Division Algebras and Quantum Theory

John C. Baez

Centre for Quantum Technologies
National University of Singapore
Singapore 117543

and

Department of Mathematics
University of California
Riverside CA 92521
email: baez@math.ucr.edu

February 9, 2011

Abstract

Quantum theory may be formulated using Hilbert spaces over any of the three associative normed division algebras: the real numbers, the complex numbers and the quaternions. Indeed, these three choices appear naturally in a number of axiomatic approaches. However, there are internal problems with real or quaternionic quantum theory. Here we argue that these problems can be resolved if we treat real, complex and quaternionic quantum theory as part of a unified structure. Dyson called this structure the 'three-fold way'. It is perhaps easiest to see it in the study of irreducible unitary representations of groups on complex Hilbert spaces. These representations come in three kinds: those that are not isomorphic to their own dual (the truly 'complex' representations), those that are self-dual thanks to a symmetric bilinear pairing (which are 'real', in that they are the complexifications of representations on real Hilbert spaces), and those that are self-dual thanks to an antisymmetric bilinear pairing (which are 'quaternionic', in that they are the underlying complex representations of representations on quaternionic Hilbert spaces). This three-fold classification sheds light on the physics of time reversal symmetry, and it already plays an important role in particle physics. More generally, Hilbert spaces of any one of the three kinds—real, complex and quaternionic—can be seen as Hilbert spaces of the other kinds, equipped with extra structure.

1 Introduction

Ever since the birth of quantum mechanics, there has been curiosity about the special role that the complex numbers play in this theory. Of course, one could also ask about the role of real numbers in classical mechanics: to a large extent, this question concerns the role of the continuum in physical theories. In quantum mechanics there is a new twist. Classically, observables are described by real-valued functions. In quantum mechanics, they are described by self-adjoint operators on a complex Hilbert space. In both theories, the expectation value of an observable in any state is *real*. But the complex numbers play a fundamental role in quantum mechanics that is not apparent in classical mechanics. Why?

The puzzle is heightened by the fact that we can formulate many aspects of quantum theory using real numbers or quaternions instead of complex numbers. There are precisely four 'normed division algebras': the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, the quaternions $\mathbb{H}$ and the octonions $\mathbb{O}$. Roughly speaking, these are the number systems extending the reals that have an 'absolute value' obeying the equation $|xy| = |x|\,|y|$. Since the octonions are nonassociative it proves difficult to formulate quantum theory based on these, except in a few special cases. But the other three number systems support the machinery of quantum theory quite nicely.

Indeed, as far back as 1934, Jordan, von Neumann and Wigner [31] saw real, complex and quaternionic quantum theory appear on an equal footing in their classification of algebras of observables.

These authors classified *finite-dimensional* algebras of observables, but later their work was generalized to the infinite-dimensional case. We review some of these classification theorems in Section 2.

At first glance, the lesson seems to be that real, complex and quaternionic quantum theory stand on an equal footing, at least in some axiomatic approaches. However, experiments seem to show that our universe is described by the complex version, not the other two. So why did nature pick the *complex* version of quantum theory?

In fact, the real and quaternionic versions of quantum theory have some 'problems', or at least striking differences from the familiar complex version. First, they lack the usual correspondence between strongly continuous one-parameter unitary groups and self-adjoint operators, which in the complex case goes by the name of Stone's Theorem. Second, the tensor product of two quaternionic Hilbert spaces is not a quaternionic Hilbert space. We discuss these issues in Section 3.

One could take these 'problems' as the explanation of why nature uses complex quantum theory, and let the matter drop there. But this would be shortsighted. Indeed, the main claim of this paper is that instead of being distinct alternatives, real, complex and quaternionic quantum mechanics are three aspects of a single unified structure. Nature did not choose one: she chose all three!

The evidence for this goes back to a very old idea in group theory: the Frobenius–Schur indicator [26, 20]. This is a way of computing a number from an irreducible unitary representation of a compact group G on a complex Hilbert space H, say

$$\rho\colon G \to \mathrm{U}(H)$$

where $\mathrm{U}(H)$ is the group of unitary operators on H. Any such representation is finite-dimensional, so we can take the trace of the operator $\rho(g^2)$ for any group element $g \in G$, and then perform the integral

$$\int\limits_G \mathrm{tr}(\rho(g^2))\, dg$$

where dg is the normalized Haar measure on G. This integral is the **Frobenius–Schur indicator**. It always equals 1, 0 or -1, and these three cases correspond to whether the representation ρ is 'real', 'complex' or 'quaternionic'. Here we are using these words in a technical sense. Remember, the Hilbert space H is complex. However:

- If the Frobenius–Schur indicator is 1, then ρ is the complexification of a representation of G on some real Hilbert space. In this case we call ρ **real**.
- If the Frobenius–Schur indicator is -1, then ρ is the underlying complex representation of a representation of G on some quaternionic Hilbert space. In this case we call ρ **quaternionic**.
- If the Frobenius–Schur indicator is 0, then neither of the above alternatives hold. In this case we call ρ **complex**.

The Frobenius–Schur indicator is an appealingly concrete way to decide whether an irreducible representation is real, quaternionic or complex. However, the significance of this threefold classification is made much clearer by another result, which focuses on the key concept of *duality*. The represention ρ has a dual ρ^*, which is a unitary representation of G on the dual Hilbert space H^*. Of course, the Hilbert space H has the same dimension as its dual, so they are isomorphic. The question thus arises: is the representation ρ isomorphic to its dual? Instead of a simple yes-or-no answer, it turns out there are three cases:

- The representation ρ is real iff it is isomorphic to its dual thanks to the existence of an invariant nondegenerate *symmetric* bilinear form $g\colon H \times H \to \mathbb{C}$.
- The representation ρ is quaternionic iff it is isomorphic to its dual thanks to the existence of an invariant nondegenerate *antisymmetric* bilinear form $g\colon H \times H \to \mathbb{C}$.

- The representation ρ is complex iff it is not isomorphic to its dual.

We recall the proof of this result in Section 4. Freeman Dyson called it the 'three-fold way' [24]. It also appears among Vladimir Arnold's list of 'trinities' [7]. So, in some sense it is well known. However, its implications for the foundations of quantum theory seem to have been insufficiently explored.

What are the implications of the three-fold way?

First, since elementary particles are often described using irreducible unitary representations of compact groups, it means that particles come in three kinds: real, complex and quaternionic! Of course the details depend not just on the particle itself, but on the group of symmetries we consider.

For example, take any spin-$\frac{1}{2}$ particle, and consider only its rotational symmetries. Then we can describe this particle using a unitary representation of SU(2) on $\mathbb{C}^2$. This representation is quaternionic. Why? Because we can think of a pair of complex numbers as a quaternion, and SU(2) as the group of unit quaternions. The spin-$\frac{1}{2}$ representation of SU(2) is then revealed to be the action of unit quaternions on the quaternions via left multiplication.

In slogan form: *qubits are not just quantum—they are also quaternionic!* More generally, all particles of half-integer spin are quaternionic, while particles of integer spin are real, as long as we consider them only as representations of SU(2).

This explains why the square of time reversal is 1 for particles of integer spin, but -1 for particles of half-integer spin. Time reversal is 'antilinear': it commutes with real scalars but anticommutes with multiplication by i. An antilinear operator whose square is 1 acts like complex conjugation. So, we automatically get such an operator on the complexification of any real Hilbert space. An antilinear operator whose square is -1 acts like the quaternion unit j, since $ij = -ji$ and $j^2 = 1$. So, we get such an operator on the underlying complex Hilbert space of any quaternionic Hilbert space.

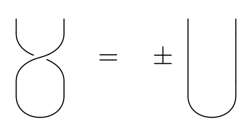
There is a nice heuristic argument for the same fact using 'string diagrams', which are a mathematical generalization of Feynman diagrams [13, 14]. Any irreducible representation of SU(2) has an invariant nondegenerate bilinear form $g\colon H \times H \to C$. This is symmetric:

$$g(v, w) = g(w, v)$$

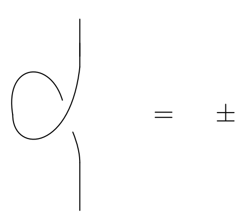
when the spin is an integer, and antisymmetric

$$g(v, w) = -g(w, v).$$

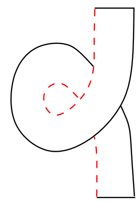
when the spin is a half-integer. So, in terms of string diagrams:



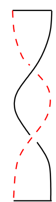
If we manipulate this equation a bit, we get:



The diagram at left shows a particle reversing its direction in time twice, so not surprisingly, the square of time reversal is ± 1 depending on whether the spin is an integer or half-integer. But we can also see more. Suppose we draw the diagram at left as a ‘ribbon’:



Then, pulling it tight, we get a ribbon with a 360 degree twist in it:



And indeed, rotating a particle by 360 degrees gives a phase of $+1$ if its spin is an integer, and -1 if its spin is a half-integer. So this well-known fact is yet another manifestation of the real/quaternionic distinction. We explain these ideas further in Section 5.

More broadly, the three-fold way suggests that complex quantum theory *contains* the real and quaternionic theories. Indeed, this can be made precise with the help of some category theory. Let $\mathrm{Hilb}_{\mathbb{R}}$, $\mathrm{Hilb}_{\mathbb{C}}$ and $\mathrm{Hilb}_{\mathbb{H}}$ stand for the categories of real, complex and quaternionic Hilbert spaces, respectively. Then there are ‘faithful functors’ from $\mathrm{Hilb}_{\mathbb{R}}$ and $\mathrm{Hilb}_{\mathbb{H}}$ into $\mathrm{Hilb}_{\mathbb{C}}$. This is a way of making precise the fact that we can treat real or quaternionic Hilbert spaces as complex Hilbert spaces equipped with extra structure. The extra structure is just a nondegenerate bilinear form $g\colon H \times H \to \mathbb{C}$, which is symmetric in the real case and antisymmetric in the quaternionic case.

Since complex quantum theory contains the real and quaternionic theories, we might be tempted to conclude that the complex theory is fundamental, with the other theories arising as offshoots. But we can also faithfully map $\mathrm{Hilb}_{\mathbb{C}}$ and $\mathrm{Hilb}_{\mathbb{H}}$ into $\mathrm{Hilb}_{\mathbb{R}}$. We can even faithfully map $\mathrm{Hilb}_{\mathbb{R}}$ and $\mathrm{Hilb}_{\mathbb{C}}$ into $\mathrm{Hilb}_{\mathbb{H}}$! We describe these faithful functors in Section 6. So, one may argue that no one form of quantum mechanics is primordial: each contains the other two! We provide details in Section 6. Then, in Section 7, we use the relation between the three forms of quantum theory to solve the problems with real and quaternionic quantum mechanics raised earlier.

2 Classifications

The laws of algebra constrain the possibilities for theories that closely resemble the quantum mechanics we know and love. Various classification theorems delimit the possibilities. Of course, theorems have hypotheses. It is easy to get around these theorems by weakening our criteria for what counts as a theory ‘closely resembling’ quantum mechanics; if we do this, we can find a large number of alternative theories. This is especially clear in the category-theoretic framework, where many theories based on convex sets of states give rise to categories of physical systems and processes sharing some features of standard quantum mechanics [15, 16, 17]. Here, however, we focus on results that pick out real, complex and quaternionic quantum mechanics as special.

Our treatment is far from complete; it is merely meant as a sketch of the subject. In particular, while we describe the Jordan algebra approach to quantum systems with finitely many degrees of freedom, and also the approach based on convex sets, we neglect the approach based on lattices of propositions. Good introductions to this enormous subject include the paper by Steirteghem and Stubbe [42] and the classic texts by Piron [37] and Varadarajan [44].

2.1 Normed division algebras

After discovering that complex numbers could be viewed as simply pairs of real numbers, Hamilton sought an algebra of 'triples' where addition, subtraction, multiplication and division obeyed most of the same rules. Alas, what he was seeking did not exist. After much struggle, he discovered the 'quaternions': an algebra consisting of expressions of the form $a1 + bi + cj + dk$ $(a, b, c, d \in \mathbb{R})$, equipped with the associative product with unit 1 uniquely characterized by these equations:

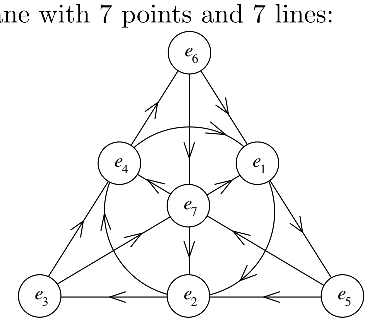
$$i^2 = j^2 = k^2 = ijk = -1.$$

He carved these equations into a bridge as soon he discovered them. He spent the rest of his life working on the quaternions, so this algebra is now called $\mathbb{H}$ in his honor.

The day after Hamilton discovered the quaternions, he sent a letter describing them to his college friend John Graves. A few months later Graves invented yet another algebra, now called the 'octonions' and denoted $\mathbb{O}$. The octonions are expressions of the form

$$a_0 1 + a_1 e_1 + a_2 e_2 + \cdots + a_7 e_7$$

with $a_i \in \mathbb{R}$. The multiplication of octonions is not associative, but it is still easily described using the **Fano plane**, a projective plane with 7 points and 7 lines:



The 'lines' here are the sides of the triangle, its altitudes, and the circle containing the midpoints of the sides. Each line contains three points, and each of these triples has a cyclic ordering, as shown by the arrows. If e_i, e_j, and e_k are cyclically ordered in this way then

$$e_i e_j = e_k, \qquad e_j e_i = -e_k.$$

Together with these rules:

- 1 is the multiplicative identity,
- $e_1, \ldots, e_7$ are square roots of -1,

the Fano plane completely describes the algebra structure of the octonions. As an exercise, we urge the reader to check that the octonions are nonassociative.

What is so great about the number systems discovered by Hamilton and Graves? Like the real and complex numbers, they are 'normed division algebras'. For us, an **algebra** will be a real vector space A equipped with a multiplication

$$\begin{array}{ccc} A \times A & \to & A \\ (x, y) & \mapsto & xy \end{array}$$

that is real-linear in each argument. We do not assume our algebras are associative. We say an algebra A is **unital** if there exists an element $1 \in A$ with

$$1a = a = a1$$

for all $a \in A$. And we define a **normed division algebra** to be a unital algebra equipped with a norm

$$|\cdot|\colon A \to [0,\infty)$$

obeying the rule

$$|ab| = |a|\,|b|.$$

The real and complex numbers are obviously normed division algebras. For the quaternions we can define the norm to be:

$$|a + bi + cj + dk| = \sqrt{a^2 + b^2 + c^2 + d^2}.$$

A similar formula works for the octonions:

$$|a_0 + a_1 e_1 + \cdots + a_7 e_7| = \sqrt{a_0^2 + \cdots + a_7^2}.$$

With some sweat, one can check that these rules make $\mathbb{H}$ and $\mathbb{O}$ into normed division algebras.

The marvelous fact is that there are no more! In an 1898 paper, Hurwitz proved that $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ are the only finite-dimensional normed division algebras [29]. In 1960, Urbanik and Wright [43] removed the finite-dimensionality condition:

Theorem 1. *Every normed division algebra is isomorphic to either $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ or $\mathbb{O}$.*

For quantum mechanics, it is important that every normed division algebra is a **$*$-algebra**, meaning it is equipped with a real-linear map

$$\begin{array}{ccc} A & \to & A \\ x & \mapsto & x^* \end{array}$$

obeying these rules:

$$(xy)^* = y^* x^*, \qquad (x^*)^* = x.$$

For $\mathbb{C}$ this map is just complex conjugation, while for $\mathbb{R}$ it is the identity map. For the quaternions it is given by:

$$(a1 + bi + cj + dk)^* = a - bi - cj - dk$$

and for the octonions,

$$(a_0 1 + a_1 e_1 + \cdots + a_7 e_7)^* = a_0 1 - a_1 e_1 - \cdots - a_7 e_7.$$

One can check in all four cases we have

$$x x^* = x^* x = |x|^2 1.$$

For the three associative normed division algebras, the $*$-algebra structure lets us set up a theory of Hilbert spaces. Let us quickly sketch how. Suppose $\mathbb{K}$ is an associative normed division algebra. Then we define a **$\mathbb{K}$-vector space** to be a right $\mathbb{K}$-module: that is, an abelian group V equipped with a map

$$\begin{array}{ccc} V \times \mathbb{K} & \to & V \\ (v,x) & \mapsto & vx \end{array}$$

that obeys the laws

$$(v+w)(x) = vx + wx, \qquad v(x+y) = vx + vy,$$
$$(vx)y = v(xy).$$

We say a map $T\colon V \to V'$ between $\mathbb{K}$-vector spaces is **$\mathbb{K}$-linear** if

$$T(vx + wy) = T(v)x + T(w)y$$

for all $v, w \in V$ and $x, y \in \mathbb{K}$. When no confusion can arise, we call a $\mathbb{K}$-linear map $T\colon V \to V'$ a **linear operator** or simply an **operator**.

The reader may be appalled that we are multiplying by scalars on the *right* here. It makes no difference except for the quaternions, which are noncommutative. Even in that case, we could use either left or right multiplication by scalars, as long as we stick to one convention. But since we write the operator T on the left in the expression $T(v)$, it makes more sense to do scalar multiplication on the right, so no symbols trade places in the law $T(vx) = T(v)x$.

There is a category of $\mathbb{K}$-vector spaces and operators between them. As usual, every vector space has a basis, and any two bases have the same cardinality, so we can talk about the dimension of a vector space over $\mathbb{K}$, and every finite-dimensional vector space over $\mathbb{K}$ is isomorphic to $\mathbb{K}^n$. Every operator $T\colon \mathbb{K}^n \to \mathbb{K}^m$ can be written as matrix:

$$(Tv)_i = \sum_j T_{ij} v_j.$$

The $*$-algebra structure of the normed division algebras becomes important when we study Hilbert spaces. We define an **inner product** on a $\mathbb{K}$-vector space V to be a map

$$\begin{array}{ccc} V \times V & \to & \mathbb{K} \\ (v, w) & \mapsto & \langle v, w \rangle \end{array}$$

that is $\mathbb{K}$-linear in the second argument, has

$$\langle v, w \rangle = \langle w, v \rangle^*$$

for all $v, w \in V$, and is **positive definite**:

$$\langle v, v \rangle \geq 0$$

with equality only if $v = 0$. (In any $*$-algebra we say $y \geq 0$ when $y = xx^*$ for some x.) As usual, an inner product gives a norm:

$$\|v\| = \sqrt{\langle v, v \rangle},$$

and we say a $\mathbb{K}$-vector space with inner product is a **$\mathbb{K}$-Hilbert space** if this norm makes the vector space into a complete metric space. Every finite-dimensional $\mathbb{K}$-Hilbert space is isomorphic to $\mathbb{K}^n$ with its standard inner product

$$\langle v, w \rangle = \sum_{i=1}^n v_i^* w_i.$$

This is fine for $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$: what about $\mathbb{O}$? It would be nice if we could think of $\mathbb{O}^n$ as an n-dimensional octonionic vector space, with scalar multiplication defined in the obvious way. Unfortunately, the law

$$(vx)y = v(xy)$$

fails, because the octonions are nonassociative! Furthermore, there are no maps $T\colon \mathbb{O}^n \to \mathbb{O}^m$ obeying

$$T(vx) = T(v)x$$

except the zero map. So, nobody has managed to develop a good theory of octonionic linear algebra.

While the octonions are nonassociative, they are still **alternative**. That is, the associative law holds whenever two of the elements being multiplied are equal:

$$(xx)y = x(xy) \qquad (xy)x = x(yx) \qquad (yx)x = y(xx).$$

Surprisingly, this is enough to let us carry out a bit of quantum theory as if $\mathbb{O}$, $\mathbb{O}^2$ and $\mathbb{O}^3$ were well-defined octonionic Hilbert spaces. The 3-dimensional case is by far the most exciting: it leads to a structure called the 'exceptional Jordan algebra'. But beyond dimension 3, it seems there is little to say.

2.2 Jordan algebras

In 1932, Pascual Jordan tried to isolate some axioms that an 'algebra of observables' should satisfy [30]. The unadorned phrase 'algebra' usually signals an associative algebra, but this not the kind of algebra Jordan was led to. In both classical and quantum mechanics, observables are closed under addition and multiplication by real scalars. In classical mechanics we can also multiply observables, but in quantum mechanics this becomes problematic. After all, given two bounded self-adjoint operators on a complex Hilbert space, their product is self-adjoint if and only if they commute.

However, in quantum mechanics one can still raise an observable to a power and obtain another observable. From squaring and taking real linear combinations, one can construct a commutative product:

$$a \circ b = \frac{1}{2}((a+b)^2 - a^2 - b^2) = \frac{1}{2}(ab + ba).$$

This product is not associative, but it is **power-associative**: any way of parenthesizing a product of copies of the same observable a gives the same result. This led Jordan to define what is now called a **formally real Jordan algebra**: a real vector space with a bilinear, commutative and power-associative product satisfying

$$a_1^2 + \cdots + a_n^2 = 0 \quad \Longrightarrow \quad a_1 = \cdots = a_n = 0$$

for all n. The last condition gives A a partial ordering: if we write $a \leq b$ when the element $b - a$ is a sum of squares, it says

$$a \leq b \text{ and } b \leq a \quad \Longrightarrow \quad a = b.$$

So, in a formally real Jordan algebra it makes sense to speak of one observable being 'greater' than another.

In 1934, Jordan published a paper with von Neumann and Wigner classifying finite-dimensional formally real Jordan algebras [31]. They began by proving that any such algebra is a direct sum of 'simple' ones. A formally real Jordan algebra is **simple** when its only ideals are $\{0\}$ and A itself, where an **ideal** is a vector subspace $B \subseteq A$ such that $b \in B$ implies $a \circ b \in B$ for all $a \in A$.

And then, they proved:

Theorem 2. *Every simple finite-dimensional formally real Jordan algebra is isomorphic to one on this list:*

- *The algebra $\mathfrak{h}_n(\mathbb{R})$ of $n \times n$ self-adjoint real matrices with the product $a \circ b = \frac{1}{2}(ab + ba)$.*
- *The algebra $\mathfrak{h}_n(\mathbb{C})$ of $n \times n$ self-adjoint complex matrices with the product $a \circ b = \frac{1}{2}(ab + ba)$.*
- *The algebra $\mathfrak{h}_n(\mathbb{H})$ of $n \times n$ self-adjoint quaternionic matrices with the product $a \circ b = \frac{1}{2}(ab + ba)$.*
- *The algebra $\mathfrak{h}_3(\mathbb{O})$ of 3×3 self-adjoint octonionic matrices with the product $a \circ b = \frac{1}{2}(ab + ba)$.*
- *The algebras $\mathbb{R}^n \oplus \mathbb{R}$ with the product*

$$(x, t) \circ (x', t') = (tx' + t'x, x \cdot x' + tt').$$

Here we say a square matrix T is **self-adjoint** if $T_{ji} = (T_{ij})^*$. For $\mathbb{K} = \mathbb{R}, \mathbb{C}$, or $\mathbb{H}$, we can identify a self-adjoint $n \times n$ matrix with an operator $T \colon \mathbb{K}^n \to \mathbb{K}^n$ that is **self-adjoint** in the sense that

$$\langle Tv, w \rangle = \langle v, Tw \rangle$$

for all $v, w \in \mathbb{K}^n$. In the octonionic case we do not know what Hilbert spaces and operators are, but we can still work with matrices. Curiously, in this case we cannot go beyond 3×3 self-adjoint matrices and still get a Jordan algebra. The 1×1 self-adjoint octonionic matrices are just the real numbers, and the 2×2 ones form a Jordan algebra that is isomorphic to a spin factor. The 3×3 self-adjoint octonionic matrices are the really interesting case: these form a 27-dimensional formally real Jordan algebra called the **exceptional Jordan algebra**.

What does all this mean for physics? The spin factors have an intriguing relation to special relativity, since $\mathbb{R}^n \oplus \mathbb{R}$ can be identified with $(n+1)$-dimensional Minkowski spacetime, and its cone of positive elements is then revealed to be none other than the future lightcone. Furthermore, we have some interesting coincidences:

- The Jordan algebra $\mathfrak{h}_2(\mathbb{R})$ is isomorphic to the spin factor $\mathbb{R}^2 \oplus \mathbb{R}$.
- The Jordan algebra $\mathfrak{h}_2(\mathbb{C})$ is isomorphic to the spin factor $\mathbb{R}^3 \oplus \mathbb{R}$.
- The Jordan algebra $\mathfrak{h}_2(\mathbb{H})$ is isomorphic to the spin factor $\mathbb{R}^5 \oplus \mathbb{R}$.
- The Jordan algebra $\mathfrak{h}_2(\mathbb{O})$ isomorphic to the spin factor $\mathbb{R}^9 \oplus \mathbb{R}$.

This sets up a relation between the real numbers, complex numbers, quaternions and octonions and the Minkowski spacetimes of dimensions 3,4,6 and 10. These are precisely the dimensions where a classical superstring Lagrangian can be written down! Far from being a coincidence, this is the tip of a huge and still not fully fathomed iceberg, which we have discussed elsewhere [9, 11, 12].

The exceptional Jordan algebra remains mysterious. Practically ever since it was discovered, physicists have looked for some application of this entity. For example, when it was first found that quarks come in three colors, Okubo and others hoped that 3×3 self-adjoint octonionic matrices might serve as observables for these exotic degrees of freedom [36]. Alas, nothing much came of this. More recently, people have discovered some relationships between 10-dimensional string theory and the exceptional Jordan algebra, arising from the fact that the 2×2 self-adjoint octonionic matrices can be identified with 10-dimensional Minkowski spacetime [23]. But nothing definitive has emerged so far.

In 1983, Zelmanov generalized the Jordan–von Neumann–Wigner classification to the infinite-dimensional case, working with Jordan algebras that need not be formally real [47]. In any formally real Jordan algebra, the following peculiar law holds:

$$(a^2 \circ b) \circ a = a^2 \circ (b \circ a).$$

Any vector space with a commutative bilinear product obeying this law is called a **Jordan algebra**. Zelmanov classified the simple Jordan algebras and proved they are all of three kinds: a kind generalizing Jordan algebras of self-adjoint matrices, a kind generalizing spin factors, and an 'exceptional' kind. For a good introduction to this work, see McCrimmon's book [34].

Zelmanov's classification theorem does not highlight the special role of the reals, complex numbers and quaternions. Indeed, when we drop the 'formally real' condition, a host of additional finite-dimensional simple Jordan algebras appear, beside those in Jordan, von Neumann and Wigner's classification theorem. While it is an algebraic *tour de force*, nobody has yet used Zelmanov's theorem to shed new light on quantum theory. So, we postpone infinite-dimensional considerations until Section 2.4, where we discuss the remarkable theorem of Solèr.

2.3 Convex cones

The formalism of Jordan algebras seems rather removed from the actual practice of physics, because in quantum theory we hardly ever take two observables a and b and form their Jordan product $\frac{1}{2}(ab + ba)$. As hinted in the previous section, it is better to think of this operation as derived from the process of *squaring* an observable, which is something we actually do. But still, we cannot help wondering: does the classification of finite-dimensional formally real Jordan algebras, and thus the special role of normed division algebras, arise from some axiomatic framework more closely tied to quantum physics as it usually practiced?

One answer involves the correspondence between states and observables. Consider first the case of ordinary quantum theory. If a quantum system has the Hilbert space $\mathbb{C}^n$, observables are described by self-adjoint $n \times n$ complex matrices: elements of the Jordan algebra $\mathfrak{h}_n(\mathbb{C})$. But matrices of this form that are nonnegative and have trace 1 also play another role. They are called **density matrices**, and they describe *states* of our quantum system: not just pure states, but also more general mixed states. The idea is that any density matrix $\rho \in \mathfrak{h}_n(\mathbb{C})$ allows us to define expectation values of observables $a \in \mathfrak{h}_n(\mathbb{C})$ via

$$\langle a \rangle = \mathrm{tr}(\rho a).$$

The map sending observables to their expectation values is real-linear. The fact that ρ is nonnegative is equivalent to

$$a \geq 0 \implies \langle a \rangle \geq 0$$

and the fact that ρ has trace 1 is equivalent to

$$\langle 1 \rangle = 1.$$

Of course, states lie in the dual of the space of observables: that much is obvious, given that the expectation value of an observable should depend linearly on the observable. The more mysterious thing is that we can identify the vector space of observables with its dual:

$$\begin{array}{ccc} \mathfrak{h}_n(\mathbb{C}) & \cong & \mathfrak{h}_n(\mathbb{C})^* \\ a & \mapsto & \langle a, \cdot \rangle \end{array}$$

using the trace, which puts a real-valued inner product on the space of observables:

$$\langle a, b \rangle = \mathrm{tr}(ab).$$

Thus, states also *correspond to* certain observables: the nonnegative ones having trace 1. We call this fact the **state-observable correspondence**.

All this generalizes to an arbitrary finite-dimensional formally real Jordan algebra A. Every such algebra automatically has an identity element [31]. This lets us define a **state** on A to be a linear functional $\langle \cdot \rangle \colon A \to \mathbb{R}$ that is **nonnegative**:

$$a \geq 0 \implies \langle a \rangle \geq 0$$

and **normalized**:

$$\langle 1 \rangle = 1.$$

But in fact, there is a one-to-one correspondence between linear functionals on A and elements of A. The reason is that every finite-dimensional Jordan algebra has a **trace**

$$\mathrm{tr} \colon A \to \mathbb{R}$$

defined so that $\mathrm{tr}(a)$ is the trace of the linear operator

$$\begin{array}{ccc} A & \to & A \\ b & \mapsto & a \circ b. \end{array}$$

Such a Jordan algebra is then formally real if and only if

$$\langle a, b \rangle = \mathrm{tr}(a \circ b)$$

is a real-valued inner product. So, when A is a finite-dimensional formally real Jordan algebra, any linear functional $\langle \cdot \rangle \colon A \to \mathbb{R}$ can be written as

$$\langle a \rangle = \mathrm{tr}(\rho \circ a)$$

for a unique element $\rho \in A$. Conversely, every element $\rho \in A$ gives a linear functional by this formula. While not obvious, it is true that the linear functional $\langle \cdot \rangle$ is nonnegative if and only if $\rho \geq 0$ in terms of the ordering on A. More obviously, $\langle \cdot \rangle$ is normalized if and only if $\mathrm{tr}(\rho) = 1$. So, states can be identified with certain special observables: namely, those observables $\rho \in A$ with $\rho \geq 0$ and $\mathrm{tr}(\rho) = 1$.

In short: whenever the observables in our theory form a finite-dimensional formally real Jordan algebra, we have an state-observable correspondence. But what is the physical meaning of the state-observable correspondence? Why in the world should *states* correspond to special *observables?*

Here is one attempt at an answer. Every finite-dimensional formally real Jordan algebra comes equipped with a distinguished observable, the most boring one of all: the identity, $1 \in A$. This is nonnegative, so if we normalize it, we get an observable

$$\rho_0 = \frac{1}{\mathrm{tr}(1)}\, 1 \in A$$

of the special kind that corresponds to a state. This state, say $\langle \cdot \rangle_0$, is just the normalized trace:

$$\langle a \rangle_0 = \mathrm{tr}(\rho_0 \circ a) = \frac{\mathrm{tr}(a)}{\mathrm{tr}(1)}.$$

And this state has a clear physical meaning: it is the **state of maximal ignorance**! It is the state where we know as little as possible about our system—or more precisely, at least in the case of ordinary complex quantum theory, the state where entropy is maximized.

For example, consider $A = \mathfrak{h}_2(\mathbb{C})$, the algebra of observables of a spin-$\frac{1}{2}$ particle. Then the space of states is the so-called 'Bloch sphere', really a 3-dimensional ball. On the surface of this ball are the pure states, the states where we know as much as possible about the particle. At the center of the ball is the state of maximum ignorance. This corresponds to the density matrix

$$\rho_0 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

In this state, when we measure the particle's spin along any axis, we have a $\frac{1}{2}$ chance of getting spin up and a $\frac{1}{2}$ chance of spin down.

Returning to the general situation, note that A always acts on its dual A^*: given $a \in A$ and a linear functional $\langle \cdot \rangle$, we get a new linear functional $\langle a \circ \cdot \rangle$. This captures the idea, familiar in quantum theory, that observables are also 'operators'. The state-observable correspondence arises naturally from this: we can get any state by acting on the state of complete ignorance with a suitable observable. After all, any state corresponds to some observable ρ, as follows:

$$\langle a \rangle = \mathrm{tr}(\rho \circ a)$$

So, we can get this state by acting on the state of maximal ignorance, $\langle \cdot \rangle_0$, by the observable $\mathrm{tr}(1)\rho$:

$$\langle \mathrm{tr}(1)\rho \circ a \rangle_0 = \frac{\mathrm{tr}(1)}{\mathrm{tr}(1)} \mathrm{tr}(\rho \circ a) = \langle a \rangle.$$

So, we see that the state-observable correspondence springs from two causes. First, there is a distinguished state, the state of maximal ignorance. Second, any other state can be obtained from the state of maximal ignorance by acting on it with a suitable observable.

While these thoughts raise a host of questions, they also help motivate an important theorem of Koecher and Vinberg [32, 46]. The idea is to axiomatize the situation we we have just described, in a way that does not mention the Jordan product in A, but instead emphasizes:

- the observable-state correspondence, and

- the fact that 'positive' observables, namely those whose observed values are always positive, form a cone.

To find appropriate axioms, suppose A is a finite-dimensional formally real Jordan algebra. Then seven facts are always true. First, the set of positive observables

$$C = \{a \in A \colon a > 0\}.$$

is a **cone**: that is, $a \in C$ implies that every positive multiple of a is also in C. Second, this cone is **convex**: if $a, b \in C$ then any linear combination $xa + (1-x)b$ with $0 \leq x \leq 1$ also lies in C. Third, it is an open set. Fourth, it is **regular**, meaning that if a and $-a$ are both in the closure $\overline{C}$, then $a = 0$. This last condition may seem obscure, but if we note that

$$\overline{C} = \{a \in A \colon a \geq 0\}$$

we see that C being regular simply means

$$a \geq 0 \quad \text{and} \quad -a \geq 0 \qquad \Longrightarrow \qquad a = 0,$$

a perfectly plausible assumption.

Next recall that A has an inner product; this is what lets us identify linear functionals on A with elements of A. This also lets us define the **dual cone**

$$C^* = \{a \in A \colon \ \forall b \in A \ \langle a, b\rangle > 0\}$$

which one can check is indeed a cone. The fifth fact about C is that it is **self-dual**, meaning $C = C^*$. This formalizes the observable-state correspondence, since it means that states correspond to special observables. All the elements $a \in C$ are positive observables, but certain special ones, namely those with $\langle a, 1\rangle = 1$, can also be viewed as states.

The sixth fact is that C is **homogeneous**: given any two points $a, b \in C$, there is a real-linear transformation $T \colon A \to A$ mapping C to itself in a one-to-one and onto way, with the property that $Ta = b$. This says that cone C is highly symmetrical: no point of C is any 'better' than any other, at least if we only consider the linear structure of the space A, ignoring the Jordan product and the trace.

From another viewpoint, however, there is a very special point of C, namely the identity 1 of our Jordan algebra. And this brings us to our seventh and final fact: the cone C is **pointed**, meaning that it is equipped with a distinguished element (in this case $1 \in C$). As we have seen, this element corresponds to the 'state of complete ignorance', at least after we normalize it.

In short: when A is a finite-dimensional formally real Jordan algebra, C is a pointed homogeneous self-dual regular open convex cone.

In fact, there is a category of pointed homogeneous self-dual regular open convex cones, where:

- An object is a finite-dimensional real inner product space V equipped with a pointed homogeneous self-dual regular open convex cone $C \subset V$.

- A morphism from one object, say (V, C), to another, say (V', C'), is a linear map $T \colon V \to V'$ preserving the inner product and mapping C into C'.

Now for the payoff. The work of Koecher and Vinberg [32, 46], nicely explained in Koecher's Minnesota notes [33], shows that:

Theorem 3. *The category of pointed homogeneous self-dual regular open convex cones is equivalent to the category of finite-dimensional formally real Jordan algebras.*

This means that the theorem of Jordan, von Neumann and Wigner, described in the previous section, also classifies the pointed homogeneous self-dual regular convex cones!

Theorem 4. *Every pointed homogeneous self-dual regular open convex cones is isomorphic to a direct sum of those on this list:*

- *the cone of positive elements in* $\mathfrak{h}_n(\mathbb{R})$,
- *the cone of positive elements in* $\mathfrak{h}_n(\mathbb{C})$,
- *the cone of positive elements in* $\mathfrak{h}_n(\mathbb{H})$,
- *the cone of positive elements in* $\mathfrak{h}_3(\mathbb{O})$,
- *the future lightcone in* $\mathbb{R}^n \oplus \mathbb{R}$.

Some of this deserves a bit of explanation. For $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{H}$, an element $T \in \mathfrak{h}_n(\mathbb{K})$ is **positive** if and only if the corresponding operator $T\colon \mathbb{K}^n \to \mathbb{K}^n$ has

$$\langle v, Tv \rangle > 0$$

for all nonzero $v \in \mathbb{K}^n$. A similar trick works for defining positive elements of $\mathfrak{h}_3(\mathbb{O})$, but we do not need the details here. We say an element $(x, t) \in \mathbb{R}^n \oplus \mathbb{R}$ lies in the **future lightcone** if $t > 0$ and $t^2 - x \cdot x > 0$. This of course fits in nicely with the idea that the spin factors are connected to Minkowski spacetimes. Finally, there is an obvious notion of direct sum for Euclidean spaces with cones, where the direct sum of (V, C) and (V', C') is $V \oplus V'$ equipped with the cone

$$C \oplus C' = \{(v, v') \in V \oplus V' \colon \; v \in C, v' \in C'\}.$$

In summary, self-adjoint operators on real, complex and quaternionic Hilbert spaces arise fairly naturally as observables starting from a formalism where the observables form a cone, and we insist on an observable-state correspondence. There is a well-developed approach to probabilistic theories that works for cones that are neither self-dual nor homogeneous: see for example the work of Barnum and coauthors [15, 16]. This has already allowed these authors to shed new light on the physical significance of self-duality [17]. But perhaps further thought on the state-observable correspondence will clarify the meaning of the Koecher-Vinberg theorem, and help us better understand the appearance of normed division algebras in quantum theory.

Finally, we should mention that in pondering the state-observable correspondence, it is worthwhile comparing the 'state-operator correspondence'. This is best known in the context of string theory [39], but it really applies whenever we have a C^*-algebra of observables, say A, equipped with a state $\langle \cdot \rangle \colon A \to \mathbb{C}$. Then the Gelfand-Naimark-Segal construction lets us build a Hilbert space H on which A acts, together with a distinguished unit vector $v \in H$ called the 'vacuum state'. The Hilbert space H is built by taking a completion of a quotient of A, so a dense set of vectors in H come from elements of A. So, we get states from certain observables. In particular, the vacuum state v comes from the element $1 \in A$.

This is reminiscent of how in our setup, the state of maximal ignorance comes from the element 1 in the Jordan algebra of observables. But there are also some differences: for example, the Gelfand-Naimark-Segal construction requires choosing a state, and it works for infinite-dimensional C^*-algebras, while our construction works for finite-dimensional formally real Jordan algebras, which have a canonical state: the state of maximum ignorance. Presumably both constructions are special cases of some more general construction.

2.4 Solèr's theorem

In 1995, Maria Pia Solèr [41] proved a result which has powerful implications for the foundations of quantum mechanics. The idea of Solèr's theorem is to generalize the concept of Hilbert space beyond real, complex and quaternionic Hilbert spaces, but then show that under very mild conditions, the only options for *infinite-dimensional* Hilbert spaces are the usual ones! Our discussion here is largely based on the paper by Holland [28].

To state Solèr's theorem, we begin by saying what sort of ring our generalized Hilbert spaces will be vector spaces over. We want rings where we can divide, and we want them to have an operation like complex conjugation. A ring is called a **division ring** if every element x has an element x^{-1} for which $xx^{-1} = x^{-1}x = 1$. It is called a **$*$-ring** if it is equipped with a function $x \mapsto x^*$ for which:

$$(x+y)^* = x^* + y^*, \qquad (xy)^* = y^* x^*, \qquad (x^*)^* = x.$$

A **division $*$-ring** is a $*$-ring that is also a division ring.

Now, suppose $\mathbb{K}$ is a division $*$-ring. As before, we define a **$\mathbb{K}$-vector space** to be a right $\mathbb{K}$-module. And as before, we say a function $T\colon V \to V'$ between $\mathbb{K}$-vector spaces is **$\mathbb{K}$-linear** if

$$T(vx + wy) = T(v)x + T(w)y$$

for all $v, w \in V$ and $x, y \in \mathbb{K}$. We define a **hermitian form** on a $\mathbb{K}$-vector space V to be a map

$$\begin{array}{ccc} V \times V & \to & \mathbb{K} \\ (v,w) & \mapsto & \langle v, w\rangle \end{array}$$

which is $\mathbb{K}$-linear in the second argument and obeys the equation

$$\langle v, w\rangle = \langle w, v\rangle^*$$

for all $v, w \in V$. We say a hermitian form is **nondegenerate** if

$$\langle v, w\rangle = 0 \quad \text{for} \quad \text{all} \quad v \in V \qquad \iff \qquad w = 0.$$

Suppose H is a $\mathbb{K}$-vector space with a nondegenerate hermitian form. Then we can define various concepts familiar from the theory of Hilbert spaces. For example, we say a sequence $e_i \in H$ is **orthonormal** if $\langle e_i, e_j\rangle = \delta_{ij}$. For any vector subspace $S \subseteq H$, we define the **orthogonal complement** $S^\perp$ by

$$S^\perp = \{w \in H\colon \ \langle v, w\rangle = 0 \quad \text{for} \quad \text{all} \quad v \in S\}.$$

We say S is **closed** if $S^{\perp\perp} = S$. Finally, we say H is **orthomodular** if $S + S^\perp = H$ for every closed subspace S. This is automatically true if H is a real, complex, or quaternionic Hilbert space.

We are now ready to state Solèr's theorem:

Theorem 5. *Let $\mathbb{K}$ be a division $*$-ring, and let H be a $\mathbb{K}$-vector space equipped with an orthomodular hermitian form for which there exists an infinite orthonormal sequence. Then $\mathbb{K}$ is isomorphic to $\mathbb{R}, \mathbb{C}$ or $\mathbb{H}$, and H is an infinite-dimensional $\mathbb{K}$-Hilbert space.*

Nothing in the assumptions mentions the continuum: the hypotheses are purely algebraic. It therefore seems quite magical that $\mathbb{K}$ is forced to be the real numbers, complex numbers or quaternions. The orthomodularity condition is the key. In 1964, Piron had attempted to prove that any orthomodular complex inner product space must be a Hilbert space [38]. His proof had a flaw which Ameniya and Araki fixed a few years later [5]. Their proof also handles the real and quaternionic cases. Eventually these ideas led Solèr to realize that orthomodularity picks out real, complex, and quaternionic Hilbert spaces as special.

Holland's review article [28] describes the implications of Solèr's results for other axiomatic approaches to quantum theory: for example, approaches using convex cones and lattices. This article is so well-written that there is no point in summarizing it here. Instead, we turn to some problems with real and quaternionic quantum theory, which can be solved by treating them as part of a larger structure—the 'three-fold way'—that also includes the complex theory.

3 Problems

Is complex quantum theory 'better' than the real or quaternionic theories? Of course it fits the experimental data better. But people have also pointed to various internal problems, or at least unfamiliar aspects, of the real and quaternionic theories, as a way to justify nature's choice of the complex theory. Here we describe two problems, both of which can be solved using the 'three-fold way'. For more on the special virtues of complex quantum theory see Hardy [27] and Vicary [45].

One problem with real and quaternionic quantum theory is that usual correspondence between one-parameter unitary groups and self-adjoint operators breaks down, or at least becomes more subtle. To see this with a minimum of distractions, *all Hilbert spaces in this section will be assumed finite-dimensional*. The same issues show up in the infinite-dimensional case, but more technical detail is required to tell the story correctly.

Suppose $\mathbb{K}$ is $\mathbb{R}, \mathbb{C}$ or $\mathbb{H}$. Let H and H' be $\mathbb{K}$-Hilbert spaces. We say an operator $T\colon H \to H'$ is **bounded** if there exists a constant $K > 0$ such that

$$\|Tv\| \leq K\,\|v\|$$

for all $v \in H$. We define the **adjoint** of a bounded operator $T\colon H \to H'$ in the usual way:

$$\langle u, T^\dagger v\rangle = \langle Tu, v\rangle$$

for all $u \in H$, $v \in H'$. It is easy to check that $T^\dagger$, defined this way, really is an operator from H' back to H. We define a $\mathbb{K}$-linear operator $U\colon H \to H'$ to be **unitary** if $UU^\dagger = U^\dagger U = 1$. Here we should warn the reader that when $\mathbb{K} = \mathbb{R}$, the term 'orthogonal' is often used instead of 'unitary'—and when $\mathbb{K} = \mathbb{H}$, people sometimes use the term 'symplectic'. For us it will be more efficient to use the same word in all three cases.

We define a **one-parameter unitary group** to be a family of unitary operators $U(t)\colon H \to H$, one for each $t \in \mathbb{R}$, such that

$$U(t + t') = U(t)U(t')$$

for all $t, t' \in \mathbb{R}$. If H is finite-dimensional, every one-parameter unitary group can be written as

$$U(t) = \exp(tS)$$

for a unique bounded operator S.

It is easy to check that the operator S is **skew-adjoint**: it satisfies $S^\dagger = -S$. In quantum theory we usually want our observables to be **self-adjoint** operators, obeying $A^\dagger = A$. And indeed, when $\mathbb{K} = \mathbb{C}$, we can write any skew-adjoint operator S in terms of a self-adjoint operator A by setting

$$S(v) = A(v)i.$$

This gives the usual correspondence between one-parameter unitary groups and self-adjoint operators.

Alas, when $\mathbb{K} = \mathbb{R}$ we have no number i, so we cannot express our skew-adjoint S in terms of a self-adjoint A. Nor can we do it when $\mathbb{K} = \mathbb{H}$. Now the problem is not a shortage of square roots of -1. Instead, there are too many—and more importantly, they do not commute! We can try to set $A(v) = S(v)i$, but this operator A will rarely be linear. The reason is that because S commutes with multiplication by j, A anticommutes with multiplication by j, so A is only linear in the trivial case $S = 0$.

Later we shall see a workaround for this problem. A second problem, special to the quaternionic case, concerns tensoring Hilbert spaces. In ordinary complex quantum theory, when we have two systems, one with Hilbert space H and one with Hilbert space H', the system made by combining these two systems has Hilbert space $H \otimes H'$. Here the tensor product of Hilbert spaces relies on a more primitive concept: the tensor product of vector spaces. The tensor product of bimodules of a noncommutative algebra is another bimodule over that algebra. But a quaternionic vector space is not a bimodule: it is only a left module of the quaternions. So, the tensor product of quaternionic Hilbert spaces is not naturally a quaternionic Hilbert space.

Unfortunately, this issue is not addressed head-on in Adler's book on quaternionic quantum theory [4]. When tensoring two quaternionic Hilbert spaces, he essentially chooses a way to make one of them into a bimodule, without being very explicit about this. There is no canonical way to make a quaternionic Hilbert space H into a bimodule. Indeed, given one way to do it, we can get many new ways as follows:

$$(xv)_{\mathrm{new}} = \alpha(x)v, \qquad (vx)_{\mathrm{new}} = vx,$$

where $v \in H$, $x \in \mathbb{H}$ and α is an automorphism of the quaternions. Every automorphism is of the form

$$\alpha(x) = gxg^{-1}$$

for some unit quaternion g, so the automorphism group of the quaternions is $\mathrm{SU}(2)/\{\pm 1\} \cong \mathrm{SO}(3)$. Thus we can 'twist' a bimodule structure on H by any element of $\mathrm{SO}(3)$, obtaining a new bimodule with the same underlying quaternionic Hilbert space.

In Bartels' work on quaternionic functional analysis, he used bimodules right from the start [18]. Later Ng [35] investigated these questions in more depth, and his paper provides a nice overview of many different approaches. Here however we shall take a slightly different tack, noting that there is a way to make the tensor product of quaternionic Hilbert spaces into a *real* Hilbert space. This fits nicely into the thesis we wish to advocate: that real, complex and quaternionic quantum mechanics are not separate theories, but fit together as parts of a unified whole.

4 The Three-Fold Way

Some aspects of quantum theory become more visible when we introduce *symmetry*. This is especially true when it comes to the relation between real, complex and quaternionic quantum theory. So, instead of bare Hilbert spaces, it is useful to consider Hilbert spaces equipped with a representation of a group. For simplicity suppose that G is a Lie group, where we count a discrete group as a 0-dimensional Lie group. Let $\mathrm{Rep}(G)$ be the category where:

- An object is a finite-dimensional complex Hilbert space H equipped with a continuous unitary representation of G, say $\rho\colon G \to \mathrm{U}(H)$.
- A morphism is an operator $T\colon H \to H'$ such that $T\rho(g) = \rho'(g)T$ for all $g \in G$.

To keep the notation simple, we often call a representation $\rho\colon G \to \mathrm{U}(H)$ simply by the name of its underlying Hilbert space, H. Another common convention, used in the introduction, is to call it ρ.

Many of the operations that work for finite-dimensional Hilbert spaces also work for $\mathrm{Rep}(G)$, with the same formal properties: for example, direct sums, tensor products and duals. This lets us formulate the **three-fold way** as follows. If an object $H \in \mathrm{Rep}(G)$ is **irreducible**—not a direct sum of other representations in a nontrivial way—there are three mutually exclusive choices:

- The representation H is not isomorphic to its dual. In this case we call it **complex**.
- The representation H is isomorphic to its dual and it is **real**: it arises from a representation of G on a real Hilbert space $H_{\mathbb{R}}$ as follows:

$$H = H_{\mathbb{R}} \otimes \mathbb{C}$$

- The representation H is isomorphic to its dual and it is **quaternionic**: it comes from a representation of G on a quaternionic Hilbert space $H_{\mathbb{H}}$ with

$$H = \text{the underlying complex representation of } \mathbb{H}_{\mathbb{H}}$$

Why? Suppose that $H \in \mathrm{Rep}(G)$ is irreducible. Then there is a 1-dimensional space of morphisms $f\colon H \to H$, by Schur's Lemma. Since $H \cong H^*$, there is also a 1-dimensional space of morphisms $T\colon H \to H^*$, and thus a 1-dimensional space of morphisms

$$g\colon H \otimes H \to \mathbb{C}$$

We can also think of these as bilinear maps

$$g\colon H \times H \to \mathbb{C}$$

that are invariant under the action of G on H.

But the representation $H \otimes H$ is the direct sum of two others: the space S^2H of symmetric tensors, and the space $\Lambda^2 H$ of antisymmetric tensors:

$$H \otimes H \cong S^2H \oplus \Lambda^2 H\,.$$

So, either there exists a nonzero g that is **symmetric**:

$$g(v,w) = g(w,v)$$

or a nonzero g that is **antisymmetric**:

$$g(v,w) = -g(w,v)\,.$$

One or the other, not both!—for if we had both, the space of morphisms $g\colon H \otimes H \to \mathbb{C}$ would be at least two-dimensional.

Either way, we can write

$$g(v,w) = \langle Jv, w\rangle$$

for some function $J\colon H \to H$. Since g and the inner product are both invariant under the action of G, this function J must commute with the action of G. But note that since g is linear in the first slot, while the inner product is not, J must be **antilinear**, meaning

$$J(vx + wy) = J(v)x^* + J(w)y^*$$

for all $v, w \in V$ and $x, y \in \mathbb{K}$.

The square of an antilinear operator is linear. Thus, J^2 is linear and it commutes with the action of G. By Schur's Lemma, it must be a scalar multiple of the identity:

$$J^2 = c$$

for some $c \in \mathbb{C}$. We wish to show that depending on whether g is symmetric or antisymmetric, we can rescale J to achieve either $J^2 = 1$ or $J^2 = -1$. To see this, first note that depending on whether g is symmetric or antisymmetric, we have

$$\pm g(v,w) = g(w,v)$$

and thus

$$\pm\langle Jv, w\rangle = \langle Jw, v\rangle.$$

Choosing $v = Jw$, we thus have

$$\pm\langle J^2 w, w\rangle = \langle Jw, Jw\rangle \ge 0.$$

It follows that $\pm J^2$ is a positive operator, so $c > 0$. Thus, dividing J by the positive square root of c, we obtain a new antilinear operator — let us again call it J — with $J^2 = \pm 1$.

Rescaled this way, J is **antiunitary**: it is an invertible antilinear operator with

$$\langle Jv, Jw\rangle = \langle w, v\rangle.$$

Now consider the two cases:

- If g is symmetric, H is equipped with a **real structure**: an antiunitary operator J with

$$J^2 = 1.$$

Furthermore J commutes with the action of G. It follows that the real Hilbert space

$$H_{\mathbb{R}} = \{x \in H\colon\; Jx = x\}$$

is a representation of G whose tensor product with $\mathbb{C}$ is H.

- If g is antisymmetric, H is equipped with a **quaternionic structure**: an antiunitary operator J with

$$J^2 = -1.$$

Furthermore J commutes with the action of G. Since the operators $I = i$, J, and $K = IJ$ obey the usual quaternion relations, H can be made into a quaternionic Hilbert space $H_{\mathbb{H}}$. There is a representation of G on $H_{\mathbb{H}}$ whose underlying complex representation is H.

Note that in both cases, g is **nondegenerate**, meaning

$$\forall v \in V \;\; g(v,w) = 0 \quad \Longrightarrow \quad w = 0.$$

The reason is that $g(v,w) = \langle Jv, w\rangle$, and the inner product is nondegenerate, while J is one-to-one.

We can describe real and quaternionic representations using either g or J. In the following statements, we do not need the representation to be irreducible:

- Given a complex Hilbert space H, a nondegenerate symmetric bilinear form $g\colon H \times H \to \mathbb{C}$ is called an **orthogonal structure** on H. So, a representation $\rho\colon G \to \mathrm{U}(H)$ is real iff it preserves some orthogonal structure g on H. Alternatively, ρ is real iff there is a real structure $J\colon H \to H$ that commutes with the action of G.

- Similarly, given a complex Hilbert space H, a nondegenerate skew-symmetric bilinear form $g\colon H \times H \to \mathbb{C}$ is called a **symplectic structure** on H. So, a representation $\rho\colon G \to \mathrm{U}(H)$ is quaternionic iff it preserves some symplectic structure g on H. Alternatively, ρ is quaternionic iff there is a quaternionic structure $J\colon H \to H$ that commutes with the action of G.

We can summarize these patterns as follows:

THE THREEFOLD WAY

complex	$H \ncong H^*$	unitary
real	$H \cong H^*$ $J^2 = 1$	orthogonal
quaternionic	$H \cong H^*$ $J^2 = -1$	symplectic

For more on how these patterns pervade mathematics, see Dyson's paper on the three-fold way [24], and Arnold's paper on mathematical 'trinities' [7]. In the next section we focus on a few applications to physics.

5 Applications

Let us consider an example: $G = \mathrm{SU}(2)$. In physics this group is important because its representations describe the ways a particle can transform under rotations. There is one one irreducible representation for each 'spin' $j = 0, \frac{1}{2}, 1, \ldots$. When j is an integer, this representation describes the angular momentum states of a boson; when j is a half-integer (meaning an integer plus $\frac{1}{2}$) it describes the angular momentum states of a fermion.

Mathematically, SU(2) is special in many ways. For one thing, all its representations are self-dual. When j is an integer, the spin-j representation is real; when j is an integer plus $\frac{1}{2}$, the spin-j representation is quaternionic. We shall see why this is true shortly, but we cannot resist making a little table that summarizes the pattern:

IRREDUCIBLE REPRESENTATIONS OF SU(2)

$j \in \mathbb{Z}$	bosonic	real	$J^2 = 1$	orthogonal
$j \in \mathbb{Z} + \frac{1}{2}$	fermionic	quaternionic	$J^2 = -1$	symplectic

By the results described in the last section, we conclude that the integer-spin or 'bosonic' irreducible representations of SU(2) are equipped with an invariant antiunitary operator J with $J^2 = 1$, and an invariant orthogonal structure. The half-integer-spin or 'fermionic' ones are equipped with an invariant antiunitary operator J with $J^2 = -1$, and an invariant symplectic structure.

Why does it work this way? First, consider the spin-$\frac{1}{2}$ representation. We can identify the group of unit quaternions (that is, quaternions with norm 1) with SU(2) as follows:

$$a + bi + cj + dk \;\mapsto\; a\sigma_0 - ib\sigma_1 - ic\sigma_2 - id\sigma_3$$

where the σ's are the Pauli matrices. This lets us think of $\mathbb{C}^2$ as the underlying complex vector space of $\mathbb{H}$ in such a way that the obvious representation of SU(2) on $\mathbb{C}^2$ gets identified with the action of unit quaternions on $\mathbb{H}$ by left multiplication. Since left multiplication commutes with right multiplication, this action is quaternion-linear. Thus, the spin-$\frac{1}{2}$ representation of SU(2) is quaternionic.

Next, note that can build all the irreducible unitary representations of SU(2) by forming symmetrized tensor powers of the spin-$\frac{1}{2}$ representation. The nth symmetrized tensor power, $S^n(\mathbb{C}^2)$, is the spin-j representation with $j = n/2$. At this point a well-known general result will help:

Theorem 6. *Suppose G is a Lie group and $H, H' \in \mathrm{Rep}(G)$. Then:*

- *H and H' are real $\implies$ $H \otimes H'$ is real.*
- *H is real and H' is quaternionic $\implies$ $H \otimes H'$ is quaternionic.*
- *H' is quaternionic and H' is real $\implies$ $H \otimes H'$ is quaternionic.*
- *H and H' are quaternionic $\implies$ $H \otimes H'$ is real.*

Proof. As we saw in the last section, H is real (resp. quaternionic) if and only if it can be equipped with an invariant antiunitary operator $J\colon H \to H$ with $J^2 = 1$ (resp. $J^2 = -1$). So, pick such an antiunitary J for H and also one J' for H'. Then $J \otimes J'$ is an invariant antiunitary operator on $H \otimes H'$, and

$$(J \otimes J')^2 = J^2 \otimes {J'}^2.$$

This makes the result obvious. ∎

In short, the 'multiplication table' for tensoring real and quaternionic representations is just like the multiplication table for the numbers 1 and -1—and also, not coincidentally, just like the usual rule for combining bosons and fermions.

Next, note that any subrepresentation of a real representation is real, and any subrepresentation of a quaternionic representation is quaternionic. It follows that since the spin-$\frac{1}{2}$ representation of $\mathrm{SU}(2)$ is quaternionic, its nth tensor power is real or quaternionic depending on whether n is even or odd, and similarly for the subrepresentation $S^n(\mathbb{C}^2)$. This is the spin-j representation for $j = n/2$. So, the spin-j representation is real or quaternionic depending on whether j is an integer or half-integer.

But what is the physical meaning of the antiunitary operator J on the spin-j representation? Let us call this representation $\rho\colon \mathrm{SU}(2) \to \mathrm{U}(H)$, where H is the Hilbert space $S^n(\mathbb{C}^2)$. Choose any element $X \in \mathfrak{su}(2)$ and let $S = d\rho(X)$. Then

$$J \exp(tS) = \exp(tS)\, J$$

for all $t \in \mathbb{R}$, and thus differentiating, we obtain

$$JS = SJ.$$

But the operator S is skew-adjoint. In quantum mechanics, we write $S = iA$ where A is the self-adjoint operator corresponding to angular momentum along some axis. Since J anticommutes with i, we have

$$JA = -AJ$$

so for every $v \in H$ we have

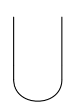
$$\langle Jv, AJv\rangle = -\langle JV, JAv\rangle = -\langle Av, v\rangle = -\langle v, Av\rangle.$$

Thus, the antiunitary operator J reverses angular momentum. Since the time-reversed version of a spinning particle is a particle spinning the opposite way, physicists call J **time reversal**.

It seems natural that time reversal should obey $J^2 = 1$, as it does in the bosonic case; it may seem strange to have $J^2 = -1$, as we do for fermions. Should not applying time reversal twice get us back where we started? It actually does, in a crucial sense: the expectation values of all observables are unchanged when we multiply a unit vector $v \in H$ by -1. Still, this minus sign reminds one of the equally curious minus sign that a fermion picks up when one rotates it a full turn. Are these signs related?

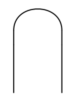
The answer appears to be yes, though the following argument is more murky than we would like. It has its origins in Feynman's 1986 Dirac Memorial Lecture on antiparticles [25], together with the well-known 'belt trick' for demonstrating that $\mathrm{SO}(3)$ fails to be simply connected (which is the reason its double cover $\mathrm{SU}(2)$ is needed to describe rotations in quantum theory.) The argument uses string diagrams, which are a generalization of Feynman diagrams. Instead of explaining string diagrams here, we refer the reader to our prehistory of n-categorical physics [13], and especially the section on Freyd and Yetter's 1986 paper on tangles.

The basic ideas apply to any Lie group G. If an irreducible object $H \in \mathrm{Rep}(G)$ is isomorphic to its dual, it comes with an invariant nondegenerate pairing $g\colon H \otimes H \to \mathbb{C}$. We can draw this as a 'cup':

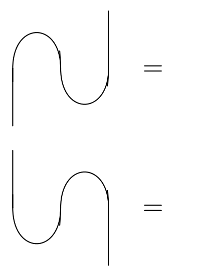


If this were a Feynman diagram, it would depict two particles of type H coming in on top and nothing going out at the bottom: since H is isomorphic to its dual, particles of this type act like their own antiparticles, so they can annihilate each other.

Since g defines an isomorphism $H \cong H^*$, we get a morphism back from $\mathbb{C}$ to $H \otimes H$, which we draw as a 'cap':

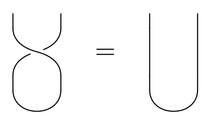


uniquely determined by requiring that the cap and cup obey the *zig-zag identities:*

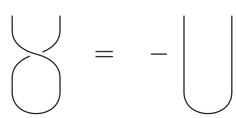


Indeed, for any $H \in \mathrm{Rep}(G)$, we have a 'cup' $H^* \otimes H \to \mathbb{C}$ and 'cap' $\mathbb{C} \to H \otimes H^*$ obeying the zig-zag identities: in the language of category theory, we say that $\mathrm{Rep}(G)$ is compact closed [13]. When H is isomorphic to its dual, we can write these using just H.

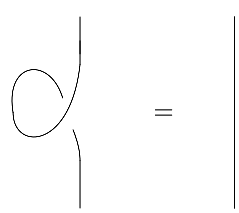
As we have seen, there are are two choices. If H is real, g is an orthogonal structure, and we can write the equation $g(w, v) = g(v, w)$ using string diagrams as follows:



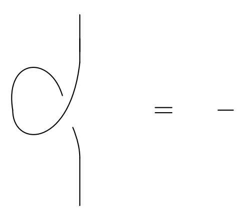
If H is quaternionic, g is a symplectic structure, and we can write the equation $g(w, v) = -g(v, w)$ as follows:



The power of string diagrams is that we can apply the same diagrammatic manipulation to both sides of an equation and get a valid new equation. So let us take the equations above, turn both sides 90° clockwise, and stretch out the string a little. When H is a real representation, we obtain:

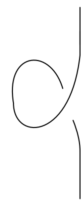


When H is quaternionic we obtain:



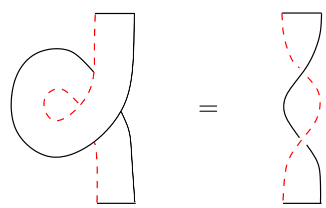
Both sides of these equations describe morphisms from H to itself. The vertical straight line at right corresponds to the identity morphism $1 \colon H \to H$. The more complicated diagram at left turns out to correspond to the morphism J^2. So, it follows that $J^2 = 1$ when the representation H is real, while $J^2 = -1$ when H is quaternionic.

While it may seem puzzling to those who have not been initiated into the mysteries of string diagrams, everything so far can be made perfectly rigorous [8]. Now comes the murky part. The string diagram at left:



looks like a particle turning back in time, going backwards for a while, and then turning yet again and continuing forwards in time. In other words, it *looks like a picture of the square of time reversal!* So, the fact that the corresponding morphism J^2 is indeed the square of time reversal when $G = \mathrm{SU}(2)$ somehow seems right.

Moreover, in the theory of string diagrams it is often useful to draw the strings as 'ribbons'. This allows us to take the diagram at left and pull it tight, as follows:



At left we have a particle (or actually a small piece of string) turning around in *time*, while at right we have a particle making a full turn in *space* as time passes. So, in the case $G = \mathrm{SU}(2)$ it seems to make sense that $J^2 = 1$ when the spin j is an integer: after all, the state vector of an integer-spin particle is unchanged when we rotate that particle a full turn. Similarly, it seems to make sense that $J^2 = -1$ when j is a half-integer: the state vector of a half-integer-spin particle gets multiplied by -1 when we rotate it by a full turn.

So, we seem to be on the brink of having a 'picture proof' that the square of time reversal must match the result of turning a particle 360 degrees. Unfortunately this argument is not yet rigorous, since we have not explained how the topology of ribbon diagrams (well-known in category theory) is connected to the geometry of rotations and time reversal. Perhaps understanding the argument better would lead us to new insights.

While our discussion here focused on the group SU(2), real and quaternionic representations of other groups are also important in physics. For example, gauge bosons live in the adjoint representation of a compact Lie group G on the complexification of its own Lie algebra; since the Lie algebra is real, this is always a real representation of G. This is related to the fact that gauge bosons are their own antiparticles. In the Standard Model, fermions are not their own antiparticles, but in some theories they can be. Among other things, this involves the question of whether the relevant spinor representations of the groups $\mathrm{Spin}(p, q)$ are complex, real ('Majorana spinors') or quaternionic ('pseudo-Majorana spinors'). The options are well-understood, and follow a nice pattern depending on the dimension and signature of spacetime modulo 8 [21]. We should emphasize that the spin groups $\mathrm{Spin}(p, q)$ are not compact unless $p = 0$ or $q = 0$, so their finite-dimensional complex representations are hardly ever unitary, and most of the results mentioned in this paper do not apply. Nonetheless, we may ask if such a representation is the complexification of a real one, or the underlying complex representation of a quaternionic one—and the answers have implications for physics.

SU(2) is not the only compact Lie group with the property that all its irreducible continuous unitary representations on complex Hilbert spaces are real or quaternionic. For a group to have this property, it is necessary and sufficient that every element be conjugate to its inverse. All compact simple Lie groups have this property except those of type A_n for $n > 1$, D_n with n odd, and E_6 (see Bourbaki [20]). For the symmetric groups S_n, the orthogonal groups $\mathrm{O}(n)$, and the special orthogonal groups $\mathrm{SO}(n)$ for $n \geq 3$, all representations are in fact real. So, there is a rich supply of real and quaternionic group representations, which leave their indelible mark on physics even when we are doing complex quantum theory.

6 Categories

As discussed so far, the three-fold way describes how certain complex representations of groups can be seen as arising from real or quaternionic representations. This gives a sense in which ordinary complex quantum theory subsumes the real and quaternionic theories. But underlying this idea is something simpler yet deeper, which arises at the level of Hilbert spaces, even before group representations enter the game. Suppose $\mathbb{K}$ is either $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, and let $\mathrm{Hilb}_{\mathbb{K}}$ be the category where:

- an object is a $\mathbb{K}$-Hilbert space;
- a morphism is a bounded $\mathbb{K}$-linear operator.

As we shall see, any one of the categories $\mathrm{Hilb}_{\mathbb{R}}$, $\mathrm{Hilb}_{\mathbb{C}}$ and $\mathrm{Hilb}_{\mathbb{H}}$ has a faithful functor to any other. This means that a Hilbert space over any one of the three normed division algebras can be seen as Hilbert space over any other, equipped with some extra structure!

At the bottom of this fact lies the chain of inclusions

$$\mathbb{R} \subset \mathbb{C} \subset \mathbb{H}.$$

Thanks to these, any complex vector space has an underlying real vector space, and any quaternionic vector space has underlying complex vector space. The same is true for Hilbert spaces. To make the underlying real vector space of a complex Hilbert space into a real Hilbert space, we take the real part of the original complex inner product. This is well-known; a bit less familiar is how the underlying complex vector space of a quaternionic Hilbert space becomes a complex Hilbert space. Here we need to take the **complex part** of the original quaternionic inner product, defined by

$$\mathrm{Co}(a + bi + cj + dk) = a + bi.$$

One can check that these constructions give functors

$$\mathrm{Hilb}_{\mathbb{H}} \to \mathrm{Hilb}_{\mathbb{C}} \to \mathrm{Hilb}_{\mathbb{R}}.$$

A bit more formally, we have a commutative triangle of homomorphisms:

$$\begin{tikzcd} \mathbb{R} \arrow[rr, "\gamma"] \arrow[dr, "\alpha"'] & & \mathbb{H} \\ & \mathbb{C} \arrow[ur, "\beta"'] & \end{tikzcd}$$

There is only one choice of the homomorphisms α and γ. There are many choices of β, since we can map $i \in \mathbb{C}$ to any square root of -1 in the quaternions. However, all the various choices of β are the same up to an automorphism of the quaternions. That is, given two homomorphisms $\beta, \beta' \colon \mathbb{C} \to \mathbb{H}$, we can always find an automorphism $\theta \colon \mathbb{H} \to \mathbb{H}$ such that $\beta' = \theta \circ \beta$. So, nothing important depends on the choice of β. Let us make a choice—say the standard one, with $\beta(i) = i$—and use that. This commutative triangle gives a commutative triangle of functors:

$$\begin{tikzcd} \mathrm{Hilb}_{\mathbb{R}} & & \mathrm{Hilb}_{\mathbb{H}} \arrow[ll, "\gamma^*"'] \arrow[dl, "\beta^*"] \\ & \mathrm{Hilb}_{\mathbb{C}} \arrow[ul, "\alpha^*"] & \end{tikzcd}$$

Now, recall that a functor $F: C \to D$ is **faithful** if given two morphisms $f, f': c \to c'$ in C, $F(f) = F(f')$ implies that $f = f'$. When $F: C \to D$ is faithful, we can think of objects of C as objects of D equipped with extra structure.

It is easy to see that the functors α^*, β^* and γ^* are all faithful. This lets us describe Hilbert spaces for a larger normed division algebra as Hilbert spaces *equipped with extra structure* for a smaller normed division algebra. None of this particularly new or difficult: the key ideas are all in Adams' book on Lie groups [3]. However, it sheds light on how real, complex and quaternionic quantum theory are related.

First we consider the extra structure possessed by the underlying real Hilbert space of a complex Hilbert space:

Theorem 7. *The functor $\alpha^*: \mathrm{Hilb}_{\mathbb{C}} \to \mathrm{Hilb}_{\mathbb{R}}$ is faithful, and $\mathrm{Hilb}_{\mathbb{C}}$ is equivalent to the category where:*

- *an object is a real Hilbert space H equipped with a unitary operator $J: H \to H$ with $J^2 = -1$;*
- *a morphism $T: H \to H'$ is a bounded real-linear operator preserving this exta structure: $TJ = J'T$.*

This extra structure J is often called a **complex structure**. It comes from our ability to multiply by i in a complex Hilbert space.

Next we consider the extra structure possessed by the underlying complex Hilbert space of a quaternionic Hilbert space. For this we need to generalize the concept of an antiunitary operator. First, given $\mathbb{K}$-vector spaces V and V', we define an **antilinear operator** $T: V \to V'$ to be a map with

$$T(vx + wy) = T(v)x^* + T(w)y^*$$

for all $v, w \in V$ and $x, y \in \mathbb{K}$. Then, given $\mathbb{K}$-Hilbert spaces H and H', we define an **antiunitary operator** $T: H \to H'$ to be an invertible antilinear operator with

$$\langle Tv, Tw \rangle = \langle w, v \rangle$$

for all $v, w \in H$.

Theorem 8. *The functor $\beta^*: \mathrm{Hilb}_{\mathbb{H}} \to \mathrm{Hilb}_{\mathbb{C}}$ is faithful, and $\mathrm{Hilb}_{\mathbb{H}}$ is equivalent to the category where:*

- *an object is a complex Hilbert space H equipped with an antiunitary operator $J: H \to H$ with $J^2 = -1$;*
- *a morphism $T: H \to H'$ is a bounded complex-linear operator preserving this extra structure: $TJ = J'T$.*

This extra structure J is often called a **quaternionic structure**. We have seen it already in our study of the three-fold way. It comes from our ability to multiply by j in a quaternionic Hilbert space.

Finally, we consider the extra structure possessed by the underlying real Hilbert space of a quaternionic Hilbert space. This can be understood by composing the previous two theorems:

Theorem 9. *The functor* $\gamma^*\colon \mathrm{Hilb}_{\mathbb{H}} \to \mathrm{Hilb}_{\mathbb{R}}$ *is faithful, and* $\mathrm{Hilb}_{\mathbb{H}}$ *is equivalent to the category where:*

- *an object is a real Hilbert space* H *equipped with two unitary operators* $J, K\colon H \to H$ *with* $J^2 = K^2 = -1$ *and* $JK = -KJ$*;*
- *a morphism* $T\colon H \to H'$ *is a bounded real-linear operator preserving this extra structure:* $TJ = J'T$ *and* $TK = K'T$*.*

This extra structure could also be called a **quaternionic structure**, as long as we remember that a quaternionic structure on a real Hilbert space is different than one on a complex Hilbert space. It comes from our ability to multiply by j and k in a quaternionic Hilbert space. Of course if we define $I = JK$, then I, J, and K obey the usual quaternion relations.

The functors discussed so far all have adjoints, which are in fact both left and right adjoints:

$$\begin{tikzcd} \mathrm{Hilb}_{\mathbb{R}} \arrow[rr, "\gamma_*"] \arrow[dr, "\alpha_*"'] & & \mathrm{Hilb}_{\mathbb{H}} \\ & \mathrm{Hilb}_{\mathbb{C}} \arrow[ur, "\beta_*"'] & \end{tikzcd}$$

These adjoints can easily be defined using the theory of bimodules [6]. As vector spaces, we have:

$$\begin{array}{rcl} \alpha_*(V) & = & V \otimes {}_{\mathbb{R}}\mathbb{C}_{\mathbb{C}} \\ \beta_*(V) & = & V \otimes {}_{\mathbb{C}}\mathbb{H}_{\mathbb{H}} \\ \gamma_*(V) & = & V \otimes {}_{\mathbb{R}}\mathbb{H}_{\mathbb{H}} \end{array}$$

In the first line here, V is a real vector space, or in other words, a right $\mathbb{R}$-module, while ${}_{\mathbb{R}}\mathbb{C}_{\mathbb{C}}$ denotes $\mathbb{C}$ regarded as a $\mathbb{R}$-$\mathbb{C}$-bimodule. Tensoring these, we obtain a right $\mathbb{C}$-module, which is the desired complex vector space. The other lines work analogously. It is then easy to make all these vector spaces into Hilbert spaces. We cannot resist mentioning that our previous functors can be described in a similar way, just by turning the bimodules around:

$$\begin{array}{rcl} \alpha^*(V) & = & V \otimes {}_{\mathbb{C}}\mathbb{C}_{\mathbb{R}} \\ \beta^*(V) & = & V \otimes {}_{\mathbb{H}}\mathbb{H}_{\mathbb{C}} \\ \gamma^*(V) & = & V \otimes {}_{\mathbb{H}}\mathbb{H}_{\mathbb{R}} \end{array}$$

But instead of digressing into this subject, which is called Morita theory [6], we wish merely to note that the functors α_*, β_* and γ_* are also faithful. This lets us describe Hilbert spaces for a smaller normed division algebra in terms of Hilbert spaces for a bigger one.

We begin with the functor α_*, which is called **complexification**:

Theorem 10. *The functor* $\alpha_*\colon \mathrm{Hilb}_{\mathbb{R}} \to \mathrm{Hilb}_{\mathbb{C}}$ *is faithful, and* $\mathrm{Hilb}_{\mathbb{R}}$ *is equivalent to the category where:*

- *an object is a complex Hilbert space* H *equipped with a antiunitary operator* $J\colon H \to H$ *with* $J^2 = 1$*;*
- *a morphism* $T\colon H \to H'$ *is a bounded complex-linear operator preserving this exta structure:* $TJ = J'T$*.*

The extra structure J here is often called a **real structure**. We have seen it already in our study of the three-fold way. It is really just a version of complex conjugation. In other words, suppose that H is a real Hilbert space. Then any element of its complexification can be written uniquely as $u + vi$ with $u, v \in H$, and then

$$J(u + vi) = u - vi.$$

Next we consider the functor β_* from complex to quaternionic Hilbert spaces:

Theorem 11. *The functor* $\beta_*\colon \mathrm{Hilb}_{\mathbb{C}} \to \mathrm{Hilb}_{\mathbb{H}}$ *is faithful, and* $\mathrm{Hilb}_{\mathbb{C}}$ *is equivalent to the category where:*

- *an object is a quaternionic Hilbert space H equipped with a unitary operator J with $J^2 = -1$;*
- *a morphism $T\colon H \to H'$ is a bounded quaternion-linear operator preserving this extra structure: $TJ = J'T$.*

The operator J here comes from our ability to multiply any vector in $\beta_*(H) = H \otimes_{\mathbb{C}} \mathbb{H}_{\mathbb{H}}$ on the left by the complex number i. This result is less well-known than the previous ones, so we sketch a proof:

Proof. Suppose H is a quaternionic Hilbert space equipped with a unitary operator J with $J^2 = -1$. Then J makes H into a right module over the complex numbers, and this action of $\mathbb{C}$ commutes with the action of $\mathbb{H}$, so H becomes a right module over the tensor product of $\mathbb{C}$ and $\mathbb{H}$, considered as algebras over $\mathbb{R}$. But this is isomorphic to the algebra of 2×2 complex matrices [3]. The matrix

$$p = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

projects H down to a complex Hilbert space $H_{\mathbb{C}}$ whose complex dimension matches the quaternionic dimension of H. By applying arbitrary 2×2 complex matrices to this subspace we obtain all of H, so $\beta_* H_{\mathbb{C}}$ is naturally isomorphic to H. $\square$

Composing α_* and β_*, we obtain the functor from real to quaternionic Hilbert spaces. This is sometimes called **quaternification**:

Theorem 12. *The functor* $\gamma_*\colon \mathrm{Hilb}_{\mathbb{R}} \to \mathrm{Hilb}_{\mathbb{H}}$ *is faithful, and* $\mathrm{Hilb}_{\mathbb{R}}$ *is equivalent to the category where:*

- *an object is a quaternionic Hilbert space H equipped with two unitary operators J, K with $J^2 = K^2 = -1$ and $JK = -KJ$;*
- *a morphism $T\colon H \to H'$ is a bounded quaternion-linear operator preserving this extra structure: $TJ = J'T$ and $TK = K'T$.*

The operators J and K here arise from our ability to multiply any vector in $\gamma_*(H) = H \otimes_{\mathbb{R}} \mathbb{H}_{\mathbb{H}}$ on the left by the quaternions j and k. Again, we sketch a proof of this result:

Proof. The operators J, K and $I = JK$ make H into a left $\mathbb{H}$-module. Since this action of $\mathbb{H}$ commutes with the existing right $\mathbb{H}$-module structure, H becomes a module over the tensor product of $\mathbb{H}$ and $\mathbb{H}^{\mathrm{op}} \cong \mathbb{H}$, considered as algebras over $\mathbb{R}$. But this tensor product is isomorphic to the algebra of 4×4 real matrices [3]. The matrix

$$p = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

projects H down to a real Hilbert space $H_{\mathbb{R}}$ whose real dimension matches the quaternionic dimension of H. By applying arbitrary 4×4 real matrices to this subspace we obtain all of H, so $\gamma_* H_{\mathbb{R}}$ is naturally isomorphic to H. $\square$

Finally, it is worth noting that some of the six functors we have described have additional nice properties:

- The categories $\mathrm{Hilb}_\mathbb{R}$ and $\mathrm{Hilb}_\mathbb{C}$ are 'symmetric monoidal categories', meaning roughly that they have well-behaved tensor products. The complexification functor $\alpha^*\colon \mathrm{Hilb}_\mathbb{R} \to \mathrm{Hilb}_\mathbb{C}$ is a 'symmetric monoidal functor', meaning roughly that it preserves tensor products.
- The categories $\mathrm{Hilb}_\mathbb{R}, \mathrm{Hilb}_\mathbb{C}$ and $\mathrm{Hilb}_\mathbb{H}$ are 'dagger-categories', meaning roughly that any morphism $T\colon H \to H'$ has a Hilbert space adjoint $T^\dagger\colon H' \to H$ such that

  $$\langle Tv, w\rangle = \langle v, T^\dagger w\rangle$$

  for all $v \in H$, $w \in H'$. All six functors preserve this dagger operation.
- For $\mathrm{Hilb}_\mathbb{R}$ and $\mathrm{Hilb}_\mathbb{C}$, the dagger structure interacts nicely with the tensor product, making these categories into 'dagger-compact categories', and the functor α^* is compatible with this as well.

For precise definitions of the quoted terms here, see our review articles [13, 14]. For more on dagger-compact categories see also the work of Abramsky and Coecke [1, 2] (who called them 'strongly compact closed'), Selinger [40], and the book *New Stuctures for Physics* [22]. The three-fold way is best appreciated with the help of these category-theoretic ideas, but we have deliberately downplayed them in this paper, to reduce the prerequisites. A more category-theoretic treatment of symplectic and orthogonal structures can be found in our old paper on 2-Hilbert spaces [8].

7 Solutions

In Section 3 we raised two 'problems' with real and quaternionic quantum theory. Let us conclude by saying how the three-fold way solves these.

We noted that on a real, complex or quaternionic Hilbert space, any continuous one-parameter unitary group has a skew-adjoint generator S. In the complex case we can write S as i times a self-adjoint operator A, which in physics describes a real-valued observable. The first 'problem' is that in the real or quaternionic cases we cannot do this.

However, now we know from Theorems 8 and 10 that real and quaternionic Hilbert spaces can be seen as a complex Hilbert space with extra structure. This solves the problem. Indeed, we have faithful functors

$$\alpha_*\colon \mathrm{Hilb}_\mathbb{R} \to \mathrm{Hilb}_\mathbb{C}$$

and

$$\beta^*\colon \mathrm{Hilb}_\mathbb{H} \to \mathrm{Hilb}_\mathbb{C}.$$

As noted in the previous section, these are 'dagger-functors', so they send skew-adjoint operators to skew-adjoint operators.

So, given a skew-adjoint operator S on a real Hilbert space H, we can apply the functor α_* to reinterpret it as a skew-adjoint operator on the complexification of H, namely $H \otimes {}_\mathbb{R}\mathbb{C}_\mathbb{C}$. By abuse of notation let us still call the resulting operator S. Now we can write $S = iA$. But resulting self-adjoint operator A has an interesting feature: its spectrum is symmetric about 0!

In the finite-dimensional case, all this means is that for any eigenvector of A with eigenvalue $c \in \mathbb{R}$, there is corresponding eigenvector with eigenvalue $-c$. This is easy to see. Suppose that $Av = cv$. By Theorem 10, the complexification $H \otimes {}_\mathbb{R}\mathbb{C}_\mathbb{C}$ comes equipped with an antiunitary operator J with $J^2 = 1$, and we have $SJ = JS$. It follows that Jv is an eigenvector of A with eigenvalue $-c$:

$$AJv = -iSJv = -iJSv = JiSv = -JAv = -cJv.$$

(In this calculation we have reverted to the standard practice of treating a complex Hilbert space as a *left* $\mathbb{C}$-module.) In the infinite-dimensional case, we can make an analogous but more subtle statement about the continuous spectrum.

Similarly, given a skew-adjoint operator S on a quaternionic Hilbert space H, we can apply the functor β^* to reinterpret it as a skew-adjoint operator on the underlying complex Hilbert space. Let us again call the resulting operator S. We again can write $S = iA$. And again, the spectrum of A is symmetric about 0. The proof is the same as in the real case: now, by Theorem 8, the underlying complex Hilbert space has equipped with an antiunitary operator J with $J^2 = -1$, but we again have $SJ = JS$, so the same calculation goes through.

The second 'problem' is that we cannot take the tensor product of two quaternionic Hilbert spaces and get another quaternionic Hilbert spaces. But Theorem 6 makes this seem like an odd thing to want! Just as two fermions make a boson, the tensor product of two quaternionic Hilbert spaces is naturally a *real* Hilbert space. After all, Theorem 8 says that a quaternionic Hilbert space can be identified with a complex Hilbert space with an antiunitary J such that $J^2 = -1$. If we tensor two such spaces, we get a complex Hilbert space equipped with an antiunitary J such that that $J^2 = 1$. Theorem 10 then says that this can be identified with a real Hilbert space.

Further arguments of this sort give four tensor product functors:

$$
\begin{array}{rcl}
\otimes\colon \mathrm{Hilb}_{\mathbb{R}} \times \mathrm{Hilb}_{\mathbb{R}} & \to & \mathrm{Hilb}_{\mathbb{R}} \\
\otimes\colon \mathrm{Hilb}_{\mathbb{R}} \times \mathrm{Hilb}_{\mathbb{H}} & \to & \mathrm{Hilb}_{\mathbb{H}} \\
\otimes\colon \mathrm{Hilb}_{\mathbb{H}} \times \mathrm{Hilb}_{\mathbb{R}} & \to & \mathrm{Hilb}_{\mathbb{H}} \\
\otimes\colon \mathrm{Hilb}_{\mathbb{H}} \times \mathrm{Hilb}_{\mathbb{H}} & \to & \mathrm{Hilb}_{\mathbb{R}}.
\end{array}
$$

We can also tensor a real or quaternionic Hilbert space with a complex one and get a complex one:

$$
\begin{array}{rcl}
\otimes\colon \mathrm{Hilb}_{\mathbb{C}} \times \mathrm{Hilb}_{\mathbb{R}} & \to & \mathrm{Hilb}_{\mathbb{C}} \\
\otimes\colon \mathrm{Hilb}_{\mathbb{R}} \times \mathrm{Hilb}_{\mathbb{C}} & \to & \mathrm{Hilb}_{\mathbb{C}} \\
\otimes\colon \mathrm{Hilb}_{\mathbb{C}} \times \mathrm{Hilb}_{\mathbb{H}} & \to & \mathrm{Hilb}_{\mathbb{C}} \\
\otimes\colon \mathrm{Hilb}_{\mathbb{H}} \times \mathrm{Hilb}_{\mathbb{C}} & \to & \mathrm{Hilb}_{\mathbb{C}}.
\end{array}
$$

Finally, of course we have:

$$
\otimes\colon \mathrm{Hilb}_{\mathbb{C}} \times \mathrm{Hilb}_{\mathbb{C}} \to \mathrm{Hilb}_{\mathbb{C}}.
$$

In short, the 'multiplication' table of real, complex and quaternionic Hilbert spaces matches the usual multiplication table for the numbers $+1, 0, -1$. This should remind us of the Frobenius–Schur indictator, mentioned in the Introduction. The moral, then, is not to fight the patterns of mathematics, but to notice them and follow them.

7.1 Acknowledgements

I thank the organizers of QPL7 for inviting me to present some of this material in Oxford on May 27th, 2010. I would also like to thank Urs Schreiber for reminding me about the state-operator correspondence, and Torsten Ekedahl, Harald Hanche-Olsen, David Roberts and others on MathOverflow for helping me with some math questions.

References

[1] S. Abramsky, Abstract scalars, loops, and free traced and strongly compact closed categories, in *Proceedings of CALCO 2005*, Lecture Notes in Computer Science **3629**, Springer, Berlin, 2005, 1–31. Also available at http://web.comlab.ox.ac.uk/oucl/work/samson.abramsky/calco05.pdf.

[2] S. Abramsky and B. Coecke, A categorical semantics of quantum protocols, available at arXiv:quant-ph/0402130.

[3] J. F. Adams, *Lectures on Lie Groups*, Benjamin, New York, 1969.

[4] S. Adler, *Quaternionic Quantum Mechanics and Quantum Fields*, Oxford U. Press, Oxford, 1995.

[5] I. Amemiya and H. Araki, A remark on Piron's paper, *Pub. Res. Inst. Math. Sci.* **2** (1966/67), 423–427.

[6] F. W. Anderson and K. R. Fuller, *Rings and Categories of Modules*, Springer, Berlin, 1998.

[7] V. I. Arnold, Symplectization, complexification and mathematical trinities, in *The Arnoldfest: Proceedings of a Conference in Honour of V.I. Arnold for His Sixtieth Birthday*, eds. E. Bierstone, B. Khesin, A. Khovanskii and J. E. Marsden, AMS, Providence, Rhode Island, 1999.

[8] J. Baez, Higher-dimensional algebra II: 2-Hilbert spaces, *Adv. Math.* **127** (1997), 125–189. Also available as arXiv:q-alg/9609018.

[9] J. Baez, The octonions, *Bull. Amer. Math. Soc.* **39** (2002), 145–205. Errata in *Bull. Amer. Math. Soc.* **42** (2005), 213. Also available as arXiv:math/0105155.

[10] J. Baez, Quantum quandaries: a category-theoretic perspective, in *Structural Foundations of Quantum Gravity*, eds. S. French, D. Rickles and J. Saatsi, Oxford U. Press, Oxford, 2006, pp. 240–265. Also available as arXiv:quant-ph/0404040.

[11] J. Baez and J. Huerta, Division algebras and supersymmetry I, in *Superstrings, Geometry, Topology, and C*-Algebras*, eds. R. Doran, G. Friedman and J. Rosenberg, *Proc. Symp. Pure Math.* **81**, AMS, Providence, 2010, pp. 65-80. Also available as arXiv:0909.0551.

[12] J. Baez and J. Huerta, Division algebras and supersymmetry II. Available as arXiv:1003.3436.

[13] J. Baez and A. Lauda, A prehistory of n-categorical physics, to appear in *Deep Beauty: Mathematical Innovation and the Search for an Underlying Intelligibility of the Quantum World*, ed. Hans Halvorson, Cambridge U. Press. Also available as arXiv:0908.2469.

[14] J. Baez and M. Stay, Physics, topology, logic and computation: a Rosetta Stone, in *New Structures for Physics*, ed. Bob Coecke, Lecture Notes in Physics **813**, Springer, Berlin, 2000, pp. 95-174. Also available as arXiv:0903.0340.

[15] H. Barnum and A. Wilce, Ordered linear spaces and categories as frameworks for information-processing characterizations of quantum and classical theory. Available as arXiv:0908.2354.

[16] H. Barnum, R. Duncan and A. Wilce, Symmetry, compact closure and dagger compactness for categories of convex operational models. Available as arXiv:1004.2920.

[17] H. Barnum, C. P. Gaebler and A. Wilce, Ensemble steering, weak self-duality, and the structure of probabilistic theories. Available as arXiv:0912.5532.

[18] T. Bartels, Functional analysis with quaternions. Available at http://tobybartels.name/notes/#quaternions.

[19] N. Bourbaki, *Elements of Mathematics*, Chapter VIII, §7, Prop. 12, Springer, Berlin, 2008.

[20] N. Bourbaki, *Elements of Mathematics*, Chapter IX, Appendix II, Prop. 4, Springer, Berlin, 2008.

[21] P. Budinich and A. Trautman, *The Spinorial Chessboard*, Springer, Berlin, 1988.

[22] B. Coecke, *New Structures for Physics*, Lecture Notes in Physics **813**, Springer, Berlin, 2000.

[23] E. Corrigan and T. J. Hollowood, The exceptional Jordan algebra and the superstring, *Comm. Math. Phys.* **122** (1989), 393–410. Also available at Project Euclid.

[24] F. Dyson, The threefold way: algebraic structure of symmetry groups and ensembles in quantum mechanics, *Jour. Math. Phys.* **3** (1962), 1199–1215.

[25] R. Feynman, The reason for antiparticles, in *Elementary Particles and the Laws of Physics: the 1986 Dirac Memorial Lectures*, Cambridge U. Press, Cambridge, pp. 1–60.

[26] F. G. Frobenius and I. Schur, Über die reellen Darstellungen der endlichen Gruppen, *Sitzungsber. Akad. Preuss. Wiss.* (1906), 186–208.

[27] L. Hardy, Quantum theory from five reasonable axioms. Available at arXiv:quant-ph/0101012.

[28] S. S. Holland Jr., Orthomodularity in infinite dimensions; a theorem of M. Solèr, *Bull. Amer. Math. Soc.* **32** (1995), 205–234. Also available as arXiv:math/9504224.

[29] A. Hurwitz, Über die Composition der quadratischen Formen von beliebig vielen Variabeln, *Nachr. Ges. Wiss. Göttingen* (1898), 309–316.

[30] P. Jordan, Über eine Klasse nichtassociativer hyperkomplexer Algebren, *Nachr. Ges. Wiss. Göttingen* (1932), 569–575.

[31] P. Jordan, J. von Neumann and E. Wigner, On an algebraic generalization of the quantum mechanical formalism, *Ann. Math.* **35** (1934), 29–64.

[32] M. Koecher, Positivitätsbereiche in $\mathbb{R}^n$, *Amer. J. Math.* **79** (1957), 575–596.

[33] M. Koecher, *The Minnesota Notes on Jordan Algebras and Their Applications*, eds. A. Krieg and S. Walcher, Lecture Notes in Mathematics **1710**, Springer, Berlin, 1999.

[34] K. McCrimmon, *A Taste of Jordan Algebras*, Springer, Berlin, 2004.

[35] C.-K. Ng, Quaternion functional analysis. Available as arXiv:math/0609160.

[36] S. Okubo, *Introduction to Octonion and Other Non-Associative Algebras in Physics*, Cambridge University Press, Cambridge, 1995.

[37] C. Piron, *Foundations of Quantum Physics*, Benjamin, New York, 1976.

[38] C. Piron, "Axiomatique" quantique, *Helv. Phys. Acta* **37** (1964), 439–468.

[39] J. Polchinski, *String Theory*, Vol. I, Section 2.9: More on states and operators, Cambridge U. Press, Cambridge, 1998.

[40] P. Selinger, Dagger compact closed categories and completely positive maps, *Proceedings of the 3rd International Workshop on Quantum Programming Languages (QPL 2005)*, Elsevier, Amsterdam, 2007, pp. 139–163. Also available at http://www.mscs.dal.ca/∼selinger/papers/#dagger.

[41] M. P. Solèr, Characterization of Hilbert spaces by orthomodular spaces, *Comm. Algebra* **23** (1995), 219–243.

[42] B. Van Steirteghem and I. Stubbe, Propositional systems, Hilbert lattices and generalized Hilbert spaces, in *Handbook of Quantum Logic and Quantum Structures: Quantum Structures*, eds. K. Engesser, D. M. Gabbay and D. Lehmann, Elsevier, Amsterdam, 2007

[43] K. Urbanik and F. B. Wright, Absolute-valued algebras, *Proc. Amer. Math. Soc.* **11** (1960), 861–866. Freely available online from the AMS.

[44] V. S. Varadrajan, *Geometry of Quantum Theory*, Springer, Berlin, 1985.

[45] J. Vicary, Completeness of dagger-categories and the complex numbers. Available as arXiv:0807.2927.

[46] E. B. Vinberg, Homogeneous cones, *Soviet Math. Dokl.* **1** (1961), 787–790.

[47] E. I. Zelmanov, On prime Jordan algebras. II, *Sibirsk Mat. J.* **24** (1983), 89–104.