Charting the 3-Sphere—An Exposition for Undergraduates

Louis Zulli

Most people are familiar with the use of latitude and longitude in locating and naming points on the surface of the Earth, which, mathematically at least, is considered to be a perfectly round two-dimensional sphere. This article is about a similar coordinate system on the 3-sphere S^3, the set of points in four-dimensional Euclidean space that lie exactly one unit from the origin. S^3 is a rich and beautiful space, and an exploration of it, even at the level of this introductory article, involves a wealth of interesting mathematics. Indeed, here are just some of the mathematical players who will appear in the brief exposition we're about to present: numbers—real, complex and quaternion; matrices—real and complex, orthogonal and unitary; linear algebra—vector spaces, subspaces, inner products, traces, eigenvalues, eigenvectors, diagonalization and the matrix exponential; group theory—conjugation and conjugacy classes; geometry—intrinsic distance, tangent spaces and the exponential map. Quite a lot of mathematics in a short expository paper about a single three-dimensional space—and we could have gone much further. In fact, the main difficulty in writing this paper was knowing when to say 'Enough!'.

Which brings me to the point of this paper. I wrote this paper as a paper to be studied by undergraduate Mathematics Majors in a Senior Seminar. Of course, I hope that this article will be read by other people in other contexts as well, but my focus while writing this paper was fixed on the Senior Seminar, and I think that this focus is reflected in the paper's style. In particular, this paper is meant to be read and discussed by a group of students, so that all might benefit from the sharing of knowledge, and so that confusions might be quickly resolved and obstacles overcome. (I also assumed that there would be a faculty member present to help the students navigate through some of the paper's more treacherous passages.) I did not attempt to make this paper either self-contained or linear. This is because few topics in mathematics are truly self-contained, and because mathematics as a whole is decidedly non-linear. Throughout the paper there are terms and facts borrowed from many branches of mathematics, and there is also the occasional tidbit, usually parenthesized, meant to entice the readers to explore new territories in the mathematical kingdom. There are also many verifications left to the readers; most of these verifications are calculations. Having said all this, I feel that I should add that I tried to organize the paper so as to assist the readers in learning the material, and that I attempted to explain, at least informally, most of the technical terms in the article. I also tried to give proofs, or sketches of proofs, of most of the claims made within. At the end of the paper is a short list of references, which I include mainly as an act of basic decency. My hope is that the readers already have their own favorite sources of mathematical knowledge, and that they will consult these favorite sources as needed. I was inspired to write this article after reading the first dozen pages of [1]. In some sense, those pages provided an outline for this paper. (After writing this article I discovered a similar treatment in Chapter 8 of [2]. A reading of that chapter would be an excellent supplement to this exposition.)

1. S^1 AND $SO(2)$. Before confronting the 3-sphere S^3 in $\mathbf{R}^4$, let's first consider the familiar 1-sphere (unit circle) S^1 in $\mathbf{R}^2$. (This circle will reappear later, as the "great circle" in S^3 containing the "Prime Meridian.") Let $z = x + iy$ be a complex number, which we will identify with the point $z = (x, y)$ in $\mathbf{R}^2$. Sometimes we'll also think of $z = (x, y)$ as being the vector $\vec{z}$ in $\mathbf{R}^2$ from the origin to (x, y). There is a well-known correspondence (bijection) between the complex numbers and certain 2×2 matrices of real numbers:

$$z = x + iy \in \mathbf{C} = \mathbf{R}^2 \leftrightarrow Z = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \in M_2(\mathbf{R}).$$

(The notational convention employed here is used throughout this article. That is, points in spaces are denoted by lowercase letters and matrices by uppercase letters. The passage from a point to a matrix representation of that point is indicated by capitalization.) This correspondence is actually a field isomorphism, meaning that it respects addition and multiplication. Via this correspondence we obtain a matrix model of C, which we can also view as a matrix model of $\mathbf{R}^2$.

In $\mathbf{R}^2$ there is an inner product, the ordinary dot product of vectors, and this inner product is quite helpful when one does geometry in $\mathbf{R}^2$. For vectors $\vec{z}_1 = (x_1, y_1)$ and $\vec{z}_2 = (x_2, y_2)$ in $\mathbf{R}^2$, $\vec{z}_1 \cdot \vec{z}_2 = \langle \vec{z}_1, \vec{z}_2 \rangle = x_1x_2 + y_1y_2$. If we treat these same vectors as complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$, then the dot product can be computed as $\langle z_1, z_2 \rangle = \frac{1}{2}(z_1\bar{z}_2 + \bar{z}_1z_2)$. What is the formula for this standard inner product in the matrix model of $\mathbf{R}^2$? It is not difficult to verify that $\langle Z_1, Z_2 \rangle = \frac{1}{2}\mathrm{tr}(Z_1Z_2^T)$ exactly corresponds to $\langle z_1, z_2 \rangle = \frac{1}{2}(z_1\bar{z}_2 + \bar{z}_1z_2)$, where Z_2^T denotes the transpose of Z_2, and $\mathrm{tr}(Z_1Z_2^T)$ denotes the trace of $Z_1Z_2^T$, the sum of the elements on the main diagonal of $Z_1Z_2^T$. Using the inner product on $\mathbf{R}^2$ we can compute the norm (length) $|z|$ of a complex number (vector) $z \in \mathbf{R}^2$. By definition, $|z|^2 = \langle z, z \rangle$. In our matrix model of $\mathbf{R}^2$ this becomes $|Z|^2 = \frac{1}{2}\mathrm{tr}(ZZ^T)$, which simplifies to $\det(Z)$, the determinant of Z, because of the special form which Z has. So $|Z|^2 = \det(Z)$.

By definition, $S^1 = \{z \in \mathbf{C}: |z| = 1\} = \{\vec{z} \in \mathbf{R}^2: |\vec{z}| = 1\}$; this is the unit circle in the plane $\mathbf{R}^2$. In the matrix model of $\mathbf{R}^2$ this corresponds to $\{Z = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}: \det(Z) = 1\}$. This set of matrices can also be described as $\{M \in M_2(\mathbf{R}): MM^T = I, \det(M) = 1\}$. A square matrix M of real numbers that satisfies the equation $MM^T = I$ is called an *orthogonal matrix*, and its determinant is necessarily 1 or -1. Those orthogonal matrices with determinant 1 are called *special orthogonal matrices*. Thus $S^1 \subseteq \mathbf{C} = \mathbf{R}^2$ corresponds to the group of 2×2 special orthogonal matrices, which is denoted $SO(2)$. Furthermore, the correspondence is a group isomorphism, where the operation in both groups is multiplication. It is worth knowing that $SO(n)$ can also be thought of as the group of orientation-preserving linear isometries of $\mathbf{R}^n$, that is, as the group of orientation-preserving rigid motions of $\mathbf{R}^n$ that fix the origin. In $\mathbf{R}^2$ the only such motions are rotations about the origin, and there is one such rotation for each point on the unit circle S^1. So perhaps it's not surprising that S^1 is isomorphic to the group $SO(2)$.

2. S^3 AND $SU(2)$. Now let $v = x + iy + jz + kt$ be a quaternion, which we will identify with the point $v = (x, y, z, t)$ in $\mathbf{R}^4$, and sometimes with the vector $\vec{v} = (v_1, v_2) \in \mathbf{C}^2$, where $v_1 = x + iy$ and $v_2 = z + it$. As above, there is a correspondence between the quaternions and certain 2×2 matrices of complex numbers. In this case, the correspondence is:

$$v = x + iy + jz + kt \in \mathbf{H} = \mathbf{R}^4 \leftrightarrow V = \begin{pmatrix} v_1 & v_2 \\ -\bar{v}_2 & \bar{v}_1 \end{pmatrix} \in M_2(\mathbf{C}),$$

where $v_1 = x + iy$ and $v_2 = z + it$. This correspondence preserves addition and multiplication; it is an isomorphism of division rings. (A division ring is a "possibly non-commutative field"—multiplication might not be commutative but all the other field axioms are satisfied.) Via this correspondence we obtain a matrix model of $\mathbf{R}^4$. In this model, the ordinary Euclidean inner product becomes $\langle V_1, V_2 \rangle = \frac{1}{2}\mathrm{tr}(V_1 \overline{V}_2^T)$, and, as above, $|V|^2 = \det(V)$.

By definition, $S^3 = \{v \in \mathbf{H}\colon |v| = 1\} = \{\vec{v} \in \mathbf{R}^4\colon |\vec{v}| = 1\}$; this is the unit 3-sphere in four-dimensional space. Note that S^3 is a (non-commutative) group via quaternion multiplication. In the matrix model of $\mathbf{R}^4$, S^3 corresponds to $\{V = \begin{pmatrix} v_1 & v_2 \\ -\bar{v}_2 & \bar{v}_1 \end{pmatrix} \in M_2(\mathbf{C})\colon \det(V) = 1\}$, which can also be described as $\{M \in M_2(\mathbf{C})\colon M\overline{M}^T = I, \det(M) = 1\}$. This matrix group is denoted $SU(2)$, and is called the two-dimensional *special unitary group*. The correspondence $S^3 \leftrightarrow SU(2)$ defined above is a group isomorphism. Because of this, we will often blur the distinction between S^3 and $SU(2)$. In particular, we will often consider points in S^3 to be matrices, labeling them with uppercase letters.

3. DISTANCE, PARALLELS AND MERIDIANS IN S^3. We can already compute the distance in $\mathbf{R}^4$ between points v and w in S^3. This is just $|v - w|$, or, in our matrix formulation, $[\det(V - W)]^{1/2}$. This distance is simply the length of a vector in $\mathbf{R}^4$ from v to w. But such a vector intersects S^3 only at v and w, and its length is not the distance in S^3 from v to w. (The distinction between these two notions of distance is easily seen in analogy: The distance across the surface of the Earth from Lisbon to Vladivostok is greater than the distance between those cities were tunneling through the Earth allowed.) How can we calculate $d_{S^3}(v, w)$, the so-called *intrinsic distance* from v to w in S^3? Consider a two-dimensional plane Π containing v, w and the origin O in $\mathbf{R}^4$. This plane will be unique unless $w = -v$, but uniqueness of Π is not required. What is $\Pi \cap S^3$, the intersection of Π and S^3 in $\mathbf{R}^4$? $\Pi \cap S^3$ lies within both Π and S^3, but it is easier to analyze within Π, where it is just the set of points in Π one unit from the origin O. Thus $\Pi \cap S^3$ is just the unit circle in Π. In S^3, the circle $\Pi \cap S^3$ is called a *great circle*, and the key fact that we need is that this circle provides the shortest path within S^3 from v to w. Accepting this as true, it is then easy to compute $d_{S^3}(v, w)$, by focusing once again on the plane Π. Since v and w lie on the unit circle in Π, the distance between them along that circle is exactly the radian measure θ of the (smaller) undirected angle vOw in Π, by the very definition of radian measure. See Figure 1. From the geometric dot product formula, we obtain $\theta = \cos^{-1}(\langle v, w \rangle)$, since v and w are unit vectors. Thus $d_{S^3}(v, w) = \cos^{-1}(\langle v, w \rangle)$, or, in our matrix formulation, $d_{S^3}(V, W) = \cos^{-1}(\frac{1}{2}\mathrm{tr}(V\overline{W}^T))$.

With this distance formula in hand, we can begin charting S^3. Since the identity matrix I corresponds to the point $(1, 0, 0, 0) \in S^3 \subseteq \mathbf{R}^4$, the positive x-axis in $\mathbf{R}^4$

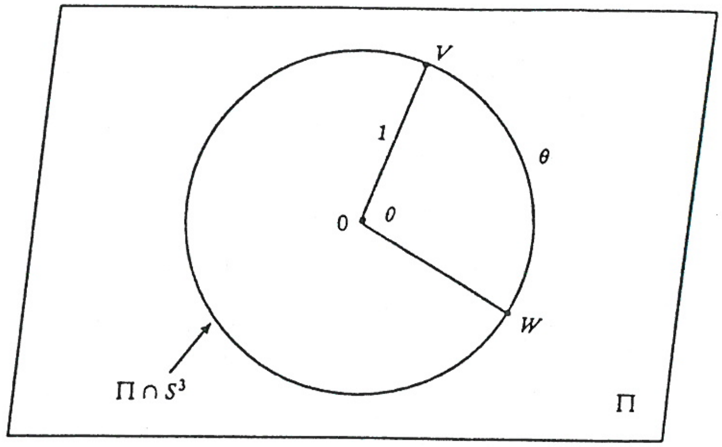

Figure 1

passes through I. We will view I as the "North Pole" in our model of S^3—not surprisingly, $-I$ will be the "South Pole." For $L \in [0, \pi]$, let $\mathscr{P}_L = \{V \in S^3\colon d_{S^3}(V, I) = L\}$. We will call $\mathscr{P}_L$ the *parallel in* S^3 *at latitude* L. Referring to the distance L as latitude is quite reasonable, since not only is L the distance in S^3 from V to I, but it's also the radian measure of angle IOV. (So we'll be measuring our latitudes down from the North Pole, rather than up and down from the Equator as is customary on Earth.) $\mathscr{P}_0 = \{I\}$, that is, I is the only point in S^3 at latitude 0, since it's the only point in S^3 whose distance from I is zero. $\mathscr{P}_\pi = \{-I\}$, although perhaps it is not immediately apparent that $-I$ is the only point in S^3 at distance π from I. We will refer to $\mathscr{P}_{\pi/2}$ as the *Equator* in S^3, and use the symbol $\mathscr{E}$ to denote this special parallel. It is helpful to have several descriptions of $\mathscr{P}_L$ at one's fingertips, in particular: $\mathscr{P}_L = \{V \in S^3\colon \frac{1}{2}\mathrm{tr}(V) = \cos(L)\}$ and $\mathscr{P}_L = \{(x, y, z, t) \in S^3 \subseteq \mathbf{R}^4\colon x = \cos(L)\}$. From the first of these two reformulations we conclude that all points on a given parallel in S^3 have the same trace, and that V lies on the Equator $\mathscr{E}$ in S^3 if and only if $\mathrm{tr}(V) = 0$. From the second reformulation we see that $\mathscr{P}_L$ is a two-dimensional sphere of radius $\sin(L)$ centered at $(\cos(L), 0, 0, 0)$ within the hyperplane $x = \cos(L)$ in $\mathbf{R}^4$. (If $L = 0$ or $L = \pi$ this sphere has radius 0, so $\mathscr{P}_0$ and $\mathscr{P}_\pi$ are indeed one point sets.) In particular, $\mathscr{E}$ is just the unit 2-sphere in the three-dimensional subspace of $\mathbf{R}^4$ spanned by the y, z and t axes—or equivalently, $\mathscr{E}$ is the set of pure imaginary unit quaternions, that is, the unit quaternions of the form $iy + jz + kt$. From the second reformulation of $\mathscr{P}_L$ we can also conclude that $S^3 = \bigcup_{L \in [0, \pi]} \mathscr{P}_L$, that is, that every point in S^3 lies on some parallel. This is because $x \in [-1, 1]$ for every point $(x, y, z, t) \in S^3 \subseteq \mathbf{R}^4$, so $x = \cos(L)$ for some $L \in [0, \pi]$. Clearly $\mathscr{P}_L \cap \mathscr{P}_{L'} = \varnothing$ if $L \neq L'$, so in fact each point in S^3 lies on a unique parallel.

Now let $E \in \mathscr{E}$. By the *meridian* $\mathscr{M}_E$ *at longitude* E we mean the set of points in S^3 of the form $\cos(\theta)I + \sin(\theta)E$, for $\theta \in [0, \pi]$. (Notice that, once again, we're deviating slightly from the convention used on Earth. To us, the longitude of a meridian in S^3 is simply the point of intersection of that meridian with the Equator. If an analogous system were used on Earth, then the meridian containing Buffalo, N.Y. would be called "Quito, Ecuador" rather than "78°W.") It is easy to check that $\mathscr{M}_E$ intersects $\mathscr{P}_L$ only at the point $\cos(L)I + \sin(L)E$, and that

$\mathscr{M}_E \cap \mathscr{M}_{E'} = \{\pm I\}$ if $E \neq E'$. To see that $S^3 = \cup_{E \in \mathscr{E}} \mathscr{M}_E$, that is, that every point $V \in S^3$ lies on at least one meridian, one could argue geometrically as follows: Choose a plane Π in $\mathbf{R}^4$ containing I, V and the origin O. As before, Π will be unique if $V \neq \pm I$, but uniqueness of Π is not required. Let $\mathbf{R}^3$ denote the three-dimensional subspace of $\mathbf{R}^4$ spanned by the y, z and t axes, so that $\mathbf{R}^3 \cap S^3 = \mathscr{E}$. Clearly $\Pi \not\subseteq \mathbf{R}^3$ since $I \in \Pi$ but $I \notin \mathbf{R}^3$. Since $\Pi \cap \mathbf{R}^3$ is necessarily a subspace of Π, either $\Pi \cap \mathbf{R}^3 = \{O\}$ or $\Pi \cap \mathbf{R}^3$ is a line through O. The first of these possibilities is eliminated by dimensional considerations, so $\Pi \cap \mathbf{R}^3$ is a line in $\mathbf{R}^3$ containing O. This line intersects $\mathscr{E}$, the unit sphere in $\mathbf{R}^3$, in two points, E and $-E$. Since $(\Pi \cap \mathbf{R}^3) \cap \mathscr{E} = (\Pi \cap S^3) \cap \mathscr{E}$, the great circle $\Pi \cap S^3$, which contains V, intersects $\mathscr{E}$ at $\pm E$. Since I and E are orthogonal points on this circle in Π, each point on the circle can be expressed as $\cos(\theta)I + \sin(\theta)E$, for some $\theta \in [0, 2\pi]$. In particular, $V = \cos(\theta)I + \sin(\theta)E$ for some $\theta \in [0, 2\pi]$. If $\theta \in [0, \pi]$ then $V \in \mathscr{M}_E$, and we're done. Otherwise, $\theta = 2\pi - \theta'$ with $\theta' \in [0, \pi]$, and substituting for θ, we obtain $V = \cos(\theta')I + \sin(\theta')(-E)$, so $V \in \mathscr{M}_{-E}$.

Although this geometric argument demonstrates that each $V \in S^3$ lies on a meridian, it may not provide the best procedure for actually calculating the point E (or $-E$) naming that meridian. In the next section we'll see another way to find the meridian in question. For the moment though, we have established that each point $V \in S^3 - \{\pm I\}$ lies on a unique parallel $\mathscr{P}_L$ and a unique meridian $\mathscr{M}_E$, and that $\mathscr{M}_E \cap \mathscr{P}_L = \{V\}$. Thus each such V has a unique latitude $L \in [0, \pi]$ and longitude $E \in \mathscr{E}$, and furthermore, no other point in S^3 has the same latitude and longitude.

4. CONJUGATION IN S^3. S^3 is a much more interesting space than either S^1 or S^2, partly because it's of a higher dimension than either of those spaces, but mostly because of its group structure, which as we shall see, meshes quite nicely with the geometry of the space. (As we saw in Section 1, the circle S^1 is also a group, but its multiplication is commutative, making the overall structure of S^1 somewhat less interesting than that of S^3. And, perhaps surprisingly, S^2 does not admit *any* group structure analogous to the structures on S^1 and S^3.) In particular, we need to understand how conjugation in S^3 interacts with the coordinate system we constructed in the previous section. To this end, let $C \in S^3$, and let $C_*: S^3 \to S^3$ denote the map $V \mapsto CVC^{-1}$, which is called *conjugation by* C. Because $C \in S^3, C\overline{C}^T = I$, so C_* is also the map $V \mapsto CV\overline{C}^T$. For each $C \in S^3, C_*$ is a group isomorphism—it's easily seen to be a homomorphism and its inverse is $(C^{-1})_*$. But, even better, C_* is an *isometry* of S^3, it preserves the distance between points in S^3. That is, for any V and W in S^3, $d_{S^3}(C_*(V), C_*(W)) = d_{S^3}(V, W)$. (This easily verified equality is a consequence of the fact that conjugation does not change the trace of a matrix.) Because conjugation is an isometry and $\mathscr{P}_L = \{V \in S^3\colon d_{S^3}(I, V) = L\}$, C_* maps $\mathscr{P}_L$ isometrically onto itself for each $L \in [0, \pi]$ and each $C \in S^3$. In fact, the collection of parallels in S^3 is precisely the set of conjugacy classes in S^3. That is, V and W are conjugate in S^3 (meaning $W = C_*(V)$ for some $C \in S^3$) if and only if V and W lie in the same parallel, that is, if and only if V and W have the same latitude in S^3. Before establishing this fact, it is helpful to expand our terminology and notation. For $L \in [0, \pi]$, let D_L denote the point $\begin{pmatrix} e^{iL} & 0 \\ 0 & e^{-iL} \end{pmatrix} \in S^3$, and let $\mathscr{D} = \{D_L\colon L \in [0, \pi]\}$. Then $\mathscr{D}$ is also the meridian $\mathscr{M}_{D_{\pi/2}}$, and we'll refer to $\mathscr{D}$ as the *Prime Meridian* in S^3. (We use the letter '$\mathscr{D}$' because all the points D_L in $\mathscr{D}$ are diagonal matrices.) Since $D_{\pi/2}$ corresponds to $(0, 1, 0, 0)$ in $\mathbf{R}^4$, the positive y-axis passes through this point, and thus $\mathscr{D}$ is contained in the unit circle $S^1 \subseteq \mathbf{R}^2$. In fact, $\mathscr{D}$ is precisely the unit semi-circle in the closed upper half-plane of $\mathbf{R}^2$. It is worth noting that, for each $L \in [0, \pi]$, $\mathscr{D}$ intersects the parallel $\mathscr{P}_L$ at the single point D_L. In particular, the latitude of D_L is L.

We claim that $\mathscr{P}_L$ is precisely the conjugacy class of D_L in S^3. We already know that every conjugate of D_L lies in $\mathscr{P}_L$. What we need to show is that each point V in $\mathscr{P}_L$ is conjugate to the diagonal matrix D_L. This can be done as follows: First off, the result is clearly true for $L = 0$ and $L = \pi$, so we'll assume that $L \in (0, \pi)$. Since V is a 2×2 matrix, the characteristic polynomial of V is $\lambda^2 - \operatorname{tr}(V)\lambda + \det(V)$. Because $V \in \mathscr{P}_L \subseteq S^3$, this becomes $\lambda^2 - (2\cos(L))\lambda + 1$, and the quadratic formula yields two distinct eigenvalues, $\lambda_1 = \cos(L) + i\sin(L) = e^{iL}$ and $\lambda_2 = \bar{\lambda}_1 = \cos(L) - i\sin(L) = e^{-iL}$. Equip the complex vector space $\mathbf{C}^2$ with its usual hermitian inner product, and choose a unit eigenvector (z_1, z_2) for V corresponding to the eigenvalue λ_1. Let $C = \begin{pmatrix} z_1 & -\bar{z}_2 \\ z_2 & \bar{z}_1 \end{pmatrix}$. Then $\det(C) = 1$ because (z_1, z_2) is a unit vector in $\mathbf{C}^2$, so $C \in S^3$. Since

$$(C^{-1}VC)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = (C^{-1}V)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = (\lambda_1 C^{-1})\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \lambda_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix},$$

$C^{-1}VC = \begin{pmatrix} \lambda_1 & \alpha \\ 0 & \beta \end{pmatrix}$, for some $\alpha, \beta \in \mathbf{C}$. But $C^{-1}VC \in S^3$, so we must have $\alpha = 0$ and $\beta = \bar{\lambda}_1$. That is, $C^{-1}VC = D_L$, or equivalently, $V = CD_LC^{-1}$, establishing the claim. (It is worth remarking that the point C in S^3 conjugating D_L to V is not unique. In fact, the set of points in S^3 which conjugate D_L to V is precisely CS^1, the coset containing C of the subgroup $S^1 \subseteq S^3$. Continued exploration along these lines leads to alternative measures of longitude in S^3, and to the intriguing equation $\mathscr{E} = S^2 = S^3/S^1$.)

Having just examined the effect of conjugation on parallels in S^3, we now turn our attention to meridians. Here there is less work to do. The basic fact is: For each $C \in S^3$ and $E \in \mathscr{E}$, $C_*(\mathscr{M}_E) = \mathscr{M}_{C_*(E)}$. That is, conjugation by C carries the meridian at longitude E isometrically onto the meridian at longitude $C_*(E)$. Thus, for each $C \in S^3$, C_* can be thought of as a "rigid rotation" of S^3 about its polar axis. See Figure 2 for a schematic picture. Putting things together, what we have

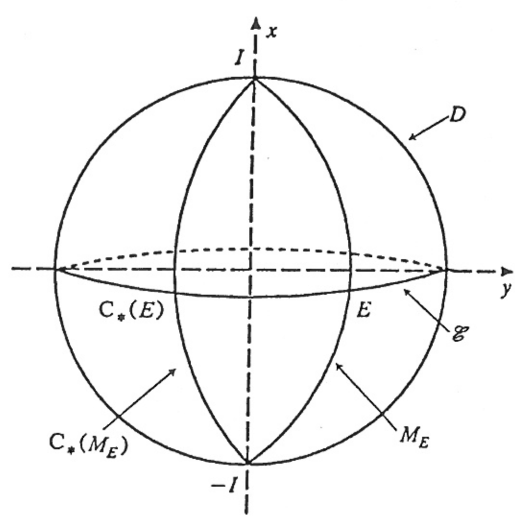

Figure 2

shown is that, for each $C \in S^3$, conjugation by C carries the point at latitude-longitude (L, E) to the point at latitude-longitude $(L, C_*(E))$.

Using conjugation we can now outline a second approach to finding the longitude of an arbitrary point $V \in S^3 - \{\pm I\}$. Let L denote the latitude of V. As was shown above, there is a C in S^3 for which $C_*(D_L) = V$. Let $E = C_*(D_{\pi/2})$. Then E has the same longitude as V, since D_L has the same longitude as $D_{\pi/2}$ and conjugation takes meridians to meridians. Furthermore, $E \in \mathscr{E}$, since $D_{\pi/2} \in \mathscr{E}$ and conjugation preserves latitudes. Thus the longitude of V is E. For an explicit formula, let $(z_1, z_2) \in \mathbf{C}^2$ be a unit eigenvector for V with eigenvalue e^{iL}. Then

$$E = \begin{pmatrix} i(z_1\bar{z}_1 - z_2\bar{z}_2) & 2iz_1\bar{z}_2 \\ 2i\bar{z}_1 z_2 & -i(z_1\bar{z}_1 - z_2\bar{z}_2) \end{pmatrix}.$$

5. THE EXPONENTIAL MAP. Let us return for a moment to the circle $S^1 \subseteq \mathbf{C} = \mathbf{R}^2$. By the *tangent space to S^1 at* 1, denoted $T_1(S^1)$, we mean the set of all vectors in $\mathbf{R}^2$ that are tangent to S^1 at 1. Although it is illustrative to draw vectors in $T_1(S^1)$ as tangent to S^1 at 1, we will consider tangent vectors as actually emanating from the origin in $\mathbf{R}^2$. Thus each vector in $T_1(S^1)$ corresponds to a vector lying in the y-axis, or, in the language of complex numbers, to an imaginary number iy. See Figure 3. Via this identification, $T_1(S^1)$ becomes a one-dimensional subspace of $\mathbf{R}^2$, and we can use the identification to define a map from $T_1(S^1)$ onto S^1. Namely, given iy representing a tangent vector to S^1 at 1, map iy to $e^{iy} \in S^1$. This aptly named *exponential map* $T_1(S^1) \to S^1$ takes the zero vector in $T_1(S^1)$ to $1 \in S^1$, takes the upward pointing tangent vector of length $\pi/2$ to $i \in S^1$, and takes the downward pointing tangent vector of length $\pi/2$ to $-i \in S^1$. Both tangent vectors of length π in $T_1(S^1)$ are mapped to $-1 \in S^1$. Putting it somewhat informally, the exponential map "wraps the tangent space $T_1(S_1)$ around S^1." In particular, it maps the interval $[-\pi i, \pi i] \subseteq T_1(S^1)$ onto S^1, and the mapping is one-to-one when restricted to $(-\pi i, \pi i)$. This map $T_1(S^1) \to S^1$ is a very special case of a

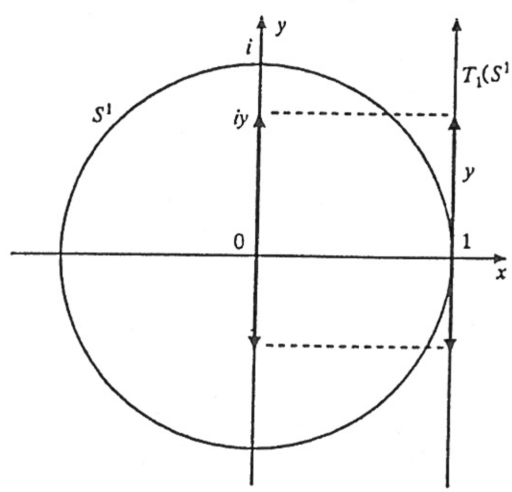

Figure 3

standard construction in differential geometry. In fact, it is possible to define an exponential map $T_p(M) \to M$ at any point p on any smooth manifold M. (Try envisioning what the exponential map would look like on the 2-sphere and on the torus.)

Now let's return to S^3. Since S^3 is a three-dimensional manifold, $T_I(S^3)$ is isomorphic to $\mathbf{R}^3$. That is, there are three linearly independent directions for tangent vectors to S^3 at I. As before we can (try to) picture vectors in $T_I(S^3)$ as being tangent to S^3 at I, but in actuality we'll treat these vectors as emanating from the origin O in $\mathbf{R}^4$. Since vectors tangent to S^3 at I are necessarily orthogonal to the vector from O to I, we can think of $T_I(S^3)$ as the collection of all vectors in $\mathbf{R}^4$ perpendicular to the x-axis, since this is the axis containing $I = (1, 0, 0, 0)$. Thus $T_I(S^3)$ can be identified with the three-dimensional subspace of $\mathbf{R}^4$ spanned by the y, z and t axes. In the language of quaternions, $T_I(S^3)$ corresponds to the pure imaginary quaternions, those quaternions of the form $iy + jz + kt$. This model allows a simple identification of $T_I(S^3)$ with our familiar $\mathbf{R}^3$—simply view the unit quaternions i, j and k as the standard unit vectors $\vec{i}$, $\vec{j}$ and $\vec{k}$ in $\mathbf{R}^3$. (Via this identification with $\mathbf{R}^3$, we see that $T_I(S^3)$ is more than just a real three-dimensional vector space, it's a "Lie algebra" over $\mathbf{R}$, where the "bracket product" of tangent vectors is just the classical cross product in $\mathbf{R}^3$.) In our matrix model of $\mathbf{R}^4$, $T_I(S^3) = \{\begin{pmatrix} iy & v_2 \\ -\bar{v}_2 & -iy \end{pmatrix} : y \in \mathbf{R},\ v_2 \in \mathbf{C}\}$, where we no longer require that the matrices have determinant 1, since tangent vectors need not be unit vectors.

Using this matrix formulation of $T_I(S^3)$, it is easy to define an exponential map $T_I(S^3) \to S^3$. It's just the map $V \mapsto \exp(V)$, where $\exp(V)$ denotes the matrix exponential of V, the same matrix exponential that is commonly introduced in linear algebra to study systems of linear differential equations. In fact, for a tangent vector $V \neq 0$ in the closed ball of radius π about the origin in $T_I(S^3)$, we claim that $\exp(V)$ is the point in S^3 at latitude $|V|$ and longitude $V/|V| \in \mathscr{E}$. (Of course, if $V = 0$ then $\exp(V) = I$.) To verify this claim, let V be such a tangent vector, and let E denote the point $V/|V| \in \mathscr{E}$. Then $\exp(V) = \exp(|V|E) = \exp(|V|CD_{\pi/2}C^{-1})$ (since all points in $\mathscr{E}$ are conjugate to $D_{\pi/2}$) $= \exp(C(|V|D_{\pi/2})C^{-1}) = C\exp(|V|D_{\pi/2})C^{-1} = CD_{|V|}C^{-1}$, where the final two equalities reflect basic properties of the matrix exponential. All that remains is to identify the latitude and longitude of $CD_{|V|}C^{-1}$, which, for notational convenience, we'll denote X. But this is easy. Since the latitude of $D_{|V|}$ is $|V|$, so is the latitude of its conjugate X. And since C_* carries $D_{\pi/2}$ to E, it must carry $D_{|V|}$ to a point in $\mathscr{M}_E$, so that the longitude of X is E. This establishes the claim made above. What it tells us is that the matrix exponential $T_I(S^3) \to S^3$ maps the closed ball of radius π about the origin in $T_I(S^3)$ onto S^3, and that the restriction of this map to the interior of that ball is one-to-one. Indeed, radii of the ball are carried isometrically onto meridians in S^3, while concentric spheres (centered at the origin) within the ball are mapped diffeomorphically onto the parallels in S^3. So, in some sense, the exponential map $T_I(S^3) \to S^3$ "wraps $T_I(S^3)$ around S^3," just as the map $iy \mapsto e^{iy}$ did in the one-dimensional case. (In fact, the exponential map $T_I(S^3) \to S^3$ is an extension of the exponential map $T_I(S^1) \to S^1$.) The inverse of the exponential map carries $S^3 - \{-I\}$ diffeomorphically onto the open ball of radius π in $T_I(S^3)$. Thus it maps $S^3 - \{-I\}$, which lies in four-dimensional space, onto an open ball in $\mathbf{R}^3$. This is analogous to mapping the surface of the Earth, which lies in three-dimensional space, onto the two-dimensional page of an atlas, via equidistant polar projection. So the exponential map provides another approach to "charting" the 3-sphere.

6. CONCLUDING REMARKS. As I remarked in the introduction, there's still much more we could do. Discussions about Lie groups, group actions and fiber bundles follow naturally from the material presented in this paper, as does a more in-depth treatment of the geometry of S^3. But for this exposition, I think that now is the right time to say 'Enough!'. I hope that this article has provided you with a better understanding (and a usable model) of the 3-sphere, and that you are inspired to pursue further some of the ideas introduced above.

REFERENCES

1. S. Akbulut and J. McCarthy, *Casson's Invariant for Oriented Homology 3-Spheres An Exposition*, Mathematical Notes 36, Princeton University Press, Princeton, 1990.
2. M. Artin, *Algebra*, Prentice Hall, Englewood Cliffs, New Jersey, 1991.
3. C. Chevalley, *Theory of Lie Groups, I*, Princeton University Press, Princeton, 1946.
4. A. Knapp, *Lie Groups, Lie Algebras, and Cohomology*, Mathematical Notes 34, Princeton University Press, Princeton, 1988.
5. S. MacLane and G. Birkhoff, *Algebra*, 2nd ed., Collier Macmillan, New York, 1979.

Department of Mathematics
Rice University
P.O. Box 1892
Houston, TX 77251
zulli@math.rice.edu

A Ph.D. dissertation is a paper of the professor written under aggravating circumstances.

Attributed by G. Pòlya to Adolph Hurwitz, as quoted in Howard W. Eves, *Mathematical Circles Revisited*, Prindle, Weber and Schmidt, Boston, 1971.

Answer to Picture Puzzle

(p. 213)

Raoul Bott