What is an Imaginary Number? The Plane and Beyond

Andrew W. Powell

Imperial College, London

Follow this and additional works at: https://scholarship.claremont.edu/jhm

Part of the Arts and Humanities Commons, and the Mathematics Commons

Recommended Citation
Andrew W. Powell, "What is an Imaginary Number? The Plane and Beyond," Journal of Humanistic Mathematics, Volume 14 Issue 2 (July 2024), pages 264-285. Available at: https://scholarship.claremont.edu/jhm/vol14/iss2/9

©2024 by the authors. This work is licensed under a Creative Commons License.
JHM is an open access bi-annual journal sponsored by the Claremont Center for the Mathematical Sciences and published by the Claremont Colleges Library | ISSN 2159-8118 | http://scholarship.claremont.edu/jhm/

The editorial staff of JHM works hard to make sure the scholarship disseminated in JHM is accurate and upholds professional ethical guidelines. However the views and opinions expressed in each published manuscript belong exclusively to the individual contributor(s). The publisher and the editors do not endorse or accept responsibility for them. See https://scholarship.claremont.edu/jhm/policies.html for more information.
What is an Imaginary Number? The Plane and Beyond

Cover Page Footnote
I would like to thank Arthur Paul Pedersen for his comments on a draft of the manuscript.

This work is available in Journal of Humanistic Mathematics: https://scholarship.claremont.edu/jhm/vol14/iss2/9
What is an Imaginary Number?  
The Plane and Beyond

Andrew Powell

*Imperial College, London, UK*
andrew.powell@imperial.ac.uk

---

**Synopsis**

In this article I argue that \( i \) is a quantity associated with the two-dimensional real number plane, whether as a vector, a bi-vector, a point or a transformation (rotation). This position provides a foundation for the complex numbers and accounts for complex numbers in some equations of applied mathematics and physics. I also argue that complex numbers are fundamentally geometrical and can be described by geometric algebra, and that moreover the meaning of complex numbers in physics varies with dimension and geometry of the manifold.

**Keywords:** complex numbers, geometric algebra, imaginary number, spacetime algebra, split complex numbers.

---

1. **Introduction**

This article is an attempt to explain what an imaginary number is, and *a fortiori* what complex numbers are. The emphasis here is mainly didactic, but there is some application to the philosophy of mathematics and physics because an imaginary number seems mysterious from an ontological and epistemological perspective, whether that is in its use in algebra or in differential equations. The primary aim is to explain what an imaginary number means when it appears in an equation.

The approach I take in this article can be traced back to Hermann Grassmann in [12] and William Kingdon Clifford in [4], but it was first articulated in a mature form by David Hestenes; see, for example, [17, 18, 19, 21].
Hestenes calls the approach geometric algebra.\footnote{The term “geometric algebra” was first used by Clifford in [4]. Mathematically an algebra is a set closed under specific computable operations, which is a narrower sense than is intended here.} The objective of geometric algebra is to find a symbolic way of reasoning about geometry that can be used as a natural language of physics.

There is a large literature on the applications of geometric algebra (see [23] for a recent survey and [8] for topic based studies), but in terms of generating understanding, there is an excellent introductory exposition in an overview of geometric algebra in [14], some informative notes in [3], two introductory text books covering algebra and calculus by Alan MacDonald [29, 30]; and for a systematic exposition of geometric algebra see [7]. In mathematics geometric algebra is often called Clifford algebra after Clifford (see [10, 36] for example), but Hestenes cautions against confusing geometric algebra used in physics with a Clifford algebra as a mathematical structure (see [22]).

In my opinion, complex analysis has a distinctive geometric character which is largely absent from real analysis, exemplified by Tristan Needham’s pioneering [33]. While it is possible to regard real analysis as a special case of differential geometry on a spatial manifold, it is rarely the case that real analysis is taught in that way. By the end of the article it should be clear that the appearance of complex analysis as distinctively geometrical is not mistaken (compare Emily Grosholtz’s [13] from a teaching of mathematics perspective).

The first four sections of this article are written for a general mathematical audience, although it would be helpful to have some knowledge of the complex numbers (including Argand diagrams and Euler’s formula), vector algebra (including the dot and cross products), linear algebra (matrices, eigenvalues and eigenvectors), and the Euclidean geometry of the plane (such as Pythagoras’s theorem and trigonometric functions). Section 5 onwards includes differential equations from physics (such as Schrödinger’s equation in quantum mechanics) and the Minkowski metric of spacetime, which requires university level physics or applied mathematics. That said, all of the physics needed for an appreciation of the later sections of this article is included in Leonard Susskind’s excellent courses entitled “The Theoretical Minimum” (videos are linked from [40]), which has an accompanying book series (see
What is an Imaginary Number? The Plane and Beyond

[41] for example). No background in geometric algebra is needed to read this article, which is the focus of the exposition. To get the most out of the article, playing with geometric algebra expressions is recommended. That is the best way to experience the power of the notation.

2. Orthodoxy

A good starting point is the definition of the imaginary number $i$: $i := \sqrt{-1}$. This is an odd definition because it is not clear why $-1$ should have a square root at all: apart from a sense of tidiness at wanting $-1$ to have a square root and the fruitfulness of the consequences of the definition, it does not seem apparent that $-1$ needs to have a square root. Historically this skepticism is why $i$ is called an imaginary number as opposed to a real number, a term attributed to René Descartes in [6]. For historical details we follow Paul Nahin’s [32], which is an excellent introduction to complex numbers as well as being an historical survey.

There are two common ways to justify the existence of $i$. The first justification is a geometrical interpretation of complex numbers in terms of the plane, discovered independently by Jean-Robert Argand in 1806 (see [1]) and Caspar Wessel in 1797 (see [43]), where a complex number $a + ib$ is interpreted as a line segment from the origin to point $(a, b)$ in the plane, for $a, b$ real numbers with $x$ coordinate $a$ and $y$ coordinate $b$, Wessel and Argand’s innovation being to view $i$ as a unit line segment perpendicular to the line segment representing 1 (see [43] for a commentary).

Figure 1: A point $(a, b)$ shown in an Argand diagram, where $a$ is shown along the $x$-axis and $b$ is shown on the $y$-axis multiplied by $i$.

This gloss of Argand and Wessel is not entirely accurate because both Argand and Wessel were working in the tradition of Euclidean geometry, where
distances are defined by means of the ratio of sides in similar triangles. What both Argand and Wessel were doing was to construct a product of two line segments as a construction in Euclidean geometry. But reference to the real numbers captures the notion of Argand diagram that is taught nowadays. The modern understanding of the Argand/Wessel view has a lot to commend it because it makes sense of the radius/angle representation of points that is represented in Euler’s formula (see Section 3 below) as well as the \((x, y)\) coordinate representation, and is (as we shall argue) ultimately a correct motivation for the complex numbers. However, this motivation is prima facie weak since it is unclear whether the two-dimensional plane in the Argand diagram is the same as the Euclidean plane (which it does turn out to be).

There is a second justification due originally to William R. Hamilton in [15], which was to treat complex numbers of the form \(a + ib\) for \(a, b\) real numbers as ordered pairs of real numbers, \(\langle a, b \rangle\), and to give them an algebra of the form:

\[
\begin{align*}
\langle a, b \rangle + \langle c, d \rangle & := \langle a + c, b + d \rangle \\
\langle a, b \rangle \times \langle c, d \rangle & := \langle a \times c - b \times d, a \times d + b \times c \rangle \\
\langle a, b \rangle - \langle c, d \rangle & := \langle a - c, b - d \rangle \\
\langle a, b \rangle \div \langle c, d \rangle & := \langle (a \times c + b \times d) \div (c^2 + d^2), (b \times c - a \times d) \div (c^2 + d^2) \rangle
\end{align*}
\]

if \(c \neq 0\) or \(d \neq 0\). It is well known that an ordered pair can be defined in set-theoretic terms, for example as \(\{\{a\}, \{a, b\}\}\), see [26]. However, it is unclear where such a strange algebra of multiplication comes from and why it should be of use to anyone. It is possible to argue that algebras are not designed to involve any thought; they are simple operations which can be mechanically computed. While there is some truth to this point of view, it avoids the issue of what motivates the algebraic operations. The algebraic approach is not a good explanation of what the complex numbers are. It may reduce the ontology of complex numbers to the real numbers and some basic set theory to construct ordered pairs of real numbers, but the algebra seems to emerge from nowhere. So let us change our emphasis to consider a more geometric approach (in the spirit of Argand and Wessel).
3. Euler’s Formula

One of the most important identities in mathematics is Euler’s formula, \( re^{i\theta} = r(\cos \theta + i \sin \theta) \), see Leonard Euler’s [9]. Aside from yielding such astonishing results as \( e^{i\pi} = -1 \), it is important because it shows that \( re^{i\theta} \) is a natural way of representing points on the circumference of a circle of radius \( r \) and an angle \( \theta \) from a fixed radius measured from the centre of the circle to the circumference (in a counter-clockwise sense).

![Figure 2: Euler’s formula as a point in an Argand diagram](image)

When you multiply two points together, you multiply the distances from the origin and add the angles, since \( (a \times e^{i\theta_1}) \times (b \times e^{i\theta_2}) = (a \times b) \times e^{i(\theta_1+\theta_2)} \).

Entities with magnitude (distance) and direction (angles) are called vectors. We seem to be saying that you can multiply two vectors together and not get the same result as any standard vector product (such as the dot product, the cross product, the wedge or the geometric product). The reason is easy to see: standard vector products are concerned with the difference in angles between two vectors rather than their sum. In fact, the vector product given by Euler’s formula can be interpreted as a form of the geometric product.

In geometric algebra (see [14] for example) the geometric product of two vectors is the sum of the dot product and the wedge product, written

\[ a \otimes b := a \cdot b + a \wedge b \tag{1} \]

where:

---

2 The easiest way to verify Euler’s formula is by using infinite power series expansions \( e^{i\theta} \), \( \cos \theta \) and \( \sin \theta \), but it is possible to show by that \( e^{i\theta} \) has the correct properties by using trigonometric identities for a sum of two angles to check that \( (\cos \theta + i \sin \theta)(\cos(\phi) + i \sin(\phi)) = \cos(\theta + \phi) + i \sin(\theta + \phi) \) and that \( e^{i0} = \cos(0) + i \sin(0) \) and \( e^{2\pi i} = 1 \) so that \( e^{i(\theta + 2\pi i)} = \cos(\theta + 2\pi) + i \sin(\theta + 2\pi) \) because \( \cos \theta = \cos(\theta + 2\pi) \) and \( \sin \theta = \sin(\theta + 2\pi) \).
Here $\theta_2$ is the angle from a fixed line to $b$, $\theta_1$ is the angle from a fixed line to $a$, $|a|$ is the length of $a$, $|b|$ is the length of $b$, and $\theta_2 - \theta_1$ is a unit area parallelogram oriented from $\theta_1$ to $\theta_2$ for $\theta_1 \neq \theta_2$.\(^3\) Thus, $a \wedge b$ is interpreted as a parallelogram of sides formed by vectors $a$ and $b$ oriented from $a$ to $b$.

\[ a \cdot b = |a| |b| \cos(\theta_2 - \theta_1); \]
\[ a \wedge b := |a| |b| \theta_2 - \theta_1 \sin(\theta_2 - \theta_1). \]

Figure 3: The wedge product of two vectors $a \wedge b$ interpreted as an oriented parallelogram.

We can see that $\theta_2 - \theta_1$ only depends on the orientation of $\theta_2$ relative to $\theta_1$ since its magnitude is 1. Let us rename a unit area counter-clockwise oriented parallelogram $i$. Later, in Section 4, $i$ will be identified with a counter-clockwise oriented unit square representing an angle of $\pi/2$, but that choice is conventional, albeit a natural choice in the context of orthogonal unit basis vectors.\(^4\) That said, it is possible to view $\theta_2 - \theta_1$ as a rotation of the plane (see Section 4), and in this context the fact that two (counter-clockwise) rotations of $\pi/2$ radians in succession, i.e. $\pi$ radians in total, map any vector $a$ to $-a$, makes the identification of $i$ with a counter-clockwise oriented unit square very natural.\(^5\) See Figure 4.

It is easily seen that the product from Euler’s formula is the geometric product of the reflection of the vector $a$ about the zero angle line with vector $b$ because such a reflection would map angle $\theta_1$ to $-\theta_1$, that is:

\[ |a| |b| (\cos(\theta_2 - (-\theta_1)) + i \times \sin(\theta_2 - (-\theta_1))) = |a| |b| e^{i(\theta_2+\theta_1)}. \]

\(^3\) If $\theta_1 = \theta_2$ then $\theta_2 - \theta_1 = 0$ and $a \wedge b = 0$.

\(^4\) If the basis vectors $b_1$ and $b_2$ are not orthogonal, then as the area of the parallelogram with sides $b_1$ and $b_2$ and angle $\theta$ between them is $|b_1| |b_2| \sin \theta$, if $|b_1| |b_2| \sin \theta = 1$, then both $b_1$ and $b_2$ cannot be unit vectors unless $\theta = \pi/2$.

\(^5\) It is possible to argue however that as $\theta_2 - \theta_1 = -\theta_1 - \theta_2$ then $\theta_2 - \theta_1 \theta_2 - \theta_1 a = -a$ for any vector $a$ if “−” is interpreted as multiplication by −1.
4. Geometric Product

The geometric product is a very interesting construct. If \( \sigma_1 \) and \( \sigma_2 \) are unit orthogonal basis vectors for a plane perpendicular to each other such that \( \sigma_1 \) rotated by \( \pi/2 \) radians counter-clockwise is \( \sigma_2 \), it is easy to see from Equations 2 and 3 that:

\[ \sigma_1 \cdot \sigma_1 = 1; \quad \sigma_2 \cdot \sigma_2 = 1; \quad \sigma_1 \cdot \sigma_2 = 0; \quad \sigma_2 \cdot \sigma_1 = 0; \]

and:

\[ \sigma_1 \wedge \sigma_1 = 0; \quad \sigma_2 \wedge \sigma_2 = 0; \quad \sigma_1 \wedge \sigma_2 = \sigma_2 \wedge \sigma_1 = \pi/2 = -\sigma_2 \wedge \sigma_1. \] (4)

Hence by Equation 1:

\[ \sigma_1 \otimes \sigma_1 = 1; \quad \sigma_2 \otimes \sigma_2 = 1; \quad \sigma_1 \otimes \sigma_2 = -\sigma_2 \otimes \sigma_1. \] (5)

It follows that by the associative property\(^6\) of \( \otimes \):

\[ (\sigma_1 \otimes \sigma_2)^2 = (\sigma_1 \otimes \sigma_2) \otimes (\sigma_1 \otimes \sigma_2) = \sigma_1 \otimes ((\sigma_2 \otimes \sigma_1) \otimes \sigma_2) \]

\(^6\) Proofs of associativity are given in [27, 35]. In terms of \( \sigma_1 \) and \( \sigma_2 \), more than two \( \otimes \) products are an order-dependent operation from left to right, where any element operates on the element immediately to its right. The easiest way to compute a \( \otimes \) product is to use the rules given in Equations 5 and to read the product \( a \otimes b \) as \( a \) operating on \( b \). For example \( (\sigma_1 \otimes \sigma_2) \otimes \sigma_2 \) is read as \( \sigma_1 \otimes (\sigma_2 \otimes \sigma_2) \), \( \sigma_2 \otimes \sigma_2 = 1 \) by Equations 5, so that \( (\sigma_1 \otimes \sigma_2) \otimes \sigma_2 = \sigma_1 \otimes (\sigma_2 \otimes \sigma_2) = \sigma_1 \) using the rule that \( \sigma_1 \otimes \vec{1} = \sigma_1 \), where \( \vec{1} \) is the multiplicative identity operator, the operator corresponding to 1.
and hence by Equations 5:

\[(\sigma_1 \otimes \sigma_2)^2 = -\sigma_1 \otimes (\sigma_1 \otimes \sigma_2) \otimes \sigma_2 = -(\sigma_1 \otimes \sigma_1) \otimes (\sigma_2 \otimes \sigma_2) = -1.\]

In other words, the geometric product of two orthogonal unit basis vectors in the plane behaves in the same way as \(i\). In view of the fact that \(\sigma_1 \wedge \sigma_2\) can be taken to be a unit counter-clockwise rotation from \(\sigma_1\) to \(\sigma_2\) (as illustrated in Figure 4 above), we see that the usage of \(i\) is consistent with Euler’s formula.

It is worth noting that the real justification of geometric algebra in two dimensional Euclidean space is that we want Pythagoras’s theorem to be true (see Figure 5 and [35]), that is:

\[|a\sigma_1 + b\sigma_2| := \sqrt{(a\sigma_1 + b\sigma_2) \otimes (a\sigma_1 + b\sigma_2)} = \sqrt{a^2 + b^2},\]  

Equation 6 only true if \(\sigma_1 \otimes \sigma_2 = -\sigma_2 \otimes \sigma_1\), which is the case if \(\sigma_1 \otimes \sigma_2\) is an oriented unit area.

\[\sqrt{a^2 + b^2}\]

\[\begin{array}{c}
\text{b} \\
\text{a}
\end{array}\]

Figure 5: Pythagoras’s theorem.

Does that mean that \(\sigma_1 \otimes \sigma_2\) is identical to \(i\)? Yes, in a way it does, because:

\[(a + b\sigma_1 \otimes \sigma_2) \otimes (c + d\sigma_1 \otimes \sigma_2) = (ac - bd) + (ad + bc)\sigma_1 \otimes \sigma_2.\]

But then a complex number is still a strange sum of a geometric product added to a scalar. We can make it a little less strange by noting that for real numbers \(a, b, c, d\):

\[(ac - bd)^2 + (ad + bc)^2 = (a^2 + b^2)(c^2 + d^2).\]

On the other hand \((\sigma_1 \otimes \sigma_2) \otimes \sigma_1 = -(\sigma_2 \otimes \sigma_1) \otimes \sigma_1\) by Equations 5, and since \(\sigma_1 \otimes \sigma_1 = 1\), \((\sigma_1 \otimes \sigma_2) \otimes \sigma_1 = -\sigma_2 \otimes (\sigma_1 \otimes \sigma_1) = -\sigma_2\). Since 1, \(\sigma_1\), \(\sigma_2\) and \(\sigma_1 \otimes \sigma_2\) forms a basis for the geometry algebra, associativity can be verified given that the basis elements are associative.
so that:
\[ |(a + b\sigma_1 \otimes \sigma_2) \otimes (c + d\sigma_1 \otimes \sigma_2)| = \sqrt{(ac - bd)^2 + (ad + bc)^2} = |a + b\sigma_1 \otimes \sigma_2| \times |c + d\sigma_1 \otimes \sigma_2|, \]  
(7)

where the length \( |a + b\sigma_1 \otimes \sigma_2| \) is \( \sqrt{a^2 + b^2} \). Equation 7 is what we would expect from Euler’s formula.

Let us consider the action of \( a + b\sigma_1 \otimes \sigma_2 \) on a vector \( c\sigma_1 + d\sigma_2 \). It is easy to see that:

\[ (a + b\sigma_1 \otimes \sigma_2) \otimes (c\sigma_1 + d\sigma_2) = (ac + bd)\sigma_1 + (ad - bc)\sigma_2. \]  
(8)

Equation 8 shows that \( a + b\sigma_1 \otimes \sigma_2 \) can be regarded as a transformation that takes vectors to vectors, although it does not exactly reproduce complex multiplication. However, we have:

\[ (a - b\sigma_1 \otimes \sigma_2) \otimes (c\sigma_1 + d\sigma_2) = (ac - bd)\sigma_1 + (ad + bc)\sigma_2, \]

which is exactly in the right form for complex numbers. What we can conclude from this is that \( a - b\sigma_1 \otimes \sigma_2 \) can be regarded as a transformation or function that scales and rotates vectors, noting that left multiplying a vector by \( a - b\sigma_1 \otimes \sigma_2 \) is the same as scaling the vector by \( \sqrt{a^2 + b^2} \) and rotating the vector by \( \tan^{-1}(-b/a) \) counter-clockwise or \( \tan^{-1}(b/a) \) clockwise. To remove the negative sign we can left multiply \( a + b\sigma_1 \otimes \sigma_2 \) by a vector, that is

\[ (c\sigma_1 + d\sigma_2) \otimes (a + b\sigma_1 \otimes \sigma_2) = (ac - bd)\sigma_1 + (ad + bc)\sigma_2. \]  
(9)

Equation 9 shows that a vector operating on a (variable) point is a complex number, which is a transformation of the plane when viewed as acting on all points of the plane. Strictly then a vector is not a transformation of a plane, but the action of a vector on a set of points is a transformation of the plane. We can phrase that in short by saying that (the action of) every two-dimensional vector (on a point) can be thought of as a scaling
and a rotation, and every scaling and rotation can be thought as a two-dimensional vector (acting on a point). This is not surprising because a point in a two-dimensional plane can be represented as either a vector or as a complex number. But complex numbers form an algebra under the geometric product whereas vectors are not closed under the geometric product, as we have seen yielding complex numbers instead under the geometric product.

It is possible to only use complex numbers to describe the geometry of the two-dimension Euclidean plane, but it is more natural to think of the two-dimensional Euclidean space as spanned by orthogonal unit vectors and having transformations given by complex numbers operating on vectors. It is worth noting that (the action of) vectors and complex numbers are linear transformations of the plane, which is to say that if \( T \) is a transformation then:

\[
T(\lambda \times x + \mu \times y) = \lambda \times T(x) + \mu \times T(y)
\]

for real numbers (or scalars) \( \lambda, \mu \) and complex numbers (or vectors) \( x, y \).

There is another common representation of transformations the plane, namely a \( 2 \times 2 \) matrix.\(^7\) A rotation and scaling matrix can be written as:

\[
\begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix}
\]

for scaling factor \( r \) and counter-clockwise rotation by angle \( \theta \). We note that matrix 10 operating on a column vector produces a column vector:

\[
\begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix}
\begin{pmatrix}
x \\
y
\end{pmatrix}
= \begin{pmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{pmatrix}.
\]

Dually, a row vector operating on matrix 10 produces a row vector:

\[
\begin{pmatrix}
x & y
\end{pmatrix}
\begin{pmatrix}
\cos \theta & \sin \theta \\
-\sin \theta & \cos \theta
\end{pmatrix}
= \begin{pmatrix} x \cos \theta - y \sin \theta & x \sin \theta + y \cos \theta \end{pmatrix}.
\]

\(^7\) The text treats rotations and scalings, but reflections have both matrix representations and geometric algebra representations. The geometric algebra representation is particularly neat. If in the plane we want to reflect vector \( v \) in unit vector \( a \), then if we write \( v = v_\parallel + v_\perp \), where \( v_\parallel \) is the component of \( v \) parallel to \( a \) and \( v_\perp \) is the component of \( v \) perpendicular to \( a \), then \( a \otimes v = a \otimes (v_\parallel + v_\perp) = |v_\parallel| + a \wedge v_\perp \), and \( a \otimes v \otimes a = |v_\parallel| a + (a \wedge v_\perp) \otimes a = v_\parallel - (v_\perp \wedge a) \otimes a = v_\parallel - (v_\perp \wedge a) \otimes a = v_\parallel - v_\perp \). Thus \( a \otimes v \otimes a \) reflects \( v \) about unit vector \( a \), see [14, 17, 35].
As a mathematical representation of a linear transformation, rotation and scaling in the plane in this case, matrices are perfectly adequate, but the role of $i$ is not immediately apparent.\(^8\)

We can conclude that in the case of the plane, $i$ is the oriented unit area plane element $\sigma_1 \otimes \sigma_2$ in the real vector space spanned by orthogonal unit vectors $\sigma_1$ and $\sigma_2$. Geometric products of two vectors are known as bi-vectors, but in two-dimensional geometric algebra, $\sigma_1 \otimes \sigma_2$ is both a bi-vector and a pseudo-scalar since its square is a scalar.

From an applied mathematics standpoint, this geometric algebra picture is sufficient to see what $i$ really is. The reason for this sufficiency is that $ae^{i\theta}$ encodes periodic plane wave motion with a fixed amplitude $|a|$. This is in contrast to $ae^\theta$, the equation of a spiral, which also describes a wave, but with an amplitude that strictly increases with $\theta$. Thus we would expect to find $i$ either in constant-amplitude wave solutions to an equation or to describe an equation where one side of the equation is related to the other by a rotation and/or scaling in a plane. $i = \cos(\pi/2) + i\sin(\pi/2) = e^{i\pi/2}$ represents a counter-clockwise rotation of $\pi/2$ radians, which is an orthogonality condition, without a scaling.

It is also possible to regard $i$ as a phase shift of $\pi/2$ in order to treat two sinusoidal waves that are $\pi/2$ out of phase as orthogonal for ease of analysis. This is so because $e^{i(\theta+\pi/2)} = ie^{i\theta}$. An example of treating a phase shift of $\pi/2$ as $i$ is the treatment of impedance in electronic engineering (see for example [24]). The resistance of an electronic circuit to alternating current is $\pi/2$ out of phase with the reactance (which comprises capacitive and inductive elements), which leads to the magnitude of the reactance being multiplied by the imaginary number $i$ (or $j$ as electronic engineers prefer).

But what about the use of $i$ in polynomial equations and the Fundamental Theorem of Algebra (due to Carl F. Gauss, see [11]) that there are $n$ complex valued solutions for $x$ in the equation $a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0 = 0$ with real number coefficients $a_{0 \leq i \leq n}$? The answer is that the Fundamental Theorem of Algebra\(^9\) shows that the solutions to real polynomial equations

\(^8\) There is an introductory book on geometrical algebra by Garret Sobczyk, [38], that is suited for those with a background in matrices and linear algebra.

\(^9\) The Fundamental Theorem of Algebra also states that all roots of polynomial equations with real number coefficients can be written as complex numbers.
can be represented in the two-dimensional real plane: complex numbers are
fundamentally two-dimensional over the real numbers, whether the zeros of
a polynomial equation are viewed as points of the plane, vectors in the plane
or linear transformations of the plane. Thus the picture of the complex num-
bers first drawn by Argand and Wessel is correct but is not a representation;
it defines the complex numbers.

The reason why complex analysis looks geometric is because it is geometric;
it is the (Euclidean) geometry of the real plane. In fact the teaching of
real analysis could learn some lessons from complex analysis. The primary
lesson would be to treat single-variable real analysis as the study of functions
which can be defined on the geometry of the Euclidean line. Multi-variable
(\(n\) say) real analysis could be thought of as the study of functions defined on
an \(n\)-dimensional Euclidean manifold with its natural geometric algebra (see
Section 5 below). That is not to say that matrices and linear transformations
do not have a role to play, as they provide powerful computational tools, but
multi-variate real analysis would be richer if it considered the geometry of
\(n\)-dimensional Euclidean manifolds.

5. Beyond the Plane

The pace of the narrative will increase from here. The remaining sections are
not comprehensive, but are designed to show the naturalness and expressive
power of geometric algebra, and to show that the imaginary number \(i\) has
richer forms in higher dimensions.

Three-dimensional geometric algebra is more complicated than two-
dimensional geometric algebra. The geometric algebra corresponding to a
real vector space spanned by orthogonal unit vectors \(\sigma_1\), \(\sigma_2\) and \(\sigma_3\) has
scalars, vectors with \(\{\sigma_1, \sigma_2, \sigma_3\}\) as a basis, bi-vectors representing oriented
planar areas with \(\{\sigma_1 \otimes \sigma_2, \sigma_2 \otimes \sigma_3, \sigma_3 \otimes \sigma_1\}\) as a basis, and a pseudo-
scalar of the form \(\sigma_1 \otimes \sigma_2 \otimes \sigma_3\) representing an oriented unit volume.

\[^{10}\text{32}^\] has a very nice construction of the complex zeros of a real cubic polynomial from
its graph, but in general complex roots cannot be determined from a standard graph in
the case of large values of the complex coefficient for real quintic polynomials (see [2]). In
general a visually-representative graph of complex zeros of a real polynomial will require 4
dimensions, as it is a mapping of complex values of the polynomial variable to the complex
values of the polynomial.
A general element in a three-dimensional geometric algebra has the form:

\[ a + b_1 \sigma_1 + b_2 \sigma_2 + b_3 \sigma_3 + c_1 \sigma_1 \otimes \sigma_2 + c_2 \sigma_2 \otimes \sigma_3 + c_3 \sigma_3 \otimes \sigma_1 + d \sigma_1 \otimes \sigma_2 \otimes \sigma_3, \]

where \( a, b_i, c_i, d \) are real number scalars, and is known as a multi-vector. In three dimensions a rotation and scaling has the form:

\[ a + b \sigma_1 \otimes \sigma_2 + c \sigma_2 \otimes \sigma_3 + d \sigma_3 \otimes \sigma_1, \quad (11) \]

and we would expect (11) to map a three-dimensional vector into a vector, that is:

\[(a + b \sigma_1 \otimes \sigma_2 + c \sigma_2 \otimes \sigma_3 + d \sigma_3 \otimes \sigma_1)(e \sigma_1 + f \sigma_2 + g \sigma_3) = (ae + bf - dg)\sigma_1 + (af - be + cg)\sigma_2 + (ag - cf + de)\sigma_3\]

if the \( \sigma_1 \otimes \sigma_2 \otimes \sigma_3 \) term \( bg + ce + df \) equals 0. The rotations and scalings of vectors in three-dimensional (Euclidean) space can be represented by the quaternions, discovered and popularized by William R. Hamilton (see [16]).

The quaternions have the form:

\[ a + bi + cj + dk, \]

where \( i, j, k \) are unit “vectors” with the property that:

\[ i^2 = j^2 = k^2 = ijk = -1. \]

Although \( i, j, k \) can be treated as unit vectors, it is unclear why they should possess these strange algebraic properties (which seem even more mysterious than \( i^2 = -1 \) in two dimensions). According to geometric algebra (see [4, 17]), \( i, j, k \) are in fact bi-vectors (that is to say oriented unit area elements) and \( i = \sigma_2 \otimes \sigma_3, j = -\sigma_3 \otimes \sigma_1, k = \sigma_1 \otimes \sigma_2 \) say (as the minus sign can be associated with \( i \) or \( j \) or \( k \)).

There is no need to stop at three dimensions. It is possible to generate a geometric algebra of Euclidean \( n \)-dimensional space by constructing a basis of \( 2^n \) distinct unit scalars, vectors, bi-vectors, and so forth up to a unit \( n \)-vector, which is also a pseudo-scalar as the geometric product \((\sigma_1 \otimes \ldots \otimes \sigma_n)^2\) is \( = 1 \) if \( n(n-1) \) is divisible by 4 and \( = -1 \) otherwise (compare [4]). Since either \( n \) or \( n-1 \) is even, it follows that \((\sigma_1 \otimes \ldots \otimes \sigma_n)^2 = 1 \) if and only if \( n \) is divisible by 4 or \( n-1 \) is divisible by 4. As in the two- and three-dimensional cases, the even grade multi-vectors, that is the unit scalars, bi-vectors, 4-vectors, and so forth form the basis of all scalings and rotations that map vectors to vectors and are known as spinors by means of a map \( M \otimes v \otimes N \) for vector \( v \) and \( M, N \) even grade multi-vectors such that \( M \otimes N = 1 \) [14, 17].
6. Imaginary Numbers in Physics: Schrödinger’s Equation

It may seem that we have answered what an imaginary number is: it is a two-dimensional number, vector, or transformation in the Euclidean plane. It turns out that for Schrödinger’s equation with Born’s Rule in quantum mechanics, this interpretation is sufficient.\(^\text{11}\) Schrödinger’s equation for a single particle in a potential field is:

\[
i\hbar \frac{\partial \psi(r, t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 \psi(r, t) + V(r)\psi(r, t).
\]

In words, it says that the average rate of spatial change (see [39]) multiplied by \(-\hbar^2/2m\), where \(\hbar\) is the reduced Planck constant obtained by dividing the Planck’s constant \(\hbar\) by \(2\pi\), and \(m\) is the mass of the particle, offset by the potential energy of the wave function is \(i\hbar\) times the magnitude and is orthogonal in the plane to the rate of change in the wave function in time.\(^\text{12}\)

While this seems a plausible physical interpretation of Schrödinger’s equation as an expression of homogeneity between space and time (modulo the potential energy and a constant scale factor), it is not very clear in which plane the average rate of spatial change and the rate of change in the wave function in time are at right angles to each other. The usual answer is to say that the plane is in fact an abstract Hilbert space, with the orthogonality being given when the inner product of two vectors equals 0. But what if the plane is defined by two new orthogonal directions? It may seem unusual to think of real space as five-dimensional, but there is no reason why every point in space could not be associated with a spatial plane that is orthogonal to the space in which the points sits. In fact, for multiple particle systems, the configuration space of particles in a system will scale as the number of particles, and the dimension of the space of the wave function is \(3n + 2\) for number of particles \(n\). However, the view of Hestenes in [19] is that a separate spatial plane is not needed and that \(i\) can be identified with a multi-vector with square \(-1\) in the physical space or, in general in the configuration space of the problem.

---

\(^{11}\) See [41] for a clear introduction to quantum mechanics and Schrödinger’s Equation, and [25] for a survey of motivations for complex numbers appearing in quantum mechanics.

\(^{12}\) Check out Grant Sanderson’s 3Blue1Brown mathematics YouTube video [https://www.youtube.com/watch?v=O85OWBJ2ayo](https://www.youtube.com/watch?v=O85OWBJ2ayo) on matrix exponentiation for a neat overview of this interpretation.
That is to say, Hestenes holds in [19] that physics should be expressed in terms of a spacetime algebra.

If we do not follow Hestenes, but treat \( i \) as a vector in a separate plane (which could be accommodated within a position that is realist about wave states, as in [31], or about finite dimensional flat Euclidean manifolds, as in [42]), it might be interesting to see what quantum mechanics will then look like. For a start it is not the dot product of a vector (with complex scalars) and its conjugate that is equal to 1; it is the geometric product of a real-valued vector with itself. If we take a vector (in the separate plane) \( v := a_1 \sigma_1 + a_2 \sigma_2 \) then \( v \otimes v = a_1^2 + a_2^2 \) by the fact that \( \sigma_1 \otimes \sigma_2 = -\sigma_2 \otimes \sigma_1 \). We can in fact develop a variation of Schrödinger’s equation based on the fact that solutions to Schrödinger’s equation should be two-dimensional vectors. Let us write Schrödinger’s equation as:

\[
H \psi = i \hbar \partial \psi / \partial t,
\]

where \( H \) is the Hamiltonian energy function and \( \psi \) is a plane vector function of position \( r \) and time \( t \). Then if we write \( \psi = \psi_1 \sigma_1 + \psi_2 \sigma_2 \), we have:

\[
(H \psi_1) \sigma_1 + (H \psi_2) \sigma_2 = \sigma_1 \otimes \sigma_2 \otimes \hbar (\sigma_1 \partial \psi_1 / \partial t + \sigma_2 \partial \psi_2 / \partial t)
\]

under the assumptions that \( H \) leaves all scalars invariant and that the Hamiltonian can be regarded as a function from vectors to vectors. Then, by equating scalar coefficients of \( \sigma_1 \) and \( \sigma_2 \), we have a pair of equations:

\[
c \psi_1 = \hbar \partial \psi_2 / \partial t, \quad d \psi_2 = -\hbar \partial \psi_1 / \partial t, \quad (12)
\]

where the operator \( H \) (the total energy) operates on vectors \( \sigma_1 \) and \( \sigma_2 \) such that \( H \sigma_1 = c \sigma_1 \) and \( H \sigma_2 = d \sigma_2 \) for real number constants \( c \) and \( d \) (which is a statement that \( \sigma_1 \) and \( \sigma_2 \) are eigenvectors of \( H \)). Solutions to Equations 12 have the form:

\[
\psi_1 = \psi_{0,1} \phi_1(r) \cos(dt/\hbar) - \psi_{0,2} \phi_1(r) \sin(dt/\hbar);
\psi_2 = \psi_{0,1} \phi_1(r) \sin(ct/\hbar) + \psi_{0,2} \phi_1(r) \cos(ct/\hbar);
\]

if \( c = d \), \( \psi \) has initial value \( \psi_0 = \psi_{0,1} \sigma_1 + \psi_{0,2} \sigma_2 \), and \( \psi_{0,1}^2 + \psi_{0,2}^2 = 1 \) and \( \psi_1^2 + \psi_2^2 = 1 \) by Born’s rule (that measurements on quantum wave functions are probabilities, and so add up to 1). There is also a more general solution of the form:

\[
\psi_1 = \phi_1(r)e^{-ict/\hbar}, \quad \psi_2 = \phi_2(r)e^{-i\theta t/\hbar},
\]
with $\phi_1^2 + \phi_2^2 = 1$ that follows the derivation of the three dimensional case below (see Equation 13). What this means is that $\psi$ is fundamentally a probability wave in space and time variables. This is plausible because if you measure a property that has a minimum value (a quantum), the value returned cannot be less than the quantum value, but may represent the unnormalized likelihood of the property being instantiated if it represents anything at all.

The problem with the approach in the last paragraph is that it is still not clear what plane $\sigma_1$ and $\sigma_2$ define and what a measurement of the wave function looks like in terms of basis vectors $\sigma_1$ and $\sigma_2$. Even worse, if $\sigma_1$ and $\sigma_2$ are spatial vectors, in the physical world with three spatial dimensions, orthogonality for any basis vector is ambiguous between two basis vectors of the other dimensions. Another approach, which is consistent with Hestenes’s [19], is to take $\sigma_1, \sigma_2, \text{and } \sigma_3$ to be a standard orthogonal unit vector basis of space. If we take a vector $v := a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$ and $i := \sigma_1 \otimes \sigma_2 \otimes \sigma_3$, then $H\psi_1\sigma_1 + H\psi_2\sigma_2 + H\psi_3\sigma_3$ can be rewritten as:

$$H\psi_1\sigma_1 + H\psi_2\sigma_2 + H\psi_3\sigma_3 =$$

$$\sigma_1 \otimes \sigma_2 \otimes \sigma_3 \otimes h(\sigma_1 \partial \psi_1 / \partial t + \sigma_2 \partial \psi_2 / \partial t + \sigma_3 \partial \psi_3 / \partial t) \quad (13)$$

In this case we cannot equate vectors, but we can see that there is a solution:

$$\psi_1 = \phi_1(r)e^{-ibt/h}, \quad \psi_2 = \phi_2(r)e^{-ict/h}, \quad \psi_3 = \phi_3(r)e^{-idt/h},$$

which naturally pairs $\sigma_1$ with $\sigma_2 \otimes \sigma_3$, $\sigma_2$ with $\sigma_3 \otimes \sigma_1$ and $\sigma_3$ with $\sigma_1 \otimes \sigma_2$ via Euler’s formula, for real number constants $b, c, d$ such that $H\sigma_1 = b\sigma_1, H\sigma_2 = c\sigma_2$ and $H\sigma_3 = d\sigma_3$ and $\sum_{i=1}^{3} \psi_i^2 = \sum_{i=1}^{3} \phi_i^2 = 1$ by Born’s rule. This works better as an approach to Schrödinger’s equation than the two-dimensional solution above, but it does suppose that any measurable property can be correlated with a spatial vector or a scalar. This is true in quantum mechanics because vectors can be measured by measuring the eigenvalues of the vector under the operator representing an observable quantity, and real number scalars can be measured directly (at least approximately).

For bi-vector valued properties (such as spin), we can measure the corresponding vector under the (Hodge Dual) correspondence $a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$ with $a_1\sigma_2 \otimes \sigma_3 + a_2\sigma_3 \otimes \sigma_1 + a_3\sigma_1 \otimes \sigma_2$ given by $v \rightarrow iv$, and we can replace the pseudo-scalar $ia$ by the scalar $a$ under the correspondence $ia \rightarrow a$. The problem with this construction is that neither $\sigma_1 \otimes \sigma_2$ and $\sigma_2 \otimes \sigma_1$ nor $\sigma_3$ and
−σ_3 are orthogonal to each other. A solution, which is inherent in the geometric algebra treatment of spinors (see [17, 14] for example), is to use half the angle between spinors or their Hodge dual vectors^{13}, which makes sense because the probability of being at π/2 radians between “spin up” and “spin down” is 1/2, which is \( \cos^2(\pi/4) \), and \( \sigma_1 \otimes \sigma_2 \) and \( \sigma_2 \otimes \sigma_1 \) are then orthogonal. Multi-particle systems can be dealt with by treating \( \phi \) as a spatially-valued function of the positions of the particles.

7. Spacetime Algebra

Another challenge is to characterize a point in spacetime. We could just assert that there is a unit vector with square −1, but this is false since all unit vectors have a square +1. For this reason the vector must be a bi-vector or a tri-vector and so forth. Let us assume that a point in spacetime has three spatial coordinates \( \langle x_i, y_i, z_i \rangle \) and a scalar time coordinate \( ct \), where \( c \) is the (scalar) speed of light in vacuum and \( x, y, z \) are real number scalars.

If we take \( i := (\sigma_0 \otimes \sigma_1 \otimes \sigma_2 \otimes \sigma_3) \) and \( i^2 = 1 \), a spacetime vector would then have the form:

\[
r = cti - (xi) \otimes \sigma_1 - (yi) \otimes \sigma_2 - (zi) \otimes \sigma_3,
\]

since we want \( c^2t^2 \) to be positive and \( x^2, y^2, z^2 \) to be negative.\(^{14}\) It follows that:

\[
r = c(t(\sigma_0 \otimes \sigma_1 \otimes \sigma_2 \otimes \sigma_3) - x(\sigma_0 \otimes \sigma_1 \otimes \sigma_2 \otimes \sigma_3) \otimes \sigma_1
\]

\[-y(\sigma_0 \otimes \sigma_1 \otimes \sigma_2 \otimes \sigma_3) \otimes \sigma_2 - z(\sigma_0 \otimes \sigma_1 \otimes \sigma_2 \otimes \sigma_3) \otimes \sigma_3.
\]

Tidying up the expression for \( r \), we have:

\[
r = cti\sigma_0 \otimes \sigma_1 \otimes \sigma_2 \otimes \sigma_3 - x\sigma_0 \otimes \sigma_2 \otimes \sigma_3 - y\sigma_0 \otimes \sigma_3 \otimes \sigma_1 - z\sigma_0 \otimes \sigma_1 \otimes \sigma_2. \quad (14)
\]

We can then show that \( r^2 = c^2t^2 - x^2 - y^2 - z^2 \), which is the Minkowski pseudo-Euclidean metric for spacetime. What this means is that space unit “vectors” are actually tri-vectors and time unit “vectors” are pseudo-scalars.

\(^{13}\) This is standard quantum mechanics; see for example [41, Equations 3.24 and 3.25] showing that the probabilities of “spin up” and “spin down” for a single particle are \( \cos^2 \theta \) and \( \sin^2 \theta \), where \( \theta \) is the angle between “spin up” and the axis of measurement for \( \sigma_3 \).

\(^{14}\) If we had \( r = cti\sigma_0 - (xi) \otimes \sigma_1 - (yi) \otimes \sigma_2 - (zi) \otimes \sigma_3 \), then we would end up with \( r = -ct\sigma_1 \otimes \sigma_2 \otimes \sigma_3 + x\sigma_0 \otimes \sigma_2 \otimes \sigma_3 + y\sigma_0 \otimes \sigma_3 \otimes \sigma_1 + z\sigma_0 \otimes \sigma_1 \otimes \sigma_2 \), where \( c^2t^2 \) and \( x^2, y^2, z^2 \) are all negative.
It is worth noting that space unit “vectors” are the Hodge dual of space unit vectors, and the Hodge dual of time unit “vectors” are scalars, which is similar to the case of quaternions in three spatial dimensions.\footnote{If the physical world admitted more than one orthogonal time axis, then time units as pseudo-scalars would be not be an adequate representation of time.}

The result in the previous paragraph relies on the fact that in four-dimensional Euclidean space, \( i^2 = +1 \). If we reverse the signature and require that \( r^2 = -c^2 t^2 + x^2 + y^2 + z^2 \), then it is not possible to generate this form from a Euclidean metric. To see this, note that the spatial coordinates \( x, y, z \) must have unit vectors \( (\sigma_1, \sigma_2 \text{ or } \sigma_3) \) associated with them because unit bi-vectors and tri-vectors square to \(-1\), and there is only one 4-vector, namely \( i \). The time coordinate \( t \) could be associated with a unit bi-vector or with a unit tri-vector. However, if the unit bi-vector was of the form \( \sigma_i \otimes \sigma_j \) for some \( i \neq j \) such that \( 0 \leq i, j \leq 3 \), then for any \( k \neq i \neq j \) (which always exists) we have \( \sigma_k \otimes (\sigma_i \otimes \sigma_j) = (\sigma_i \otimes \sigma_j) \otimes \sigma_k \), so that cross-terms do not cancel. If the unit tri-vector was of the form \( \sigma_i \otimes \sigma_j \otimes \sigma_k \) for \( k \neq i \neq j \) then
\[
\sigma_i \otimes (\sigma_i \otimes \sigma_j \otimes \sigma_k) = (\sigma_i \otimes \sigma_j \otimes \sigma_k) \otimes \sigma_i,
\]
so again cross-terms do not cancel. As these are the only possibilities, it follows that it is not possible to generate \( r^2 = -c^2 t^2 + x^2 + y^2 + z^2 \) from a Euclidean metric.

The conclusion is that spacetime has a natural Minkowski geometry associated with the condition \( i^2 = +1 \), which was discovered by James Cockle in [5] and known as the hyperbolic or split complex numbers in the case of a two-dimensional space [37], and the split quaternions in the case of three dimensional space [34]. The development of a curved spacetime follows in the standard way by assuming that every neighborhood in the spacetime can be transformed in a neighborhood dependent way to a flat Minkowski geometry (see for example [28]).

8. Equations of Physics

It should be apparent that there is no reason in general why elements of a geometric algebra should not appear explicitly in the equations of physics. Hestenes treats Dirac’s equation for electron dynamics in this way, see [20, 21], and Equation 13 is a rewriting of Schrödinger’s equation. We have seen that spin terms are the even elements of a geometric algebra, which is to say they are of the form \( a \sigma_i \otimes \sigma_j \) for \( 0 \leq i, j \leq 3 \) in spacetime and \( a \) a real
number, while force terms are vectors, \( b \sigma_i \) for \( b \) a real number. If all mass particles have spin, and acceleration can be represented by the second time derivative of Equation 14, then we would expect to be able to express the relationship with force as:

\[
(\sum_{i=0}^{3} \sum_{j=0}^{3} a_i \sigma_i \otimes \sigma_j) \otimes ((\partial^2 x / \partial t^2) \sigma_0 \otimes \sigma_2 \otimes \sigma_3 \\
+ (\partial^2 y / \partial t^2) \sigma_3 \otimes \sigma_1 + (\partial^2 z / \partial t^2) \sigma_0 \otimes \sigma_1 \otimes \sigma_2) = \sum_{i=0}^{3} b_i \sigma_i.
\]

Since this is not a physics article, and the purpose of this section is to emphasize that the equations of physics should be expressible as differential equations in the terms of geometric algebra, the merits of this idea will not be discussed further here. However, we cannot omit a mention of the triumph of geometric algebra in formulating Maxwell’s equations in classical electromagnetism as a single equation:

\[
(\partial_t + \nabla) F = 0 \quad (15)
\]

for time derivative operator \( \partial_t \), (spatial) gradient operator \( \nabla \), and electromagnetic field \( F = E + B \), where the electric field \( E \) is a vector field and the magnetic field \( B \) is a bi-vector field [30].

9. Conclusion

In this article I have argued that \( i \) is a quantity associated with the two-dimensional real number plane, as a vector, a bi-vector, a point, or a transformation (rotation). This position provides a foundation for the complex numbers and accounts for complex numbers in some equations of applied mathematics and physics. I have also argued that complex numbers are fundamentally geometrical and can be described by geometric algebra, and that moreover the meaning of complex numbers in physics varies with dimension and geometry of the manifold.

References


[36] Pierre Renaud. Clifford Algebras Lecture Notes on Applications in Physics. [https://hal.archives-ouvertes.fr/hal-03015551v2](https://hal.archives-ouvertes.fr/hal-03015551v2), January 2022.


