Understanding quaternions and the Dirac belt trick

Mark Staley

University of Ontario Institute of Technology, 2000 Simcoe Street North, Oshawa, Ontario, Canada

E-mail: staleymd@gmail.com

Received 3 January 2010, in final form 14 February 2010
Published 10 March 2010
Online at stacks.iop.org/EJP/31/467

Abstract
The Dirac belt trick is often employed in physics classrooms to show that a $2\pi$ rotation is not topologically equivalent to the absence of rotation whereas a $4\pi$ rotation is, mirroring a key property of quaternions and their isomorphic cousins, spinors. The belt trick can leave the student wondering if a real understanding of quaternions and spinors has been achieved, or if the trick is just an amusing analogy. The goal of this paper is to demystify the belt trick and to show that it suggests an underlying four-dimensional parameter space for rotations that is simply connected. An investigation into the geometry of this four-dimensional space leads directly to the system of quaternions, and to an interpretation of three-dimensional vectors as the generators of rotations in this larger four-dimensional world. The paper also shows why quaternions are the natural extension of complex numbers to four dimensions. The level of the paper is suitable for undergraduate students of physics.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Every student learns that complex numbers are necessary to solve certain algebraic equations such as $x^2 + 1 = 0$. The student also learns that complex numbers may be represented as two-dimensional vectors lying in the so-called Argand plane, with the x-axis representing the real numbers and the y-axis representing the pure imaginary numbers. The geometric interpretation of complex numbers is supported by the observation that the addition of two complex numbers represents the addition of two vectors, and the product of two unit-length complex numbers represents a sequence of two rotations.

Upon seeing the connection between complex numbers and two-dimensional geometry, the curious mind is apt to wonder whether there is an extension applicable to three-dimensional geometry. During the 19th century, this germ of an idea was lodged in the minds of many
people including William Rowan Hamilton. In 1843 Hamilton discovered an algebra of three-dimensional rotations that was based on a new set of objects he called *quaternions* (van der Waerden 1976) provides a good history). So impressed was he with these new objects that he spent the remaining 22 years of his life working out their properties (Crowe 1967). After Hamilton died his quaternions were mostly abandoned, leading the biographer E T Bell to suggest that the great man had been the victim of a mono-maniacal delusion (Bell 1937).

Developments of the 20th century have largely vindicated Hamilton. Firstly, historians have shown that Gibbs and Heaviside were heavily influenced by Hamilton’s work when they developed modern vector analysis (Stephenson 1966, Crowe 1967, Silva and Martins 2002). Secondly, quaternions are now widely used in the computer graphics and aerospace industries to reduce the computational costs associated with performing rotations on vectors (Kuipers 1999). And thirdly *spinors*, those fundamental building blocks of modern particle physics, are isomorphic to quaternions (Kronsbein 1967). On this third point it is interesting to note that Hamilton’s 19th century belief that quaternions held the key to understanding the universe has its modern parallel in the view of some physicists that spinors are more fundamental than space-time vectors (Bohm 1971, Penrose and Rindler 1986).

While spinors and quaternions are undoubtedly important and useful, their geometrical properties can be difficult to understand. Take for example the Dirac belt trick. The trick is motivated by the observation that when a spinor (such as an electron) is rotated by $2\pi$ its quantum mechanical wavefunction reverses sign, which has observable implications (Silverman 1980). A second rotation of $2\pi$ restores the wavefunction back to its original form. Dirac came up with the belt trick as a way to demonstrate this effect using an everyday three-dimensional object, and it is often used in undergraduate physics classrooms. The trick shows that a $2\pi$ rotation is not equivalent to no rotation although a $4\pi$ rotation is. A belt is held fixed at one end while the other end is rotated (twisted) through an angle of $4\pi$ about an axis parallel to its length. It is shown that the belt can be untwisted without any further rotation by simply passing one end of the belt under the rest of the belt (see section 3). Hence, a $4\pi$ rotation is topologically equivalent to no rotation.

The belt trick can leave the student wondering if a real understanding of quaternions and spinors has been achieved, or if the trick is just an amusing analogy. And deeper study of quaternions and spinors leads to further questions, such as why do they have four components (or two complex components), whereas ordinary spatial geometry gets by with three? Or why do rotations in three-dimensional space become similarity transformations when expressed in the language of quaternions? There are several approaches to answering these questions, the most common of which is based on group theory. One can define spinor transformations using the $SU(2)$ group and show that these are isomorphic to (and form a double cover of) $SO(3)$, the three-dimensional rotation group (see, for example, Joshi (1982)). Other approaches use geometrical pictures to aid in understanding. For example there is an approach based on projective geometry (Bohm 1971, Frescura and Hiley 1981), and another approach based on the ‘vector plus flag picture’ (Misner et al 1973).

This paper presents a geometrical approach to understanding complex numbers and quaternions that is based on equating their elements with generators of rotations. The unit imaginary ‘$i$’ in complex number theory can be interpreted as the generator of rotations in two dimensions. In four dimensions there are six generators of rotations, and these can be decomposed into two sets of generators each consisting of three elements corresponding to the hyper-complex numbers in Hamilton’s system of quaternions. This paper will show that each of these triplets can be interpreted as the basis of a coordinate system in three dimensions,
similar to the way that a point in the Argand plane can be used to represent a vector in two
dimensions. One of these triplets forms the basis of a right-handed coordinate system while
the other forms the basis of a left-handed coordinate system.

The paper is organized as follows. Section 2 presents an interpretation of complex numbers
in terms of two-dimensional rotations that allows for an extension to higher dimensions.
Section 3 discusses three-dimensional geometry and the Dirac belt trick. The purpose of
this section is to demystify the belt trick and to motivate the development of quaternions.
Section 4 describes the algebraic and geometrical properties of quaternions and shows how
ordinary three-dimensional geometry is reclaimed. Finally, section 5 concludes.

2. Two-dimensional geometry

Consider the rotation of a vector in two dimensions:

$$v' = Rv,$$

where

$$R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad v = \begin{pmatrix} x \\ y \end{pmatrix}.$$ 

Now rewrite $R$ as

$$R = I \cos \theta + i \sin \theta,$$

where

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$ 

The following multiplication table may be used to combine rotation operations:

$$\begin{array}{c|cc}
I & i & -I \\
i & i & -I \\
I & -I & I \\
\end{array}$$

(1)

The above algebra is isomorphic to the system of complex numbers, with $I$ standing for 1 and
$i$ standing for the unit imaginary $\sqrt{-1}$. Hence, rotations can be represented using complex
numbers.

Recall that a rotation can be constructed out of a large sequence of infinitesimal rotations
using a so-called generator. We can now show that the symbol $i$ is the generator of rotations
in two dimensions. Let us say we want to apply two consecutive rotations to a vector. Call the
first rotation $R_1 = \cos \theta_1 I + \sin \theta_1 i$ and the second $R_2 = \cos \theta_2 I + \sin \theta_2 i$. Using table (1) and
some simple trigonometric identities we can show that $R_1 R_2 = \cos(\theta_1 + \theta_2) I + \sin(\theta_1 + \theta_2)i$.

So to combine rotations we simply add the angles together. By extension, a rotation by a
given angle $\theta$ is equivalent to a sequence of $N$ rotations each of angle $\theta/N$. Taking the limit
as $N \to \infty$ we have

$$R = I \cos \theta + i \sin \theta = \lim_{N \to \infty} \left( I + i \frac{\theta}{N} \right)^N,$$

(2)

where we have used the property that $\cos(\alpha) \to 1$ as $\alpha \to 0$ and $\sin(\alpha) \to \alpha$ as $\alpha \to 0$.

The quantity inside the parentheses on the right-hand side of equation (2) is of the form of an
infinitesimal rotation, and the generator is $i$.

Let us now investigate the parameter space of two-dimensional rotations. Figure 1 shows
one way to parameterize rotations using a single dimension corresponding to the angle $\theta$. We
may note that the angles $\pi$ and $-\pi$ are equivalent in this representation. Figure 2 shows how
to construct an alternative parameter space by increasing the number of dimensions to 2 using \( I \) and \( i \). The coordinates of a point representing a rotation are constrained to lie on the unit circle.

We can introduce scaling operations into our algebra by admitting complex numbers having non-unit magnitudes. A general operation consisting of a rotation and a scaling operation can be written as

\[
R = r (I \cos \theta + i \sin \theta), \quad r > 0,
\]

which can be represented in figure 2 by a point lying anywhere in the plane (i.e., it is no longer constrained to lie on the unit circle). We can interpret the plane in figure 2 as the Argand plane.

A point in the Argand plane can be used to represent a rotation/scaling operation, but it can also be used to represent a two-dimensional vector. This duality can be made clearer by noting that the unit basis vectors (in column format) are related to \( I \) and \( i \) as follows:

\[
\begin{pmatrix} 1 \\ 0 \end{pmatrix} = I \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} = i \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]

which means that an arbitrary vector \((a, b)\) (here in row format) can be obtained by rotating and scaling the unit vector \((1, 0)\) using \(aI + bi\). This implies that we can rotate \((a, b)\) by the angle \(\theta\) and scale it by a factor of \(r\) by simply multiplying \(aI + bi\) by \(r(\cos \theta I + \sin \theta i)\) and reading off the components of the resulting complex number. Hence, we can dispense with vectors altogether and use complex numbers to represent both vectors and rotation/scaling operations. Note that for pure rotations the parameter ‘\(r\)’ is equal to 1. We will see later that a similar formalism holds for quaternions.

The goal of the following two sections will be to extend the above reasoning to three and four dimensions. In doing so we will endeavor to create something like a multi-dimensional
extension of the Argand plane. That is, we will wish to build a parameter space that can act as a Euclidean vector space. One useful property of Euclidean spaces that can be used to test if a parameter space is admissible for such purposes is that it be *simply connected*. A space is simply connected if one can continuously deform any path connecting two end points into any other path connecting the same two end points without moving the end points themselves. A special case is where the starting and ending points are the same, in which case the path is a loop. If one can shrink the loop down to a point without moving the starting/stopping points, then the space is simply connected. The Argand plane is an example of a simply connected space.

3. Three-dimensional rotations

The rotation of a vector $v$ in three dimensions about an axis of rotation $u$ (of unit norm) by an angle $\theta$ is given by

$$v' = v_\parallel + v_\perp \cos \theta + (u \times v_\perp) \sin \theta$$

where $v_\parallel = u(u \cdot v)$ is the component of $v$ in the direction of $u$, and $v_\perp = v - u(u \cdot v)$ is the component of $v$ perpendicular to $u$ (i.e. in the plane of rotation).

Consider the parameter space for three-dimensional rotations. Any given rotation can be represented by a unit vector $u$ and an angle $\theta$ which takes on values between $-\pi$ and $\pi$. Multiplying the set of unit vectors $\{u\}$ by the parameter $\theta$ we have a solid sphere of radius $\pi$, which represents the parameter space of rotations. This parameter space is similar to that shown in figure 1 in the sense that a rotation about any axis $u$ through an angle of $\pi$ is equivalent to a rotation about that same axis through an angle of $-\pi$. Recall that in the case of two-dimensional rotations it was possible to extend the dimensionality of the parameter space from 1 to 2, and so construct an alternative parameter space in two dimensions that could be mapped in a one-to-one manner with the underlying vector space (see figure 2). In the case of three-dimensional rotations our parameter space is already of dimension 3, so any attempt to create a similar ‘Argand-space’ would require an increase in the number of dimension beyond 3. Before we attempt to go down that road, let us for the moment confine ourselves to three dimensions and continue to explore the properties of the parameter space of rotations as envisioned so far.

To simplify the discussion even further, let us restrict the dimensionality of the parameter space to 2, in which case the parameter space is a *disc* of radius $\pi$, representing rotations about two axes. The centre of the disc corresponds to no rotation and is the starting point for any sequence of rotations. A rotation of angle $\theta$ about an axis $u$ can be represented in the disc by drawing a line segment of length $\theta$ extending from the centre of the disc in the direction $u$. A rotation of angle $\pi$ is represented by a line segment extending from the centre of the disc to the perimeter of the disc (see the examples in figure 3).

Another way to study the parameter space of rotations about two axes is to use a belt. The twists of a belt can serve to keep track of the path followed by a series of rotations. To see how this is so, lay a belt flat on a table. You may verify that it is possible to bend or twist the belt about any axis in the plane of the table, but it is impossible to bend it about the vertical axis without ripping it apart2. Hence a belt can be twisted about two axes only. Now hold one end of the belt with your left hand (the end with no buckle) and twist the other end (the end with the buckle) about an axis parallel to the length of the belt. You will see that the various orientations of the buckle are recorded as varying amounts of twist at different points.

2 In reality you can deform a belt about the vertical axis, but you will see that this deformation is accomplished by a succession of deformations about axes lying in the plane of the table.
along the length of the belt. If you twist the belt by $\pi$ radians, you will see something like the picture in figure 3(a).

So we now have two ways of recording the path in parameter space taken by any sequence of rotations about two axes. Figure 3 illustrates the correspondence between various twists of the belt and points in the parameter disc. In figure 3(a) the belt is twisted by an angle $\pi$ about the axis parallel to the length of the belt. In this case the axis of rotation is in a direction from left to right. This is represented in the disc as a line segment of length $\pi$ extending left to right from the centre of the disc to the perimeter of the disc. In figure 3(b) the belt is twisted by an angle $\pi$ about an axis in the plane of the table that is perpendicular to the length of the belt. In that case the belt is coiled by half a revolution. This is represented in the disc as a line segment of length $\pi$ extending from the centre of the disc up to the top of the perimeter of the disc.

Figure 4 depicts a series of $2\pi$ rotations. Recall that a rotation of $\pi$ radians about some axis is equivalent to a rotation of $-\pi$ radians around the same axis, which means that any two points located on opposite sides of the parameter disc are equivalent. The way to represent a $2\pi$ rotation using the parameter disc is to first draw a line from the centre of the disc to a point on the perimeter of the disc, then jump to the opposite end of the disc and continue drawing the line back to the centre of the disc. For example, figure 4(a) shows how a $2\pi$ rotation about the axis parallel to the length of the belt is represented in the disc. Note that the starting and ending points of the parameter path are both represented as points located at the centre of the disc, and these correspond to the starting and ending orientations of the belt buckle.

As shown by the discs in figure 4, parameter paths corresponding to $2\pi$ rotations can be deformed into one another by moving the perimeter points such that they are always located
Deformation of $2\pi$ rotations. Parameter paths in the disc corresponding to $2\pi$ rotations can be deformed into one another by moving the perimeter points such that they are always located at opposite ends of the disc. There is no need to move the starting and ending points at the centre of the disc. These deformations correspond to the movements of the belt in which the buckle is kept in a fixed orientation. In the above sequence, a $2\pi$ rotation is deformed into a $-2\pi$ rotation about the same axis. The transition from (b) to (c) requires that the buckle be passed through the centre of the coil.

Figure 4. Deformation of $2\pi$ rotations. Parameter paths in the disc corresponding to $2\pi$ rotations can be deformed into one another by moving the perimeter points such that they are always located at opposite ends of the disc. There is no need to move the starting and ending points at the centre of the disc. These deformations correspond to the movements of the belt in which the buckle is kept in a fixed orientation. For example, to change the orientation from that

at opposite ends of the disc. There is no need to move the starting and ending points at the centre of the disc. These deformations correspond to movements of the belt in which the buckle is kept in a fixed orientation. For example, to change the orientation from that
shown in figure 4(a) to that of 4(b), one need only move the belt buckle towards the left. The twist then changes into a coil. The transition from 4(b) to 4(c) requires that the belt buckle be passed under the belt, i.e. through the centre of the coil. The final move consists of pulling the belt buckle back towards the right. This sequence of moves shows that a rotation of $2\pi$ about one axis can be continuously deformed into a rotation of $2\pi$ about another axis, or into a rotation of $-2\pi$ about the original axis, without changing the end points.

An important observation is that it is not possible to deform a belt that has been twisted by $2\pi$ into a flat belt without changing the orientation of the buckle. Similarly, it is not possible to deform any of the paths shown in the discs in figure 4 into a single point at the centre of the disc. This means that our parameter space for rotations is not simply connected. Recalling the discussion at the end of section 2, our parameter space does not appear to be a good candidate for representing three-dimensional Euclidean vectors. We cannot construct a three-dimensional ‘Argand-space’ out of our disc.

We are now in a position to see why a rotation by $4\pi$ is in some sense equivalent to no rotation. Twist the belt by $4\pi$ about the axis parallel to the belt. Then, keeping the first $2\pi$ twist intact, deform the second $2\pi$ twist as shown in figure 4. The result is a $2\pi$ twist followed by a $-2\pi$ twist, which cancels the first. Normally this belt trick is done by passing the entire twisted belt under itself, but some experimentation shows that such a maneuver is equivalent to the one just described.

The lesson from the above exercise is that if we wish to build a simply connected parameter space for rotations, we should keep in mind that from any given starting point all paths corresponding to a $2\pi$ rotation should correspond to the same end point. And a rotation of $4\pi$ is equivalent to no rotation, and so can be thought of as a closed loop in a simply connected parameter space. There remains the question of whether a $2\pi$ rotation is equivalent to no rotation. Our everyday experience tells us that the answer to that question is yes, but there remains the possibility that we will have to give up that notion when constructing a simply connected parameter space. If that is the case, then a $2\pi$ rotation and a zero rotation will be located at different points within the parameter space, and the mapping between points of the parameter space and points in the underlying Euclidean vector space will not be as straightforward as was the case in two dimensions.

Recalling our earlier discussion suggesting the need to expand the number of dimensions, we are immediately led to wonder if a sphere could be used to represent rotations about two axes. Taking the Earth as an example, we could use the north pole as the starting point for rotations, with the direction of each line of longitude representing the direction of the axis of rotation. The lines of latitude would represent the angles of rotation. A $2\pi$ rotation might correspond to a line of longitude circling the Earth and ending back at the north pole. One can immediately see that this scheme does not work. The lines of longitude all cross each other at the south pole, implying that all rotations by $\pi$ are equivalent to one another, which is obviously not true. However, a small modification leads to a workable parameter space. All one needs to do is to equate the south pole with $2\pi$ rotations. This may be accomplished by setting the actual angle of rotation to be twice the angle depicted by the line of latitude. Now all $2\pi$ rotations are represented by the same point at the south pole, all paths connecting the north and south poles are continuously deformable into one another (the parameter space is simply connected) and a $4\pi$ rotation (which takes us back to the north pole) is equivalent to no rotation. Note that a circle around the globe can be shrunk down to a point. Extending this geometrical insight, we should be able to parameterize three-dimensional rotations in a simply connected manner using a four-dimensional hyper-sphere. So let us turn to a study of four-dimensional Euclidean geometry.
4. Quaternions

Generalizing from the two-dimensional geometry described in section 2, there are six planes of rotation in four-dimensional space and hence six generators of rotations:

\[ T_{12} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad T_{23} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \]

\[ T_{13} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad T_{24} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \]

\[ T_{14} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad T_{34} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \]

As they stand these generators do not form a closed algebra. But we may combine them to form two sets of generators, \([i, j, k]\) and \([l, m, n]\), each of which forms a closed algebra as follows:

\[ i = T_{12} + T_{34} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \]

\[ j = T_{13} - T_{24} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \]

\[ k = T_{23} + T_{14} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \]

and

\[ l = T_{12} - T_{34} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \]

\[ m = T_{13} + T_{24} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \]
\[ n = T_{23} - T_{14} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}. \]

If we add the identity matrix \( I \) to our list we have the following multiplication tables:

\[
\begin{array}{cccc}
I & i & j & k \\
i & I & j & k \\
j & j & I & -I \\
k & k & -j & I \\
\end{array}
\]

\[
\begin{array}{cccc}
I & l & m & n \\
l & I & m & -n \\
m & m & I & -l \\
-l & n & m & I \\
\end{array}
\]

Note that each element of the set \( \{i, j, k\} \) commutes with each element of \( \{l, m, n\} \), e.g. \( il = li \) etc. The left table above can be summarized as

\[ i^2 = j^2 = k^2 = ijk = -I, \]

which is isomorphic to Hamilton’s famous quaternion equation, with \( I \) standing for 1 and \( \{i, j, k\} \) standing for Hamilton’s triplet of imaginary numbers. A similar quaternion equation holds for \( \{l, m, n\} \).

Consider now an arbitrary quaternion \( R = wI + xi + yj + zk \). This can always be written in the form

\[ R = r \left[ \cos \theta I + \sin \theta (u_i i + u_j j + u_k k) \right], \quad r > 0, \quad u_x^2 + u_y^2 + u_z^2 = 1. \]

The quantity in the curly bracket can be rewritten as

\[ \cos \theta I + \sin \theta (u_i i + u_j j + u_k k) = \lim_{N \to \infty} \left[ I + \frac{\theta}{N} (u_i i + u_j j + u_k k) \right]^N, \]

which is analogous to equation (2) and can be proved in a similar manner. So we may say that \( \{i, j, k\} \) are the generators of rotations in our four-dimensional vector space.

The mapping between quaternions and vectors in four dimensions can be demonstrated using the following basis vectors:

\[
\begin{array}{ccc}
I = I & i = i & j = j \\
l = l & m = m & n = n \\
\end{array}
\]

\[
\begin{array}{ccc}
\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\
\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \\
\end{array}
\]

\[ ^3 \text{Another representation of quaternions is found in the triplet \((-i\sigma_x, -i\sigma_y, -i\sigma_z)\) where } \sigma \text{ are the Pauli spin matrices. Using the } 2 \times 2 \text{ representation of } i \text{ from section 2, this representation is the same as } (n, m, -l). \text{ For further discussion of the relationship between quaternions and spinors, see Kronsbein (1967).} \]
Given this duality of four-dimensional vectors and quaternions one may consider 
\[(w, x, y, z)\] as either representing a vector in four dimensions or representing a 
rotation and a scaling operation. This duality property was known to Hamilton and 
was emphasized by his successor Tait who considered it a key property of quaternions (Silva and 
Martins 2002). It is in the sense of this duality that one may consider quaternions to be the 
natural extension of complex numbers.

We now turn to the question of how a system of four-dimensional rotations can give 
rise to a system of three-dimensional geometry. Consider the transformation of a 
quaternion \(Q\) by some other quaternion \(R\):

\[\Psi = Q \Phi.\]

Now imagine another unit quaternion \(R\) being applied to both \(\Phi\) and \(\Psi\):

\[\Phi' = R \Phi, \quad \Psi' = R \Psi.\]  \((6)\)

Since \(R\) is of unit magnitude \((r = 1\) in equation \((4)\)), it preserves the norms of \(\Phi\) and \(\Psi\). 
We want to enquire as to the new relationship between \(\Phi'\) and \(\Psi'\). That is, to find the new 
quaternion \(Q'\) such that

\[\Psi' = Q' \Phi'.\]  \((7)\)

Let \(R = aI + bi + cj + dk\) where \(a^2 + b^2 + c^2 + d^2 = 1\), and define \(R^\dagger \equiv aI - bi - cj - dk\). The 
following property is readily verified:

\[R^\dagger R = RR^\dagger = I.\]

It then follows from \((6)\) and \((7)\) that

\[Q' = RQR^\dagger\]

which is a similarity transformation.

Defining \(Q = wI + xi + yj + zk\), \(Q' = w' I + x'i + y'j + z'k\), \(\psi = (x, y, z)\), \(\psi' = (x', y', z')\) and 
using the definition of \(R\) in equation \((4)\) with \(r = 1\), one may show that

\[w' = w\]

and

\[v' = v_1 + v_\perp \cos 2\theta + (u \times v_\perp) \sin 2\theta,\]  \((8)\)

where \(v_1 = u(u \cdot v)\) and \(v_\perp = v - u(u \cdot v)\). This is none other than equation \((3)\) but with 
twice the angle!

And so we see how the geometry of three dimensions is reclaimed. The portion of 
the quaternion \(xi + yj + zk\) (called a 'pure quaternion' by Hamilton) is transformed under 
rotations exactly like an ordinary three-dimensional polar vector, while the 'scalar' part of 
the quaternion \(wI\) remains unchanged. The factor of 2 appearing with the angle \(\theta\) is consistent 
with the image of the parameter space hypothesized in the previous section. The angle \(\theta\) in 
equation \((4)\) can be interpreted as the line of latitude circling the four-dimensional hypersphere 
(of radius \(r\)), and the unit vector \((u_x, u_y, u_z)\) can be interpreted as the direction of longitude 
away from the north pole of the hypersphere. Note that since the elements of \([i, j, k]\) commute 
with the elements of \([l, m, n]\), quaternions built out of \([i, j, k]\) are invariant with respect to 
similarity transformations based on \([l, m, n]\), and vice versa. So there are actually two separate 
three-dimensional worlds within the four-dimensional system that we have constructed. The 
relationship between these two three-dimensional worlds is considered next.

A pure quaternion \(xi + yj + zk\) behaves like a polar vector under rotation but it does 
not behave like a polar vector under reflection. To illustrate, consider the special case of a
quaternion that is the product of two pure quaternions: \( Q = (xi + yj + zk)(ui + vj + wk) \). Consider the reflection of \( xi + yj + zk, ui + vj + wk \) and \( Q \) through a mirror oriented in the \( x-y \) plane. If \( xi + yj + zk \) acts like a polar vector, then upon reflection it should become \( xi + yj - zk \). Similarly for \( ui + vj + wk \). But the sign of the \( k \)th component of \( Q' = (xi + yj - zk)(ui + vj - wk) \) does not become flipped, so our treatment of reflections is not consistent. Silva and Martins (2002) avoid this inconsistency by treating pure quaternions as axial vectors. Equation (5) suggests another way to define reflections for quaternions. The unit vectors \([i, j, k]\) form the basis of a right-handed coordinate system while the unit vectors \([l, m, n]\) form the basis of a left-handed coordinate system (through reflection in the \( x-y \) plane). By extension, the mirror image of \([i, j, k]\) should be \([l, m, n]\), which leads to a consistent transformation of \( Q \). And so we see that the two three-dimensional worlds discussed in the previous paragraph are located on opposite sides of the looking glass.

5. Conclusion

The purpose of this paper has been to develop an understanding of complex numbers and quaternions by focusing on the properties of the parameter space of rotations and on the mapping between these parameter spaces and Euclidean vector spaces. In two dimensions, a complex number can represent a vector in the Argand plane, or it can represent a rotation. In four dimensions a similar duality gives rise to quaternions. One can reclaim three-dimensional geometry from quaternions by analysing the effects of similarity transformations. The six generators of rotations in four dimensions decompose into two sets of three-dimensional generators, and these sets are related to each other by a parity transformation.

Acknowledgments

The author would like to thank E Poisson, D Staley and two anonymous referees for helpful suggestions on improving the manuscript.

References

Bell E T 1937 Men of Mathematics (New York: Simon and Schuster)
Bohm D 1971 Space-time geometry as an abstraction from spinor ordering Perspectives in Quantum Theory ed W Yourgrau and A van der Merwe (Cambridge, MA: MIT Press)
Coddens G 2002 Spinor approach to the rotation and reflection groups Eur. J. Phys. 23 549–64
Crowe M 1967 A History of Vector Analysis (Notre Dame, IN: University of Notre Dame Press)
Frescura F and Hiley B 1981 Geometric interpretation of the Pauli spinor Am. J. Phys. 49 152–7
Girard P 1984 The quaternions group and modern physics Eur. J. Phys. 5 25–32
Joshi A W 1982 Elements of Group Theory for Physicists 3rd edn (New York: Wiley)
Kronberin J 1967 Kinematics-quaternions-spinors—and Pauli’s spin matrices Am. J. Phys. 35 335–42
Kuipers J 1999 Quaternions and Rotation Sequences (Princeton, NJ: Princeton University Press)
Misner C, Thorne K and Wheeler J 1973 Gravitation (San Francisco: Freeman)
Penrose R and Rindler W 1986 Spinors and Space-Time (Cambridge: Cambridge University Press)
Silva C and Martins R 2002 Polar and axial vectors versus quaternions Am. J. Phys. 70 958–63
Silverman M 1980 The curious problem of spinor rotation Eur. J. Phys. 1 116–22
Stephenson R J 1966 Development of vector analysis from quaternions Am. J. Phys. 34 194–201
van der Waerden B L 1976 Hamilton’s discovery of quaternions Math. Mag. 49 227–34