

It is important to note that the projection π is invariant under the given diffeomorphism, i.e. it is the same whether or not we apply the diffeomor-

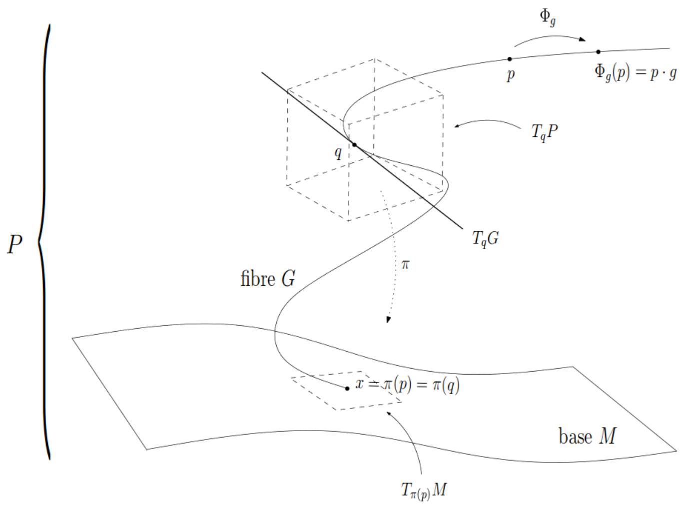

Figure 5: Visualisation of a principal bundle

phism to it. That is to say.

$$\pi(R_g(p)) = \pi(pe^{i\theta}) = \pi(p) \tag{24}$$

Where we have used the fact that each point on a given circle in $\mathbf{S}^3$ gets sent via the Hopf map to the same point in $\mathbf{S}^2$.[3]

10.1 Principal Connections

Another topic we can revisit with regards to principal bundles is that of connections. Namely, we can talk about a special case of the Ehresmann connection called a *principal connection* for a principal bundle P. Just as we did for the Ehresmann connection, we start by decomposing the tangent space of P, T_pP into two complementary subspaces. In an identical way to the previous case, we define $V_pP = \mathrm{Ker}(d_p\pi)$, and specify that:

[3] Another way of putting this, is that we can think of the inverse Hopf map's property of mapping a point to a circle in the following way.

$$\pi^{-1}(m) = (m, g) \quad \forall\, m \in M$$

Where $(m, g) \in \mathbf{S}^3$ and g can be any point on $\mathbf{S}^1$.

$$T_pP = V_pP \oplus H_pP$$

We then additionally require a condition on H_pP which incorporates the group G of a principal bundle into this definition. Namely, we have the following.

$$dR_g(H_pP) = H_{pg}$$

Intuitively we would expect this to be correct. Recall that we defined R_g to act on P in such a way that $R_g(p) = pg$. It naturally follows that if we take the Jacobian of R_g, i.e. dR_g, it would have the same effect on a subspace of the tangent space T_pP. It is also worth noting that this is a similar property to the one we derived for the Ehresmann connection, where $d_p\pi(H_pP) = T_{\pi(p)}\mathbf{S}^2$.

Recall in section 9 that due to the way we defined V_pP, V_pP is tangent to the fibre $\mathbf{S}^1$. We can show this is consistent with $\mathbf{S}^1$ as a part of the principal bundle, i.e. that V_pP is tangent to the diffeomorphism R_g defined by $\mathbf{S}^1$ (at the identity). We do this by using the complex number form of $\mathbf{S}^3 = \{\mathbf{z} \in \mathbb{C}^2 : \|\mathbf{z}\| = 1\}$ which we used at the start of this section. Recall that we found V_pP to be

$$V_pP = \mathrm{Ker}(d_p\pi) = \mathrm{span}\left\{\begin{pmatrix} b \\ -a \\ -d \\ c \end{pmatrix}\right\}$$

To show how a vector in $\mathbf{S}^3$ is equivalent to one in $\mathbb{C}^2$, we use the quaternions $\mathbb{H}$ since they are isomorphic to $\mathbf{S}^3$. Thus the above vector in the span is equal to the following with respect to the quaternionic basis $(1, i, j, k)$.

$$\begin{aligned}
\begin{pmatrix} b \\ -a \\ -d \\ c \end{pmatrix} &= b - ai - dj + ck \\
&= 1(b - ai) + k(c - di)
\end{aligned}$$

Here we can see that if we let k correspond to the usual $i \in \mathbb{C}$, this is the form an element of $\mathbb{C}^2$ takes, since it is a complex linear combination of the basis $(1, k)$. Thus using this notation, V_pP can be rewritten as follows.

$$V_pP = \text{span}\left\{ \begin{pmatrix} b \\ c \end{pmatrix}, i \begin{pmatrix} -a \\ -d \end{pmatrix} \right\}$$

Note that the span is defined to be a real linear combination of vectors, so i is separate from this. Thus we are free to multiply both vectors by -1 say, to get.

$$\begin{aligned}
V_pP &= \text{span}\left\{ \begin{pmatrix} -b \\ -c \end{pmatrix}, i \begin{pmatrix} a \\ d \end{pmatrix} \right\} \\
&= \text{span}\left\{ i \left(i \begin{pmatrix} b \\ c \end{pmatrix}, \begin{pmatrix} a \\ d \end{pmatrix} \right) \right\} \\
&= \text{span}\{ i(z_1, z_2) \}
\end{aligned}$$

Where $z_1, z_2 \in \mathbb{C}$. Recall that we defined the diffeomorphism for the principal bundle to be $R_g(p) = (z_1, z_2)e^{i\theta}$. The tangent to this is

$$\frac{d}{d\theta}((z_1, z_2)e^{i\theta}) = i(z_1, z_2)e^{i\theta}$$

Thus, evaluated at the identity, i.e. when $\theta = 0$, we have that the tangent to $R_g(p)$ is $i(z_1, z_2)$. Since V_pP is the span of this, we do indeed get the result that V_pP is tangent to R_g at the identity.

10.2 Holonomy on a Principal Bundle

Previously when we talked about holonomy on the sphere, we used classical differential geometry to do so, and we defined the covariant derivative so that we could find a condition for a vector to be parallel transported along the sphere. Now that we have defined the notion of a principal bundle, we can instead use our principal connection that we have just defined, to specify that instead of parallel transporting the vectors directly, that they be transported along the fibres of the bundle. We can do this since, as we mentioned in section 9, V_pP is tangent to the fibres, whilst H_pP is orthogonal to them. Thus as we will see later in this section, we can specify that the lift, which is a path in the tangent bundle P, be horizontal. An idea of what a horizontal lift looks like is shown in figure 6 from [6, p. 16] below. Note how in the diagram, $\dot{q}(t)$ is always in the horizontal plane denoted $H_{q(t)}P$.

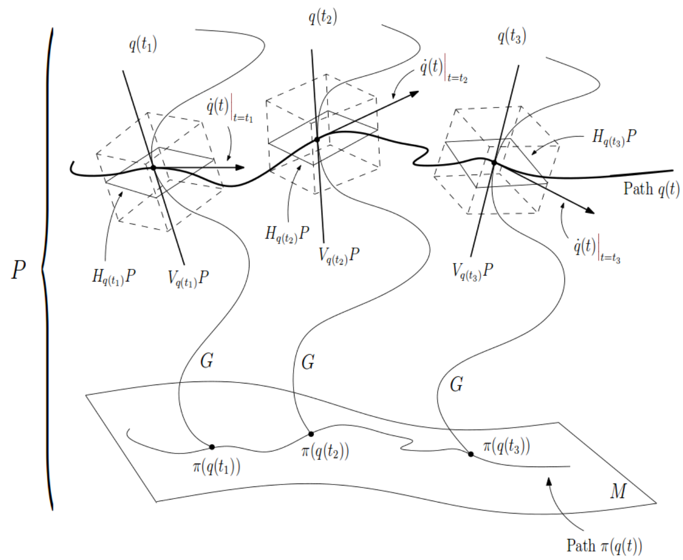

Figure 6: Diagram of a horizontal lift $q(t)$ through the principal bundle

In order to specify that a path is horizontal, we first need to define the idea of a connection 1-form. Now we know V_pP is tangent to the fibres, which are copies of $\mathbf{S}^1$. Note also that the Lie algebra $\mathfrak{g}$ of a Lie group G is equivalent to the tangent space to G by definition. In this case, the Lie group is $G = \mathbf{S}^1$ (can be shown that this is a *Lie* group), and it can be shown that $\mathfrak{g} = i\mathbb{R}$. Thus there is a one-to-one correspondence between V_pP and the Lie algebra $\mathfrak{g}$ of $\mathbf{S}^1$, i.e. $V_pP \cong i\mathbb{R}$. Since V_pP is a subset of the entire tangent space T_pP, this means that there is a linear map $T_pP \to i\mathbb{R}$, which we will call the connection 1-form.

Let $p \in P$ and $\mathbf{v} \in T_pP$. Then the canonical connection 1-form $\omega_p : T_pP \to i\mathbb{R}$ is a G-equivariant differential 1-form defined as follows.

$$\omega_p(\mathbf{v}) = p^\dagger \mathbf{v}$$

Where † is the adjoint operator (the conjugate transpose). It can be shown, e.g. in [6, p. 27], that the image of this map is indeed $i\mathbb{R}$. It is easy to show at least that the outcome is a scalar, because we are multiplying a 1 x n vector by an n x 1 vector. G-equivariant means that ω_p is equivariant with regards to $G = \mathbf{S}^1$ i.e. $\forall\ g \in G$ and $\forall\ \mathbf{v} \in T_pP$, $\omega_p(\mathbf{v} \cdot g) = \omega_p(\mathbf{v}) \cdot g$. The proof of this is essentially trivial since g is a scalar, i.e.

$$\omega_p(\mathbf{v} \cdot g) = p^\dagger(\mathbf{v} \cdot g) = p^\dagger(\mathbf{v}) \cdot g = \omega_p(\mathbf{v}) \cdot g$$

Now since we know that $V_pP \cong i\mathbb{R}$, we have that $\omega_p(V_pP) = i\mathbb{R}$. Recall that $T_pP = V_pP \oplus H_pP$, thus since $\omega_p : T_pP \to i\mathbb{R}$, we automatically have the result that $\omega_p(H_pP) = 0$. That is equivalently the following.

$$H_pP = \mathrm{Ker}(\omega_p)$$

Thus we now have a way of specifying that a lift is horizontal, by referring to this requirement.

Now, due to the fact we have used the generalised concept of a principal bundle (aside from assuming that $G = \mathbf{S}^1$) up to this point. We are actually able to use for the rest of this section the generalised Hopf bundle which we mentioned at the end of section 7.3:

$$\mathbf{S}^1 \hookrightarrow \mathbf{S}^{2n+1} \xrightarrow{\pi} \mathbb{CP}^n$$

Suppose we want to talk about holonomy on $\mathbb{CP}^n$, a generalisation of the sphere. If we let $\gamma(t)$ be a curve in $\mathbb{CP}^n$, then there should be a curve in the principal bundle $\mathbf{S}^{2n+1}$ that is mapped to this curve by π. In fact there exists a whole equivalence class of curves in $\mathbf{S}^{2n+1}$, such that for each $\xi(t)$ in the class, we have:

$$\pi(\xi(t)) = \gamma(t)$$

We call any such $\xi(t)$ a *lift* of the curve $\gamma(t)$. The fact that we have an equivalence class of curves being mapped to the same curve γ is analogous to the property we observed in the Hopf fibration of multiple points in $\mathbf{S}^3$ corresponding to the same point in $\mathbf{S}^2$.

We can use property (24) in this case, if we generalise it to $\mathbf{S}^{2n+1}$.[4] That is to say we can use the fact π is invariant under the diffeomorphism R_g, and multiply ξ on the right by $e^{i\theta}$. Thus any two lifts ξ of γ are part of the same equivalence class if they differ by a factor of $e^{i\theta}$. In this sense θ is called the *phase*, and it is a measure of the separation of two different lifts. This allows us to rewrite the above equation as follows.

$$\pi(\xi(t)e^{i\theta(t)}) = \gamma(t)$$

Let $q(t) = \xi(t)e^{i\theta(t)}$. Then by varying $\theta(t)$, $q(t)$ can be equal to any particular ξ in the equivalence class. We can think of $q(t)$ as being a function of θ, with the outcome being the lifts of γ.

As mentioned earlier, we want to achieve what we did in the section about holonomy but now with regards to $\mathbb{CP}^n$, i.e. we want to *parallel* transport the vectors around $\mathbb{CP}^n$ and see whether they are changed or not, and if so, how they are changed. However we can do this in a different way to before by using the connection form, by specifying that the lift of the curve from $\mathbb{CP}^n$ to $\mathbf{S}^{2n+1}$ is *horizontal*. We define this to mean that $q(t) \in \mathbf{S}^{2n+1}$ is horizontal if:

[4] i.e. let $\xi(t) \in \mathbb{C}^{n+1} \cap \mathbf{S}^{2n+1}$ so that we can apply the diffeomorphism we defined. We also assume that π has the same property it does in the $\mathbf{S}^3$ bundle of mapping circles in $\mathbf{S}^{2n+1}$ to points in $\mathbb{CP}^n$.

$$\dot{q}(t) \in H_{q(t)}$$

In other words we want the tangent vector of our lift of γ to belong to the horizontal space of $\mathbf{S}^{2n+1}$. Using our requirement for a vector to belong to H_pP, i.e. $H_{q(t)} = \mathrm{Ker}(\omega_{q(t)})$, this is equivalent to the following.

$$\begin{aligned} & \omega_{q(t)}(\dot{q}(t)) = 0 \\ \Leftrightarrow \quad & q(t)^\dagger \dot{q}(t) = 0 \end{aligned}$$

We want $\theta(t)$ to be such that the lift is horizontal, thus we want to solve the above equation for $\theta(t)$. Thus we have.

$$\begin{aligned} q(t)^\dagger \dot{q}(t) &= e^{-i\theta}\xi^\dagger(e^{i\theta}\dot{\xi} + ie^{i\theta}\dot{\theta}\xi) \\ &= \xi^\dagger\dot{\xi} + i\dot{\theta}\xi^\dagger\xi \\ &= \omega_\xi(\dot{\xi}) + i\dot{\theta} \qquad (\xi \in \mathbf{S}^{2n+1} \Rightarrow \xi^\dagger\xi = 1) \\ &= 0 \\ \Rightarrow \dot{\theta} &= i\omega_\xi(\dot{\xi}) \end{aligned}$$

Therefore we are able to find a requirement of $\theta(t)$ for $q(t)$ to be horizontal. From this, it is possible to calculate the change in $\theta(t)$ along the curve, which we will denote by $\Delta\theta$, by simply integrating our expression for $\dot{\theta}$. Suppose we have a closed curve on $\mathbb{CP}^n$, i.e. $\gamma(t+T) = \gamma(t)$ (where T is a fixed number). Then we find that

$$\Delta\theta = \theta(T) - \theta(0) = \int_0^T i\omega_\xi(\dot{\xi})\, dt \qquad (25)$$

What we can deduce from this is that the change in θ is usually non-zero, even when the lift corresponds to a closed curve γ (i.e. one that returns to its original position). That is, for a lift $q(t) = \xi(t)e^{i\theta(t)}$ to remain horizontal, it will actually be rotated by a *holonomy angle* $\Delta\theta$. Thus we recover a similar result to the one we found at the end of the section on holonomy.

Namely that a curve (previously a vector), that is transported horizontally (previously parallel) and returned to its original position, will end up being rotated by an angle.

11 Hermitian Matrices

In the section about holonomy, we looked at the idea of a vector being parallel transported on the sphere, and ending up rotated by an angle due to the curvature of the sphere. We then did a similar thing for the principal bundle in the previous section. In this chapter we consider another analogue to this idea, where we take a parameter-dependent Hermitian matrix, and consider how, as the parameters travel along a path, the eigenvectors of this matrix are rotated by an angle.

A specific example of this in quantum mechanics is the *Berry* phase, which we will examine in more detail in a later section.

Let $H(\mathbf{R}(\mathbf{t}))$ be an n x n complex Hermitian matrix, where $\mathbf{R}(t) = (R_1(t), R_2(t), ..., R_{n^2}(t))$ is a vector of n^2 time-dependent real parameters (there are n^2 parameters since H has dimension n^2 over $\mathbb{R}$). The eigenvectors $\xi(\mathbf{R}(t)) \in \mathbb{C}^n$ of H are given by the usual eigenvector equation.

$$H\xi = \lambda\xi$$

Where $\lambda(\mathbf{R}(t)) \in \mathbb{R}$ denotes the corresponding eigenvalues. Since the lengths of the eigenvectors do not matter, we can assume ξ to be of unit length, i.e. let $\xi \in \mathbf{S}^{2n-1} \subset \mathbb{C}^n$. Then we observe that we can multiply ξ by $e^{i\theta}$ for any $\theta \in \mathbb{R}$ (since $e^{i\theta}$ is just a scalar in $\mathbb{C}^n$) and it will still be an eigenvector in $\mathbf{S}^{2n-1}$. Thus we can once again introduce $q(t)$ to denote the variation of these eigenvectors as follows.

$$q(t) = e^{i\theta(t)}\xi(t), \quad q(0) = \xi(\mathbf{R}(0)) \tag{26}$$

This is identical to when we used property (24) in section 10.2. Therefore from here it is possible to calculate the phase shift $\theta(t)$ in the eigenvectors of a Hermitian matrix in the same way we did before, by letting the parameters vary along a curve, C say, in parameter space. In the next section we will be looking at a specific case of this, when we consider variations around diabolic points.

11.1 Diabolic Points

When $H(\mathbf{R})$ has a single eigenvalue with algebraic multiplicity 2 at a point $\mathbf{R}$, we call this point a *diabolic* point. We wish to consider variations of the eigenvectors around these points, in order to calculate the phase changes. For this section we will work with 2 x 2 matrices instead of the general n x n case in order to simplify things. A complex 2 x 2 Hermitian matrix with so-called double eigenvalue d must take the following form.[5]

$$\begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix}, \qquad d \in \mathbb{R}$$

For convenience, we can set $d = 0$, so that the diabolic point is at $\mathbf{R} = \mathbf{0}$ in parameter space. To consider variation centred around this point, we want our eigenvalues to be $\pm r$ for some $r \in \mathbb{R}$. In order to do this, we vary the parameters $\mathbf{R}(\mathbf{t})$ that H is dependent on in such a way that we satisfy this restriction on the eigenvalues. Suppose we take a Hermitian matrix of the following form.

$$H = \begin{pmatrix} a & b + ic \\ b - ic & -a \end{pmatrix}, \qquad a, b, c \in \mathbb{R} \tag{27}$$

Here the parameters we are varying are $a(t), b(t)$ and $c(t)$ to maintain the symmetry (however in general a 2 x 2 Hermitian matrix would be dependent upon 4 parameters). Then it can be shown that the eigenvalues of this matrix are $\pm\sqrt{a^2 + b^2 + c^2} = \pm r$, which is what we required.

[5]This is shown in appendix A.3

Aside: As it turns out, this matrix can be obtained via a sum of the Pauli matrices. i.e.

$$\begin{aligned} H &= a\sigma_3 + b\sigma_1 + c\sigma_2 \\ &= a\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \end{aligned}$$

All of the Pauli matrices $\{I, \sigma_1, \sigma_2, \sigma_3\}$ are also Hermitian, i.e. they are all equal to their conjugate transpose, which is easy to see. Another interesting point about the Pauli matrices, is that their span is isomorphic to the space of quaternions $\mathbb{H}$. This can be shown by the following isomorphism relations of the basis elements.

$$1 \mapsto I,\ i \mapsto i\sigma_3,\ j \mapsto i\sigma_2,\ k \mapsto i\sigma_1$$

Essentially the reason behind this is they possess the same multiplication relations. For instance $ij = k$ corresponds to $i\sigma_3 i\sigma_2 = -(-i\sigma_1) = i\sigma_1$, and $i^2 = -1$ corresponds to $(i\sigma_3)^2 = -I$.

For ease of computation, we can re-parameterise the matrix in terms of spherical coordinates (noting that the eigenvalue r is consistent with the parameter r due to the above). We have

$$\begin{aligned} a &= r\cos\phi \\ b &= r\sin\phi\cos\theta \\ c &= r\sin\phi\sin\theta \quad 0 \le \theta < 2\pi,\ 0 < \phi < \pi \end{aligned}$$

Thus we can compute the eigenvectors ξ_1 and ξ_2 of H with respect to r, θ and ϕ. We solve $(H - rI)\xi = 0$ to get

$$\xi_{1,2} = \begin{pmatrix} a \pm r \\ b - ic \end{pmatrix} = \begin{pmatrix} r(\cos\phi \pm 1) \\ r\sin\phi e^{-i\theta} \end{pmatrix}$$

Naturally, we can choose to normalise these eigenvectors. Note that we can rewrite ξ_1 as follows.

$$\xi_1 = 2r\cos(\phi/2)\begin{pmatrix} \cos(\phi/2) \\ \sin(\phi/2)e^{-i\theta} \end{pmatrix}$$

Where we have used the double angle formulae, $\cos 2\phi + 1 = 2\cos^2\phi$ and $\sin 2\phi = 2\cos\phi\sin\phi$. Similarly using $\cos\phi - 1 = -2\sin^2\phi$, we can see that

$$\xi_2 = 2r\sin(\phi/2)\begin{pmatrix} -\sin(\phi/2) \\ \cos(\phi/2)e^{-i\theta} \end{pmatrix}.$$

So in normalised form, the two eigenvectors are as follows.

$$\xi_1 = \begin{pmatrix} \cos(\phi/2) \\ \sin(\phi/2)e^{-i\theta} \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} -\sin(\phi/2) \\ \cos(\phi/2)e^{-i\theta} \end{pmatrix}$$

Note that we talk about *normalised* in the sense of the complex inner product, i.e. $\|\xi_1\|^2 = \xi_1 \cdot \overline{\xi_1}$. So for instance we have

$$\|\xi_1\|^2 = \begin{pmatrix} \cos(\phi/2) \\ \sin(\phi/2)e^{-i\theta} \end{pmatrix} \cdot \begin{pmatrix} \cos(\phi/2) \\ \sin(\phi/2)e^{i\theta} \end{pmatrix} = \cos^2(\phi/2) + \sin^2(\phi/2) = 1$$

And similarly for ξ_2. Recall from section 10.2 that the condition for the lift of a curve to be horizontal (which was equivalent to parallel transporting a vector) was $q^\dagger \dot{q} = 0$. We want to apply this to the eigenvectors of H, so to do this we first of all need to calculate $\dot{\xi}_1$ and $\dot{\xi}_2$.

$$\begin{aligned} \dot{\xi}_1 &= \frac{1}{2}\dot{\phi}\begin{pmatrix} -\sin(\phi/2) \\ \cos(\phi/2)e^{-i\theta} \end{pmatrix} - i\dot{\theta}\begin{pmatrix} 0 \\ \sin(\phi/2)e^{-i\theta} \end{pmatrix} \\ \dot{\xi}_2 &= -\frac{1}{2}\dot{\phi}\begin{pmatrix} \cos(\phi/2) \\ \sin(\phi/2)e^{-i\theta} \end{pmatrix} - i\dot{\theta}\begin{pmatrix} 0 \\ \cos(\phi/2)e^{-i\theta} \end{pmatrix} \end{aligned}$$

Then we can calculate $\xi_1^\dagger\dot{\xi}_1$ and $\xi_2^\dagger\dot{\xi}_2$. We get the following.

$$\xi_1^\dagger\dot{\xi}_1 = \left(\cos(\phi/2), \sin(\phi/2)e^{i\theta}\right)\left(\frac{1}{2}\dot{\phi}\begin{pmatrix} -\sin(\phi/2) \\ \cos(\phi/2)e^{-i\theta} \end{pmatrix} - i\dot{\theta}\begin{pmatrix} 0 \\ \sin(\phi/2)e^{-i\theta} \end{pmatrix}\right)$$

$$
\begin{aligned}
&= 0 - i\dot{\theta}\sin^2(\phi/2) \\
&= \frac{i}{2}\dot{\theta}(\cos\phi - 1) \\
\xi_2^\dagger \dot{\xi}_2 &= -\frac{i}{2}\dot{\theta}(\cos\phi + 1)
\end{aligned}
$$

We use the same substitution as we used in (26) (i.e. $q(t) = e^{i\gamma(t)}\xi$). Then following the same method we used in section 10.2, we obtain

$$
\begin{aligned}
q^\dagger \dot{q} &= \xi^\dagger \dot{\xi} + i\dot{\gamma} \\
&= 0 \\
\Rightarrow \dot{\gamma} &= -\frac{1}{2}\dot{\theta}(1 \pm \cos\phi).
\end{aligned}
$$

Thus we obtain an equation for $\dot{\gamma}$, which allows us to calculate the phase shift γ explicitly if we know θ and ϕ. i.e. we have

$$
\gamma(t) = -\frac{1}{2}\oint_C \dot{\theta}(1 \pm \cos\phi)dt
$$

Where $C(t) = \mathbf{R}(t)$ is the curve describing the motion of the parameters. Since we are considering variation around the diabolic point, this curve surrounds the diabolic point. To illustrate how holonomy happens in this instance, we could for example look at the same curve we used in section 4 (i.e. a curve of constant latitude on the sphere, where $\phi(t) = \phi_0, \theta(t) = t$). Then we would get (if we assume $\gamma(0) = 0$)

$$
\gamma(t) = -\frac{1}{2}(1 \pm \cos\phi_0)t.
$$

In this case the holonomy angle if we went around a full rotation on the sphere is $\gamma(2\pi) - \gamma(0) = -\pi(1 \pm \cos(\phi_0))$. This is nonzero for any $\phi_0 \neq 0$ or π, which is what we would expect, since it coincides with the first example.

11.2 Symmetric Matrices

Suppose we want to restrict the above case to real Hermitian matrices H only. Clearly a real Hermitian matrix is just a symmetric matrix, since the conjugate of a real number is itself, so we would effectively be examining real symmetric matrices. Starting off from a general 2 x 2 real symmetric matrix H, assuming we have a double eigenvalue, we have the exact same form as before.

$$\begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix}, \qquad d \in \mathbb{R}$$

This is apparent when you consider the proof in the appendix works exactly the same for a real matrix, since $\bar{b}$ just becomes b. We once again choose to set $d = 0$ to make things easier, and then we consider variation around this point, using a similar approach to before.

$$\begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \qquad a, b, d \in \mathbb{R}$$

Clearly this time, we are somewhat restricted in that the off-diagonal entries cannot vary by $\pm ic$. The eigenvalues of this matrix are now just $\pm\sqrt{a^2 + b^2} = \pm r$, so we have one less degree of freedom than the first time. Then we once again swap a and b for a different coordinate system, but now we use polar coordinates instead of spherical coordinates. i.e.

$$\begin{aligned} a &= r\cos\phi \\ b &= r\sin\phi \quad 0 \leq \phi < 2\pi \end{aligned}$$

We can then solve $(H - rI)\xi = 0$ to get the eigenvectors.

$$\xi_{1,2} = \begin{pmatrix} a \pm r \\ b \end{pmatrix} = \begin{pmatrix} r(\cos\phi \pm 1) \\ r\sin\phi \end{pmatrix}.$$

Note that the eigenvectors take the same form as before, except for the missing factor of $e^{i\theta}$. These eigenvectors can be normalised in the same way to get:

$$\xi_1 = \begin{pmatrix} \cos(\phi/2) \\ \sin(\phi/2) \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} -\sin(\phi/2) \\ \cos(\phi/2) \end{pmatrix}$$

From which we can compute $\dot{\xi}_1$ and $\dot{\xi}_2$ as follows.

$$\dot{\xi}_1 = \frac{1}{2}\dot{\phi} \begin{pmatrix} -\sin(\phi/2) \\ \cos(\phi/2) \end{pmatrix}, \quad \dot{\xi}_2 = -\frac{1}{2}\dot{\phi} \begin{pmatrix} \cos(\phi/2) \\ \sin(\phi/2) \end{pmatrix}$$

Thus in this case, due to the missing θ terms when we differentiate, we end up with the result that $\xi_1^\dagger \dot{\xi}_1 = \xi_2^\dagger \dot{\xi}_2 = 0$. As it turns out, this does not necessarily end up with a trivial result however, since we have yet to compute $q^\dagger \dot{q}$. In fact we find that

$$\begin{aligned} q^\dagger \dot{q} &= i\dot{\gamma} = 0 \\ \Rightarrow \dot{\gamma} &= 0 \\ \Rightarrow \gamma(t) &= c \end{aligned}$$

By inspection, we can see that $\xi_1(\phi+2\pi) = -\xi_1(\phi)$ and $\xi_2(\phi+2\pi) = -\xi_2(\phi)$. Thus since $e^{i\pi} = -1$, we can see that the phase shift $\gamma = \pi$ around a diabolic point, a result that was first discovered by Longuet-Higgins in [8].

A diagram of what this looks like is shown in figure 7(b) below from [5].

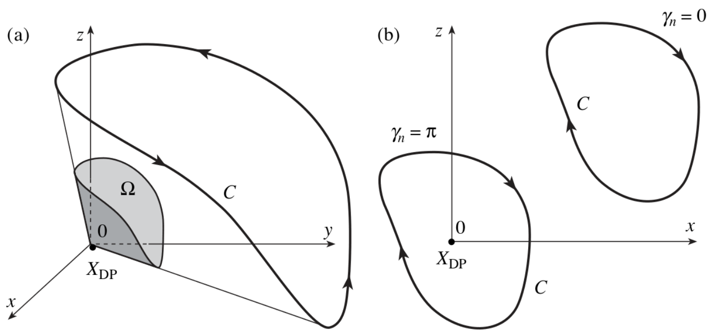

Figure 7: Values of γ for closed curves around a diabolic point for (a) complex and (b) real Hermitian matrices

We can see that if C goes round the diabolic point, denoted by X_{DP}, the value of $\gamma = \pi$ but if C lies outside the diabolic point, $\gamma = 0$.

For figure 7(a) which we covered in the previous section, γ is shown to depend upon the solid angle Ω that C creates around the diabolic point. Note that the parameter space has 3 dimensions in this case, and only 2 dimensions in case (b), which is consistent with what we have done.

11.3 Relation to 3-sphere

At the start of this section we introduced hermitian matrices as a separate concept to the rest of the report. Now I would like to explain the link between them and what we have studied previously, specifically how 2 x 2 hermitian matrices are related to the 3-sphere. Once again we use the complex number form of $\mathbf{S}^3$, i.e $\mathbf{S}^3 = \{\mathbf{z} = (z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}$. We note that in this form, it is possible to form a 2 x 2 matrix from the components of $\mathbf{z}$, z_1 and z_2, as follows.

$$\rho = \mathbf{z}\mathbf{z}^\dagger = \begin{pmatrix} z_1\overline{z_1} & z_1\overline{z_2} \\ z_2\overline{z_1} & z_2\overline{z_2} \end{pmatrix}$$

This matrix, referred to as the *density matrix* in [7], is in fact Hermitian (since $\overline{z_1\overline{z_2}} = z_2\overline{z_1}$), and it also has unit trace (since $z_1\overline{z_1}+z_2\overline{z_2} = |z_1|^2+|z_2|^2 = 1$). Furthermore it can be shown that $\rho^2 = \rho$ since

$$\begin{aligned} \rho^2 &= (\mathbf{z}\mathbf{z}^\dagger)(\mathbf{z}\mathbf{z}^\dagger) \\ &= \mathbf{z}(\mathbf{z}^\dagger\mathbf{z})\mathbf{z}^\dagger \\ &= \mathbf{z}\mathbf{z}^\dagger \quad (\mathbf{z}^\dagger\mathbf{z} = (\overline{z_1}, \overline{z_2}) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1\overline{z_1} + z_2\overline{z_2} = 1) \\ &= \rho \end{aligned}$$

To show how this matrix is essentially equivalent to the one we looked at, consider the form we first analysed in equation (27) before we decided to set the eigenvalue $d = 0$. i.e. we have

$$H = \begin{pmatrix} d+a & b+ic \\ b-ic & d-a \end{pmatrix}, \qquad a,b,c,d \in \mathbb{R}$$

To satisfy the unit trace and $\rho^2 = \rho$ requirements, we can choose to set $d = \frac{1}{2}$, and to scale all the parameters by $\frac{1}{2}$ to get

$$H = \frac{1}{2}\begin{pmatrix} 1+a & b+ic \\ b-ic & 1-a \end{pmatrix}, \qquad a,b,c \in \mathbb{R}$$

Then provided we make a further assumption that $\mathbf{R} = (a,b,c) \in \mathbf{S}^2 \subset \mathbb{R}^3$, we can quite easily see that H does satisfy these qualities. For instance we have

$$H^2 = \frac{1}{4}\begin{pmatrix} 1+2a+a^2+b^2+c^2 & 2(b+ic) \\ 2(b-ic) & 1-a+a^2+b^2+c^2 \end{pmatrix} = H$$

Thus we can think of H as a parameterisation of the density matrix ρ.

12 Berry Phase

We continue from where we left off in section 11, where we had an n x n Hermitian matrix $H(\mathbf{R}(t))$ and eigenvectors $\xi(\mathbf{R}(t))$ (which we generalised as $q(t)$). This section closely follows [4] in content, and is a quantum mechanical example of what we did with holonomy in previous sections. Note that in this circumstance, H refers to the *Hamiltonian* (the total energy) of the system, the eigenvectors ξ are the *eigenstates*, and the corresponding eigenvectors $\lambda(t)$ are the energy levels at each given eigenstate. We also assume that the evolution is cyclic, i.e. $\mathbf{R}(0) = \mathbf{R}(T)$ for some final time T, which means that $H(\mathbf{R}(0)) = H(\mathbf{R}(T))$, so H takes the same form at the start that it does at the end. As before, the parameters travel along the closed curve $C : t \in [0,T] \mapsto \mathbf{R}(t)$, and so by extension the eigenvectors are described by this curve.

The evolution of our eigenstates $q(t)$ in a quantum mechanical system are given by the Schrödinger equation.

$$i\hbar\frac{d}{dt}q = Hq \tag{28}$$

Recall our expression we had for q in equation (26) in terms of ξ. We can substitute this into (28), to obtain an equation which is in terms of $\xi(t)$ and $\theta(t)$.

$$\begin{aligned} i\hbar e^{i\theta}(i\dot{\theta}\xi + \dot{\xi}) &= He^{i\theta}\xi \\ &= \lambda e^{i\theta}\xi \end{aligned}$$

Where we have used that ξ is an eigenvector of H. By multiplying on the left by $\xi^\dagger e^{-i\theta(t)}$, and using the fact that $\xi^\dagger\xi = 1$, we get:

$$\begin{aligned} -\hbar\dot{\theta} + i\hbar\xi^\dagger\dot{\xi} = \lambda \\ \Rightarrow \dot{\theta} = -\frac{1}{\hbar}\lambda + i\xi^\dagger\dot{\xi} \end{aligned}$$

We assume we are on a loop, i.e. $\xi(0) = \xi(T)$, so we can integrate this equation from $t = 0$ to $t = T$ to get an equation for $\theta(t)$. Set $\theta(0) = 0$ and we have:

$$\theta(T) = -\frac{1}{\hbar}\int_0^T \lambda\, dt + i\int_0^T \xi^\dagger\dot{\xi}\, dt$$

At this point it's worth noting that the second integral is equivalent to what we found in equation (25), if we were to assume $\theta(0) = 0$ in that case. The first term only comes about as a result of the introduction of the Schrödinger equation, but otherwise the two equations are identical.

By recalling that ξ is a function of $\mathbf{R}(t) = (R_1(t), R_2(t), ..., R_{n^2}(t))$, i.e. it is not an explicit function of t, we can apply the chain rule and write out $\dot{\xi}$ in terms of $\mathbf{R}(t)$ as follows.

$$\dot{\xi} = \frac{d\xi}{dt}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n^2} \frac{\partial \xi}{\partial R_i} \frac{dR_i}{dt} \\
&= \nabla_{\mathbf{R}} \xi \cdot \frac{d\mathbf{R}}{dt}
\end{aligned}
$$

Where $\nabla_{\mathbf{R}}\xi$ denotes the grad of ξ, a vector, i.e. $\nabla_{\mathbf{R}}\xi = \left(\frac{\partial \xi}{\partial R_1}, \frac{\partial \xi}{\partial R_2}, \ldots, \frac{\partial \xi}{\partial R_{n^2}}\right)^T$. Thus we are able to use that $\frac{d\mathbf{R}}{dt}dt = d\mathbf{R}$, and the fact that $\mathbf{R}(t)$ travels along a closed curve C is parameter space, to rewrite the second integral as a line integral along C. We also make the choice of initial conditions $\xi(\mathbf{R}(0)) = \xi(\mathbf{R}(T))$.

$$
\theta(T) = -\frac{1}{\hbar}\int_0^T \lambda \, dt + i \oint_C \xi^\dagger \nabla_{\mathbf{R}} \xi \cdot d\mathbf{R} \tag{29}
$$

This expression can be split into two parts, the first part called the *dynamical* phase and the second called the *Berry* phase. The dynamical phase corresponds to the usual expression of the phase accumulated by a system in a state with energy λ for a time T [4, p. 4]. As previously mentioned, the Berry phase corresponds to the extra phase factor acquired that is effectively the holonomy of the system.

12.1 Properties of Berry Phase

It can be shown that the Berry phase is independent of the speed at which the closed curve C is traversed provided the total time is the same, i.e. the phase factor will be the same if say, the system travels faster at the start and slower at the end. To show this, we can insert the substitution $t \mapsto \tau(t)$ into the integral and show that it is equivalent. i.e.

$$
\begin{aligned}
i\int_0^T \xi^\dagger(\mathbf{R}(t))\frac{d}{dt}\xi(\mathbf{R}(t))\, dt &= i\int_{\tau(0)}^{\tau(T)} \xi^\dagger(\mathbf{R}(\tau))\frac{d\tau}{dt}\frac{d}{d\tau}\xi(\mathbf{R}(\tau))\, \frac{dt}{d\tau} d\tau \\
&= i\int_{\tau(0)}^{\tau(T)} \xi^\dagger(\mathbf{R}(\tau))\frac{d}{d\tau}\xi(\mathbf{R}(\tau))\, d\tau
\end{aligned}
$$

Then the integral is the same, provided we assume $\tau(0) = 0$ and $\tau(T) = T$, which should be the case since we are just rescaling the time in between.

The Berry phase is also invariant, up to a multiple of 2π, of the phase of our eigenvectors. Say for instance we chose $\xi' = e^{i\gamma(t)}\xi$ instead of ξ. Then we would have for the Berry phase

$$\begin{aligned} i\int_0^T \xi'^{\dagger}\frac{d}{dt}\xi'\,dt &= i\int_0^T \overline{e^{i\gamma(t)}}\xi^{\dagger}\frac{d}{dt}(e^{i\gamma(t)}\xi)\,dt \\ &= i\int_0^T e^{-i\gamma(t)}\xi^{\dagger}(e^{i\gamma(t)}\frac{d}{dt}\xi + i\frac{d}{dt}\gamma(t)e^{i\gamma(t)}\xi)\,dt \\ &= i\int_0^T \xi'^{\dagger}\frac{d}{dt}\xi'\,dt - \int_0^T \frac{d}{dt}\gamma(t)\,dt \\ &= i\int_0^T \xi'^{\dagger}\frac{d}{dt}\xi'\,dt + \gamma(0) - \gamma(T) \end{aligned}$$

Where we have used that $\xi^{\dagger}\xi = 1$. Now note that since $\xi(\mathbf{R}(0)) = \xi(\mathbf{R}(T))$ by our assumption at the start of this section, and similarly $\xi'(\mathbf{R}(0)) = \xi'(\mathbf{R}(T))$. This means that $e^{i\gamma(0)} = e^{i\gamma(T)}$, because $e^{i\gamma(t)}$ is the phase difference between these two eigenvectors, and they are the same at the start as they are at the end. Thus $\gamma(0) - \gamma(T) = 0$ (modulo 2π) and so the Berry phase is indeed independent of how we scale the eigenvectors (up to modulo 2π). This property is called *gauge invariance*.

One further thing we can do with the Berry phase is transform it into a surface integral from a line integral. This is most easily demonstrated when the parameter space is 3-dimensional, by using Stokes' theorem, although it is possible in higher dimensions. First we note Stokes' theorem for a surface Ω where $\partial\Omega = C$, the curve in parameter space.

$$\oint_C \mathbf{F}\cdot d\mathbf{R} = \int_\Omega \nabla\times\mathbf{F}\cdot d\mathbf{S}$$

Then we can transform the Berry phase used in equation (29) to get

$$i\int_\Omega \nabla\times\xi^{\dagger}\nabla\xi\cdot d\mathbf{S}$$

A Appendix

A.1 Proof that the Hopf Map is a Submersion

To prove that the Hopf map is indeed a submersion, we have to show that its Jacobian ($d_p\pi : T_p\mathbf{S}^3 \to T_{\pi(p)}\mathbf{S}^2$) is surjective. The Jacobian, which we also wrote down in (23) is as follows.

$$\begin{pmatrix} dw_1 \\ dw_2 \\ dw_3 \end{pmatrix} = 2 \begin{pmatrix} a & b & -c & -d \\ d & c & b & a \\ -c & d & -a & b \end{pmatrix} \begin{pmatrix} da \\ db \\ dc \\ dd \end{pmatrix}$$

Where the Jacobian is the 3 x 4 matrix. One way to show that it is surjective is to compute the minors of the matrix, namely the Jacobian is surjective if at least one 3x3 minor is non-zero. The formula for the minors are as follows.

$$m_1 = \begin{vmatrix} a & b & -c \\ d & c & b \\ -c & d & -a \end{vmatrix}, \quad m_2 = \begin{vmatrix} a & b & -d \\ d & c & a \\ -c & d & b \end{vmatrix},$$

$$m_3 = \begin{vmatrix} a & -c & -d \\ d & b & a \\ -c & -a & b \end{vmatrix}, \quad m_4 = \begin{vmatrix} b & -c & d \\ c & b & a \\ d & -a & b \end{vmatrix}$$

Through some lengthy calculations, we can compute these minors. We get:

$$\begin{aligned} m_1 &= ab^2 + a^3 + ac^2 + ad^2 \\ m_2 &= b^3 + a^2b + bc^2 + bd^2 \\ m_3 &= -a^2c - b^2c - cd^2 - c^3 \\ m_4 &= -a^2d - b^2d - d^3 - c^2d \end{aligned}$$

Then we notice the following:

$$
\begin{aligned}
m_1^2 + m_2^2 + m_3^2 + m_4^2 &= a^2(a^2+b^2+c^2+d^2)^2 + b^2(a^2+b^2+c^2+d^2)^2 \\
&\quad + c^2(-a^2-b^2-c^2-d^2)^2 + d^2(-a^2-b^2-c^2-d^2)^2 \\
&= 1
\end{aligned}
$$

Where we have used the fact that $a^2 + b^2 + c^2 + d^2 = 1$. Thus we have shown that indeed at least one minor must be non-zero. Therefore the matrix is surjective, and the Hopf map is a submersion.

A.2 Proof that $T_p\mathbf{S}^3 = \text{span}\{J_1\mathbf{n}, J_2\mathbf{n}, J_3\mathbf{n}\}$

First of all, notice that for any vector $\mathbf{n} \in \mathbf{S}^3 \subset \mathbb{R}^4$, $\mathbf{n}$ is normal to $\mathbf{S}^3$, since it is a vector pointing from the origin to the surface of $\mathbf{S}^3$. Thus it stands to reason that the tangent space of $\mathbf{S}^3$ is spanned by 3 orthogonal vectors in $\mathbb{R}^4$ that are orthogonal to $\mathbf{n}$ (orthogonality implies they form a linearly independent set), since $\mathbf{S}^3$ is a surface in $\mathbb{R}^4$. This can apply to any 3 such vectors in $\mathbb{R}^4$, but in particular it applies to the set $\{J_1\mathbf{n}, J_2\mathbf{n}, J_3\mathbf{n}\}$. To show this, consider the inner product of each of these vectors with $\mathbf{n}$. For $i = 1, 2, 3$, we have

$$
\begin{aligned}
\langle \mathbf{n}, J_i\mathbf{n} \rangle &= \langle J_i^\dagger \mathbf{n}, \mathbf{n} \rangle \quad \text{(using the definition of the adjoint operator)} \\
&= \langle J_i^T \mathbf{n}, \mathbf{n} \rangle \quad \text{(since we are in } \mathbb{R}^4\text{)} \\
&= -\langle J_i \mathbf{n}, \mathbf{n} \rangle \quad \text{(due to the fact } J_i \text{ is antisymmetric)} \\
&= -\langle \mathbf{n}, J_i \mathbf{n} \rangle \quad \text{(since the inner product is symmetric)} \\
&= 0
\end{aligned}
$$

So the set $\{J_1\mathbf{n}, J_2\mathbf{n}, J_3\mathbf{n}\}$ is orthogonal to $\mathbf{n}$, and it remains to show that the set is orthogonal to itself. We have for $i \neq j$

$$
\begin{aligned}
\langle J_i\mathbf{n}, J_j\mathbf{n} \rangle &= \langle \mathbf{n}, J_i^T J_j \mathbf{n} \rangle \\
&= -\langle \mathbf{n}, J_i J_j \mathbf{n} \rangle
\end{aligned}
$$

$$
\begin{aligned}
&= -\sum_{k=1}^{3} \epsilon_{ijk} \langle \mathbf{n}, J_k \mathbf{n} \rangle \quad \text{(where } \epsilon_{ijk} \text{ is a scalar)} \\
&= 0 \quad \text{(by the above result)}
\end{aligned}
$$

It can also be shown they are orthonormal since $\|J_i\mathbf{n}\| = 1$ for $i = 1, 2, 3$. Thus we have proven that $T_p\mathbf{S}^3 = \text{span}\{J_1\mathbf{n}, J_2\mathbf{n}, J_3\mathbf{n}\}$.

A.3 Form of a Hermitian Matrix with a Double Eigenvalue

The general form of a 2 x 2 complex Hermitian matrix H is the following.

$$
H = \begin{pmatrix} a & \bar{b} \\ b & d \end{pmatrix}, \qquad a, d \in \mathbb{R}, b \in \mathbb{C}
$$

The eigenvalues of H are solutions of the characteristic equation, which is

$$
\begin{aligned}
(a - \lambda)(d - \lambda) - |b|^2 &= 0 \\
\lambda^2 - (a + d)\lambda + ad - |b|^2 &= 0 \\
\Rightarrow 2\lambda &= a + d \pm \sqrt{(a - d)^2 + 4|b|^2}
\end{aligned}
$$

Therefore H only has a double eigenvalue when the square root is zero, i.e. when $(a - d)^2 + 4|b|^2 = 0$, which means that $a = d$ and $b = 0$. Thus the form in section 11.1 is chosen. In such a case, the eigenvalue λ of H is d.

References

[1] D. W. Lyons. An Elementary Introduction to the Hopf Fibration, *Mathematics Magazine* **76**, pp. 87-98, 2003.

[2] A. J. Hanson. *Visualizing Quaternions*, Morgan Kaufmann, San Francisco, CA, 2006.

[3] R. H. Cushman and L. M. Bates. *Global Aspects of Classical Integrable Systems*, Birkhäuser, Basel, 1997.

[4] S. Filipp. *New Aspects of the Quantum Geometric Phase*, Dissertation, Faculty of Technical Physics, University of Vienna, 2006.

[5] A. A Mailybaev, O. N. Kirillov, and A. P. Seyranian. Berry Phase around Degeneracies, *Doklady Mathematics* **73**, pp. 129-133, 2006.

[6] R. Way. *Dynamics in the Hopf Bundle, the Geometric Phase and Implications for Dynamical Systems*, Ph.D. thesis, University of Surrey, 2008.

[7] H. Urbantke. Two-level Quantum Systems: States, Phases, and Holonomy, *American Journal of Physics* **59**, pp. 503-509, 1991.

[8] H. C. Longuet-Higgins. The Intersection of Potential Energy Surfaces in Polyatomic Molecules *Proc. R. Soc. London A.* **344**, pp. 147-156, 1975.