The geometrical meaning of spinors as a key to make sense of quantum mechanics

Gerrit Coddens

The geometrical meaning of spinors as a key to make sense of quantum mechanics

Gerrit Coddens

Laboratoire des Solides Irradiés,
CEA-DSM-IRAMIS, CNRS UMR 7642, Ecole Polytechnique,
28, Route de Saclay, F-91128-Palaiseau CEDEX, France

17th March 2021

Abstract. This paper aims at explaining that the key to understanding quantum mechanics (QM) is a perfect geometrical understanding of the spinor algebra that is used in its formulation. Spinors occur naturally in the representation theory of certain symmetry groups. The spinors that are relevant for QM are those of the homogeneous Lorentz group SO(3,1) in Minkowski space-time $\mathbb{R}^4$ and its subgroup SO(3) of the rotations of three-dimensional Euclidean space $\mathbb{R}^3$. In the three-dimensional rotation group the spinors occur within its representation SU(2). We will provide the reader with a perfect intuitive insight about what is going on behind the scenes of the spinor algebra. We will then use the understanding acquired to derive the free-space Dirac equation from scratch proving that it is a description of a statistical ensemble of spinning electrons in uniform motion, completely in the spirit of Ballentine's statistical interpretation of QM. This is a mathematically rigorous proof. Developing this further we allow for the presence of an electromagnetic field. We can consider the result as a reconstruction of QM based on the geometrical understanding of the spinor algebra. By discussing a number of problems in the interpretation of the conventional approach, we illustrate how this new approach leads to a better understanding of QM.

PACS. 02.20.-a Group Theory

1 Introduction

1.1 Three famous quotes

Richard Feynman [1] (recipient of the Nobel prize of physics in 1965), is notorious for having stated:

□ *"I think that I can safely say that nobody understands quantum mechanics".*

On the other hand Michael Atiyah (winner of the Fields medal in 1966) is not less notorious for having stated:

□ *"No one fully understands spinors. Their algebra is formally understood but their general significance is mysterious. In some sense they describe the "square root" of geometry and, just as understanding the square root of -1 took centuries, the same might be true of spinors"* [2].

□ *" . . . the geometrical significance of spinors is still very mysterious. Unlike differential forms, which are related to areas and volumes, spinors have no such simple explanation. They appear out of some slick algebra, but the geometrical meaning is obscure . . . "* [3].

There is an obvious analogy here. Both towering scientists express their dismay with something they consider not to be properly understood. At least a part of Feynman's problems might be directly due to the fact that Atiyah's problems are integrally copied into QM as a consequence of the use the latter makes of spinors. Therefore understanding spinors is a prerequisite for understanding QM. After reading the quotes it seems obvious that solving the problems mentioned could really be a tall order. However, there is a tiny hole through which we can worm our way towards a new vantage point, offering a different angle of approach that allows to solve the problem of the meaning of spinors. Pointing this out and developing the ideas then further is what the present paper is all about. Spinors are part of the group representation theory of the homogeneous Lorentz group in Minkowski space-time $\mathbb{R}^4$ and of the rotation groups in

the Euclidean vector spaces $\mathbb{R}^n$. The whole of QM is written in the language of such spinors, i.e. a language of symmetry.

1.2 Nobody understands spinors

Many people have difficulties in apprehending the concept of spinors. In search for enlightenment, the reader will discover that it is very hard to find a clear definition of what a spinor is in the literature. Cartan e.g. states in his monograph [4]: *"A spinor is a kind of isotropic vector"*. Using a terminology "a kind of" can hardly be considered as a valid part of a clear definition. And a literature search reveals that this is an ever recurring theme. In all the various presentations I was able to consult, one just develops the algebra and states at the end of it that certain quantities introduced in the process are spinors. This is completely at variance with the usual practice, where the definition of a concept *precedes* the theorem about that concept. This way of introducing spinors leaves us without any clue as to what is going on behind the scenes, e.g. in the form of a conceptual mental image of what a spinor is supposed to be. What we are hitting here are actually manifestations of the state of affairs described by Michael Atiyah in the two quotes reproduced in Subsection 1.1.

What is going on here? In algebraic geometry, geometry and algebra go hand in hand. We have a geometry, an algebra and a dictionary in the form of a one-to-one correspondence that translates the algebra into the geometry and *vice versa*. As may transpire from what Atiyah says, the problem with the spinor concept is thus that in the approaches which are presented in textbooks the algebra and the geometry have not been developed in parallel. It is all "algebra first". We have only developed the algebra and neglected the geometry and the dictionary. The approach has even been so asymmetrical that we are no longer able to guess the geometry from the algebra.

It is perhaps worth formulating here a provocative question. Spinors occur in the representation SU(2) of the three-dimensional rotation group in $\mathbb{R}^3$. As it uses spinors, which seem particularly difficult to understand, SU(2) appears to be a mystery representation of the three-dimensional rotation group. Now here is the question: How on Earth can it be that there is something mysterious about the three-dimensional rotation group? Is it not mere Euclidean geometry? This seems to suggest that there might be something simple that we have overlooked and that has escaped our attention.

We will see that this is indeed true. In the first part of this article we will restore the balance between the algebra and the geometry by providing the reader also with the geometry and the dictionary. This way, he will be able to clearly understand the concept of spinors in SU(2). The reader will see that the strategy followed to solve the riddle what the square root of a vector might mean is somewhat analogous to the one that solves the puzzle what the square root of -1 means, as will be discussed in detail in Subsection 2.6. We will define the spinor concept in its own right and show afterwards that one can define an isomorphism that allows to interpret a spinor as "squaring to a vector".

We will thus try to build the theory of spinors starting from geometry. This way the underlying ideas will become clear in the form of "visual" geometrical clues. This will suffice for what the reader will need to know about spinors in the rotation and the Lorentz groups for applications in QM. When the reader will have understood the ideas that are underlying the geometrical approach to spinors, he should in principle be able to design or complete the proofs of this approach himself by applying these ideas. With our apologies to the mathematicians amongst the readership, we will therefore not strive for a formal perfection of our presentation. Our presentation may in this respect be considered as clumsy or deficient from the viewpoint of mathematical rigor, but as explained above mathematically rigorous presentations have their own inconvenience, viz. that they may render it very difficult to perceive the underlying ideas. Our aim is not to give a perfect formal account of the mathematical theory. Such accounts were already written more than hundred years ago. The aim of the paper is to provide new geometrical insight in the theory, something even mathematicians might value, and confer to the reader all of this insight he needs. As the ultimate goal is to obtain a better understanding of QM, I just cannot afford getting the reader lost on the journey by a harshly formal presentation. We want to render the ideas so clear and utterly obvious that the reader will become fluent enough to derive all further developments himself without any substantial difficulty. The self-learning that will intervene in carrying out this exercise will certainly help him much better to become acquainted with the subject matter than reading and mechanically checking the algebra of an exhaustive and formally perfect account of it in a book.

□ Remark 1. We can take advantage of the second quote of Aityah to point out that it will be shown that there are two completely distinct algebras at stake in the Clifford algebra on which spinor algebra is based: one for the group elements and one for vectors and multi-vectors. A same algebraic expression in the two algebras can thereby represent two completely distinct geometrical objects, e.g. a reflection with respect to a plane and a unit vector. These two algebras should therefore not be confused. The algebra for vectors and multivectors comprises what is called the exterior algebra. The differential forms mentioned by Atiyah are a language to deal with this exterior algebra. Clifford algebra is another such language. The differential forms are anti-symmetric multivectors. The spinors belong to the other algebra and represent group elements.

□ Remark 2. It has become fashionable to express QM in the language of geometrical algebra, based on the work of Hestenes [5]. However, Hestenes adopts the Clifford algebra as God-given. It conveniently descends from heaven and some of its results seem to follow by magic from thin air, just by adopting some stunning rules, e.g. that we can sum objects of different dimensions.

What we need and will develop is an approach that digs deeper into the mathemetical foundations and underbuilds also the Clifford algebra by constructing it from scratch such that it can be seen where it comes from. Despite the lesser elegance this may entail for the presentation, this additional insight is absolutely necessary to fully understand QM. The complex number i is not a generator for rotations as Hestenes claims. He also eludes answering the question what it means to sum objects of different dimensions, despite the fact that this is a totally legitimate question (see Subsection 2.8).

The development of the spinor theory which will be given in this article is an improvement of our presentation of spinors given in Chapter 3 of reference [6]. There is of course some overlap with reference [6] but not everything is systematically reproduced here. There is a substantial overlap with the HAL archive deposit [7] which also gives the full details about the generalization of the group theory of spinors to SO(n), which we are not reproducing here based on considerations about length and context.

1.3 Nobody understands Quantum Mechanics

Feynman's statement reflects an unprecedented, very unpleasant situation in physics. We find it enlightening to formulate the problem of the meaning of QM exactly in the same terms as the problem of the meaning of spinors. As a matter of fact, QM provides us with a complete set of algebraic rules to calculate and predict the outcome of experiments with staggering precision. But the reverse side of the medal is that nobody knows what this algebra means, i.e we do not know what the corresponding geometry and the dictionary are. This situation of a complete divorce between algebra and geometry has been summarized in a poignant way by Mermin by introducing the catchphrase that what one ought to do is to just "shut up and calculate" [8]. This is a source of frustration (for physicists) and distress (for students). Are we really condemned to spend a lifetime in physics calculating as a headless chicken?

□ Remark 1. The example often cited to illustrate the degree of precision quantum theory can reach is the comparison (see e.g. [9], p.162) between the experimentally measured value 0.001 159 652 180 73(2) and the theoretically calculated value 0.001 159 652 181 643(764) for $(g-2)/2$, where g is the anomalous g-factor of the electron.

In view of this nagging lack of understanding, some people have recently proposed to reconstruct QM from scratch [10]. The present paper proposes such a reconstruction in a way that is perhaps totally different from what a physicist might expect, because it starts the journey by digging into the mathematics of spinors, and then derives the Dirac equation from scratch with the rigor of a mathemical proof. This can then serve as a crystal-clear and mystery-free starting point for trying to make sense of the meaning of QM.

A first justification for this claim is the following argument. Establishing the geometrical meaning of spinors in mathematics comes logically prior to any possible application of spinors in QM. At the time we start doing the physics, the geometrical meaning is already established as a mathematical fact beyond any further discussion. Mathematics can only be right or wrong and everybody can check if the correspondence between the geometry and the algebra laid down in the dictionary proposed is correct or otherwise. Therefore using the geometrical meaning of spinors to interpret the formalism of QM which is written in the language of spinors is immune to any questioning because it is pure mathematics and situated outside the scope of a debate in physics. Furthermore, the geometrical meaning we will propose does not alter or affect the algebra. It is just added as perfectly fitting new insight such that the algebra used in our new approach to QM remains the same as in the traditional approach to QM. The new approach will therefore automatically reproduce the agreement of the theory with the experimental data that was obtained in the traditional approach and it will therefore be an unassailable reconstruction of QM.

A second justification for the claim is the analysis we were able to make of a number of quantum paradoxes considered to elude any intuitive explanation, as we will discuss later. The most convincing case is in my opinion the solution of the paradox of the Stern-Gerlach experiment [11]. Its analysis does not only validate the reconstruction by showing that it permits really to come to grasps with the counter-intuitive results of this experiment. The more rigorous and general new approach also lays bare a number of limitations and intellectual cracks in the standard approach. Within the new framework all these disturbing little wrinkles can be spotted and ironed out.

We hope that together with [11] this paper can provide a decent introduction to this new approach to QM based on the understanding of spinors. The two papers could constitute a solid starting basis for further study of my other results and of the foundations of QM in general.

1.4 Remarks about style and notation

The style of the present paper may look very informal but there is a strong commitment behind this choice of presentation. In fact, due to a concern of absolute rigor, the presentations by mathematicians are in general so formal that it is for laymen completely impossible to make sense of them. The chilling effect of this formal abstraction has been described by Dieudonné [12]. Such austere presentations might be all right for mathematicians but other people than mathematicians may need to use their theories. We hope that a pleasant, less highbrow presentation can be a good

trade-off between rigor and intuition that will be accessible to as broad a community as possible. The impenetrability of the original publications may tempt people who need to use the mathematics into trying to develop parallel *ad hoc* interpretations and it is at this point that overinterpretations and errors can creep in with dire consequences. This has happened many times in standard QM and we will have to point out a few of such mishaps in the present paper.

Let us now spell out a number notations and conventions we will use. We will note by $\mathrm{F}(A,B)$ the set of all mappings from the set A to the set B. We will note by $\mathrm{L}(V,W)$ the set of all linear mappings from the vector space V to the vector space W. They correspond thus to $k \times m$ matrices if $\dim V = k$ and $\dim W = m$. One often notes $\mathrm{L}(\mathbb{R}^n,\mathbb{R}^n)$ as $\mathrm{M}_n(\mathbb{R})$ in the literature, while one notes $\mathrm{L}(\mathbb{C}^n,\mathbb{C}^n)$ as $\mathrm{M}_n(\mathbb{C})$. The notation SU(2) refers to the special unitary group of dimension 2. It is the group of complex 2×2 matrices $\mathbf{M}$ which satisfy the conditions $\det \mathbf{M} = 1$ (special) and $\mathbf{M}^\dagger = \mathbf{M}^{-1}$ (unitary). We will see that it is a representation of the rotation group in $\mathbb{R}^3$.

The n-dimensional rotation group in $\mathbb{R}^n$, the matrix group $\mathrm{SO}(n)$ that represents it in $\mathbb{R}^n$ and the corresponding matrix group we construct in this paper and [7] are strictly spoken three different mathematical objects that are linked by group isomorphisms. But these isomorphisms justify the *abus de language* to treat these mathematical objects as identical. For convenience we will note the n-dimensional rotation group in $\mathbb{R}^n$ by its most intuitive representation $\mathrm{SO}(n)$. This way we will speak about the spinors of $\mathrm{SO}(n)$ although in reality they are not concepts that occur in the $n \times n$ matrix representation $\mathrm{SO}(n)$, but in the representation that acts on a subset $\mathscr{W} \subset \mathbb{C}^{2^\nu}$, constructed in this paper and [7]. Here $\nu = \lfloor \frac{n}{2} \rfloor$, where $\lfloor \cdot \rfloor$ is the integer part. We will use this notation ν throughout the paper. The quantity ν enters naturally into the discussions as will become clear when we go along with developing the presentation. The notation $\mathrm{SL}(2,\mathbb{C})$ stands for the special linear group of 2×2 complex matrices $\mathbf{M}$ with $\det \mathbf{M} = 1$. We will encounter it as a representation of the homogeneous Lorentz group in Section 3.

As will be explained, the rotation groups $\mathrm{SO}(n)$ can be obtained as a subgroup of a larger group generated by reflections. We will call the group elements obtained by an even number of reflections rotations, and call the group elements obtained from an odd number of reflections reversals. Reflections are special reversals. General reversals are products of a reflection and a rotation.

2 Spinors in the rotation groups SO(n)

2.1 Methodology

To develop the theory of spinors for the rotation groups $\mathrm{SO}(n)$, we will start from a simple specific case and then see how we can generalize it. We will this way discover and take the ideas and the difficulties one by one, while in a general abstract approach many of the underlying ideas may become hidden. The simplest case in point is the representation SU(2) for the rotations in $\mathbb{R}^3$. We understand very well the formalism for SO(3). We rotate vectors, written as 3×1 column matrices by multiplying them to the left by 3×3 rotation matrices. It is natural to expect that the same philosophy will apply in SU(2) and to attempt to make sense of SU(2) by analogy with what happens in SO(3). But as we will see such heuristics are a deadlock. It is the blind spot of our unawareness about this deadlock which impedes us figuring out what spinors are about. The spinors, which are the 2×1 matrices on which the 2×2 SU(2) rotation matrices are operating do not correspond to images of vectors of $\mathbb{R}^3$ or $\mathbb{C}^3$.

2.2 Preliminary caveat: Spinors do not build a vector space

2.2.1 Summing spinors is a priori not defined

As we will see, spinors in SU(2) do not build a vector space but a curved manifold. This is almost never clearly spelled out. A consequence of this is that physicists believe that the linearity of the Dirac equation (and the Schrödinger equation) implies the superposition principle in QM, which is wrong because the spinors are not building a vector space. In this respect Cartan stated that physicists are using spinors like vectors. This confusion plays a major rôle in one of the meanest paradoxes of QM, *viz.* the double-slit experiment [13].

It is important to point out that within the context of pure group theory it is even a transgression to make linear combinations of rotation matrices in SO(3). A linear combination of rotation matrices will in general no longer be a rotation matrix. Within $\mathrm{L}(\mathbb{C}^3,\mathbb{C}^3)$ or $\mathrm{L}(\mathbb{R}^3,\mathbb{R}^3)$ we can nevertheless try to find a meaning for such linear combinations, because the matrices are operating on elements of a vector space $\mathbb{R}^3$ or $\mathbb{C}^3$, yielding again elements of the same vector space $\mathbb{R}^3$ or $\mathbb{C}^3$. The matrix group SO(3) is embedded within the matrix group $\mathrm{L}(\mathbb{R}^3,\mathbb{R}^3)$. The linear combinations of the matrices in $\mathrm{L}(\mathbb{R}^3,\mathbb{R}^3)$ can then be interpreted by falling back on the meaning of linear combinations of vectors in the image space. But in SU(2) this will not be possible as the spinors are not building a vector space (see Remark 1 in Subsection 2.3 below).

The caveat we are introducing here is actually much more general. In group representation theory one introduces purely formal expressions $\sum_j c_j \mathbf{D}(g_j)$ which build the so-called group ring [14]. This happens e.g. when we construct so-called all-commuting operators, which are also called Casimir operators [15]. Here $\mathbf{D}(g_j)$ are the representation matrices of the group elements $g_j \in G$ of the group $(G, \circ)$ with operation $\circ$ and c_j are elements of a number field $\mathbb{K}$, which can e.g. be $\mathbb{R}$ or $\mathbb{C}$. This is purely formal as in the definition of a representation $\mathbf{D}$ we define the operation $\mathbf{D}(g_j)\mathbf{D}(g_k) = \mathbf{D}(g_j \circ g_k)$, but we do not define the operation $\sum_j c_j \mathbf{D}(g_j)$ as corresponding to $\mathbf{D}(\sum_j c_j g_j)$, for the very simple reason that $\sum_j c_j g_j$ is in general not defined. Only the operation $\circ$ has been defined. A good text book should insist thus on the fact that introducing $\sum_j c_j \mathbf{D}(g_j)$ is purely formal [15] in the sense that it is pure algebra without geometrical counterpart. To illustrate this, we could ask the question what the meaning of the sum of two permutations p and q:

$$\begin{pmatrix} 1 & 2 & \cdots & j & \cdots & n \\ p_1 & p_2 & \cdots & p_j & \cdots & p_n \end{pmatrix} + \begin{pmatrix} 1 & 2 & \cdots & j & \cdots & n \\ q_1 & q_2 & \cdots & q_j & \cdots & q_n \end{pmatrix}, \tag{1}$$

in the permutation group S_n is supposed to be. To illustrate this further, imagine the group $(G, \circ)$ of moves of a Rubik's cube. It is obvious in this example that $g_j \circ g_k$ is defined while $g_j + g_k$ is not. Giving geometrical meaning to $g_j + g_k$ requires introducing new definitions. This will be done in Subsection 2.5. As we will see it can be done by introducing sets. E.g. we can define $g_j + g_k$ as the set $\{g_j, g_k\}$. This way, we can give a meaning to expressions of the type $\sum_j c_j g_j$, with $c_j \in \mathbb{N}$. Giving meaning to $\sum_j c_j g_j$, with $c_j \in \mathbb{C}$ will require further efforts. We will dwell further on this issue of making linear combinations of spinors in Subsection 2.5.

2.2.2 Ideals

A concept that is very instrumental in reminding us of the no-go zone of linear combinations of spinors is the concept of an ideal. The spinors $\boldsymbol{\phi}$ of $\mathrm{SO}(n)$ build a set $\mathscr{I}$ such that for all rotation matrices $\mathbf{R}$ (which work on them by left multiplication), $\mathbf{R}\boldsymbol{\phi}$ also belongs to the set: $\forall \boldsymbol{\phi} \in \mathscr{I}, \forall \mathbf{R} \in G : \mathbf{R}\boldsymbol{\phi} \in \mathscr{I}$. One summarizes this by stating that $\mathscr{I}$ is a left ideal. Here G can stand for SU(2) or $\mathrm{SO}(n)$. The crucial point is that this does not imply that the set of spinors would be a vector space, such that: $\neg(\forall \boldsymbol{\phi}_1 \in \mathscr{I}, \forall \boldsymbol{\phi}_2 \in \mathscr{I}, \forall c_1 \in \mathbb{C}, \forall c_2 \in \mathbb{C} : c_1\boldsymbol{\phi}_1 + c_2\boldsymbol{\phi}_2 \in \mathscr{I})$. For the group SO(3) we can point out easily two trivial ideals, which are topologically disconnected, *viz.* the proper rotations and the reversals (which include the reflections), because it is impossible to change a left-handed frame into a right-handed frame by a proper rotation.

2.3 Construction of SU(2): the geometrical meaning of spinors

The idea behind the meaning of a 2×1 spinor of SU(2) is that we will no longer rotate vectors, but that we will "rotate" rotations. To explain what we mean by this, we start from the following diagram for a group G:

$\circ$	g_1	g_2	g_3	$\cdots$	g_j	$\cdots$	
g_1	$g_1 \circ g_1$	$g_1 \circ g_2$	$g_1 \circ g_3$	$\cdots$	$g_1 \circ g_j$	$\cdots$	
g_2	$g_2 \circ g_1$	$g_2 \circ g_2$	$g_2 \circ g_3$	$\cdots$	$g_2 \circ g_j$	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		
g_k	$g_k \circ g_1$	$g_k \circ g_2$	$g_k \circ g_3$	$\cdots$	$g_k \circ g_j$	$\cdots$	$\Longleftarrow T_{g_k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		

(2)

This diagram tries to illustrate a table for group multiplication. Admittedly, we will not be able to write down such a table for an infinite group, but we will only use it to render more vivid the ideas. Such a table tells us everything about the group we need to know: One can check on such a table that the group axioms are satisfied, and one can do all the necessary calculations. For the rotation group, we do not need to know how the rotations work on vectors. We might need to know how they work on vectors to construct the table, but once this task has been completed, we can forget about the vectors. The infinite table in Eq. 2 defines the whole group structure. When we look at one line of the table - the one flagged by the arrow - we see that we can conceive a group element g_k in a hand-waving way as a "function" $g_k : G \to G$ that works on the other group elements g_j according to: $g_k : g_j \to g_k(g_j) = g_k \circ g_j$. We can thus identify g_k with a function. More rigorously, we can say that we represent the group element g_k by a group automorphism $T_{g_k} \in \mathrm{F}(G, G) : g_j \to T_{g_k}(g_j) = g_k \circ g_j$. A rotation operates in this representation not on a vector, but on other rotations. We "turn rotations" instead of vectors. This is a construction which always works: The automorphisms of a group G are themselves a group that is isomorphic to G, such that they can be used to represent G.

It can be easily seen that this idea about the meaning of a spinor is true. As we will show below in Eq. 8, the general form of a rotation matrix $\mathbf{R}$ in SU(2) is:

$$\mathbf{R} = \begin{bmatrix} \xi_0 & -\xi_1^* \\ \xi_1 & \xi_0^* \end{bmatrix}, \tag{3}$$

A 2×1 spinor $\boldsymbol{\phi}$ can then be seen to be just a stenographic notation for a 2×2 SU(2) rotation matrix $\mathbf{R}$ by taking its first column $\hat{\mathrm{c}}_1(\mathbf{R})$:

$$\mathbf{R} = \begin{bmatrix} \xi_0 & -\xi_1^* \\ \xi_1 & \xi_0^* \end{bmatrix} \rightarrow \boldsymbol{\phi} = \hat{\mathrm{c}}_1(\mathbf{R}) = \begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix}, \tag{4}$$

This is based on the fact that the first column of $\mathbf{R}$ contains already the whole information about $\mathbf{R}$ and that $\mathbf{R}_1\hat{\mathrm{c}}_1(\mathbf{R}) = \hat{\mathrm{c}}_1(\mathbf{R}_1\mathbf{R})$. Instead of $\mathbf{R}' = \mathbf{R}_1\mathbf{R}$, we can write then $\boldsymbol{\phi}' = \mathbf{R}_1\boldsymbol{\phi}$ without any loss of information. We can alternatively also use the second column $\hat{\mathrm{c}}_2(\mathbf{R})$ as a shorthand and as a (so-called) conjugated spinor. (In [11] it is explained that $\hat{\mathrm{c}}_2(\mathbf{R})$ corresponds to a reversal. But in this paper we will hardly pay attention to the conjugated spinors. We will almost all of the time focus our attention on the first column as the represention of a rotation). We have this way discovered the well-defined geometrical meaning of a spinor. As already stated, it is just a group element. This is all that spinors in SU(2) are about. Spinors code group elements. Within SU(2), 2×2 rotation matrices operate on 2×1 spinor matrices. These spinor matrices represent themselves the rotations that are "rotated". Explaining that a spinor in SU(2) is a rotation is in our opinion far more illuminating than describing it as the square root of an isotropic vector according to the textbook doctrine. It is this insight that breaks the deadlock of our incomprehension. We will explain the textbook relationship between spinors and square roots of isotropic vectors in Subsection 2.7.

Stating that a spinor in SU(2) is a rotation is actually an *abus de langage*. A spinor is, just like a 3×3 SO(3) rotation matrix, an unambiguous representation of a rotation within the group theory. But due to the isomorphism we can merge the concepts and call the matrix or the spinor a rotation, in complete analogy with what we proposed in Subsection 1.4. For didactical reasons we can consider a spinor as conceptually equivalent to a system of "generalized coordinates" for a rotation.

We should not be surprised by the removal of the vectors from the formalism in favour of the group elements themselves as described above. Group theory is all about this kind of abstraction. We try to obtain general results from just a few abstract axioms for the group elements, without bothering about their intuitive meaning in a more specific context of a practical realization. And as far as representations are concerned, we do not have to get back to a specific context. We always have a representation at hand in the form of group automorphisms. This is a well-known fact, but in its general abstract formulation this fact looks indeed very abstract. Here we can see what this abstract representation in terms of automorphisms intuitively means in the context of the specific example of the rotation group. The idea is then no longer abstract: We can identify the 2×2 matrices $\mathbf{R}$ of SU(2) with the group automorphisms T_{g_k}, and the 2×1 rotation matrices $\boldsymbol{\phi}_j$ with the group elements g_j, such that $g_j \rightarrow g_k \circ g_j = T_{g_k}(g_j)$ is algebraically represented by: $\boldsymbol{\phi}_j \rightarrow \mathbf{R}\boldsymbol{\phi}_j$.

□ Remark 1. From this it must be already obvious that spinors in SU(2) do not build a vector space as we stressed in Subsection 2.2. The three-dimensional rotation group is not a vector space but a curved manifold (because the group is non-Abelian). We cannot try to find a meaning for a linear combination $\sum_j c_j\mathbf{R}_j$ of SU(2) matrices $\mathbf{R}_j$, in analogy to what we can do with 3×3 matrices in SO(3), where we can fall back on the fact that 3×1 matrices of the image space correspond to elements of a vector space $\mathbb{R}^3$ or $\mathbb{C}^3$. The reason for this is that the spinors $\boldsymbol{\phi}_j$ do not build a vector space, such that we cannot define $\sum_k c_k\mathbf{R}_k$ by falling back on some definition for $\sum_j c_j\boldsymbol{\phi}_j$ in the image space. And the very reason why we cannot define $\sum_j c_j\boldsymbol{\phi}_j = \sum_j c_j\hat{\mathrm{c}}_1(\mathbf{R}_j) = \hat{\mathrm{c}}_1(\sum_j c_j\mathbf{R}_j)$, is that we cannot define $\sum_j c_j\mathbf{R}_j$. In trying to define linear combinations of SU(2) matrices or spinors we hit thus a vicious circle from which we cannot escape. Furthermore, the relation between spinors and vectors of $\mathbb{R}^3$ is not linear as may have already transpired from Atiyah's statement cited above and as we will explain below (see Subsection 2.7). This frustrates all attempts to find a meaning for a linear combination of spinors in SU(2) based on the meaning of the linear combination with the same coefficients in SO(3). Trying to make sense of linear combinations of spinors is therefore an impasse.

□ Remark 2. We can extrapolate [7] the idea that the representation theory "rotates rotations rather than vectors" to SO(n), such that we will then obtain a good geometrical intuition for the group theory. If we could also extrapolate to SO(n) the idea that spinors are group elements, we would then obtain a very good intuition for spinors that is generally valid. We could then e.g. also understand why spinors constitute an ideal $\mathscr{I}$. The ideal would then just be the group and the group is closed with respect to the composition of rotations.

□ Remark 3. Unfortunately things are not that simple and we will not be able to realize this dream. The idea that spinors are just rotations gives us a very nice intuition for them in SU(2). But the interpretation in SU(2) of a single column matrix as a shorthand for the whole information needed to define a group element unambiguously is not correct in general. A first example of a case where the column matrices cannot be identified with group elements is the representation SL(2,$\mathbb{C}$) of the homogeneous Lorentz group. In fact, defining an element of the homogeneous Lorentz group requires specifying six independent real parameters. That information cannot possibly be present in a single 2×1 column of the 2×2 representation matrix. A second

example is the representation given by the Clifford algebra of SO(n). Characterizing an element of the rotation group SO(n) of $\mathbb{R}^n$ requires specifying $n(n-1)/2$ independent real parameters (see the discussion about the *Vielbein* in paragraph 2.7.1). The complete information about these $n(n-1)/2$ independent real parameters cannot always be crammed into the complex $2^\nu \times 1$ column matrices used in the representations, because there is a small set of values of n for which $n(n-1)/2 > 2^{\nu+1}$. The information about a group element contained in a column matrix is in these cases thus forcedly partial. These examples show that the identification between group elements and the column matrices we call spinors anticipated here is not true in general. The general meaning of a spinor can thus not be that it is a group element. What the general meaning could be becomes then less clear such that one has to consider like Cartan isotropic vectors, representing oriented planes, as discussed in [7].

However, due to the fact that we are forced to introduce a superposition of two states in order to derive the Dirac equation (see Subsection 4.1), the 4×1 column matrices used in the Dirac theory will again contain all the information about the group elements. For the applications in QM we can therefore maintain the idea that a spinor is a group element! Furthermore, what we stated in the previous remark does not imply that we do not understand the algebra of the representation. In fact, the $2^\nu \times 2^\nu$ representation matrices of the group SO(n) do represent the group elements. For the application in QM this means that we will really completely understand the formalism. In the approach to the general case SO(n) the main idea will thus be to consider the formalism just as a formalism of rotation matrices and the column matrices as auxiliary sub-quantities which encase only a subset of the complete information about group elements.

□ Remark 4. We must point out that we do not know with certainty to which extend Atiyah wanted to be general when he talked about "the square root of geometry". We think that what Atiyah had in mind was based on Eq. 29 and Eq. 57, rather than making a general statement for SO(n). We can see from Eq. 29 and Eq. 57 that the terminology "square root" used by Atiyah is only a loose metaphor, and in the generalization of the approach to groups of rotations in $\mathbb{R}^n$, with $n > 3$, the metaphor will become even more loose [7]. For SO(n) the ideas can be based on the developments in Subsection 2.6 where we point out a quadratic relationship between vectors and spinors, that is generally valid.

But for the moment we want to explore the idea of a single-column spinor that contains the complete information about a rotation in SU(2), where the intuitively attractive idea that a column spinor represents a group element is viable. It remains to explain under which form the information about the rotation is wrapped up inside this column matrix. This is done in several steps.

2.4 Generating the group from reflections

The first step is deciding that we will generate the whole group of rotations and reversals from reflections, based on the idea that a rotation of SO(3) is the product of two reflections as explained in Fig. 1. We therefore need to cast a reflection into the form of a 2×2-matrix. The coordinates of the unit vector $\mathbf{a} = (a_x, a_y, a_z)$, which is the normal to the reflection plane that defines the reflection A, should be present as parameters within the reflection matrix $\mathbf{A}$ but we do not know how. Therefore we decompose heuristically the matrix $\mathbf{A}$ that codes the reflection A defined by $\mathbf{a}$ linearly as $a_x\tau_x + a_y\tau_y + a_z\tau_z$, where τ_x, τ_y, τ_z are unknown matrices, as summarized in the following diagram:

$$\begin{array}{ccc} \text{unit vector } \mathbf{a} = (a_x, a_y, a_z) \in \mathbb{R}^3 & \xrightarrow[\text{reflection } A]{\mathbf{a} \text{ defines a}} & 2 \times 2 \text{ complex reflection matrix } \mathbf{A} \\ \Big\downarrow \text{definition} & & \Big\downarrow \text{Dirac's heuristics} \\ \mathbf{a} = a_x\mathbf{e}_x + a_y\mathbf{e}_y + a_z\mathbf{e}_z & \xleftarrow[\text{decompositions}]{\text{analogy of}} & \mathbf{A} = \underbrace{a_x\tau_x + a_y\tau_y + a_z\tau_z}_{\text{noted as } \mathbf{a}\cdot\boldsymbol{\tau}} \end{array} \tag{5}$$

If we know the matrix τ_x, this will tell us where and with which coefficients a_x pops up in $\mathbf{A}$. The same applies *mutatis mutandis* for τ_y and τ_z. The matrices τ_x, τ_y, τ_z, we use to code this way reflection matrices within $\mathbb{R}^3$, can be found by expressing isomorphically through $\mathbf{AA} = \mathbf{a}^2\mathbb{1} = \mathbb{1}$ what defines a reflection, viz. that the reflection operator A is idempotent. We find out that this can be done provided the three matrices simultaneously satisfy the six conditions $\tau_\mu\tau_\nu + \tau_\nu\tau_\mu = 2\delta_{\mu\nu}\mathbb{1}$, i.e. provided we take e.g. the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$ for τ_x, τ_y, τ_z. Here $\mathbb{1}$ is the 2×2 unit matrix.

□ Remark 1. Physicists among the readers will recognize that this construction is *algebraically* completely analogous to the one that introduces the gamma matrices in the Dirac equation. However, geometrically it is entirely different. Dirac's approach aims at taking the "square root of the Klein-Gordon equation". It searches thus a way to write vectors, e.g. the four-vector $(E, c\mathbf{p})$ as a linear expression that permits to interpret it as the square root of a quadratic form, e.g. $E^2 - c^2\mathbf{p}^2$, as e.g. explained by Deheuvels [16]. Dirac's derivation is therefore taking place within the context of the algebra of vectors and multivectors. Our approach consists in finding the expression for reflections. Our derivation takes thus place within the algebra of group elements. The two approaches do not define the same geometrical objects and not the same algebras.

□ Remark 2. We may note that to an extent, the fact that our heuristics work is a kind of a fluke, because the fact that the reflection matrix is linear in a_x, a_y, a_z within SU(2) is special and not general. It is typical of the spinor-based representations

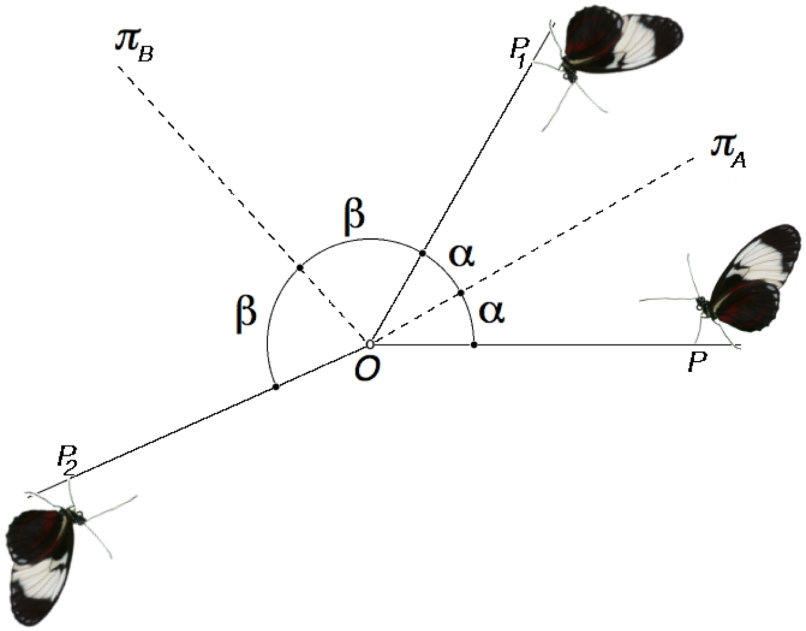

Fig. 1. A rotation R in $\mathbb{R}^3$ as the product of two reflections A and B defined by their reflection planes π_A and π_B. The planes π_A and π_B in $\mathbb{R}^3$ intersect along a straight line ℓ defined by $\ell = \{\, \mathbf{r} \in \mathbb{R}^3 \parallel (\exists\lambda \in \mathbb{R})(\mathbf{r} = \lambda\mathbf{n}) \,\}$. The plane of the figure is taken perpendicular to the line ℓ and intersects ℓ in the point O. We use the names π_A and π_B of the planes to label their intersections with the plane of the figure. The position vector OP of the point P to be reflected, is at an angle α with respect to π_A. We call $A(P) = P_1$ and $B(P_1) = P_2$. The position vector OP_1 is at angle β with respect to π_B. The angle between π_A and π_B is then $\alpha + \beta$. As can be seen from their operations on the Heliconius butterfly, reflections have negative parity, but the product of two reflections conserves the parity. The product of the two reflections is therefore a rotation $R = B \circ A$, with axis ℓ and rotation angle $2(\alpha + \beta)$. Only the relative angle $\alpha + \beta$ between π_A and π_B appears in the final result, not its decomposition into α and β. Hence, the final result will not be changed when we turn the two planes together as a whole around ℓ keeping $\alpha + \beta$ fixed (After [6]).

we present in this paper. A counter-example is the expression for a reflection matrix $\mathbf{A}$ in SO(3), which is quadratic in the parameters a_x, a_y, and a_z:

$$\mathbf{A} = \mathbb{1} - 2 \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \otimes \begin{bmatrix} a_x & a_y & a_z \end{bmatrix}. \tag{6}$$

Writing $\mathbf{A}$ this way permits to verify immediately algebraically that it corresponds to $\mathbf{v} \to A(\mathbf{v}) = \mathbf{v} - 2(\mathbf{a} \cdot \mathbf{v})\,\mathbf{a}$. Writing $\mathbb{1}$ as $(a_x^2 + a_y^2 + a_z^2)\,\mathbb{1}$ in Eq. 6 shows that the expression is purely quadratic. This is due to the fact that vectors in SO(3) are rank-2 tensor products of the spinors of SU(2) as we will discuss in this paper. We may also note that we have defined the reflection matrices without defining a "vector space" on which they would be working. They are defined *en bloc*, and it is this aspect that saddles us up with the problem of the meaning of the column matrices, called spinors, which occur in the formalism. We are used to qualify such column matrices as column vectors, but as we pointed out spinors are not vectors. It is thus no longer natural to break down the square matrices into columns. The complete information resides in the block of the square matrix. When we break up that block into columns the information contained in a column may be partial, and perhaps the question what a column then means might be a ill-conceived (see [7], pp.22-23). This complies with the idea expressed in Remark 3 of Subsection 2.3.

We discuss in Subsection 2.9 of [7] that the solution $(\tau_x, \tau_y, \tau_z) = (\sigma_x, \sigma_y, \sigma_z)$ is not unique and that there are many other possible choices. But we follow here the tradition to adopt the choice of the Pauli matrices. The reflection matrix $\mathbf{A}$ is thus given by:

$$A \to \mathbf{A} = a_x\sigma_x + a_y\sigma_y + a_z\sigma_z = \begin{bmatrix} a_z & a_x - \imath a_y \\ a_x + \imath a_y & -a_z \end{bmatrix} \hat{=} \mathbf{a}\cdot\boldsymbol{\sigma}. \tag{7}$$

The symbol $\hat{=}$ serves here to warn that the notation $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ is a purely conventional shorthand for $a_x\sigma_x + a_y\sigma_y + a_z\sigma_z$. It does not express a true scalar product involving $\mathbf{a}$, but just exploits the mimicry with the expression for a scalar product to introduce the shorthand.

By expressing a rotation as the product of two reflections, one can then derive the well-known Rodrigues formula:

$$\mathbf{R}(\mathbf{n},\varphi) = \mathbf{BA} = \begin{bmatrix} b_z & b_x - \imath b_y \\ b_x + \imath b_y & -b_z \end{bmatrix} \begin{bmatrix} a_z & a_x - \imath a_y \\ a_x + \imath a_y & -a_z \end{bmatrix}$$

$$= \cos(\varphi/2)\,\mathbb{1} - \imath\sin(\varphi/2)\,[\,\mathbf{n}\cdot\boldsymbol{\sigma}\,], \tag{8}$$

for a rotation by an angle φ around an axis defined by the unit vector $\mathbf{n}$. To derive this result it suffices to consider two reflections A (with matrix $[\mathbf{a}\cdot\boldsymbol{\sigma}]$) and B (with matrix $[\mathbf{b}\cdot\boldsymbol{\sigma}]$) whose planes contain $\mathbf{n}$, and which have an angle $\varphi/2$ between them. Using the algebraic identity $[\mathbf{b}\cdot\boldsymbol{\sigma}]\,[\mathbf{a}\cdot\boldsymbol{\sigma}] = (\mathbf{b}\cdot\mathbf{a})\,\mathbb{1} + \imath(\mathbf{b}\wedge\mathbf{a})\cdot\boldsymbol{\sigma}$ yields then the desired result. There is an infinite set of such pairs of planes, and which precise pair one chooses from this set does not matter.

Starting from Eq. 8 it is easy to check that each rotation matrix has the form given by Eq. 3 and therefore belongs to SU(2). Conversely each element of SU(2) is a rotation matrix. We can now appreciate also why SU(2) is a double covering of SO(3). Consider the matrix product:

$$\mathbf{BA} = \begin{bmatrix} b_z & b_x - \imath b_y \\ b_x + \imath b_y & -b_z \end{bmatrix} \begin{bmatrix} a_z & a_x - \imath a_y \\ a_x + \imath a_y & -a_z \end{bmatrix}, \tag{9}$$

in the derivation of the Rodrigues equation in Eq. 8. Imagine that we keep A fixed and increase the angle $\vartheta = \varphi/2$ between the reflection planes π_A and π_B of A and B from $\vartheta = 0$ onwards. Of course φ is the angle of the rotation $R = BA$. This means that the reflection plane π_B with normal vector $\mathbf{b}$ that defines B is rotating. In the matrix product that occurs in Eq. 9, the numbers in the matrix $\mathbf{A}$ would remain fixed, while the numbers in the matrix $\mathbf{B}$ would be continuously changing like the digits that display hundredths of seconds on a wrist watch. When the starting value of the angle $\vartheta = \varphi/2$ between the reflection planes π_B and π_A is zero, the reflection planes are parallel, $\pi_B \parallel \pi_A$, and the starting value of $\mathbf{b}$ is $\mathbf{b} = \mathbf{a}$. When $\vartheta = \varphi/2$ reaches the value π, the rotating reflection plane π_B will have come back to its original position parallel to the fixed reflection plane π_A, and the resulting rotation BA will correspond to a rotation over an angle $\varphi = 2\vartheta = 2\pi$.

As far as group elements are concerned, we have thus made a full turn both of the reflection B and the rotation BA when π_B will have made a turn of $\vartheta = \pi$ in $\mathbb{R}^3$. This is because we need to rotate a plane in $\mathbb{R}^3$ only over $\vartheta = \pi$ to bring it back to its original position. The consequence of this is that we can define any plane π_U (or reflection U) always equivalently by two normal unit vectors $\mathbf{u}$ and $-\mathbf{u}$. These full turns of B and $R = BA$ within the group must be parameterized with a "group angle" $\varphi_G = 2\pi$ if we want to express the periodicity within the rotation group in terms of trigonometric functions. However, for the normal vector $\mathbf{b}$ which we have used to define B and which belongs to $\mathbb{R}^3$ this is different. For $\vartheta = \varphi/2 = 0$, its starting value is $\mathbf{b} = \mathbf{a}$. For $\vartheta = \varphi/2 = \pi$, its value has become $\mathbf{b} = -\mathbf{a}$, such that we obtain $\mathbf{R} = -\mathbb{1}$ in Eq. 9. There is nothing wrong with that because both the normal vectors $\mathbf{b} = \mathbf{a}$ and $\mathbf{b} = -\mathbf{a}$ define the same plane $\pi_B \parallel \pi_A$. Each group element g is thus represented by two matrices $\mathbf{G}$ and $-\mathbf{G}$. As the group elements B and $R = BA$ have recovered their initial values we have $\varphi_G = 2\pi$. In general, we have $\varphi_G = 2\vartheta = \varphi$. Only after a rotation over a "group angle" $\varphi_G = 4\pi$, which corresponds to a rotation of π_B over an angle $\vartheta = \varphi/2 = 2\pi$ will we obtain the values $BA = \mathbb{1}$ and $\mathbf{b} = \mathbf{a}$.

□ Remark 3. It is often presented as a mystery of QM that one must turn the wave function over $\varphi_G = \varphi = 4\pi$ before we obtain the starting configuration ($\vartheta = \varphi/2 = 2\pi$) again. There is even a beautiful neutron experiment that has been performed to provide physical proof for the truth of this fact to physicists [17]. We can see from a proper understanding of the group theory that this is quite trivial and that it is a mathematical rather than physical truth. Most textbooks mystify this subject matter by invoking topological arguments. We explain this link with topology in [6], Subsection 3.11.2 and Fig. 3.5, where we compare a full turn on the group with a full turn on a Moebius ring. This link is thus conceptually very clear and simple. However, in the illustration of this topological argument by Feynman [18], Dirac [19] or Misner *et al.* [20] the connection between the topological argument and the physical model is hard to see. It is e.g. very difficult to follow how disentangling the threads in the work of Misner *et al.* would make the point.

□ Remark 4. Representing a rotation as the product of two reflections is convenient for calculating the product of two rotations. Consider two rotations $\mathbf{R}_1(\mathbf{n}_1,\varphi_1)$ and $\mathbf{R}_2(\mathbf{n}_2,\varphi_2)$. Call π the plane defined by $\mathbf{n}_1$ and $\mathbf{n}_2$. Call π_1 the plane of

the reflection that defines $\mathbf{R}_1$ as $\pi \circ \pi_1$ and π_2 the plane of the reflection that defines $\mathbf{R}_2$ as $\pi_2 \circ \pi$. It follows then that $\mathbf{R}_2 \circ \mathbf{R}_1 = \pi_2 \circ \pi \circ \pi \circ \pi_1 = \pi_2 \circ \pi_1$.

2.5 Fleshing out the caveat: A superposition principle for spinors?

2.5.1 An SU(2)-specific approach

In Subsection 2.2 we issued the warning that spinors can a priori not be summed. We can now illustrate how the procedure of summing spinors is geometrically obscure. Consider a rotation R_1 over an angle φ around the axis defined by the unit vector $\mathbf{n}$, and a rotation R_2 over an angle ϑ around the axis defined by the unit vector $\mathbf{m}$. Using Eq. 8, we have then:

$$\begin{aligned} \boldsymbol{\phi}_1 &= \hat{c}_1(\mathbf{R}_1) = \begin{bmatrix} \cos(\varphi/2) - \imath n_z \sin(\varphi/2) \\ -\imath n_x \sin(\varphi/2) + n_y \sin(\varphi/2) \end{bmatrix}, \\ \boldsymbol{\phi}_2 &= \hat{c}_1(\mathbf{R}_2) = \begin{bmatrix} \cos(\vartheta/2) - \imath m_z \sin(\vartheta/2) \\ -\imath m_x \sin(\vartheta/2) + m_y \sin(\vartheta/2) \end{bmatrix}. \end{aligned} \tag{10}$$

Summing $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ as though they were vectors is algebraically perfectly feasible. We obtain:

$$\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 = \begin{bmatrix} \cos(\varphi/2) + \cos(\vartheta/2) - \imath(n_z \sin(\varphi/2) + m_z \sin(\vartheta/2)) \\ -\imath((n_x \sin(\varphi/2) + m_x \sin(\vartheta/2)) + (n_y \sin(\varphi/2) + m_y \sin(\vartheta/2)) \end{bmatrix}. \tag{11}$$

But what does the result mean geometrically? The quantity $\boldsymbol{\phi} = \boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$ cannot represent a rotation because $\boldsymbol{\phi}^\dagger \boldsymbol{\phi} \neq 1$. It is therefore not a true spinor. It corresponds obviously to $\hat{c}_1(\mathbf{R}_1 + \mathbf{R}_2)$, and as explained in Remark 1 of Subsection 2.3 we cannot interpret $\mathbf{R}_1 + \mathbf{R}_2$ the way we can interpret a sum of rotation matrices in SO(3), because the spinors do not build a vector space. To interpret $\mathbf{R}_1 + \mathbf{R}_2$ we would need an interpretation of sums of spinors, and to interpret sums of spinors we would need an interpretation of sums of rotation matrices. Therefore, when we try to transpose the ideas from SO(3) to SU(2), we end up running in circles.

But suppose now that we try to normalize the result in Eq. 11 to 1 as physicists do routinely. The result will then remain a linear combination of spinors but it is now a special one, whereby the coefficients used in the linear combination preserve the normalization. One must then find a rationale to explain what the geometrical idea behind such a procedure could be. Mind in this respect that we have no idea about the geometrical meaning of $\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$ in the first place. How do we justify defining a procedure on a quantity that is undefined? The procedure remains thus geometrically impenetrable, and we have rendered the situation worse. We have now *concealed* the fact that there are conceptual problems with making linear combinations of spinors, because the final quantity obtained is now (almost always) algebraically identical to a true spinor. Let us prove this. To normalize $\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$ according to the Hermitian norm we calculate:

$$(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2)^\dagger (\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2) = 2\,[\,1 + \cos(\varphi/2)\cos(\vartheta/2) + (\mathbf{n} \cdot \mathbf{m})\,\sin(\varphi/2)\sin(\vartheta/2)\,]. \tag{12}$$

Here:

$$\cos(\Omega/2) = \cos(\varphi/2)\cos(\vartheta/2) + (\mathbf{n} \cdot \mathbf{m})\,\sin(\varphi/2)\sin(\vartheta/2), \tag{13}$$

allows for a geometrical interpretation: Ω is the rotation angle of the product rotation $R_2 \circ R_1$ as shown e.g. in Appendix C of the monograph of Jones [21]. We are already running into trouble here, because it is certainly conceivable that $1 + \cos(\Omega/2) = 0$. The result $\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$ is then zero, such that it cannot be normalized to 1. This happens e.g. when we define $\mathbf{R}_2(\mathbf{m}, \vartheta)$ by: $\mathbf{m} = \mathbf{n}$ and $\vartheta = \varphi + 2\pi$. This is actually the only way this can happen, because $\boldsymbol{\phi}_1 = -\boldsymbol{\phi}_2$ implies $\mathbf{R}_1 = -\mathbf{R}_2$, such that $\mathbf{m} = \mathbf{n}$ and $\vartheta = \varphi + 2\pi$. This example is the absolute proof for the fact that the sum of two spinors is not a spinor. Let us now continue carrying out the algebra keeping this in mind and check whether there could be other problems. Writing the sum $\mathbf{R}_1 + \mathbf{R}_2$ in the form of the Rodrigues equation Eq. 8 makes it clear that the vector:

$$\mathbf{v} = \mathbf{n}\sin(\varphi/2) + \mathbf{m}\sin(\vartheta/2), \tag{14}$$

plays a prominent rôle in the algebra. Let us now assume that $1 + \cos(\Omega/2) \neq 0$ and calculate the result of normalizing the purely formal algebraic sum $\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$ to 1. This yields:

$$\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 \rightsquigarrow \frac{1}{\sqrt{2(1 + \cos(\Omega/2))}} \begin{bmatrix} \cos(\varphi/2) + \cos(\vartheta/2) - \imath v_z \\ -\imath(v_x + \imath v_y) \end{bmatrix}. \tag{15}$$

Let us now try to identify the right-hand side with a spinor $\boldsymbol{\psi}$ representing a rotation $R(\mathbf{u},\alpha)$ over an angle α around an axis defined by the unit vector $\mathbf{u}$:

$$\boldsymbol{\psi} = \begin{bmatrix} \cos(\alpha/2) - \imath \sin(\alpha/2) u_z \\ -\imath \sin(\alpha/2)(u_x + \imath u_y) \end{bmatrix}. \tag{16}$$

Very obviously, the rotation angle α must then be given by:

$$\begin{aligned} \alpha &= 2 \arccos \left[\frac{\cos(\varphi/2) + \cos(\vartheta/2)}{\sqrt{2(1+\cos(\Omega/2))}} \right] \\ &= 2 \arccos \left[\frac{\cos(\varphi/2) + \cos(\vartheta/2)}{2\cos(\Omega/4)} \right]. \end{aligned} \tag{17}$$

But we must check whether this is a meaningful expression. The rotation angle α will only be defined if $|\cos(\varphi/2) + \cos(\vartheta/2)| \leq \sqrt{2(1+\cos(\Omega/2))}$. To check this, we square both sides and rewrite 2 as $\cos^2(\varphi/2)+\sin^2(\varphi/2)+\cos^2(\vartheta/2)+\sin^2(\vartheta/2)$. We obtain then the inequality:

$$\begin{aligned} &\cos^2(\varphi/2) + \cos^2(\vartheta/2) + 2\cos(\varphi/2)\cos(\vartheta/2) \\ &\leq \sin^2(\varphi/2) + \cos^2(\varphi/2) + \sin^2(\vartheta/2) + \cos^2(\vartheta/2) \\ &+2\cos(\varphi/2)\cos(\vartheta/2) + 2\,(\mathbf{n}\cdot\mathbf{m})\,\sin(\varphi/2)\sin(\vartheta/2), \end{aligned} \tag{18}$$

where we have used the definition of $\cos(\Omega/2)$. Simplification leads to:

$$0 \leq \sin^2(\varphi/2) + \sin^2(\vartheta/2) + 2\,(\mathbf{n}\cdot\mathbf{m})\,\sin(\varphi/2)\sin(\vartheta/2) = \mathbf{v}^2, \tag{19}$$

such that the inequality is indeed satisfied. It must be noted now that $|\mathbf{v}|$ can be larger than 1 (but not larger than 2). It is therefore *a priori* not obvious that we can identify:

$$\frac{\mathbf{v}}{\sqrt{2(1+\cos(\Omega/2))}} = \sin(\alpha/2)\,\mathbf{u}, \tag{20}$$

where $\mathbf{u} \parallel \mathbf{v}$ is a unit vector. But the calculations that occur in the simplification from Eq. 18 to Eq. 19 show that $\mathbf{v}^2 = 2(1+\cos(\Omega/2)) - |\cos(\varphi/2) + \cos(\vartheta/2)|^2$, such that we have indeed $|\mathbf{v}| \leq \sqrt{2(1+\cos(\Omega/2))}$. We can thus calculate the unit vector $\mathbf{u} \parallel \mathbf{v}$ from:

$$\mathbf{u} = \frac{\mathbf{n}\sin(\varphi/2) + \mathbf{m}\sin(\vartheta/2)}{\sqrt{\sin^2(\varphi/2) + \sin^2(\vartheta/2) + 2\,(\mathbf{n}\cdot\mathbf{m})\,\sin(\varphi/2)\sin(\vartheta/2)}}. \tag{21}$$

While the normalized sum of two spinors can this way be interpreted in terms of a well-defined rotation $R(\mathbf{u},\alpha)$, it is not obvious what this is kind of operation $(R_1, R_2) \rightarrow R(\mathbf{u},\alpha)$ is then supposed to mean geometrically. The meaning of the unit vector $\mathbf{u}$ is at least algebraically clear as the sum of two wedge products. But the definition of the rotation angle α looks impenetrable.

A superposition principle for spinors, i.e. summing and making linear combinations of them with a wave picture in mind, as physicists routinely do, is thus an all but self-evident procedure. Within the initial set of underlying ideas this procedure is *a priori* geometrically meaningless despite its misleading apparent algebraic simplicity. Interpreting a sum of spinors as presented in this paragraph is actually a conceptual impasse, because the sum can be zero. The use of the superposition principle in physics requires therefore a supplementary geometrical justification. That this caveat is not futile at all can be appreciated from the fact that it is the very introduction of the superposition principle which transforms the spinor formalism, which is in essence purely geometrical and classical, into a much less obvious Hilbert space formalism of QM. One of the mysterious creatures we introduce this way is Schrödinger's cat. This need for a justification of the superposition principle is further directly related to the conceptual difficulties encountered under the form of the so-called particle-wave duality in QM. And in interference we become directly confronted with the fact that the sum of two spinors can be zero when $1 + \cos(\Omega/2) = 0$ as outlined above. This leads to severe conceptual difficulties.

2.5.2 General group-theoretical approach

We may note that the *ad hoc* attempt to interpret the meaning of an element of the group ring presented in paragraph 2.5.1 is specific to SU(2). It does not solve the problem of the meaning of an element of the group ring for the permutations in Eq. 1 or for moves of the Rubik's cube. And even within SU(2) it fails as it is meaningless. We will refrain from interpreting the group theory in terms of a vector space $\mathbb{C}^2$ and proceed as always on the basis of purely group-theoretical considerations. Formal sums of group elements and group rings occur all the time in the group theory (see e.g. [15]). In this context, one encounters e.g. *formal* identities:

$$g \circ (h_1 + h_2 \cdots h_n) = (h_1 + h_2 \cdots h_n) \circ g. \tag{22}$$

Here $g, h_1, h_2, \cdots h_n$ are all group elements. In fact, all this expresses is an identity for sets:

$$g \circ \{h_1, h_2, \cdots h_n\} = \{h_1, h_2, \cdots h_n\} \circ g. \tag{23}$$

From a purely group-theoretical viewpoint, we can thus interpret sums of group elements in terms of sets. The interpretation is *naturally provided by the group theory.* Here the coefficients in the linear combinations are all equal to one. We can extend this idea further and allow for integer values. We could e.g. imagine that we have a collection of 3000 Rubik's cubes whereby 2000 of the cubes have the configuration of group element g_1 and 1000 the configuration of group element g_2. We could then note the collection as $2000g_1 + 1000g_2$, or in terms of frequencies as $(2/3)g_1 + (1/3)g_2$. In QM we will note this collection as $\sqrt{2/3}g_1 + \sqrt{1/3}g_2$. It based on the fact that spinors $\boldsymbol{\psi}$ in SU(2) satisfy the identity $\boldsymbol{\psi}^\dagger\boldsymbol{\psi} = 1$, such that if we want to count objects, e.g. electrons, which all carry just one spinor with them in order to describe their state, then we must do it by counting $\boldsymbol{\psi}^\dagger\boldsymbol{\psi} = 1$. We must postpone the in-depth discussion of this to paragraph 4.5.1 when we will have derived the Dirac equation.

2.6 A parallel formalism for vectors

By construction, the representation SU(2) contains for the moment, (as we explained) *deliberately*, only group elements. Of course, it would be convenient if we were also able to calculate the action of the group elements on vectors. This is our next step. We can figure out how to do this based on the fact that we have already used a unit vector $\mathbf{a}$ to define a reflection A and its corresponding reflection matrix $\mathbf{A}$. Inversely, the reflection A also defines $\mathbf{a}$ up to a sign, such that there exists a one-to-one correspondence between reflections A and the two-member sets of unit vectors $\{\mathbf{a}, -\mathbf{a}\}$ (and the corresponding two-member sets of reflection matrices $\{\mathbf{A}, -\mathbf{A}\}$). This one-to-one correspondence between two-member sets of vectors and reflections will actually impose the formalism for vectors upon us. We can consider that a reflection A and its parameter set $\{\mathbf{a}, -\mathbf{a}\}$ are conceptually the same thing.

When a reflection travels around the group, the two-member set of vectors $\{\mathbf{a}, -\mathbf{a}\}$ will travel together with it. Let us explain what we mean by the informal term "traveling" here. In SO(3), a vector $\mathbf{v} \in \mathbb{R}^3$ has a 3×1 representation matrix $\mathbf{V}$. It is transformed by a group element g with 3×3 representation matrix $\mathbf{G}$ into another vector $\mathbf{v}' = g(\mathbf{v}) \in \mathbb{R}^3$: we just calculate the 3×1 representation matrix $\mathbf{V}'$ of $g(\mathbf{v})$ as $\mathbf{V}' = \mathbf{G}\mathbf{V}$. The vector $\mathbf{v}$ travels this way under a group action to another vector $\mathbf{v}'$. The point we want to make is that in SU(2), things are not as simple. Under the action of a group element g with matrix representation $\mathbf{G}$ a reflection A will not travel to another reflection A'.

Let G be the group generated by the reflections. The subgroup of pure rotations $G_+ \subset G$ is the subset obtained from an even number of reflections. The subset $G_- \subset G$ obtained from an odd number of reflections is not a subgroup. It contains the reflections and the reversals. Reflections are of course geometrical objects of a different type than reversals and pure rotations. This transpires also from the fact that a reflection is defined by a unit vector $\mathbf{a} \in \mathscr{S}_2$, where $\mathscr{S}_2$ is the unit sphere in $\mathbb{R}^3$. It is thus defined by two independent real parameters while rotations and reversals are defined by three independent real parameters. Group elements $g_1 \in G$ and $g_2 \in G$ are of the same geometrical type if they are related by a similarity transformation: $\exists g \in G : g_2 = g \circ g_1 \circ g^{-1}$. They have then the same group character.

In general, a new group element gA obtained by operating with an arbitrary group element $g \in G$ on the reflection A will no longer be a reflection that can be associated with a unit vector like it was the case for A, because in general gA can be of a different geometrical type than A. Group elements that transform a reflection A into an other reflection B, are the identity element $AA = \mathbb{1}$ and rotations R that can be written as $R = BA$. For this to be possible the rotation axis of R must belong to the reflection plane π_A of A. In other words the reflections do not travel according to the general rule $A \to gA$.

In order to transform a reflection A *always* into another reflection, we must use a similarity transformation: $A \to gAg^{-1}$. Hence, if B and A are reflections, defined by the unit vectors $\mathbf{b}$ and $\mathbf{a}$ then there exists a group element $g \in G$, such that $B = gAg^{-1}$ and $\mathbf{b} = g(\mathbf{a})$. Hence, if A is a reflection operating on $\mathbf{r} \in \mathbb{R}^3$ then the similar reflection B

that operates on $g(\mathbf{r}) \in \mathbb{R}^3$ will be represented by $g \circ A \circ g^{-1}$. The reflection plane π_B and normal $\mathbf{b}$ of this reflection B will have the same angles with respect to $g(\mathbf{r})$ as π_A and $\mathbf{a}$ with respect to $\mathbf{r}$. We can move thus this way the reflection A in $\mathbf{r}$ around to group elements B in $g(\mathbf{r})$, and of course the parameter set $\{\mathbf{a}, -\mathbf{a}\}$ will travel with it from $\mathbf{r}$ to $g(\mathbf{r})$ to a parameter set $\{\mathbf{b}, -\mathbf{b}\} = \{g(\mathbf{a}), -g(\mathbf{a})\}$. The ambiguity between $\{\mathbf{a}, -\mathbf{a}\}$ and $\{\mathbf{b}, -\mathbf{b}\}$ is also carried along. For the representation matrices of reflections we have thus:

$$\{[\,\mathbf{b}\cdot\boldsymbol{\sigma}\,], -[\,\mathbf{b}\cdot\boldsymbol{\sigma}\,]\} \equiv \mathbf{B} = \mathbf{GAG}^{-1} \equiv \mathbf{G}\,\{[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,], -[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]\}\,\mathbf{G}^{-1} \text{ if } g \in G, \tag{24}$$

whereby we allow for the ambiguity in the sign of $\mathbf{b}$, because Eq. 24 is not a transformation law for vectors, but for reflections and their associated two-member sets of vectors.

Of course the idea would be that $g(\mathbf{a}) = \mathbf{b}, \forall g \in G_+$ and $g(\mathbf{a}) = -\mathbf{b}, \forall g \in G_-$, but the combined presence of $\mathbf{G}$ and $\mathbf{G}^{-1}$ does not permit to reproduce the change of sign in the formalism, because it has been for designed for group elements, not for vectors. This is very clear for $A(\mathbf{a}) = -\mathbf{a}$, while in the formalism $\mathbf{A}\,[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]\,\mathbf{A}^{-1} = [\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$, which is the correct calculation for $A = A \circ A \circ A^{-1}$. On the other hand, a vector $\mathbf{b}$ that is perpendicular to $\mathbf{a}$ is characterized by $[\,\mathbf{b}\cdot\boldsymbol{\sigma}\,][\,\mathbf{a}\cdot\boldsymbol{\sigma}\,] = -[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,][\,\mathbf{b}\cdot\boldsymbol{\sigma}\,]$.

To see this, consider the rotation R that transforms $\mathbf{e}_x$ into $\mathbf{a}$ and $\mathbf{e}_y$ into $\mathbf{b}$. For the reflections σ_x and σ_y we have $\sigma_x\sigma_y = -\sigma_x\sigma_y$. The similarity transformation based on $\mathbf{R}$ will transform σ_x into the reflection A with matrix representation $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ and σ_y into the reflection B with matrix representation $[\,\mathbf{b}\cdot\boldsymbol{\sigma}\,]$. Applying the similarity transformation to $\sigma_x\sigma_y = -\sigma_x\sigma_y$ proves then the identity. Therefore $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,][\,\mathbf{b}\cdot\boldsymbol{\sigma}\,][\,\mathbf{a}\cdot\boldsymbol{\sigma}\,] = -[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,][\,\mathbf{a}\cdot\boldsymbol{\sigma}\,][\,\mathbf{b}\cdot\boldsymbol{\sigma}\,] = -[\,\mathbf{b}\cdot\boldsymbol{\sigma}\,]$, while the vector $\mathbf{b} \in \pi_A$ belongs to the reflection plane and should not change sign under the reflection A.

We see thus that in all cases, we get the sign of the reflected vector wrong. We can thus lift the ambiguity and treat the vectors correctly by introducing the sign by brute force:

$$\begin{aligned}
[\,\mathbf{b}\cdot\boldsymbol{\sigma}\,] &= +\mathbf{G}\,[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]\,\mathbf{G}^{-1}, \text{ if } g \in G_+, \\
[\,\mathbf{b}\cdot\boldsymbol{\sigma}\,] &= -\mathbf{G}\,[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]\,\mathbf{G}^{-1}, \text{ if } g \in G_-.
\end{aligned} \tag{25}$$

In doing so, we quit the formalism for group elements and enter a new formalism for vectors. The transition is enacted by conceiving and elaborating the idea that we can use the matrix $\mathbf{A} = [\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ also as the representation of the unit vector $\mathbf{a}$, since the matrix $\mathbf{A}$ contains the components of the vector $\mathbf{a}$ and the reflection A defines $\mathbf{a}$. To get rid of the ambiguity about the signs of the vectors that exist within the definition of the reflection matrices, it suffices to use $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ as a representation for a unit vector $\mathbf{a}$, and to introduce the rule that $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ is transformed according to:

$$[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,] \rightarrow [\,R(\mathbf{a})\cdot\boldsymbol{\sigma}\,] = -\mathbf{R}\,[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]\,\mathbf{R}^{-1} \text{ under } \textit{reflections } R \in G_-. \tag{26}$$

This will be further justified below. The transformation under other elements $g \in G$ is then obtained by using the decomposition of g into reflections. We have this way developed a parallel formalism for the matrices $\mathbf{A}$, wherein $\mathbf{A}$ takes now a different meaning, *viz.* that of a representation of a unit vector $\mathbf{a}$ and obeys a different kind of transformation algebra, that is no longer linear but quadratic in the transformation matrices. This idea can be generalized to a vector $\mathbf{v}$ of arbitrary length v, which is then represented by $\mathbf{V} = v_x\sigma_x + v_y\sigma_y + v_z\sigma_z$. In fact, the scalar v is a group invariant, because the rotation group is defined as the group that leaves v invariant. We have then $\mathbf{V}^2 = -(\det \mathbf{V})\mathbb{1} = v^2\mathbb{1}$.

This idea that within SU(2) a vector $\mathbf{v} \in \mathbb{R}^3$ is represented by a matrix $\mathbf{v}\cdot\boldsymbol{\sigma}$ according to the isomorphism:

$$\mathbf{v} = v_x\mathbf{e}_x + v_y\mathbf{e}_y + v_z\mathbf{e}_z \longleftrightarrow v_x\sigma_x + v_y\sigma_y + v_z\sigma_z = \begin{bmatrix} v_z & v_x - iv_y \\ v_x + iv_y & -v_z \end{bmatrix} \hat{=}\, \mathbf{v}\cdot\boldsymbol{\sigma}, \tag{27}$$

was introduced by Cartan [4]. It is a definition that makes it possible to do calculations on vectors. In reading Cartan one could get the impression that we have the leisure to introduce this definition at will. In reality, it is not a matter of mere definition. While introducing the idea as a definition would not lead to errors in the formalism, it would nevertheless be a false presentation of the state of affairs, because it is no longer at our discretion to define things at will. As we can see from the reasoning above, the definition is entirely forced upon us by the one-to-one correspondence between sets of unit vectors $\{\mathbf{a}, -\mathbf{a}\}$ and reflections A.

We cannot stress enough that even if reflections $A \in \mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$ and unit vectors $\mathbf{a} \in \mathbb{R}^3$ are both represented by the same 2×2 matrix $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$, they are obviously completely different quantities, belonging to completely different spaces $\mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$ and $\mathbb{R}^3$ and completely different algebras.

Using $(\mathbf{v}_1 + \mathbf{v}_2)^2 - \mathbf{v}_1^2 - \mathbf{v}_2^2 = 2\,\mathbf{v}_1\cdot\mathbf{v}_2$, one can derive from the rule $\mathbf{V}^2 = v^2\mathbb{1}$ that $\mathbf{V}_1\mathbf{V}_2 + \mathbf{V}_2\mathbf{V}_1 = 2\,(\mathbf{v}_1\cdot\mathbf{v}_2)\,\mathbb{1}$, which can be seen as an alternative definition of the parallel formalism for vectors. As anticipated above, we can use this result to check the correctness of the rule of Eq. 26 geometrically. It suffices in this respect to observe that the reflection A, defined by the unit vector $\mathbf{a}$, transforms $\mathbf{v}$ into $A(\mathbf{v}) = \mathbf{v} - 2(\mathbf{v}\cdot\mathbf{a})\,\mathbf{a}$. Expressed in the matrices this yields: $\mathbf{V} \rightarrow -\mathbf{AVA}$.

We see that the transformation law for vectors $\mathbf{v}$ is quadratic in $\mathbf{A}$ in contrast with the transformation law for group elements g, which is linear: $\mathbf{G} \rightarrow \mathbf{AG}$. *Vectors transform thus quadratically as rank-2 tensor products of spinors, whereas spinors transform linearly.* This gives us a full understanding of the relationship between vectors and spinors. It is much easier to understand this relationship in the terms used here, vectors are quadratic expressions in terms of spinors, than in the equivalent terms used by Atiyah, spinors are square roots of vectors.

□ Remark 1. This solution is analogous to the solution proposed by Gauss, Wessel and Argant to solve the problem of the meaning of $\imath = \sqrt{-1}$. As described on p. 118 of reference [12], one defines first $\mathbb{C}$ as $\mathbb{R}^2$, with two operations + and × defined by $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$ and $(x_1, y_1) \times (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1)$. One then shows that $(\mathbb{R}, +, \times)$ is isomorphic to $(\mathscr{R}, +, \times)$ where $\mathscr{R} = \{ (x, y) \in \mathbb{C} \parallel y = 0 \} \subset \mathbb{C}$. This permits to identify $\mathbb{R} \equiv \mathscr{R}$ and justifies introducing the notations $1 \equiv (1, 0) \in \mathbb{R}$, $\imath \equiv (0, 1)$ and $(x, y) \equiv x + \imath y$. One can prove then that $\imath^2 \equiv (0, 1)^2 = (-1, 0) \equiv -1$.

The fact that this solution for the riddle what the meaning of a spinor is, has escaped attention is due to the fact that spinors are in general introduced based on the construction proposed in Eq. 29 below. This construction emphasizes the fact that a spinor is a kind of square root of a vector at the detriment of the notion developed here, that a vector is a rank-2 expression in terms of spinors. But these relations between spinors and vectors are a property that constitutes only a secondary notion, which is not really instrumental in clarifying the concept of a spinor. The essential and clarifying notion in SU(2) is that a spinor corresponds to a rotation.

The reader will notice that the definition $\mathbf{V} = \mathbf{v}\cdot\boldsymbol{\sigma}$ with $\mathbf{V}^2 = v^2\mathbb{1}$ is analogous to Dirac's way of introducing the gamma matrices to write the energy-momentum four-vector as $E\gamma_t + c\mathbf{p}\cdot\boldsymbol{\gamma}$ and postulating $(E\gamma_t + c\mathbf{p}\cdot\boldsymbol{\gamma})^2 = (E^2 - c^2p^2)\mathbb{1}$. In other words, it is the metric that defines the whole formalism, because we are considering groups of metric-conserving transformations (as the definition of a geometry in the philosophy of Felix Klein's Erlangen program).

For more information about the calculus on the rotation and reversal matrices, we refer the reader to reference [6]. Let us just mention that as a reflection A works on a vector $\mathbf{v}$ according to $\mathbf{V} \rightarrow -\mathbf{AVA} = -\mathbf{AVA}^{-1}$, a rotation $R = BA$ will work on it according to $\mathbf{V} \rightarrow \mathbf{BAVAB} = \mathbf{RVR}^{-1} = \mathbf{RVR}^{\dagger}$. The identity $\mathbf{R}^{-1} = \mathbf{R}^{\dagger}$ explains, in an alternative way, why the representation we end up with is SU(2).

In summary, there are two parallel formalisms in SU(2), one for the vectors and one for the group elements. In both formalisms a matrix $\mathbf{V} = \mathbf{v}\cdot\boldsymbol{\sigma}$ can occur but with different meanings. In a formalism for group elements, $\mathbf{v}$ fulfills the rôle of the unit vector $\mathbf{a}$ that defines the reflection A, such that we must have $|\mathbf{v}| = 1$, and then the reflection matrix $\mathbf{V} = \mathbf{A}$ transforms according to: $\mathbf{A} \rightarrow \mathbf{GA}$ under a group element g with matrix representation $\mathbf{G}$. The new group element represented by $\mathbf{GA}$ will then in general no longer be a reflection that can be associated with a unit vector like it was the case for $\mathbf{A}$. In a formalism of vectors, $|\mathbf{v}|$ can be different from 1 and the matrix $\mathbf{V}$ (that represents now a vector) transforms according to: $\mathbf{V} \rightarrow \mathbf{GVG}^{-1} = \mathbf{GVG}^{\dagger}$. Here $\mathbf{GVG}^{\dagger}$ can be associated again with a vector.

We cannot emphasize enough that the vector formalism is a *parallel* formalism that is different from the one for reflections because the reflections defined by $\mathbf{a}$ and $-\mathbf{a}$ are equivalent, while the vectors $\mathbf{a}$ and $-\mathbf{a}$ are not. We have here two concepts that are algebraically identical but not geometrically and this is the source of a lot of confusion. The folklore that one must rotate a wave function by 4π to obtain the same wave function again is part of that confusion. The reflection operator $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ is a thing that is entirely different from the unit vector $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$, even if their expressions are algebraically identical. By rotating a reflection plane over an angle π we obtain the same reflection, while it takes rotating over an angle 2π to obtain the same vector $\mathbf{a}$.

□ Remark 2. Both in the representation matrices $\mathbf{A} = [\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ for reflections A and $\mathbf{V} = [\,\mathbf{v}\cdot\boldsymbol{\sigma}\,]$ for vectors $\mathbf{v}$, the quantities σ_x, σ_y, σ_z are the three Pauli matrices. In the representation $(\mathbf{e}_j \leftrightarrow \sigma_j = [\,\mathbf{e}_j\cdot\boldsymbol{\sigma}\,])$ defined by Eq. 27, the Pauli matrices $\sigma_x, \sigma_y, \sigma_y$ are just the images, *i.e.* the coding of the three basis vectors $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$. As clearly indicated in the diagram of Eq. 5, $\boldsymbol{\sigma}$ is a shorthand for the triple $(\sigma_x, \sigma_y, \sigma_z)$. The use of the symbol $\hat{=}$ serves to draw the attention to the fact that the notation $[\,\mathbf{v}\cdot\boldsymbol{\sigma}\,]$ is a purely conventional shorthand for $v_x\sigma_x + v_y\sigma_y + v_z\sigma_z$ which codes the vector $\mathbf{v}$ within the formalism. It is thus analogous to writing $v_x\mathbf{e}_x + v_y\mathbf{e}_y + v_z\mathbf{e}_z$ pedantically as: $(v_x, v_y, v_z) \cdot (\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$. The danger of using the convenient shorthand $[\,\mathbf{v}\cdot\boldsymbol{\sigma}\,]$ is that it conjures up the image of a scalar product, while there is no scalar product whatsoever.

The fact that $[\,\mathbf{v}\cdot\boldsymbol{\sigma}\,]$ represents the vector $\mathbf{v}$, and that the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$ just represent the basis vectors $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$, was clearly stated by Cartan, but physicists have nevertheless *hineininterpretiert* the vector $-\frac{\hbar q}{2m_0c}\,[\,\mathbf{B}\cdot\boldsymbol{\sigma}\,]$ as a scalar product $\mathbf{B}\cdot\boldsymbol{\mu}$ in the theory of the anomalous g-factor for the electron. Here $\boldsymbol{\mu}$ would be the magnetic dipole of the electron and $-\mathbf{B}\cdot\boldsymbol{\mu}$ its potential energy with the magnetic field $\mathbf{B}$. In reality $\mathbf{B}\cdot\boldsymbol{\sigma}$ just expresses the magnetic-field pseudo-vector $\mathbf{B}$. The quantity $\frac{\hbar}{2}\boldsymbol{\sigma}$ can never represent the spin, because it is already defined in Euclidean geometry before we apply this geometry to the physics where we want to consider spin. This reveals that physicists do not only use spinors like vectors: They also use vectors like scalars. We have fully discussed and tidied up this problem in [11] where we have proposed a better interpretation of the Stern-Gerlach experiment.

□ Remark 3. A similar confusion arises in the definition of the helicity of the neutrino [9], pp.105-106, Eq. 5.30, [22]. It is defined as $\frac{\hbar}{2}[\,\mathbf{u}\cdot\boldsymbol{\sigma}\,]$, and claimed to be "the projection" of the "spin" $\frac{\hbar}{2}\boldsymbol{\sigma}$ on the unit vector $\mathbf{u} = \mathbf{p}/|\mathbf{p}|$. This is again a confusion between the shorthand notation $[\,\mathbf{u}\cdot\boldsymbol{\sigma}\,]$ for the representation of the vector $\mathbf{u}$ and a true scalar product. As just mentioned, in reality $[\,\mathbf{u}\cdot\boldsymbol{\sigma}\,]$ just represents the unit vector $\mathbf{u}$. The factor $\frac{\hbar}{2}$ has been added only due to the confusion and the belief that $\frac{\hbar}{2}\boldsymbol{\sigma}$ would then be the spin operator, while the true spin operator is $\frac{\hbar}{2}[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]$. There is absoluty no reference to spin whatsoever in the operator $[\,\mathbf{u}\cdot\boldsymbol{\sigma}\,]$. The definition leads to a confusing discussion about the difference between helicity and chirality in

textbooks. This example shows that physicists cannot deny that they have considered $[\,\mathbf{u}\cdot\boldsymbol{\sigma}\,]$ and $[\,\mathbf{B}\cdot\boldsymbol{\sigma}\,]$ as a true scalar products.

2.7 The quadratic relation between vectors and spinors

2.7.1 Isotropic vectors

We will illustrate the quadratic relationship between spinors and vectors further in what we can consider as the final step in the construction of the formalism. We can picture a rotation R by a rotated triad of three basis vectors $\mathbf{e}'_x = R(\mathbf{e}_x)$, $\mathbf{e}'_y = R(\mathbf{e}_y)$, and $\mathbf{e}'_z = R(\mathbf{e}_z)$. This is a 1-1-correspondence. Triads visualize rotations and *vice versa*. This is a second important idea, that can be carried over to the general case of SO(n): We can code group elements by identifying them with a rotated basis of $\mathbb{R}^n$, a so-called *Vielbein*. This is a German word meaning "many legs", and the idea is that each basis vector is a leg. The first unit vector of the *Vielbein* of $\mathbb{R}^n$ corresponds to $n-1$ independent real parameters due to the normalization condition. The second unit vector corresponds to $n-2$ independent real parameters due to the normalization and the orthogonality conditions. The third unit vector corresponds to $n-3$ independent real parameters, etc... This shows that the *Vielbein* or a rotation in $\mathbb{R}^n$ corresponds to $n(n-1)/2$ independent real parameters as we claimed previously in Remark 3 of Subsection 2.3.

In SU(2) we can code the basis triad within an isotropic vector $\mathbf{e}'_x + \imath\mathbf{e}'_y = (x, y, z) \in \mathbb{C}^3$. This is also a 1-1-correspondence. From $(x, y, z) \in \mathbb{C}^3$ we can get $\mathbf{e}'_x$ and $\mathbf{e}'_y$ back by taking real and imaginary parts, while $\mathbf{e}'_z = \mathbf{e}'_x \wedge \mathbf{e}'_y$. We can represent thus a rotation by an isotropic vector, a vector whose square is 0.

□ Remark 1. It is often stated in this respect that an isotropic vector has zero length and that it is orthogonal to itself. This is however based on the wrong notion that the extrapolation to $\mathbb{C}^3$ of the Euclidean norm, $|\cdot|_E$ defined by: $\forall(x, y, z) \in \mathbb{R}^3$, $|(x, y, z)|_E = \sqrt{x^2 + y^2 + z^2}$, would still be a correct norm function for $(x, y, z) \in \mathbb{C}^3$. The correct norm to be used for $(x, y, z) \in \mathbb{C}^3$ is the Hermitian norm $|\cdot|_H$ defined by: $\forall(x, y, z) \in \mathbb{C}^3 : |(x, y, z)|_H = \sqrt{xx^* + yy^* + zz^*}$.

□ Remark 2. Presented this way, this idea may look like a stroke of genius. But in reality, it is just the consequence of embedding $\mathbb{R}^{2n}$ within $\mathbb{C}^{2n}$. We can thus embed $\mathbb{R}^4$ within $\mathbb{C}^4$. Instead of the basis of the mutually orthogonal unit vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, $\mathbf{e}_4$ of $\mathbb{R}^4$ as a basis for $\mathbb{C}^4$ one can use a coordinate transformation and use the alternative orthogonal basis $\boldsymbol{\varepsilon}_1 = \mathbf{e}_1 + \imath\mathbf{e}_2$, $\boldsymbol{\varepsilon}_1^* = \mathbf{e}_1 - \imath\mathbf{e}_2$ and $\boldsymbol{\varepsilon}_2 = \mathbf{e}_3 + \imath\mathbf{e}_4$, $\boldsymbol{\varepsilon}_2^* = \mathbf{e}_3 - \imath\mathbf{e}_4$ for $\mathbb{C}^4$ (see paragraph 4.6.1 of [7]). This basis can also be normalized using the Hermitian norm. The subspace spanned by $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ suffices to define the complete *Vielbein* of $\mathbb{R}^4$ and is isomorphic to $\mathbb{C}^2$. The space $\mathbb{R}^3$ is a subspace of $\mathbb{R}^4$, and once we have defined it, it becomes this way possible to treat also $\mathbb{R}^3$ in terms of $\mathbb{C}^2$. This is the reason why will end up with a formalism SU(2). The use of isotropic vectors is thus just a consequence of introducing $\boldsymbol{\varepsilon}_1 = \mathbf{e}_1 + \imath\mathbf{e}_2$, but the idea becomes somewhat concealed by the fact that we work with $\mathbb{R}^3$ instead of $\mathbb{R}^4$, such that we do not have $\boldsymbol{\varepsilon}_2 = \mathbf{e}_3 + \imath\mathbf{e}_4$ to tip us off.

The reference triad is coded by the isotropic vector $(x_0, y_0, z_0) = \mathbf{e}_x + \imath\mathbf{e}_y = (1, \imath, 0)$,with representation matrix:

$$\mathbf{M}_0 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}. \tag{28}$$

Consider now the rotation matrix $\mathbf{R}$ from Eq. 3. Under the rotation $\mathbf{R}$, the isotropic vector (x_0, y_0, z_0) with matrix $\mathbf{M}_0$ will be transformed to the isotropic vector (x, y, z) with 2×2 representation matrix $\mathbf{M} = x\sigma_x + y\sigma_y + z\sigma_z$. This rotated isotropic vector (x, y, z) codes the rotated triad and thus also the rotation $\mathbf{R}$. The representation matrix $\mathbf{M} = \mathbf{R}\mathbf{M}_0\mathbf{R}^{-1}$ is given by:

$$\mathbf{M} = \begin{bmatrix} z & x - \imath y \\ x + \imath y & -z \end{bmatrix} = 2\begin{bmatrix} -\xi_0\xi_1 & \xi_0\xi_0 \\ -\xi_1\xi_1 & \xi_0\xi_1 \end{bmatrix}$$

$$= \sqrt{2}\begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix} \otimes [\,-\xi_1, \xi_0\,]\sqrt{2} = 2\,[\,\boldsymbol{\chi} \otimes \dot{\boldsymbol{\psi}}^\dagger\,]. \tag{29}$$

As for an isotropic vector we have $x^2 + y^2 + z^2 = 0$, it follows that $\det(\mathbf{M}) = 0$. This implies that the columns of the matrix $\mathbf{M}$ are proportional. Also the lines of $\mathbf{M}$ are proportional. This is the reason why we can write $\mathbf{M}$ as a tensor product as done in Eq. 29, introducing the column "spinor" $\boldsymbol{\chi}$ and the conjugated row "spinor" $\dot{\boldsymbol{\psi}}^\dagger$. We are putting here the words spinor between quotes, because for the moment it is not yet obvious that they correspond to the same concept as the one we introduced above. We will address this issue very soon. The notation $\dot{\boldsymbol{\psi}}^\dagger$ just serves to distinguish row spinors $\dot{\boldsymbol{\psi}}^\dagger$ from column spinors $\boldsymbol{\chi}$. We will explain below the reason for this rather complicated looking notation $\dot{\boldsymbol{\psi}}^\dagger$. The square roots $\sqrt{2}$ are introduced for normalization purposes. There is some possibility of

confusion with the terminology here. From the purely algebraic point of view of matrix algebra, we could call these spinor quantities column "vectors" and row "vectors", but from the geometrical point of view, spinors are not vectors, because they code rotations, and rotations do not build a vector space.

□ Remark 3. When we will try to generalize the formalism to SO(n), we will no longer be able to factorize the matrix of an isotropic vector as done here. For a matrix $\mathbf{M}$ of rank $\rho > 2$ we can no longer conclude from $\det \mathbf{M} = 0$ that there exist $\rho \times 1$ matrices $\boldsymbol{\chi}$ and $1 \times \rho$ matrices $\dot{\boldsymbol{\psi}}^\dagger$ such that $\mathbf{M} = \boldsymbol{\chi} \otimes \dot{\boldsymbol{\psi}}^\dagger$, because this would imply that all columns of $\mathbf{M}$ are proportional and all rows of $\mathbf{M}$ are proportional, while it suffices that only two columns and two rows of $\mathbf{M}$ are proportional.

For the moment we can see how for the specific case of SU(2), the gimmick $\mathbf{M} = \boldsymbol{\chi} \otimes \dot{\boldsymbol{\psi}}^\dagger$ permits us to "halve" the formalism. In fact, the isotropic vector that codes the rotation transforms under rotations *quadratically* according to $\mathbf{M} \rightarrow \mathbf{RMR}^{-1} = \mathbf{RMR}^\dagger = \mathbf{R}\,[\,\boldsymbol{\chi} \otimes \dot{\boldsymbol{\psi}}^\dagger\,]\,\mathbf{R}^\dagger$, with multiplications on both sides. We could obtain the same result by stipulating that we must transform $\boldsymbol{\chi} \rightarrow \mathbf{R}\boldsymbol{\chi}$ and $\dot{\boldsymbol{\psi}}^\dagger \rightarrow \dot{\boldsymbol{\psi}}^\dagger \mathbf{R}^\dagger$. Now a spinor $\boldsymbol{\phi}$ that contains the same information as a rotation matrix transforms *linearly* according to $\boldsymbol{\phi} \rightarrow \mathbf{R}\boldsymbol{\phi}$, with only left multiplications. On the other hand, an isotropic vector contains the same information as a rotation matrix, because it codes the triad.

Let us now show that the "spinor" formalism for the isotropic vector is algebraically identical to the spinor formalism for the rotations, such that $\boldsymbol{\chi}$ is indeed algebraically a spinor. The reference triad is coded by the isotropic vector $(x_0, y_0, z_0) = \mathbf{e}_x + \imath \mathbf{e}_y = (1, \imath, 0)$, leading to:

$$\mathbf{M}_0 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} = \sqrt{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes [\,0, 1\,]\sqrt{2} \quad \Rightarrow \quad \boldsymbol{\chi}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ \dot{\boldsymbol{\psi}}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{30}$$

This reference triad corresponds to the identity matrix. The corresponding spinor $\boldsymbol{\phi} = \hat{\mathrm{c}}_1(\mathbb{1})$, is indeed equal to $\boldsymbol{\chi}_0$, such that we have checked that the formalism based on multiplying $\boldsymbol{\chi}_0$ to the left according to $\boldsymbol{\chi}_0 \rightarrow \boldsymbol{\chi} = \mathbf{R}\boldsymbol{\chi}_0$ is just identical to the formalism based on multiplying $\boldsymbol{\phi}$ according to $\boldsymbol{\phi} \rightarrow \mathbf{R}\boldsymbol{\phi}$, such that $\boldsymbol{\chi} = \boldsymbol{\phi}$, while $\dot{\boldsymbol{\psi}}$ corresponds to the conjugated spinor.

To summarize, it is not possible in SU(2) to build a linear representation based on vectors because vectors are of rank two in terms of spinor quantities, but is possible to build a linear representation based on spinors by "halving" the formalism. We could also proceed by only right multiplications on $\dot{\boldsymbol{\psi}}^\dagger$ according to $\dot{\boldsymbol{\psi}}^\dagger \rightarrow \dot{\boldsymbol{\psi}}^\dagger \mathbf{R}^\dagger$, but that would be completely equivalent. The conjugated spinor $\dot{\boldsymbol{\psi}}$ transforms like $\boldsymbol{\chi}$, by left multiplication by $\mathbf{R}$, and gives rise to the second column of the matrix in Eq. 3. It contains the same information as $\boldsymbol{\chi}$. Using $\dot{\boldsymbol{\psi}}$ instead of $\dot{\boldsymbol{\psi}}^\dagger$ allows us then also to limit ourselves to calculations that contain only left multiplications. In other words, in the notation $\dot{\boldsymbol{\psi}}^\dagger$, the symbol † is supposed to flag that it is transformed by right multiplication by $\mathbf{R}^\dagger$, while the dot is used to distinguish quantities $\dot{\boldsymbol{\psi}}$ from quantities $\boldsymbol{\chi}$, showing that the quantities $\dot{\boldsymbol{\psi}}$ originally have entered the formalism under the form of row spinors $\dot{\boldsymbol{\psi}}^\dagger$. Whereas the formalism $\mathbf{M} \rightarrow \mathbf{RMR}^{-1}$ was not linear in the parameters of the rotation matrix $\mathbf{R}$, halving the formalism to $\boldsymbol{\phi} \rightarrow \mathbf{R}\boldsymbol{\phi}$ has rendered it linear.

Because a rotation depends only on three independent real parameters, we can normalize these spinors to 1 such that $\xi_0\xi_0^* + \xi_1\xi_1^* = 1$. In fact, the normalization is a consequence of the fact that the matrix in Eq. 3 belongs to SU(2). The spinor contains thus exactly three independent parameters that characterize a rotation (e.g. the three Euler angles, or a rotation axis defined by a unit vector $\mathbf{n}$ and a rotation angle φ). From these spinors and using the identity $\xi_0\xi_0^* + \xi_1\xi_1^* = 1$ we can calculate backwards to (x, y, z). The result is:

$$x = \xi_0^2 - \xi_1^2, \quad y = \imath(\xi_0^2 + \xi_1^2), \quad z = -2\xi_0\xi_1. \tag{31}$$

From this we can recover the basis vectors $\mathbf{e}'_x(x_1, y_1, z_1)$, $\mathbf{e}'_y(x_2, y_2, z_2)$:

$$\begin{array}{ll} x_1 = \frac{1}{2}(\xi_0^2 - \xi_1^2 + \xi_0^{*2} - \xi_1^{*2}), & y_1 = \frac{\imath}{2}(\xi_0^2 + \xi_1^2 - \xi_0^{*2} - \xi_1^{*2}), \\ z_1 = -(\xi_0\xi_1 + \xi_0^*\xi_1^*), & \\ & \\ x_2 = \frac{\imath}{2}(-\xi_0^2 + \xi_1^2 + \xi_0^{*2} - \xi_1^{*2}), & y_2 = \frac{1}{2}(\xi_0^2 + \xi_1^2 + \xi_0^{*2} + \xi_1^{*2}), \\ z_2 = (\xi_0\xi_1 - \xi_0^*\xi_1^*). & \end{array} \tag{32}$$

and from this finally $\mathbf{e}'_z = (x_3, y_3, z_3) = \mathbf{e}'_x \wedge \mathbf{e}'_y$:

$$x_3 = \xi_0\xi_1^* + \xi_0^*\xi_1, \quad y_3 = \imath(\xi_0\xi_1^* - \xi_0^*\xi_1), \quad z_3 = \xi_0\xi_0^* - \xi_1\xi_1^*. \tag{33}$$

We can also calculate ξ_0 and ξ_1 from x, y and z, and this leads to the expressions introduced by Cartan:

$$\begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix} = \begin{bmatrix} \pm\sqrt{\frac{x - \imath y}{2}} \\ \\ \pm\sqrt{\frac{-x - \imath y}{2}} \end{bmatrix}. \tag{34}$$

This shows how the reference triad of basis vectors is expressed within a spinor. Similar expressions can be derived to show e.g. how the three Euler angles are expressed within a spinor. The Rodrigues formula shows how the rotation axis $\mathbf{n}$ and the rotation angle φ are expressed within the spinor.

□ Remark 4. In many textbooks, spinors are introduced on the basis of this algebra for the isotropic vector, putting the emphasis on halving the formalism. It is this approach that leads to the idea that a spinor is the square root of a vector, based on the fact that the isotropic vector appears as a tensor product of two spinors in Eq. 29. This tensor product is not a pure square, because the spinors $\boldsymbol{\chi}$ (a rotation) and $\dot{\boldsymbol{\psi}}$ (a reversal) are not identical, such that calling the spinor the square root of the vector is only a loose informal description. The presence of the square roots in Eq. 34 can inspire here also the idea that a spinor is the "square root" of a vector. Finally, the Rodrigues equation Eq. 8 can also be expressed as $\mathbf{R}(\mathbf{n},\varphi) = \frac{1}{2}\, e^{-\imath\varphi/2}\,(\mathbb{1} + [\,\mathbf{n}\cdot\boldsymbol{\sigma}\,]) + \frac{1}{2}\, e^{+\imath\varphi/2}\,(\mathbb{1} - [\,\mathbf{n}\cdot\boldsymbol{\sigma}\,])$. Within this algebraic form the presence of $\varphi/2$ in the exponentials leads also to the idea of a "square root". But we can appreciate from our approach that in SU(2) the true meaning of a spinor is not that it is "a kind of isotropic vector" as stated by Cartan, but just a rotation. In generalizing this idea, we can change the definition of a spinor to make it just a group element rather than a column matrix. The isotropic vector is merely a secondary tool to express this idea through quite ingenious "slick algebra". The basic idea that a spinor is a rotation is much simpler and developing it requires much less ingenuity.

□ Remark 5. In reference [6], pp.63-66, we discuss also the way SU(2) is introduced in textbooks based on a stereographic projection. We show that this method is in reality conceptually flawed because it only considers the basis vector $\mathbf{e}'_z$, which cannot represent the complete information about a rotation. A rotation of $\mathbf{e}_z$ to $\mathbf{e}'_z$ does not define a unique rotation, as one can afterwards still rotate the basis triad freely around $\mathbf{e}'_z$ over a rotation angle φ.

□ Remark 6. Many a physicist will be used to the concept of infinitesimal generators used to define the Lie algebra. In this context, the infinitesimal generators pick up algebraic expressions that are algebraically identical to those for the reflection matrices. We must point out that this algebraic identity is a mere coincidence. The definitions of the Pauli matrices in terms of reflection matrices and in terms of infinitesimal generators are conceptually completely different. One should indeed already feel rather puzzled by the fact that due to the algebraic identity a reflection operator appears to be related to an infinitesimal rotation. The solution of this riddle becomes obvious by considering rotations or Lorentz transformations in $\mathbb{R}^4$. We have then four reflection operators, while there are six infinitesimal generators, such that the two concepts are now clearly seen not to be equivalent. The four reflection operators have four-dimensional vector symmetry and are true generators for the rotation group. The infinitesimal generators have six-dimensional tensor symmetry. They are a vector basis for the six-dimensional tangent space to the Lie group. This explains also why the infinitesimal generators for SU(3) cannot be found by following the strategy outlined in Subsection 2.4.

□ Remark 7. The set of all isotropic vectors of $\mathbb{C}^3$ is the isotropic cone $\mathscr{C}$. Biedenharn [23] evokes the relation between a spinor and an isotropic vector $(x,y,z) \in \mathscr{C}$. There is only one element $(x,y,z) = (0,0,0) \in \mathscr{C}$ that belongs to real space $\mathbb{R}^3$. Biedenharn concludes from this observation that spinors certainly cannot be objects that rotate in physical space. This is very obviously not true and the confusion is due to the notation (x,y,z) which suggests that the isotropic vector could be a set of *position coordinates*, while it is obvious from the development that the isotropic vector is meant to be a set of *rotation coordinates*.

2.7.2 Real unit vectors

Eq. 29 is the reason why one says that a spinor is a square root of a vector. We can see that this is only very approximately true as the two spinors $\boldsymbol{\chi}$ and $\dot{\boldsymbol{\psi}}$ are different. There is a relation between spinors and vectors that illustrates in a much more direct and less artificial way how vectors are "squares" of spinors. Consider a rotation R with matrix $\mathbf{R}$ that turns the reference triad. The vector $\mathbf{e}'_z = R(\mathbf{e}_z)$ of the rotated reference triad in Eq. 33 can be expressed as:

$$[\,\mathbf{e}'_z\cdot\boldsymbol{\sigma}\,] = 2\,\boldsymbol{\chi}\otimes\boldsymbol{\chi}^\dagger - \mathbb{1}. \tag{35}$$

In fact,

$$[\,\mathbf{e}_z\cdot\boldsymbol{\sigma}\,] + \mathbb{1} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} = \sqrt{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \end{bmatrix}\sqrt{2}. \tag{36}$$

Under the rotation $\mathbf{R}$ this transforms to:

$$\begin{aligned}[\,\mathbf{e}'_z\cdot\boldsymbol{\sigma}\,] + \mathbb{1} &= \mathbf{R}\,(\,[\,\mathbf{e}_z\cdot\boldsymbol{\sigma}\,] + \mathbb{1}\,)\mathbf{R}^{-1} \\ &= \sqrt{2}\begin{bmatrix} \xi_0 \\ \xi_1 \end{bmatrix} \otimes \begin{bmatrix} \xi_0^* & \xi_1^* \end{bmatrix}\sqrt{2}.\end{aligned} \tag{37}$$

Where we have used $\mathbf{R}^{-1} = \mathbf{R}^\dagger$ and $\mathbf{R}\,\mathbb{1}\mathbf{R}^{-1} = \mathbb{1}$ to obtain the desired result. With respect to this identity, introducing the isotropic vectors to argue that vectors are rank-2 quantities in terms of spinors is thus rather a step away from a truly illuminating conceptual understanding of the quadratic relationship. It makes everything more difficult and less clear. We can illustrate this relation between a vector and its spinor in SU(2). We represent the vector by its spherical coordinates (θ, ϕ) as follows:

$$[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,] = \begin{bmatrix} \cos\theta & \sin\theta e^{-\imath\phi} \\ \sin\theta e^{\imath\phi} & -\cos\theta \end{bmatrix}. \tag{38}$$

Note that we use ϕ and φ as different symbols in this article. The same applies for θ and ϑ. The rotation required to rotate $\mathbf{e}_z$ to $\mathbf{a}$ along a great circle has axis $\mathbf{n} = (\cos(\phi+\pi/2), \sin(\phi+\pi/2), 0)$ and angle θ. The angle of rotation is counterclockwise when we look at it from the point $(\cos(\phi+\pi/2), \sin(\phi+\pi/2), 0)$. The rotation is thus expressed by:

$$\mathbf{R} = \begin{bmatrix} \cos(\theta/2) & -\imath\sin(\theta/2)e^{-\imath(\phi+\pi/2)} \\ -\imath\sin(\theta/2)e^{\imath(\phi+\pi/2)} & \cos(\theta/2) \end{bmatrix}. \tag{39}$$

One can then check that $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,] = \mathbf{R}\,[\,\mathbf{e}_z\cdot\boldsymbol{\sigma}\,]\,\mathbf{R}^\dagger$, and that:

$$[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,] = 2\,\boldsymbol{\chi}\otimes\boldsymbol{\chi}^\dagger - \mathbb{1}, \quad \text{with:} \quad \boldsymbol{\chi} = \begin{bmatrix} \cos(\theta/2) \\ -\imath\sin(\theta/2)e^{\imath(\phi+\pi/2)} \end{bmatrix}. \tag{40}$$

The spinor $\boldsymbol{\chi}$ that we can associate with $\mathbf{a}$ is thus the rotation required to turn $\mathbf{e}_z$ to $\mathbf{a}$. We can also write $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ as:

$$[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,] = \boldsymbol{\chi}\otimes\boldsymbol{\chi}^\dagger - \dot{\boldsymbol{\psi}}\otimes\dot{\boldsymbol{\psi}}^\dagger \tag{41}$$

This is based on:

$$[\,\mathbf{e}_z\cdot\boldsymbol{\sigma}\,] = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \end{bmatrix}. \tag{42}$$

The various column spinors we obtain are the columns of the rotation matrix. The line spinors are their Hermitian conjugates. The conjugated spinors can be obtained by considering:

$$[\,\mathbf{e}_z\cdot\boldsymbol{\sigma}\,] - \mathbb{1} = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} = -\sqrt{2}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \end{bmatrix}\sqrt{2}. \tag{43}$$

Under the rotation $\mathbf{R}$ this transforms to:

$$\begin{aligned} [\,\mathbf{e}'_z\cdot\boldsymbol{\sigma}\,] - \mathbb{1} &= \mathbf{R}\,(\,[\,\mathbf{e}_z\cdot\boldsymbol{\sigma}\,] - \mathbb{1}\,)\mathbf{R}^{-1} \\ &= -\sqrt{2}\begin{bmatrix} -\xi_0^* \\ \xi_1^* \end{bmatrix} \otimes \begin{bmatrix} -\xi_0 & \xi_1 \end{bmatrix}\sqrt{2}, \end{aligned} \tag{44}$$

such that:

$$[\,\mathbf{e}'_z\cdot\boldsymbol{\sigma}\,] = \mathbb{1} - 2\,\dot{\boldsymbol{\psi}}\otimes\dot{\boldsymbol{\psi}}^\dagger. \tag{45}$$

The conjugated spinor is thus the alternative spinor obtained by taking the second column of the rotation matrix. We may note that representation matrices of all the basis vectors are linked by a similarity transformation to $[\,\mathbf{e}_z\cdot\boldsymbol{\sigma}\,]$ such that they all have eigenvalues 1 and -1.

2.8 Justifying the introduction of a Clifford algebra

The author has figured out the whole contents of the present paper from scratch because he found the textbook presentations impenetrable. The author has also not studied books on Clifford algebra in depth, such that some works may well provide the motivation we will try to give here, and which we were not able to spot in textbooks. Our criticism is based on the observation that very often mathematical objects that algebraically look identical are in reality entirely different geometrical objects. We have seen that we can introduce representations $[\,\mathbf{v}\cdot\boldsymbol{\sigma}\,]$ for vectors $\mathbf{v}\in\mathbb{R}^3$ into the formalism by extrapolating the meaning of the algebra of the representations $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ of reflection operators $A\in \mathrm{L}(\mathbb{R}^3,\mathbb{R}^3)$. We have seen how confusing $A\in \mathrm{L}(\mathbb{R}^3,\mathbb{R}^3)$ and $\mathbf{a}\in\mathbb{R}^3$ through the algebraic identity of their representation matrices $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ can trap us into a conceptual impasse of trying to give geometrical meaning

to mindless algebra. This is not the end of the story. Whereas it is meaningful in the group theory to consider the product $R = BA$ of two reflections B and A and the corresponding representation matrix $\mathbf{R} = [\,\mathbf{b}\cdot\boldsymbol{\sigma}\,][\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$, it is *a priori* not defined what the purely formal product of two vectors $\mathbf{v}_1$ and $\mathbf{v}_1$ defined by $[\,\mathbf{v}_2\cdot\boldsymbol{\sigma}\,][\,\mathbf{v}_1\cdot\boldsymbol{\sigma}\,]$ is supposed to mean. Here again, entirely different geometrical objects are represented by identical algebraic expressions. We have learned definitions for $\mathbf{v}_1\cdot\mathbf{v}_2$ and for $\mathbf{v}_1 \wedge \mathbf{v}_2$, but not for $[\,\mathbf{v}_2\cdot\boldsymbol{\sigma}\,][\,\mathbf{v}_1\cdot\boldsymbol{\sigma}\,]$. But inspection of the algebra reveals that:

$$[\,\mathbf{v}_2\cdot\boldsymbol{\sigma}\,][\,\mathbf{v}_1\cdot\boldsymbol{\sigma}\,] = (\mathbf{v}_1\cdot\mathbf{v}_2)\,\mathbb{1} + \imath\,[\,(\mathbf{v}_2 \wedge \mathbf{v}_1)\cdot\boldsymbol{\sigma}\,], \tag{46}$$

an algebraic identity we used in deriving Eq. 8. We recognize here the familiar quantities $\mathbf{v}_1\cdot\mathbf{v}_2$ and $\mathbf{v}_1 \wedge \mathbf{v}_2$. Whereas this kind of algebra is meaningful for reflection matrices, it is *a priori* not meaningful for vectors. It can be given a meaning *a posteriori* in terms of vectors, at the risk of introducing confusion by ignoring the fact that the vector formalism is a parallel formalism, as we clearly outlined from the outset. Based on this confusion one can obtain then a formalism whereby one sums quantities that are not of the same type, by writing expressions of the type:

$$\mathbf{v}_2 \vee \mathbf{v}_1 = \mathbf{v}_2\cdot\mathbf{v}_1 + \mathbf{v}_2 \wedge \mathbf{v}_1, \tag{47}$$

as a shorthand for Eq. 46. What Clifford algebra does is defining *mano militare* that such expressions are meaningful as an algebra on multi-vectors. In general such a definition is introduced out of the blue. By focussing on the purely algebraic part of the formalism, it is possible to confuse the vectors $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$ and the reflection matrices $[\,\mathbf{a}\cdot\boldsymbol{\sigma}\,]$. This has several inconveniences. First of all, it is puzzling for the reader to understand where this idea comes from, because the algebra adds quantities of different symmetries and dimensions. All at once one teaches him that from now on one can add kiwis and bananas, while one has told him before during his whole life that this is not feasible. Moreover, this is done tacitly, as though this would not be problem at all. Nothing is done to ease away the bewilderment of a critical reader. One only laconically teaches him how to get used to it without asking further questions. One just rolls out the algebra such that reader can learn to imitate it mindlessly. As this is rather easy the reader will quickly become acquainted with it such that the justified initial questions will be silenced. But it takes a algebraic shortcut to the full geometrical explanation by exploiting algebraic coincidences.

The second problem is that after the introduction of the definition of the Clifford algebra with its cuisine of adding kiwis and bananas, all the geometry of the rotations seems to follow effortless from this definition in an extremely elegant way. This gives the impression that everything is derived by magic from thin air, which really leaves one left wondering. In fact, the only vital ingredient needed to obtain this powerful and elegant formalism seems to be the impenetrable slight of hand of adding kiwis and bananas.

For sure, our presentation looks somewhat more cumbersome and less elegant than the approach where one takes off from the definition of the Clifford algebra in grand style. But that elegant grand style is only a short-cut to the detailed explanation, and is obtained by sweeping some more tedious parts under the carpet. The strong point of our approach is that it provides the detailed geometrical motivation for the complete Clifford algebra. An interesting feature that also exists in our approach is that we can consider all kinds of products:

$$[\,\mathbf{a}_1\cdot\boldsymbol{\sigma}\,][\,\mathbf{a}_2\cdot\boldsymbol{\sigma}\,]\cdots[\,\mathbf{a}_m\cdot\boldsymbol{\sigma}\,] = \mathbf{a}_1 \vee \mathbf{a}_2 \vee \cdots \vee \mathbf{a}_m. \tag{48}$$

The worked-out expressions correspond contain hyper-parallelopepids and other quantities of various dimensions (that can be symmetrical or anti-symmetrical). The symmetry is signaled by the presence or absence of a factor $\imath$. These quantities transform under a rotation $\mathbf{R}$ to:

$$\mathbf{R}[\,\mathbf{a}_1\cdot\boldsymbol{\sigma}\,]\mathbf{R}^{-1}\mathbf{R}\,[\,\mathbf{a}_2\cdot\boldsymbol{\sigma}\,]\,\mathbf{R}^{-1}\cdots\mathbf{R}\,[\,\mathbf{a}_m\cdot\boldsymbol{\sigma}\,]\,\mathbf{R}^{-1} = \mathbf{R}\,[\,\mathbf{a}_1\cdot\boldsymbol{\sigma}\,][\,\mathbf{a}_2\cdot\boldsymbol{\sigma}\,]\cdots[\,\mathbf{a}_m\cdot\boldsymbol{\sigma}\,]\,\mathbf{R}^{-1}. \tag{49}$$

That we can rotate all these quantities within a unique formalism is thus not an asset of Clifford algebra that would not exist in our exploratory approach. We see that by formalizing the algebra for the sake of elegance we can obtain a very abstract formulation whereby we loose completely sight of the clear geometrical ideas. Mathematicians would argue that this does not matter. But the problem is that now confusion reigns. And when the cat is away, the mice will play. The abstraction eases extrapolating the algebra in a meaningless way beyond the limits defined by its geometrical meaning, e.g. by introducing linear combinations of spinors. From that point on the framework may now contain some well hidden logical nonsense, as taking linear combinations of spinors is not a granted procedure. The structure that results from this transgression is the very elegant Hilbert space formalism of QM. This is now highly abstract and any obvious link with the original geometrical meaning has been completely flushed. This favours an attitude where calculating becomes much more important than thinking. As matter of fact, in QM the leitmotiv has become to "shut up and calculate". And after hiding away this way the whole geometrical meaning of the formalism, a physicist may enter the room and ask: I have a beautiful formalism that grinds out theoretical predictions which agree with the experimental data to unprecedented precision, but I just cannot figure out what it means.

2.9 Construction of a basis of reflection matrices for $\mathbb{R}^n$

We now want to indicate how one can generalize the methods described in this Section to SO(n), with $n > 3$, briefly outlining how the reflection matrices of $\mathbb{R}^n$ are defined by generalizing the approach explained in Section 2. The full details are give in [7]. It has to be pointed out that defining a rotation in $\mathbb{R}^n$ will require in general more than two reflections for $n > 3$ [24].

We start from the rotation group SO(3) of $\mathbb{R}^3$ and the 2×2 Pauli matrices. They satisfy $\sigma_j\sigma_k + \sigma_k\sigma_j = 2\delta_{jk}\,\mathbb{1}$. We will be able to proceed in steps whereby at each step we can add two basis vectors while we double the size of the representation matrices. In other words SO(4) and SO(5) will be represented by 4×4 matrices, SO(6) and SO(7) by 8×8 matrices, and so on. In general SO(n) will thus be represented be $2^\nu \times 2^\nu$ matrices. The whole procedure can be proved by Peano induction. The procedure echoes the procedure for Pauli matrices at the block level. The reason for increasing the size of the representation matrices from $2^\nu \times 2^\nu$ to $2^{\nu+1} \times 2^{\nu+1}$ is that there are no further representation matrices available for introducing new basis vectors with the set of $2^\nu \times 2^\nu$ matrices. If we note the $2^\nu \times 2^\nu$ matrices which present the $2\nu + 1$ basis vectors $\mathbf{e}_j$ of $\mathbb{R}^{2\nu+1}$ as γ_j, and the $2^{\nu+1} \times 2^{\nu+1}$ matrices which present the $2\nu + 3$ basis vectors $\mathbf{e}_k$ of $\mathbb{R}^{2\nu+3}$ as ζ_k the algorithm based on Peano induction is given by:

$$\zeta_j = \begin{bmatrix} \gamma_j & \\ & -\gamma_j \end{bmatrix}, \quad \zeta_{2\nu+2} = \begin{bmatrix} & \mathbb{1} \\ \mathbb{1} & \end{bmatrix}, \quad \zeta_{2\nu+3} = \begin{bmatrix} & -\imath\mathbb{1} \\ \imath\mathbb{1} & \end{bmatrix}. \tag{50}$$

Proving all this by Peano induction is straightforward. This is all typical of what we stated in Subsection 1.2, viz. that what we have explained should put the reader into a position wherein he can now effortlessly carry out this generalization himself.

3 Spinors in the homogeneous Lorentz group

As students we learn the theory of special relativity by studying boosts along the x-axis. Such collinear boosts form a group noted as SO(1,1) which is abelian. This approach does not prepare us for the additional difficulties which occur in the homogeneous Lorentz group SO(3,1), which is non-abelian and allows for boosts allowed in all directions of $\mathbb{R}^3$. The difficulty stems from the fact that the composition of two non-collinear boosts is no longer a simple boost but the product of a boost and a rotation. Consequently SO(3) is a subgroup of the homogeneous Lorentz group. It is then obvious that the group is non-abelian. A general element of the group depends on six independent real parameters, three for the boosts and three for the rotations. This number of six follows also from the rule $n(n-1)/2$ derived from the number of independent real parameters needed to specify the tetrad of the four basis vectors or the so-called *Vierbein*.

As space-time is four-dimensional we have $\nu = 2$ and we need 4×4 reflection matrices. The conditions to expresses that these reflection matrices must be idempotent are now different due to the fact that the metric is now given by $c^2t^2 - x^2 - y^2 - z^2$, such that we must now find matrices γ_μ that satisfy $\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2g_{\mu\nu}\mathbb{1}$. Here $g_{\mu\nu}$ are the elements of the metric tensor and $\mathbb{1}$ is the 4×4 unit matrix. Hence $g_{tt} = 1$, $g_{xx} = g_{yy} = g_{zz} = -1$ and all other elements are zero. We may note that the reflection matrix γ_x squares to $-\mathbb{1}$, while the reflection matrix γ_t squares to $\mathbb{1}$, but this is not a real problem because the representation is a double covering whereby both $\mathbb{1}$ and $-\mathbb{1}$ represent the identity element. Rather than following the algorithm given in Subsection 2.9 to determine the gamma matrices we will adopt here the so-called Cartan-Weyl representation:

$$\begin{aligned} &\gamma_x = \begin{bmatrix} & \sigma_x \\ -\sigma_x & \end{bmatrix}, \quad \gamma_y = \begin{bmatrix} & \sigma_y \\ -\sigma_y & \end{bmatrix}, \quad \gamma_z = \begin{bmatrix} & \sigma_z \\ -\sigma_z & \end{bmatrix}, \\ &\gamma_t = \begin{bmatrix} & \mathbb{1} \\ \mathbb{1} & \end{bmatrix}, \qquad \gamma_5 = \begin{bmatrix} \mathbb{1} & \\ & -\mathbb{1} \end{bmatrix}. \end{aligned} \tag{51}$$

Here $\mathbb{1}$ stands again for the 2×2 matrix. We will thus use the same symbol for different objects, but within a given context there will be no confusion. We have added γ_5 because it is also often used. We will use sans-serif characters to note 4×4 matrices. A unit four-vector (a_t, a_x, a_y, a_z) and a reflection A defined by the unit vector (a_t, a_x, a_y, a_z) are thus represented by the matrix:

$$\mathsf{A} = \begin{bmatrix} & a_t\mathbb{1} + \mathbf{a}\cdot\boldsymbol{\sigma} \\ a_t\mathbb{1} - \mathbf{a}\cdot\boldsymbol{\sigma} & \end{bmatrix}$$

$$= \begin{bmatrix} & & a_t + a_z & a_x - \imath a_y \\ & & a_x + \imath a_y & a_t - a_z \\ a_t - a_z & -a_x + \imath a_y & & \\ -a_x - \imath a_y & a_t + a_z & & \end{bmatrix}. \tag{52}$$

The proper homogeneous Lorentz transformations are obtained from an even number of reflections. The transformations obtained from an odd number of reflections will be called reversals or improper Lorentz transformations. In the Cartan-Weyl representation, proper Lorentz transformations are therefore block diagonal, while the reversals have a block structure along the secondary diagonal. A column vector of a representation matrix of a homogeneous Lorentz transformation will therefore contain only two complex entries, which is not sufficient to completely characterize the six real independent parameters needed to define the transformation completely. We see thus that if we consider a spinor as a column matrix then it does not specify a group element. But as we shall show below, to be able to derive the Dirac equation one must introduce a superposition state of a proper Lorentz transformation and a reversal, and the consequence of this will be that a column matrix contains again all the information about a group element.

The matrices $\mathbf{V} = v_t\mathbb{1} + \mathbf{v}\cdot\boldsymbol{\sigma}$ and $\mathbf{V}^\star = v_t\mathbb{1} - \mathbf{v}\cdot\boldsymbol{\sigma}$, where $(v_t, \mathbf{v})$ is a four-vector that is no longer of unit length are used in the so-called SL(2,$\mathbb{C}$) representations of the Lorentz group (Note that the symbol $\star$ used here is not the symbol for complex conjugation $*$). They are obtained one from another by the parity transformation $\mathbf{v}| - \mathbf{v}$. They are related by $\mathbf{V}^{\star\star} = \mathbf{V}$, $\mathbf{V}\mathbf{V}^\star = (\det\mathbf{V})\,\mathbb{1}$ and $\det(\mathbf{V}) = \det(\mathbf{V}^\star) = v_t^2 - \mathbf{v}^2$. They are thus each others inverses when $\det(\mathbf{V}) = 1$. The entries of $\mathbf{V}$ are the minors of those of $\mathbf{V}^\star$ and *vice versa*. Furthermore $\mathbf{V}^\dagger = \mathbf{V}$. The scalar product $v_t w_t - \mathbf{v}\cdot\mathbf{w}$ of two four-vectors $(v_t, \mathbf{v})$ and $(w_t, \mathbf{w})$ is given by $\frac{1}{2}(\mathbf{V}^\star\mathbf{W} + \mathbf{W}^\star\mathbf{V})$. For a product of two reflection matrices we have:

$$\mathsf{AB} = \begin{bmatrix} & \mathbf{A} \\ \mathbf{A}^\star & \end{bmatrix}\begin{bmatrix} & \mathbf{B} \\ \mathbf{B}^\star & \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{B}^\star & \\ & \mathbf{A}^\star\mathbf{B} \end{bmatrix},$$
$$\mathsf{BA} = \begin{bmatrix} & \mathbf{B} \\ \mathbf{B}^\star & \end{bmatrix}\begin{bmatrix} & \mathbf{A} \\ \mathbf{A}^\star & \end{bmatrix} = \begin{bmatrix} \mathbf{B}\mathbf{A}^\star & \\ & \mathbf{B}^\star\mathbf{A} \end{bmatrix}. \tag{53}$$

We have then $\mathsf{ABBA} = \det(\mathbf{AB})\,\mathbb{1}$. Let us now choose $\det(\mathbf{AB}) = 1$, and call $\mathbf{L} = \mathbf{AB}^\star$. We have then $\det(\mathbf{L}) = 1$ such that $\mathbf{L} \in$ SL(2,$\mathbb{C}$). Concomitantly $\mathbf{L}^{-1} = \mathbf{B}\mathbf{A}^\star$. Furthermore: $(\mathbf{A}^\star\mathbf{B})^\dagger = \mathbf{B}^\dagger\mathbf{A}^{\star\dagger} = \mathbf{B}\mathbf{A}^\star = \mathbf{L}^{-1}$ such that $(\mathbf{A}^\star\mathbf{B}) = \mathbf{L}^{\dagger-1}$ and finally, $(\mathbf{B}^\star\mathbf{A}) = \mathbf{L}^\dagger$, such that:

$$\mathsf{L} = \mathsf{AB} = \begin{bmatrix} \mathbf{L} & \\ & \mathbf{L}^{\dagger-1} \end{bmatrix}, \quad \mathsf{L}^{-1} = \mathsf{BA} = \begin{bmatrix} \mathbf{L}^{-1} & \\ & \mathbf{L}^\dagger \end{bmatrix}. \tag{54}$$

Let us now introduce the notation for $\mathbf{L} \in$ SL(2,$\mathbb{C}$):

$$\mathbf{L} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \det\mathbf{L} = ad - bc = 1. \tag{55}$$

We have then:

$$\mathbf{L} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad \mathbf{L}^\dagger = \begin{bmatrix} a^* & c^* \\ b^* & d^* \end{bmatrix},$$
$$\mathbf{L}^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \quad \mathbf{L}^\dagger = \begin{bmatrix} d^* & -c^* \\ -b^* & a^* \end{bmatrix}. \tag{56}$$

We can this way make all the calculations for the Lorentz group by restricting the use of the matrices $\mathbf{L}$ to SL(2,$\mathbb{C}$).

□ Remark 1. As explained below, all rotations and all boosts can be represented by SL(2,$\mathbb{C}$) matrices and thus also all their products. The representation SL(2,$\mathbb{C}$) contains thus all the orthochronous Lorentz transformations. The 2×2 matices $\mathbf{L}$, with $\det\mathbf{L} = -1$ are topologically disconnected from the matrices of SL(2,$\mathbb{C}$), which indicates that they constitute the antichronous Lorentz transformations. A negative determinant can only be obtained by combining reflections with unit vectors $\pm(a_t, \mathbf{a})$ and $\pm(0, \mathbf{b})$. The simplest example is the product of $(1, \mathbf{0})$ and $(0, \mathbf{b})$. The first reflection is the time reversal T while the second reflection does not compensate for it, such that the product is indeed antichronous. Boosts are obtained from two reflections with unit vectors of the type $\pm(a_t, \mathbf{a})$, rotations are obtained from two reflections with unit vectors of the type $\pm(0, \mathbf{a})$. A

composition of two non-collinear boosts cannot be obtained from two reflections, in conformity with the remark we made about SO(n) at the beginning of Subsection 2.9.

□ Remark 2. The SL(2,ℂ) representations have one drawback, viz. that the identity element, $\mathbf{e}_t$ and the time reversal operation T are all represented by $\mathbb{1}$. This ambiguity can be a source of errors. In the Cartan-Weyl representation the ambiguity is lifted because $\gamma_t \neq \mathbb{1}$. The matrix γ_t represents then T in the group algebra and $\mathbf{e}_t$ in the multivector algebra.

□ Remark 3. In SU(2), Eq. 29 shows how we can represent an isotropic vector (which has "zero length") as a "square" of a spinor. We can do exactly the same in SL(2,ℂ) for four vectors $\mathbf{v}_j$ of "zero length" with representation matrices $\mathbf{V}_j$. In SU(2) the spinor contained all the information about the rotated reference frame, and thus also all the information about the rotation. It was the first column of the rotation matrix. In SL(2,ℂ), the images $\mathbf{L}\mathbf{V}_j\mathbf{L}^\dagger$ of these four vectors of zero length represent now the full information about the transformed reference frame and thus about the Lorentz transformation $\mathbf{L}$ (in Eq. 55). The four vectors of zero length are:

$$\begin{aligned}
&\mathbf{v}_1 = \mathbf{e}_t + \mathbf{e}_z, \quad \mathbf{V}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{V}_1' = \begin{bmatrix} 2aa^* & 2ac^* \\ 2ca^* & 2cc^* \end{bmatrix} = \sqrt{2}\begin{bmatrix} a \\ c \end{bmatrix} \otimes \begin{bmatrix} a^* & c^* \end{bmatrix}\sqrt{2} \\
&\mathbf{v}_2 = \mathbf{e}_x + \imath\mathbf{e}_y, \quad \mathbf{V}_2 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{V}_2' = \begin{bmatrix} 2ab^* & 2ad^* \\ 2b^* & 2cd^* \end{bmatrix} = \sqrt{2}\begin{bmatrix} a \\ c \end{bmatrix} \otimes \begin{bmatrix} b^* & d^* \end{bmatrix}\sqrt{2} \\
&\mathbf{v}_3 = \mathbf{e}_x - \imath\mathbf{e}_y, \quad \mathbf{V}_3 = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}, \quad \mathbf{V}_3' = \begin{bmatrix} 2ba^* & 2bc^* \\ 2da^* & 2dc^* \end{bmatrix} = \sqrt{2}\begin{bmatrix} b \\ d \end{bmatrix} \otimes \begin{bmatrix} a^* & c^* \end{bmatrix}\sqrt{2} \\
&\mathbf{v}_4 = \mathbf{e}_t - \mathbf{e}_z, \quad \mathbf{V}_4 = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{V}_4' = \begin{bmatrix} 2bb^* & 2bd^* \\ 2db^* & 2dd^* \end{bmatrix} = \sqrt{2}\begin{bmatrix} b \\ d \end{bmatrix} \otimes \begin{bmatrix} b^* & d^* \end{bmatrix}\sqrt{2}
\end{aligned} \tag{57}$$

Note that the spinors that occur in $\mathbf{V}_1'$ or in $\mathbf{V}_4'$ are only determined by the values of $\mathbf{V}_1'$ or $\mathbf{V}_4'$ up to a phase factor such that $\mathbf{V}_1'$ and $\mathbf{V}_4'$ do not yield the full information about the reference tetrad, while the two spinors which occur in the theoretical expressions do contain the full information. These two spinors correspond to the columns of $\mathbf{L}$. These spinors can be considered as square roots of vectors in the sense given by Atiyah.

As vectors must be transformed by similarity transformations $\mathsf{V} \rightarrow \mathsf{L}\mathsf{V}\mathsf{L}^{-1}$, the vectors in the two SL(2,ℂ) representations will transform according to $\mathbf{V} \rightarrow \mathbf{L}\mathbf{V}\mathbf{L}^\dagger$ (which is well known, see e.g. [21], p.174, Eq. 9.39) and $\mathbf{V}^\star \rightarrow \mathbf{L}^{\dagger-1}\mathbf{V}^\star\mathbf{L}^{-1}$. As the SL(2,ℂ) matrices are subject to the condition that their determinant is equal to 1, they can contain thus exactly the six independent real parameters needed to specify a Lorentz transformation. We can use the transformation property $\mathbf{V} \rightarrow \mathbf{L}\mathbf{V}\mathbf{L}^\dagger$ to calculate backwards the SL(2,ℂ) transformation matrix of a boost $\mathbf{B}(\mathbf{v})$ with velocity $\mathbf{v} = v\mathbf{u}$:

$$\mathbf{B}(\mathbf{v}) = \sqrt{\frac{\gamma+1}{2}}\mathbb{1} - \sqrt{\frac{\gamma-1}{2}}[\,\mathbf{u}\cdot\boldsymbol{\sigma}\,]. \tag{58}$$

The rotation matrices are taken over from SU(2) which is embedded in SL(2,ℂ). For a boost $\mathbf{B}$ we have thus $\mathbf{B}^\dagger = \mathbf{B}$ while for a rotation $\mathbf{R}$ we have $\mathbf{R}^\dagger = \mathbf{R}^{-1}$. The calculations are really lengthy and tedious, but with the aid of the formalism of SL(2,ℂ) we can prove that the composition of two non-collinear boosts $\mathbf{B}(\mathbf{v}_2)\mathbf{B}(\mathbf{v}_1)$ is the product $\mathbf{R}(\mathbf{s},\alpha)\mathbf{B}(\mathbf{v})$ of a rotation and a boost, as we stated at the beginning of this section:

$$\mathbf{B}(\mathbf{v}_2)\mathbf{B}(\mathbf{v}_1) = \left[\,\cos(\alpha/2)\mathbb{1} - \imath\sin(\alpha/2)\,[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\right]\left[\sqrt{\frac{\gamma+1}{2}}\mathbb{1} - \sqrt{\frac{\gamma-1}{2}}[\,\mathbf{u}\cdot\boldsymbol{\sigma}\,]\right]. \tag{59}$$

Here $\mathbf{u}_1 = \mathbf{v}_1/v_1$, $\mathbf{u}_2 = \mathbf{v}_2/v_2$, $\mathbf{u} = \mathbf{v}/v$. We note the unit vector perpendicular to the plane defined by $\mathbf{u}_1$ and $\mathbf{u}_2$ and which defines the rotation axis as $\mathbf{s}$. The rotation angle is α. The angle between $\mathbf{u}_1$ and $\mathbf{u}_2$ is called θ, such that: $\mathbf{u}_1\cdot\mathbf{u}_2 = \cos\theta$ and $\mathbf{u}_1 \wedge \mathbf{u}_2 = \sin\theta\,\mathbf{s}$. The rotation and the boost are then defined by:

$$\gamma = \gamma_1\gamma_2(1 + \frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}) \tag{60}$$

$$\sin(\alpha/2) = \sin\theta\sqrt{\frac{(\gamma_1-1)(\gamma_2-1)}{2\left[\,1+\gamma_1\gamma_2\left(1+\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)\right]}} \tag{61}$$

$$\cos(\alpha/2) = \sqrt{\frac{(\gamma_2+1)(\gamma_1+1)}{2\left[\,1+\gamma_1\gamma_2\left(1+\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)\right]}} + \sqrt{\frac{(\gamma_2-1)(\gamma_1-1)}{2\left[\,1+\gamma_1\gamma_2\left(1+\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}\right)\right]}}\cos\theta \tag{62}$$

$$\mathbf{v} = \frac{\mathbf{v}_1+\mathbf{v}_2}{1+\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}} + \frac{1}{c^2}\frac{\gamma_1}{\gamma_1+1}\frac{\mathbf{v}_1\wedge(\mathbf{v}_1\wedge\mathbf{v}_2)}{1+\frac{\mathbf{v}_1\cdot\mathbf{v}_2}{c^2}}. \tag{63}$$

There exists also an equation $\mathbf{B}(\mathbf{v}_2)\mathbf{B}(\mathbf{v}_1) = \mathbf{B}(\mathbf{w})\mathbf{R}(\mathbf{s},\varphi)$ with a reverse order of boost and rotation. The ordeal of going through similar tedious calculations as for the first identity can now be avoided by taking the Hermitian conjugate of this first identity $\mathbf{B}(\mathbf{v}_2)\mathbf{B}(\mathbf{v}_1) = \mathbf{R}(\mathbf{s},\alpha)\mathbf{B}(\mathbf{v})$ which is: $\mathbf{B}(\mathbf{v}_1)\mathbf{B}(\mathbf{v}_2) = \mathbf{B}(\mathbf{v})\mathbf{R}^{-1}(\mathbf{s},\alpha)$ and then carrying out the substitution $(\mathbf{v}_1,\mathbf{v}_2)|(\mathbf{v}_2,\mathbf{v}_1)$. There is a whole monograph by Ungar dedicated to the calculation of compositions of Lorentz transformations within the homogeneous Lorentz group [25], i.e. the hyperbolic geometry of the Lorentz group. Ungar also introduces the concept of gyrovectors. The fact that non-collinear boosts lead to Lorentz transformations that are no longer pure boosts leads to Thomas precession when a spinning particle follows an orbit [26], [27].

Using the same reasoning within the Dirac representation as used in SU(2) it is easy to see that the rotation which corresponds to the SU(2) matrices $\pm\mathbf{R}(\mathbf{s},\varphi)$ is now represented by the two 4×4 matrices with the 2×2 block structure:

$$\mathsf{R} = \pm\begin{bmatrix} \mathbf{R}(\mathbf{s},\varphi) & \\ & \mathbf{R}(\mathbf{s},\varphi) \end{bmatrix}. \tag{64}$$

This follows also from Eq. 54 and $\mathbf{R}^\dagger = \mathbf{R}^{-1}$. The $\pm$ sign stil occurs here because $\det(\pm\mathbf{R}) = 1$. The matrices we use are still constituting a double-covering of the proper Lorentz group.

4 Spinor-based approach to quantum mechanics

4.1 The Dirac equation from scratch

The following derivation of the free-space Dirac equation from scratch has been discussed in the monograph [6], especially in pp. 153-168, with additions scattered over various papers (the Appendix of [28], pp.1-2 of [29]). For this reason we provide here the complete derivation in a presentation that in our opinion is also more clear.

We start from the assumption that an electron at rest spins with an angular frequency ω_0 around a fixed spin-axis defined by a unit vector $\mathbf{s} \in \mathbb{R}^3$. We have no theoretical justification for this assumption. We were just led by curiosity and introduced it *ex nihilo*. We have no other justification for it than that we are able to derive the Dirac equation from it (and some other assumptions like $\hbar\omega_0/2 = m_0c^2$ we will introduce below) by a rigorous mathematical proof.

□ Remark 1. Lorentz objected that the electron cannot spin because this assumption cannot explain the magnetic moment of the electron. In fact, if all the charge of the electron were put on its equator and made to travel at the speed of light by the spinning motion, then this would still not be sufficient to produce the magnetic moment. Moreover the present experimentally established upper bound for the radius of the electron is much smaller than the value for the electron radius Lorentz adopted. But this argument by Lorentz does not hold sway since, as we explained in [11], the magnetic moment is not due to a current loop. The algebraic expressions $-\frac{q}{2m_0}(\mathbf{B}\cdot\hat{\mathbf{L}})\,\mathbb{1}$ for the normal and $-\frac{\hbar q}{2m_0}[\,\mathbf{B}\cdot\boldsymbol{\sigma}\,]$ for the anomalous Zeeman effect have completely different symmetries in the Clifford algebra, because the dot product in the normal Zeeman effect is a true *scalar* product, while the dot product in the anomalous Zeeman effect is a shorthand, which expresses a *vector*. Here $\hat{\mathbf{L}}$ represents the three angular-momentum operators. The interaction responsible for the anomalous Zeeman effect appears in the algebra as a direct coupling between a point charge and the magnetic field, not as the interaction of a current loop with the magnetic field. For the electron the value $g = 2$ is almost exact, which is further support for the thesis that its magnetic moment is not produced by current loops. The exchange mechanism of Heisenberg and Majorana based on the Coulomb interaction and the exclusion principle shows that the magnetic moment of the electron can be explained without current loops.

□ Remark 2. The neutron has a g-factor of -3.826 085 45(90) despite the fact that it "has no charge". The reason for this is that it is made up of positive and negative quarks. The charges of these quarks cancel but the magnetic moments induced by the current loops of the quarks add up because the opposite charges are circulating in opposite senses. This observation alone would already have been sufficient to point out that even within a paradigm of current loops Lorentz's objection was not completely waterproof.

We can express the assumed spinning motion with the aid of the Rodrigues equation Eq. 8 by replacing $\varphi = \omega_0\tau$. Here τ is the proper time. The rest mass of the electron will be noted as m_0. This describes now the spinning motion of an object, e.g. a particle or a top. This is analogous to the way we use $\mathbf{r}(t)$ in Newtonian mechanics to describe the motion of an object along its orbit.

□ Remark 3. There is a tedious technicality involved with the definition of the spin axis and the unit vector $\mathbf{s}$ parallel to it. SU(2) treats rotations, while we are dealing here with a spinning motion. The axis of a spinning object is not exactly the same concept as the axis of a rotation. This is discussed in [6] pp. 129-148. The unit vector $\mathbf{s}$ that defines the spinning motion will coincide e.g. with the physical symmetry axis of a spinning top. We can e.g. imagine that we rotate ourselves with respect to the spinning top. The spin axis will then still be the symmetry axis. The spin axis transforms thus as a vector because all we can do as human beings by moving corresponds to vector transformations. The spin vector $\mathbf{v}$ is thus a vector, and will be transformed according to $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \rightarrow \mathbf{R}[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\mathbf{R}^{-1}$ in SU(2) or more generally $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \rightarrow \mathbf{L}[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\mathbf{L}^\dagger$ in SL(2,$\mathbb{C}$). On the other hand, one speaks about the axis of a rotation $\mathbf{R}_0$, which implies that it is associated with the group element and will therefore transform according to $\mathbf{R}_0 \rightarrow \mathbf{R}\mathbf{R}_0$. The unit vector $\mathbf{n}$ that we can draw parallel to the axis of a rotation does therefore not

transform as a vector. As human beings we cannot bring about a transformation $\mathbf{R}_0 \rightarrow \mathbf{R}\mathbf{R}_0$ by physical motion (compare with Subsection 2.6).

The time derivative of $\mathbf{R}(\mathbf{s}, \omega_0\tau)$ yields:

$$\frac{d\mathbf{R}}{d\tau} = -\imath(\omega_0/2)[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\,\mathbf{R}, \quad \text{and:} \quad \frac{d\boldsymbol{\chi}}{d\tau} = -\imath(\omega_0/2)[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\,\boldsymbol{\chi}, \tag{65}$$

where the 2×1 spinor $\boldsymbol{\chi}$ is the first column of $\mathbf{R}(\mathbf{s}, \omega_0\tau)$. In order to derive Eq. 65 from Eq. 8 it has been assumed that $\frac{d\mathbf{s}}{d\tau} = 0$. Considering that $\mathbf{s}$ can also vary leads to much more complicated equations with extra terms. Hence, we have introduced the underlying assumption that the orientation of the spin axis remains fixed. We must thus remember in the further derivation of the Dirac equation below that it is only valid for an electron with a fixed orientation of its spin axis. The case of a precessing spin axis is *a priori* not covered by this derivation. The Dirac equation can thus not be used to study precession, a limitation beyond guessing in Dirac's derivation of the equation.

Eq. 65 is defined for a single spinning top or electron at some unspecified position $\mathbf{r}_0 \in \mathbb{R}^3$. The function $\mathbf{R}$ is thus a function of the variable τ but not of the variable $\mathbf{r} = (x, y, z)$ for the position. It can thus not be differentiated with respect to x, y or z. The spinor $\boldsymbol{\chi}$ corresponding to $\mathbf{R}$ is thus not a spinor wave function $\boldsymbol{\chi} \in \mathrm{F}(\mathbb{R}^4, \mathbb{C}^2)$, but a function $\boldsymbol{\chi} \in \mathrm{F}(\mathbb{R}, \mathbb{C}^2)$. It is here that we must adopt Ballentine's statistical interpretation [30] in order to introduce a wave function $\boldsymbol{\chi} \in \mathrm{F}(\mathbb{R}^4, \mathbb{C}^2)$. The wave function will describe a statistical ensemble of imaginary non-interacting electrons who are all in the identical state $\mathbf{R}(\mathbf{s}, \omega_0\tau)$ at time τ with the same phase angle. As we want the probability that an electron is in a certain position $\mathbf{r}$ to be the same for all $\mathbf{r} \in \mathbb{R}^3$, we must put one imaginary electron at each position $\mathbf{r} \in \mathbb{R}^3$. This leads to the replacement of $\mathbf{R} \in \mathrm{F}(\mathbb{R},\mathrm{SU}(2))$: $\tau \rightarrow \mathbf{R}(\mathbf{s}, \omega_0\tau)$ in Eq. 65 by a function $\mathbf{R} \in \mathrm{F}(\mathbb{R}^4,\mathrm{SU}(2))$: $\tau \rightarrow \mathbf{R}(\mathbf{r}, \tau)$. This has been discussed in full detail in Appendix B of [11]. The equation remains the same, but the meaning of $\mathbf{R}$ has changed, because its spatial domain has been changed from one point in $\mathbb{R}^3$ to the whole of $\mathbb{R}^3$. It allows now an electron to be anywhere with equal probability.

Simultaneity of electron positions within the wave function does not imply simultaneity in the real world. It only reflects a *simultaneity of description* of the possible events (see also Subsection 4.3). The wave function is a mathematical tool that will be used to calculate probabilities in experiments at any moment we want in time. The velocity of the imaginary electrons is zero. The function $\mathbf{R}$ obeys thus the Heisenberg uncertainty relation because the velocity is exactly known while the position is completely unknown. But this has nothing to do with physics, it is a merely mathematical construction. It is based on the idea that it is rather convenient to describe the electron over whole Euclidean space $\mathbb{R}^3$ or whole Minkowski space-time $\mathbb{R}^4$. The function $\mathbf{R}$ describes now a statistical ensemble of spinning electrons in uniform motion (which is here rest). The modified Eq. 65 can now be lifted to the Cartan-Weyl representation using Eq. 64. We can then write the following differential equation $\forall(\mathbf{r}, \tau) \in \mathbb{R}^4$:

$$\begin{aligned}\begin{bmatrix} & \frac{d}{d\tau}\mathbb{1} \\ \frac{d}{d\tau}\mathbb{1} & \end{bmatrix}\begin{bmatrix} \mathbf{R} & \\ & \mathbf{R} \end{bmatrix} &= -\imath\frac{\omega_0}{2}\begin{bmatrix} & [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \\ [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] & \end{bmatrix}\begin{bmatrix} \mathbf{R} & \\ & \mathbf{R} \end{bmatrix} \\ &= -\imath\frac{\omega_0}{2}\begin{bmatrix} & [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \\ -[\,-\mathbf{s}\cdot\boldsymbol{\sigma}\,] & \end{bmatrix}\begin{bmatrix} \mathbf{R} & \\ & \mathbf{R} \end{bmatrix}.\end{aligned} \tag{66}$$

This expresses that the spin vector $\mathbf{s} = \mathbf{a} \wedge \mathbf{b} \in \mathbb{R}^3$ (defined in in Eq. 8) is an axial vector, which does not change sign under a parity transformation P, because P flips the signs of $\mathbf{a}$ and $\mathbf{b}$ simultaneously. If we had used a block-diagonal expression for $\mathbf{s}$, the identity could never have turned out correctly because the four-potential is a four-vector and the global block structures must match. This lifts the differential equation in the SU(2) representation to the Dirac representation. The electrons are for the moment at rest. We will now use covariance to put them all into the same uniform motion with non-zero velocity $v < c$. Under a general Lorentz transformation L we have:

$$\begin{aligned}&\mathsf{L}\begin{bmatrix} & \frac{d}{d\tau}\mathbb{1} \\ \frac{d}{d\tau}\mathbb{1} & \end{bmatrix}\mathsf{L}^{-1}\cdot\mathsf{L}\begin{bmatrix} \mathbf{R} & \\ & \mathbf{R} \end{bmatrix} \\ &= -\imath\frac{\omega_0}{2}\mathsf{L}\begin{bmatrix} & [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \\ -[\,-\mathbf{s}\cdot\boldsymbol{\sigma}\,] & \end{bmatrix}\mathsf{L}^{-1}\cdot\mathsf{L}\begin{bmatrix} \mathbf{R} & \\ & \mathbf{R} \end{bmatrix}.\end{aligned} \tag{67}$$

The signs that go with $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]$ follow from the fact that for an electron at rest we must obtain twice Eq. 65. This evidences that $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]$ is an axial vector. We define:

$$\mathsf{L}\begin{bmatrix} \mathbf{R} & \\ & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \phi & \\ & \phi^{\dagger-1} \end{bmatrix} = \Phi. \tag{68}$$

We have used here the general structure Φ derived in Eq. 54. The result of carrying out the Lorentz transformations L in Eq. 67 is: :

$$\begin{aligned}&\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1}-c[\,\nabla\cdot\boldsymbol{\sigma}\,] \\ \frac{\partial}{\partial t}\mathbb{1}+c[\,\nabla\cdot\boldsymbol{\sigma}\,] & \end{array}\right]\Phi \\ &=-\imath\frac{\omega_0}{2}\left[\begin{array}{cc} & s_t\mathbb{1}+[\,\mathbf{s}'\cdot\boldsymbol{\sigma}\,] \\ -[\,s_t\mathbb{1}-[\,\mathbf{s}'\cdot\boldsymbol{\sigma}\,]\,] & \end{array}\right]\Phi. \end{aligned} \tag{69}$$

This corresponds to the fact that the true general form of the four-gradiant is $(\frac{d}{dct},-\nabla)$ and the true general form of the spin vector $(s_t,\mathbf{s})$. We can in the rest frame replace $(\frac{d}{dc\tau},\mathbf{0})$ by the four-gradient because $(\frac{d}{dc\tau},-\nabla)$ yields in the rest frame the same result on the wave function as $(\frac{d}{dc\tau},\mathbf{0})$. Note that the operations $(\frac{d}{dct},-\nabla)$ on Φ are defined, while they are not defined for the function $\mathbf{R}$ in Eq. 65. Eq. 69 can also be written as:

$$[\,\gamma_t\frac{\partial}{\partial t}-c\nabla\cdot\boldsymbol{\gamma}\,]\,\Phi=-\imath\frac{\omega_0}{2}\,\gamma_5\,[\,s_t\gamma_t+\mathbf{s}\cdot\boldsymbol{\gamma}\,]\Phi, \tag{70}$$

where we have dropped the accent on $\mathbf{s}'$, in order to write the covariant equation in its standard form. We introduce the notation:

$$\mathsf{S}=\left[\begin{array}{cc} & s_t\mathbb{1}+[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \\ -[\,s_t\mathbb{1}-[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\,] & \end{array}\right]\quad\Rightarrow\quad\mathsf{S}^2=\mathbb{1}, \tag{71}$$

and define:

$$\Psi=(\mathbb{1}+\mathsf{S})\Phi, \tag{72}$$

such that:

$$\mathsf{S}\Psi=\mathsf{S}(\mathbb{1}+\mathsf{S})\Phi=(\mathsf{S}+\mathbb{1})\Phi=\Psi. \tag{73}$$

The reason why we introduce this mixed state Ψ is that the operator S corresponding to the spin axis in Eq. 69 is a reflection, while Φ is a Lorentz transformation. We have thus a reversal $\mathsf{S}\Phi$ on the right-hand side. We can never simplify this to Φ to obtain the Dirac equation, because a reversal can never become equal to a proper Lorentz transformation (actually multiplied by a constant m_0c^2). In the Cartan-Weyl representation the algebra makes this very obvious because the block structures of a proper Lorentz transformation and a reversal do not even match. We must therefore replace the pure state Φ by a mixed state Ψ. This mixed state corresponds conceptually to a set as explained in paragraph 2.5.2.

□ Remark 2. This can be further illustrated by the analogous problem in SU(2), which is that we can never simplify $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\mathbf{R}$ to $\mathbf{R}$ because $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\mathbf{R}$ is a reversal and $\mathbf{R}$ is a rotation. In general we will also not be able to simplify $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\boldsymbol{\chi}$ to $\boldsymbol{\chi}$. In the analogy $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]$ is the counterpart of S and $\boldsymbol{\chi}$ is the counterpart of Φ. To obtain a simplifying identity $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\boldsymbol{\psi}=\boldsymbol{\psi}$ leading to a Dirac-like equation we can consider the set $\mathscr{A}=\{\boldsymbol{\chi},[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\boldsymbol{\chi}\}$, which corresponds to the mixed state $\boldsymbol{\psi}=\boldsymbol{\chi}+[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\boldsymbol{\chi}$. For the set $\mathscr{A}$ we have then $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\mathscr{A}=\mathscr{A}$ and for the corresponding mixed state $[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\boldsymbol{\psi}=\boldsymbol{\psi}$. In the analogy the mixed state $\boldsymbol{\psi}$ is the counterpart of the mixed state Ψ.

This fact that the wave function Ψ is now a mixed state stresses once more that the wave function describes a statistical ensemble. The covariance of Eq. 73 follows from:

$$\mathsf{L}\mathsf{S}\mathsf{L}^{-1}\cdot\mathsf{L}(\mathbb{1}+\mathsf{S})\mathsf{L}^{-1}\cdot\mathsf{L}\Phi=\mathsf{L}(\mathsf{S}+\mathbb{1})\mathsf{L}^{-1}\cdot\mathsf{L}\Phi\quad\Rightarrow\quad\mathsf{S}'(\mathbb{1}+\mathsf{S}')\Phi'=(\mathsf{S}'+\mathbb{1})\Phi'. \tag{74}$$

As we have assumed that $\mathbf{s}$ does not vary with time:

$$\left[\begin{array}{cc} & \frac{d}{d\tau}\mathbb{1} \\ \frac{d}{d\tau}\mathbb{1} & \end{array}\right]\left[\begin{array}{cc} & [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \\ -[-[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\,] & \end{array}\right]=\mathbf{0}. \tag{75}$$

By covariance we have then:

$$\mathsf{L}\left[\begin{array}{cc} & \frac{d}{d\tau}\mathbb{1} \\ \frac{d}{d\tau}\mathbb{1} & \end{array}\right]\mathsf{L}^{-1}\cdot\mathsf{L}\left[\begin{array}{cc} & [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \\ -[-[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\,] & \end{array}\right]\mathsf{L}^{-1}=\mathsf{L}\,\mathbf{0}\,\mathsf{L}^{-1}=\mathbf{0}, \tag{76}$$

such that:

$$\begin{aligned}&\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1}-c[\,\nabla\cdot\boldsymbol{\sigma}\,] \\ \frac{\partial}{\partial t}\mathbb{1}+c[\,\nabla\cdot\boldsymbol{\sigma}\,] & \end{array}\right]\mathsf{S} \\ =\;&\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1}-c[\,\nabla\cdot\boldsymbol{\sigma}\,] \\ \frac{\partial}{\partial t}\mathbb{1}+c[\,\nabla\cdot\boldsymbol{\sigma}\,] & \end{array}\right]\left[\begin{array}{cc} & s_t\mathbb{1}+[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,] \\ -[\,s_t\mathbb{1}-[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\,] & \end{array}\right] \\ =\;&\mathbf{0}. \end{aligned} \tag{77}$$

Hence:

$$\begin{aligned}&\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1} - c[\ \nabla\cdot\boldsymbol{\sigma}\ ] \\ \frac{\partial}{\partial t}\mathbb{1} + c[\ \nabla\cdot\boldsymbol{\sigma}\ ] & \end{array}\right](\mathbb{1}+\mathsf{S})\Phi \\ &= (\mathbb{1}+\mathsf{S})\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1} - c[\ \nabla\cdot\boldsymbol{\sigma}\ ] \\ \frac{\partial}{\partial t}\mathbb{1} + c[\ \nabla\cdot\boldsymbol{\sigma}\ ] & \end{array}\right]\Phi. \end{aligned} \tag{78}$$

Using Eq. 69 this leads to:

$$\begin{aligned}&\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1} - c[\ \nabla\cdot\boldsymbol{\sigma}\ ] \\ \frac{\partial}{\partial t}\mathbb{1} + c[\ \nabla\cdot\boldsymbol{\sigma}\ ] & \end{array}\right](\mathbb{1}+\mathsf{S})\Phi \\ &= -\imath\frac{\omega_0}{2}\,(\mathbb{1}+\mathsf{S})\mathsf{S}\,\Phi = -\imath\frac{\omega_0}{2}\,\mathsf{S}\,(\mathbb{1}+\mathsf{S})\Phi. \end{aligned} \tag{79}$$

With the aid of Eq. 73 this implies:

$$\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1} - c[\ \nabla\cdot\boldsymbol{\sigma}\ ] \\ \frac{\partial}{\partial t}\mathbb{1} + c[\ \nabla\cdot\boldsymbol{\sigma}\ ] & \end{array}\right]\Psi = -\imath\frac{\omega_0}{2}\,\mathsf{S}\,\Psi = -\imath\frac{\omega_0}{2}\,\Psi. \tag{80}$$

In summary:

$$\left[\begin{array}{cc} & \frac{\partial}{\partial t}\mathbb{1} - c[\ \nabla\cdot\boldsymbol{\sigma}\ ] \\ \frac{\partial}{\partial t}\mathbb{1} + c[\ \nabla\cdot\boldsymbol{\sigma}\ ] & \end{array}\right]\Psi = -\imath\frac{\omega_0}{2}\,\Psi. \tag{81}$$

which after substituting $\hbar\omega_0/2 = m_0c^2$, yields the celebrated Dirac equation:

$$\left[\begin{array}{cc} & -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\mathbb{1} + \frac{c\hbar}{\imath}[\ \nabla\cdot\boldsymbol{\sigma}\ ] \\ -\frac{\hbar}{\imath}\frac{\partial}{\partial t}\mathbb{1} - \frac{c\hbar}{\imath}[\ \nabla\cdot\boldsymbol{\sigma}\ ] & \end{array}\right]\Psi = m_0c^2\,\Psi. \tag{82}$$

We have already discussed this substitution in [11]. It implies that the whole rest energy of the electron corresponds to the kinetic energy of its spinning motion. We introduce this assumption for the sole reason that it permits to recover the Dirac equation. It is an equation that marries quantum mechnics ($E = h\nu$) with special relativity ($E = m_0c^2$) in a very simple identity. This is thus really a completely rigorous derivation of the Dirac equation from scratch. Due to he definition of Φ in Eq. 68, Ψ is here defined as:

$$\Psi = (\mathbb{1}+\mathsf{S})\Phi = \left[\begin{array}{cc} \phi & (\mathbb{1}s_t + [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,])\phi^{\dagger-1} \\ -(\mathbb{1}s_t - [\,\mathbf{s}\cdot\boldsymbol{\sigma}\,])\phi & \phi^{\dagger-1} \end{array}\right]. \tag{83}$$

The 4×2 matrix corresponding the first block column of Ψ corresponds to Eq. 5.52 on p.163 of [6]. In order to highlight the transformation properties of this 4×2 matrix under Lorentz transformations it can be further elaborated to yield Eq. 5.58 on p. 166 of [6]. The columns of Ψ are called bispinors because they combine two columns of SL(2,$\mathbb{C}$) matrices.

4.2 Consequences

From the Dirac equation we can derive the Schrödinger and the Pauli equation, such that the new approach offers a broad platform to deal with QM. The Dirac equation has here been derived with the rigor of a mathematical proof while the traditional equations were obtained by educated guessing from the de Broglie ansatz which has also been conjectured. The advantage of our approach resides in the fact that we now know exactly on what kind of assumptions the derivation of the Dirac equation has been based. This is not the case in the traditional approach and as the wave mechanics are able to describe many stunning experimental results which seem to defy any attempt of common-sense explanation, one can become convinced that the Dirac equation must be based on some very magical unknown quantum axioms. The problem is that this opens the door to an infinite set of fazing assumptions. There have been many attempts at interpreting QM, e.g. the many-worlds interpretation [31], Cramer's transactional interpretation [32] and Bohm's approach [33] just to mention some of them. And indeed, many of the interpretations carry some exotic perfume of quantum magic floating around them, with assumptions that run contrary to daily-life experience, like the existence of many parallel worlds or signalling backwards in time. This is something we wanted to avoid at all price in our approach, because such assumptions cannot be proved or contradicted by direct telltale experiments. The result is that we are left wondering what we should think about them and it remains anybody's guess which of these assumptions could be physically the more acceptable.

It is therefore quite sobering to learn that the derivation of the Dirac equation presented here is entirely classical. This raises the question where the quantum magic then comes from. This requires of course a very detailed discussion

(see below). It could be that the secret is somewhere hidden in the way we use this equation. It is for this reason important to be sure that we absolutely understand all the mathematics, which is why it is convenient that we have built up everything from scratch. It should permit to know if there are problems in the way we use the equation by a very thorough investigation on a case-by-case basis. We have therefore tried to analyze a number of quantum paradoxes in order to figure out if their explanation might require some magic after all (see [11] and below in paragraph 4.5.2). In any case the entirely classical derivation prompts for caution. We might be overinterpreting some things and we should try to avoid introducing axioms that are too remote from the possibility of experimental testing.

As we have pointed out Dirac's historical derivation has been carried out within the algebra of the vectors and multivectors (i.e. the exterior algebra), such that its real meaning remained hidden. We have derived it here within the algebra of the group elements, which permits us to see that the Dirac equation describes a statistical ensemble of spinning electrons in uniform motion. In the traditional derivation based on the multivector algebra we cannot even imagine that it is missing this crucial point and that the real stage for the scene is the algebra of group elements. Consequently, the possibility that the electron could spin is firmly denied by the standard dogma. One of the reasons for this might have been Lorentz's objection. Another reason could be Biedenharn's argument we mentioned in Remark 7 of paragraph 2.7.1. Our approach is thus at variance with the standard dogma but it cannot be attacked on the basis of this fact, because it is an alternative approach to QM that leads to the same algebraic results as the traditional approach and therefore is in a completely equivalent agreement with experimental results. The surplus values of our approach are the mathematical rigor and its conceptual clarity.

The operator identies $\hat{E} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}$ and $\hat{\mathbf{p}} = \frac{\hbar}{\imath}\nabla$ play a crucial rôle in the traditional derivations of the Schrödinger and Dirac equations. They have been obtained by guessing, as mentioned above. The identities $\hat{E} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}$ and $\hat{\mathbf{p}} = \frac{\hbar}{\imath}\nabla$ do not play any rôle at all in our derivation of the Dirac equation. They are just a corollary of the proof. The lack of rigor and clarity in the procedure to define quantum operators, even within the context of the scalar Schrödinger equation, has been pointed out by Messiah [34] who evoked that the correct forms are obtained in a process of trial and error. The extrapolation of the definition of these identities from the context of a scalar wave function for the Schrödinger equation to the context of spinor wave functions in the Pauli and Dirac equations is also far from self-evident, as e.g. a spinor $\boldsymbol{\chi}$ of SU(2) can contain two different angular frequencies $\pm\omega_0$. When we apply the operator $\hat{E} = -\frac{\hbar}{\imath}\frac{\partial}{\partial t}$ to such a spinor $\boldsymbol{\chi}$ we do no longer obtain the neat result $E\boldsymbol{\chi}$ but $E[\,\mathbf{s}\cdot\boldsymbol{\sigma}\,]\,\boldsymbol{\chi}$. To overcome this problem we have been forced to introduce mixed states $\boldsymbol{\psi}$. This has been further discussed in Remark 2 of paragraph 4.1 and in [11]. In the treatment of precession in a magnetic field this problem becomes worse because the angular frequencies can now take two different absolute values as discussed in the treatment of the Stern-Gerlach experiment in [11].

In our approach, negative frequencies are just related to inverting the sense of the spinning motion, because we have not introduced anti-particles at any stage in the derivation. There are no anti-particles in the geometry we have used. We have further discussed the interpretation of negative frequencies in the traditional approach on p.7 of [11]. The negative frequencies correspond to positive energies $E = h|\omega_0/2|$.

As already mentioned, we are unwittingly responsible ourselves for the Heisenberg uncertainty relation between position and momentum because it is introduced by the mathematical construction of the wave function.

Within the scope of the present derivation of the Dirac equation it is not possible to consider fermions with zero rest mass propagating at the speed of light c, due to the fact that the derivation starts from considering a spinning particle in its rest frame, while a particle propagating at the speed of light does not have a rest frame. As Dirac guessed his equation he could not suspect this limitation of the domain of validity of his equation. This led him to the assumption that neutrinos were fermions, that could be described by the Dirac equation whereby the rest mass is put to zero. This conception of the neutrino as a fermion with zero rest mass and travelling at the speed of light c has prevailed for a long time. It is finally the results about the neutrino oscillations from Super Kamiokande [35] which have invalidated Dirac's neutrino theory, while they are in agreement with our derivation of the Dirac equation.

Our derivation of the Dirac equation has been made entirely within the framework of special relativity such that it does not contain any incompatibility between special relativity and QM. However, we constructed the wave function starting from a Lorentz frame at rest in flat Minkowski space-time and then used covariance to introduce Lorentz frames in uniform motion with respect to this original frame. This is how the wave function corresponds to the description of a statistical ensemble of spinning electrons in uniform motion. This is a *global* approach and it is not at all obvious how one can generalize this procedure to curved space-time manifolds with *local* frames of different velocities.

It must be stressed that all the electrons within the statistical ensemble described by the wave function are moving at a well-defined uniform speed $v < c$. The phase of the wave function $(Et - \mathbf{p}\cdot\mathbf{r})/\hbar$ is a scalar invariant whose value in the rest frame is $m_0c^2\tau/\hbar$. The two expressions are just related by a boost. The value $m_0c^2\tau/\hbar$ is obtained from $\omega_0\tau/2$ by applying the identity $\hbar\omega_0/2 = m_0c^2$. The boost with velocity v transforms τ into $\tau = \gamma(t - vx/c^2)$ (if we take $\mathbf{v}$ parallel to the x-axis). If we wanted to interpret this expression for the proper time in terms of a wave moving in space we could rewrite it as $-(\gamma v/c^2)(x - c^2t/v)$ and conclude that the phase velocity of the wave is $c^2/v > c$. This is actually indeed correct. The phase velocity in a frame at rest is infinite and just corresponds to the velocity of the signal that would be needed to synchronize all the clocks in the reference frame up to infinite distance. All clocks will then have the same phase. In a frame moving with the velocity v this infinite velocity becomes c^2/v. In other words,

c^2/v is nothing more than the slope of the time axis in a Minkowski space-time diagram. The superluminal phase velocities are just due to the introduction of a Lorentz frame wherein all clocks have been synchronized up to infinite distance. The clocks we use in the wave function are the spinning motions of the electrons and we have synchronized all their phases by the construction of the wave function. As the electrons are all traveling at the speed $v < c$, there is no need for introducing wave packets to build an electron with a speed given by the group velocity $v_g = \frac{d\omega}{dk} < c$, because all electrons are already traveling at the speed $v < c$. In other words, no wave packets are needed, and the electron can remain a point particle.

The wave function is not a matter wave in the way it has been originally conjectured. The introduction of the wave packets leads to other problems because the speeds of the waves which contribute to the wave packet are different such that the wave packet spreads out with time and the "particle aspect" of the wave packet is lost. This prompts speculations about a collapse of the wave function as the result of a measurement. It further prompts theoretical considerations about solitons. It is a whole concatenation of problems resulting from the overinterpretation in terms of matter waves of a purely mathematical wave function. Of course, we have been duped by the fact that historically it went in reverse order. The matter waves were conjectured first and it is only later that the wave functions became the spinors and probability amplitudes of QM. We must finally note that the wave function we have constructed is completely coherent. We propose some thoughts about this choice in Subsection 4.3.

There is no collapse of the wave function. The wave function is not a matter wave and the electron is not a wave packet, a real physical entity that could physically collapse, but a purely mathematical tool. The purpose of the wave function is not to describe single events but probabilities of events for a large statisical ensemble. We can state this with confidence in the new approach because it has just been constructed that way. For a dice you could define a probability function $f \in \mathrm{F}(\mathscr{V}, \mathbb{R})$, where $\mathscr{V} = \{1, 2, 3, 4, 5, 6\}$ and $\forall x \in \mathscr{V} : f(x) = \frac{1}{6}$, or somewhat artificially a wave function $\psi \in \mathrm{F}(\mathscr{V}, \mathbb{R})$, such that $\forall x \in \mathscr{V} : \psi(x) = \frac{1}{\sqrt{6}}$, $f(x) = |\psi(x)|^2$. That function f and the wave function ψ do not collapse when somebody throws a dice and obtains 5. The only thing that has collapsed is that person's prior lack of knowledge about the outcome for the dice.

4.3 Why we can use coherent-source boundary conditions for an incoherent source

What one does in solving a wave equation and imposing the boundary conditions is assuming that the source is coherent. We could e.g. consider plane waves impinging on a set-up, and then everywhere on the plane of the source, we would attribute the same phase to the particle by the boundary conditions. This is not self-evident because the source may be incoherent. This shows that in this kind of problem the phase itself is not important and that it is the phase difference built up starting from the source which counts. That is then the reason why we can act as though the source were coherent. It is this phase relation imposed at the source which applies then to all particles. In fact, we can put all initial phases equal to zero and keep in mind the error this induces. Then we can correct for this error again at the end. (If this were to be incorrect because e.g. an electron has undergone a spin flip during its trajectory, then the coherence of the wave solution would have been lost anyway). As what we finally measure is $|\boldsymbol{\psi}|^2 = \boldsymbol{\psi}^\dagger\boldsymbol{\psi}$ or $|\psi|^2 = \psi^*\psi$ for each particle, this will not change the experimental results. What counts in the end is the amplitude rather than the phase of the wave and that the interactions that occur after leaving the source are not incoherent such that they do not provoke decoherence [28].

□ Remark 1. This may sound absurd, because when you have two particles with dynamics described by ψ_1 and ψ_2 (e.g. within the context of a Schrödinger equation), then correcting their phases by $\psi_1 \to \psi_1 e^{\imath\alpha_1}$ and $\psi_2 \to \psi_2 e^{\imath\alpha_2}$, might destroy the interference in $\psi_1 + \psi_2$: $|\psi_1 + \psi_2|^2 \neq |\psi_1 e^{\imath\alpha_1} + \psi_2 e^{\imath\alpha_2}|^2$. However, the idea that you should add up two electron spinors $\psi_1 + \psi_2$ and then square the result, is wrong and never applies.

Each particle must be counted in its own right, because in the build-up of an interference pattern the particles appear as dots on a detector screen one by one [36]. Hence we must always determine the quantities of single particles according to $|\psi_1|^2 + |\psi_2|^2$, never according to $|\psi_1 + \psi_2|^2$. When there is an interference pattern the weights of the functions ψ_1 and ψ_2 of two individual particles obey and comply already with the number of particles implied by the intensity of the interference term $|\psi_L + \psi_R|^2$ that one would write down according to the textbook rule, where R and L could e.g. refer to the left and the right slit in the double-slit experiment, see Eq. 4 of Ref. [13]. E.g. if there were destructive interference $\psi_L(\mathbf{r}) + \psi_R(\mathbf{r}) = 0$ there would be just no particles j with spinor ψ_j present in $\mathbf{r}$, such that the weight coefficients $c_L(\mathbf{r})$ and $c_R(\mathbf{r})$ in the set described by $c_L\psi_L + c_R\psi_R$ must satisfy $c_L(\mathbf{r}) = c_R(\mathbf{r}) = 0$ rather than $c_L(\mathbf{r}) = -c_R(\mathbf{r}) \neq 0$. This requires a very detailed discussion (see [13] and paragraph 4.5.1).

If the particles are electrons and the wave functions are spinors of SU(2), the idea remains the same but the argument must be written differently, because the phases are rather changed by some rotations $\boldsymbol{\psi}_1 \to \mathbf{R}_1\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2 \to \mathbf{R}_2\boldsymbol{\psi}_2$. The rest of the argument remains *mutatis mutandis* the same because for the SU(2) matrices $\mathbf{R}_j$ we have $\mathbf{R}_j^\dagger\mathbf{R}_j = \mathbb{1}$. We can even consider the case that the spins of the electrons emitted by the source are not aligned, but this is more elaborate. In fact, we must then reconstruct the SU(2) matrices $\mathbf{D}(g)$ from their corresponding spinors $\boldsymbol{\psi}(g)$,

develop the proof on the SU(2) matrices, and then switch back to the corresponding spinors in order to be able to derive the final result. The reason for this is that we can write a similarity transformation for the SU(2) matrices, but not for the spinors, because we cannot multiply a 2×1 spinor to the right with a 2×2 SU(2) matrix. The procedure to perform a similarity transformation on a spinor is visualized in the following equation and commuting diagram:

$$\boldsymbol{\psi}(g) = \begin{bmatrix} u \\ v \end{bmatrix} \leftrightarrow \mathbf{D}(g) = \begin{bmatrix} u & -v^* \\ v & u^* \end{bmatrix}$$

$$\begin{array}{ccc} \boldsymbol{\psi}(g) & & \boldsymbol{\psi}(h) \\ \downarrow & & \uparrow \\ \mathbf{D}(g) & \xrightarrow[\mathbf{D}(g)\rightarrow\mathbf{D}(h)=\mathbf{R}\,[\,\mathbf{D}(g)\,]\,\mathbf{R}^{-1}]{\text{simililarity transformation}} & \mathbf{D}(h) \end{array} \tag{84}$$

Within the context of the Dirac equation the analogous correspondence between the 4×1 spinors and the 2×2 representation matrices of SL(2,$\mathbb{C}$) is given by Eqs. (4.9) and (5.58) in [6]. The relation between the 2×2 representation matrices of SL(2,$\mathbb{C}$) and the 4×4 representation matrices in the Cartan representation of the Dirac formalism is also given in [6]. A change of axis $\mathbf{s}(g) \rightarrow \mathbf{s}(h)$, embodied by a corresponding change of group elements $g \rightarrow h$, is obtained by a similarity transformation $\mathbf{D}(g(0)) \rightarrow \mathbf{D}(h(0)) = \mathbf{R}\,[\,\mathbf{D}(g(0))\,]\,\mathbf{R}^{-1}$, based on a rotation $\mathbf{R}$. This will then evolve with time to $\mathbf{D}(h(t)) = \mathbf{R}\,[\,\mathbf{D}(g(t))\,]\,\mathbf{R}^{-1}$ (as proved in [6], pp. 310-313), whose spinor $\boldsymbol{\psi}(h(t))$ will comply now with the boundary conditions for a coherent source. The correction in the end to recover the correct incoherent-source value $\mathbf{D}(g(t)) = \mathbf{R}^{-1}\,[\,\mathbf{D}(h(t))\,]\,\mathbf{R}$ is then just the reverse similarity transformation. And then we will see that the spinors $\boldsymbol{\psi}(g(t))$ and $\boldsymbol{\psi}(h(t))$ will yield the same result $[\,\boldsymbol{\psi}(h(t))\,]^{\dagger}[\,\boldsymbol{\psi}(h(t))\,] = [\,\boldsymbol{\psi}(g(t))\,]^{\dagger}[\,\boldsymbol{\psi}(g(t))\,] = 1$. The main idea which makes this all work is that in a coherent process the spin axis does not change due to the interaction with the measuring device, else the process would be incoherent. A change of $\mathbf{s}$ would be accompanied with a corresponding change in the measuring device due to conservation laws. In the double-slit experiment this would e.g. permit to know through which slit the particle has gone. We may note that to our knowledge the issue that in general the source will be incoherent has never been addressed in examples of solving Schrödinger, Pauli or Dirac equations in textbook examples. The source has always been tacitly implied to be coherent by the choice of the boundary conditions. This has been then a kind of blind spot. The problem cannot be solved without a proper understanding of spinors and of the relationship between the particles and the waves.

4.4 The minimal substitution

The minimal substitution was introduced in classical mechanics to calculate orbits. It was validated by comparing the results of the calculations performed with it in the Lagrange-Hamilton formalisms with those of traditional Newtonian mechanics. It was then proved to work also for relativistic orbits. It is a kind of startling that traditional QM explains you first that it is entirely different from classical mechanics and that we must stop thinking about orbits, and then introduces the minimal substitution without any comment as though it would be self-evident.

We can propose the following justification for it. Just as the equation $E^2 - c^2\mathbf{p}^2 = (m_0c^2)^2$ and the four-vector $(E, c\mathbf{p}) = m_0c^2(\gamma, \gamma\mathbf{v}/c)$ can serve to define a global Lorentz transformation for the wave function in the free-space Dirac equation, $(E - qV)^2 - c^2(\mathbf{p} - q\mathbf{A})^2 = (m_0c^2)^2$ and the four-vector $(E - qV, c(\mathbf{p} - q\mathbf{A}))$ can be used to define a field of local Lorentz transformations for the wave function of the Dirac equation in the presence of an electromagnetic potential. This substitution is also covariant.

4.5 Discussion

4.5.1 The Born rule, Schrödinger's cat, the particle-wave duality and the double-slit experiment

In solving the Schrödinger and Dirac equations we impose boundary conditions. This shows that the probabilities we use are conditional. The condition is the experimental set-up. The probabilities do indeed depend on the experimental set-up as Bohr has claimed. The boundaries we use are idealized. E.g. the walls of the slits are considered as perfectly planar while on a microscopic scale they must present some roughness and they are made of molecules or atoms. Following Einstein we could imagine here lots of hidden variables which do not reside within the particles but are located within the set-up itself (see paragraph 4.5.2). The viewpoints of Einstein and Bohr are therefore not entirely mutually exclusive.

It is now time to complete the work we started in paragraph 2.5.2. We have for the moment considered one electron in each point $\mathbf{r} \in \mathbb{R}^3$. We could consider that there is a larger number of electrons $N \in \mathbb{N}$ in each point. Following the ideas expressed in paragraph 2.5.2 we would thus have to consider $N\mathbf{R}$ to start with, as each electron would have its copy of the matrix $\mathbf{R}$ or spinor $\boldsymbol{\psi}$ attached to it. But as the spinors of SU(2) statisfy the condition $\boldsymbol{\psi}^\dagger\boldsymbol{\psi} = 1$ we must calculate the number of electrons by using the quantity $\boldsymbol{\psi}^\dagger\boldsymbol{\psi} = 1$. This leads then to the rule that we must use $\sqrt{N}\boldsymbol{\psi}$. We can extrapolate this idea to the case that we use probablity densities and then also normalize the wave function. It may represent a lot of effort to write all this down in a mathematically tidy way, but the rationale presented here gives us a justification for the Born rule.

The construction of mixed states as corresponding to sets gives us a simple interpretation for the wave function of Schrödingers cat. The wave function does not describe a cat that is half dead and half alive, but an ensemble of cats whereby half of the cats are dead and half of them alive. In this approach, the algebra cannot be taken literally. We must refrain from insisting on carrying out the algebra to the very end, by treating spinors like vectors, as in the example in paragraph 2.5.2 where we obtained $\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 = \mathbf{0}$. The mixed state must be considered as a juxtaposition, rather than as a real algebraic sum. It would be preferable to write $\{c_1\boldsymbol{\phi}_1, c_2\boldsymbol{\phi}_2, \cdots, c_p\boldsymbol{\phi}_p\}$ rather than $c_1\boldsymbol{\phi}_1 + c_2\boldsymbol{\phi}_2 + \cdots + c_p\boldsymbol{\phi}_p$, because carrying out the algebra by brute force anyway is just wrong mathematics and the notation $\{c_1\boldsymbol{\phi}_1, c_2\boldsymbol{\phi}_2, \cdots, c_p\boldsymbol{\phi}_p\}$ just sticks to the real meaning.

However, this solution of refraining from brute-force algebra raises a severe issue in the case of interference, e.g. in the double-slit experiment, where $\boldsymbol{\psi}_L + \boldsymbol{\psi}_R = \mathbf{0}$ leads indeed to a zero physical intensity, suggesting that the procedure should be taken seriously anyway. Interpreting $\boldsymbol{\psi}_L + \boldsymbol{\psi}_R = \mathbf{0}$ in terms of sets is here certainly not meaningful. It would imply that the union of two non-empty sets would be empty. We have therefore proposed in Subsection 4.3 that the individual particles must already comply with the intensity pattern.

In fact, based on the way we constructed the wave function we can also propose a solution for the conondrum of the particle-wave duality. What behaves as a wave is the wave function, i.e. the statistical ensemble of the electrons that have been used in the experiment. It is this wave as a whole that can flow as a dense fluid of non-interacting electrons through both slits in a double-slit experiment. The fluid is virtual and has been created by the simultaneity of description we have used in defining the wave function in Subsection 4.1. The single electrons behave as particles and each of them goes through a single slit. The idea that an electron can go through both slits simultaneously and interfere with itself is of course related to the concept of wave packets. We have explained that the superluminal phase velocities do not justify introducing wave packets.The electrons are detected as points on a detector screen and are therefore always particles.

Still there seems to be a contradiction that puts us on a tight rope, because we need to explain the interference pattern. This apparent contradiction can be solved as follows. The boundary conditions we impose on the wave equations are non-local, because the macroscopic set-up of the experiment is non-local. The wave function itself is also non-local because in its construction we have synchronized clocks in a Lorentz frame up to infinite distance. We are trying thus to solve a differential equation with non-local boundary conditions that will lead to a non-local solution. We could first try to find all solutions by making the calculations as though we took the sums in the superpostion state seriously, although we know that we we should not for the reasons described above and which all trace back to the undeniable fact that we cannot add spinors like vectors. In other words, we cheat and ignore the taboo. This way we could set up a pool of all possible solutions for the differential equation ignoring the existence of any taboo. The correct solutions which do respect the taboo will then also be present within this pool of all possible solutions.

One can check afterwards which solutions in the pool allow for an alternative interpretation in terms of spinors or sets of spinors. We can this way get away with our unethical behaviour of cheating by basing ourselves on the accomplished fact that we have found a solution that has stood the test. An example of this procedure for finding an alternative interpretation is given in reference [13], where the calculation $\boldsymbol{\psi}_L + \boldsymbol{\psi}_R = \mathbf{0}$, which would be valid for vectors, is not valid for spinors, as has been shown above in paragraph 2.5.1. One can then argue that the algebra used to obtain the solution $\boldsymbol{\psi}_L + \boldsymbol{\psi}_R = \mathbf{0}$ is logically flawed for spinors but valid for finding the pool of solutions of the differential equation for vectors. The vector solutions are found by using a Huyghens' principle for the solution of differential equations of a certain type [37,38]. This Huyghens' principle is completely devoid of any physical meaning. It is just a mathematical method that has been proved to work for certain types of differential equations. This transpires already in Kirchhoff's elaboration [39] of the Huyghens' principle for electromagnetic waves, whereby one is forced to accept that they can travel sometimes backwards in space again. Feynman's description is even more eloquent [40] with *"photons going faster or slower than the conventional speed of light, electrons going backwards in time".* But we should not be amazed that we need such unphysical and non-local aspects as going backwards in space and in time, or traveling faster than light, because we must find an overall solution which satisfies all the non-local boundary conditions. All that we must keep in mind is that it is not real. When one has solved the problem at one boundary one must still solve it at other boundaries and this really might require accepting unphysical propagation for the waves in the Huyghens' principle. But the result of the correct mathematical procedure is with high precison $\psi_L + \psi_R$, as e.g. shown by Feynman's all-histories elaboration of the Huyghens' principle.

We have worked out these ideas for the double-slit expeiment [13], which contains much more than the incomplete discussion we can present here. Let us remind that we have derived the Dirac equation by rigorous mathematics. It leads to a differential equation that we can solve rigorously. That is a first mathematically rigorous track. This track reproduces the experimental results. Next we follow a second mathematical track. We use probability calculus and we come to the conclusion that the solution of the differential equation does not comply with our intuition, which tells us that we should just measure the sum of the probabilities observed in two single-slit experiments. It is very important to realize at this point that the paradox is here no longer between the physics and the mathematics but between two different mathematical tracks. *The solution of the paradox must thus reside entirely within the mathematics such that no quantum magic can be involved!*

Solving the mathematical paradox takes more than we have developed up to now. An essential rôle is played by the fact that it is impossible to know through which slit the particle has gone [13]. This is because coherent interactions do not leave any information behind on the crime scene of their passage through the set-up, such that we cannot possibly know the exact trajectory the electron has taken, because no information has been created. The answer to the question through which slit the particle has traveled is *undecided*, just like the question if the fifth postulate of Euclidean geometry is true cannot be answered on the basis of the first four postulates because they do not contain the required information. It is as giving a list of commercial items loaded into a ship and then asking what the age of the captain is.

The information that would allow to answer the question "which way" does not exist. That is what defines the conditional probabilities and the boundary conditions imposed by the double-slit set-up. We should not mix up conditional probabilities from different set-ups because they can define incompatible conditional probabilities. Undecided events and their conditional probabilities cannot be treated in terms of the conditional probabilities from other set-ups which contain and create entirely different information to the extent that the question "which way" can now be decided. Such other set-ups just define different conditional probabilities because each set-up has its own conditional probabilities. The conditional probability that we do not know through which slit the electron has gone cannot be constructed from the conditional probabilities where we know that the electron has gone through a given slit. For sure the electron has gone through one of the two slits, but the information is not available. Respecting this caveat is exactly what the mathematics do when you solve the differential equation with the boundary conditions of the double-slit experiment. When we use intuitive probability calculus we are trampling this caveat, by only considering the locality of the interactions. Perhaps because our macroscopic intuition is entirely based on incoherent processes where the answers to questions are always decided.

The explanation offered must be considered as cogent to the extent that there is no other way to avoid completely illogical conclusions, e.g. that at the quantum level other rules of probability calculus would prevail. This is wrong because we have already pointed out that the solution of the paradox does not reside within the physics. We must also reject it because we should not accept the defaitist philosophy that we cannot think straight. We should not loose our head and ressort to quantum magic in moments of adversity but stay cool and remain convinced that another, purely logical explanation *must* exist.

The Schrödinger equation has been used with full generality for all kinds of different particles, like He atoms, electrons, C_{60} molecules. The tacit assumption that all these different types of particles obey the same equation has been adopted as self-evident. In reality it would have to be proved for each type of particle and that it works with such generality must be considered as a fluke. Nature could have just been different.

4.5.2 Conclusion

Where do we stand now? What we have learned is that the Dirac equation can be derived without introducing extraordinary assumptions, such as many worlds or signalling back in time. One might now reason as follows. We have a spinor formalism that we understand. With this spinor formalism a huge corpus of calculations has been constituted which reproduce the experimental results correctly. Hence all that remains to be done is to use the geometrical interpretation to know what the calculations mean.

However, this could be too optimistic. The problem is that during all these years it has not been known what the formalism means. Not knowing what the formalism means can be full of boobytraps. It can lead to overinterpretations or transgressions of the domain of validity of the calculations. There may remain problems in the way we have used the Dirac equation. The reader can consult [11] to check that there are a lot of occasions where things can just go wrong. This lays bare the intellectual limits of the punch line "shut up and calculate" with its force-feeding of black-box algebra. It presents a state of affairs that is not satisfactory as satisfactory. We can therefore only try to apply the spinor formalism with the new understanding in order to check if we can figure out an explanation for a given experiment. This boils down to case-by-case investigations. For each experiment this can remain very difficult in its own right.

We have undertaken a survey to check if the solution of some quantum paradoxes may not require quantum magic after all. We have already mentioned this to a certain extent in Subsection 1.3 on pp.7-9 of [11]. We have treated

a number of problems and come to the conclusion that they do not require quantum magic. We have been able to derive the Dirac equation starting from some very clear and simple assumptions. We have solved the paradox of the particle-wave duality (see Subsection 4.5.1 and [13,29]). We have solved the problem of the paradox of Schrödinger's cat (see Subsection 4.5.1, Subsection 2.3.2 of [7] and [28]). We have offered an explanation for the double-slit experiment in [13,29] (which can be further enriched by Subsection 4.3 or the Appendix of [28] to deal with incoherent sources). We have explained why the phase velocity of the wave function is c^2/v. We have explained the Stern-Gerlach experiment in [11]. The work of Hansen and Ravndal [41] already explains tunneling such that it has no longer to be addressed anymore. The problem of entanglement and hidden variables has been settled by many authors starting with Kupczynski back in 1987 [42] who have pointed out an error in the derivation of the Bell inequalities. This error is closely related to the one we encountered in the discussion of the double-slit experiment, viz. that we should not combine conditional probabilities defined by different experimental set-ups. In both cases QM teaches us is that in following our common-sense intuition about probabilty calculus we expose ourselves to committing subtle, subliminal logical errors, which can be really hard to spot, such that it looks *almost* as though we cannot think straight.

4.5.3 Epilogue

We are now reaching the end of a very long journey. I started with three quotes and I think I could end with a fourth one by Murray Gell-Mann [43] recipient of the 1969 Nobel prize in physics:

□ *"Niels Bohr brainwashed a whole generation of theorists into thinking that the job* [interpreting quantum theory] *was done 50 years ago".*

Bohr has brainwashed more than one generation. For students in the past QM may have looked frustrating and even ugly. The main message from this paper is that we should never again accept something like the Copenhagen interpretation, with all its internal contradictions. We must break away from it, in the words of Dieudonné [12]: "Pour l'honneur de l'esprit humain" (For the sake of the honor of the human spirit). I hope that for future students QM may from now on look more like an enthralling poem of very beautiful mathematics.

References

1. R.P. Feynman, in *The Character of Physical Law*, Messenger Lectures, Cornell University Podcast (1964), printed by the BBC, Modern Library (1965)
2. G. Farmelo, in *The strangest Man. The hidden Life of Paul Dirac: Quantum Genius*, Faber and Faber (2009), p. 430.
3. A. Pais, M. Jacob, D.I. Olive, and M. Atiyah, in *Paul Dirac - The Man and his Work*, Cambridge University Press, Cambridge (1998), pp. 113-114.
4. E. Cartan, in *The Theory of Spinors*, Dover, New York, (1981).
5. D. Hestenes, Zitterbewegung in Radiative Processes, in *D. Hestenes and A. Weingartshofer (eds.), The Electron, New Theory and Experiment*, Fundamental Theories of Physics, Springer (1991), p.21-36
6. G. Coddens, in *From Spinors to Quantum Mechanics*, Imperial College Press, London (2015).
7. G. Coddens, Spinors for everyone, https://hal.archives-ouvertes.fr/cea-01572342v1.
8. N. David Mermin, What's Wrong with this Pillow, *Physics Today*, (April 1989), page 9,
9. F. Halzen and A.D. Martinn in *Quarks and Leptons, An introductory Course in Modern Particle Physics*, John Wiley and Son, (1984). The values quoted in the main text have been taken from: D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse, G.; Cavity Control of a Single-Electron Quantum Cyclotron: Measuring the Electron Magnetic Moment. Phys.Rev. **A. 83** 052122 (2011); T. Aoyama, M. Hayakawa, T. Kinoshita, M. Nio, Tenth-Order QED Contribution to the Electron g-2 and an Improved Value of the Fine Structure Constant, Phys. Rev. Lett. **109**, 111807 (2012).
10. P. Ball, Quantum Theory Rebuilt From Simple Physical Principles, Quanta Magazine (2017), https://www.quantamagazine.org/quantum-theory-rebuilt-from-simple-physical-principles-20170830/.
11. G. Coddens, The Exact Theory of the Stern-Gerlach Experiment and Why it Does Not Imply that a Fermion Can Only Have Its Spin Up or Down, Symmetry **13**,134 (2021).
12. J. Dieudonné, in *Pour l'honneur de l'esprit humain - les mathématiques aujourd'hui*, Hachette, Paris (1987).
13. G. Coddens, A solution of the paradox of the double-slit experiment, https://hal.archives-ouvertes.fr/cea-01459890v3.
14. B.L. van der Waerden, in *Group Theory and Quantum Mechanics*, Springer, Berlin-Heidelberg (1974), translated from *Die gruppentheoretische Methode in der Quantenmechanik*, Springer, Berlin (1932).
15. See e.g. B.E. Sagan, in *The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions*, Springer Graduate Texts in Mathematics, **203**, Second Edition, Springer, New York (2001), who states clearly (on p.7) that the linear combinations are purely *formal*. Instead of a group ring [14], Sagan calls the set of purely formal linear combinations a G-module, and the corresponding calculus the group algebra.

16. R. Deheuvels in *Tenseurs et Spineurs*, Presses Universitaires de France, Paris (1993), p.232.
17. H. Rauch, A. Zeilinger, G. Badurek, A. Wilfing, W. Bauspiess, and U. Bonse, Verification of coherent spinor rotation of fermions, *Phys. Lett.* **54A**, 425-427 (1975).
18. R.P. Feynman, and S. Weinberg, in *Elementary Particles and the Laws of Physics*, Cambridge University Press, New York (1987).
19. M. Staley, Understanding quaternions and the Dirac belt trick, *Eur. J. Phys.* **31**, 467 (2010).
20. C.W. Misner, K.S. Thorne, and J.A. Wheeler, in *Gravitation*, Freeman, San Francisco (1970).
21. H.F. Jones, in *Groups, Representations and Physics*, Adam Hilger, Bristol (1990).
22. P. Marmier and E. Sheldon in *The Physics of Nuclei and Particles, Vol. I*, Academic Press, New York (1969), p.341.
23. K.C. Biedenharn and J.D. Louck , in *Angular Momentum in Quantum Mechanics, Theory and Application*, Encyclopedia of Mathematics and its Applications, Vol. 8, Addison-Wesley, Reading, MA (1981), and Cambridge University Press (1985).
24. J. Gallier in *Geometric Methods and Applications for Computer Science and Engineering*, Texts in Applied Mathematics **38**, 2nd edition, Springer (2011), especially Chapter 8 about the Cartan-Dieudonné theorem.
25. A.A. Ungar in *Beyond the Einstein Addition Law and its Gyroscopic Thomas Precession: The Theory of Gyrogroups and Gyrovector Spaces*, Kluwer, New York, Fundamental Theories of Physics, **117**, (2002).
26. J.A. Rhodes, and M.D. Semon, Relativistic velocity space, Wigner rotation and Thomas precession, *Am. J. Phys.* **72**, 945 (2004).
27. G. Coddens, On magnetic monopoles, the anomalous g-factor of the electron and the spin-orbit coupling in the Dirac Theory, https://hal-cea.archives-ouvertes.fr/cea-01269569.
28. G. Coddens, A linearly polarized electromagnetic wave as a swarm of photons half of which have spin -1 and half of which have spin +1, https://hal.archives-ouvertes.fr/hal-02636464v3.
29. G. Coddens, A proposal to get some common-sense intuition for the paradox of the double-slit experiment, https://hal.archives-ouvertes.fr/cea-01383609v5.
30. L.E. Ballentine, in *Quantum Mechanics, A Modern Development, 2nd edition*, World Scientific, Singapore (1998).
31. H. Everett, "Relative State" Formulation of Quantum Mechanics, *Rev. Mod. Phys.* **29**, 454-462 (1957).
32. J. Cramer, The transactional interpretation of quantum mechanics, *Rev. Mod. Phys.* **58** 795-798 (2009).
33. D. Bohm, A Suggested Interpretation of Quantum Theory in Terms of "Hidden" Variables I, *Phys. Rev.* **85**, 166-179 (1952).
34. A. Messiah in *Quantum Mechanics, Volume I*, North Holland, Amsterdam, 1965.
35. Y. Fukuda et al. (Super-Kamiokande Collaboration), Evidence for Oscillation of Atmospheric Neutrinos, *Phys. Rev. Lett.* **81**, 1562-1567 (1998).
36. A. Tonomura, J. Endo, T. Matsuda and T. Kawasaki, *Am. J. Phys.* **57**, 117 (1989); see also M.P. Silverman in *More than one Mystery, Explorations in Quantum Interference*, Springer, p. 3 (1995).
37. J.J. Duistermaat in *Huygens' Principle 1690-1990: Theory and Applications*, H. Blok, H. Ferwerds and H.K. Kuiken eds., Elsevier, (1992), p. 273;
38. J. Hadamard in *Le problème de Cauchy et les Equations aux Dérivés partielles linéaires hyperboliques*, reprint of the 1923 publication, Dover, New York, (1952).
39. R.S. Longhurst, in *Geometrical and Physical Optics, 2nd edition*, Longman Group, London (1967).
40. R.P. Feynman in *QED: The Strange Theory of Light and Matter*, Princeton University Press (1985).
41. A. Hansen and F. Ravndal, Klein's Paradox and Its Resolution, *Phys. Scripta* **23**, 1036 (1981).
42. M. Kupczynski, Closing the Door on Quantum Nonlocality, Entropy 2018, 20(11), https://doi.org/10.3390/e20110877 and the many references therein.
43. M. Gell-Mann, Is nature simple? In *D. Huff, & O. Prewett, (Eds.) The nature of the physical universe, 1976 Nobel Conference*, p. 29, New York: Wiley Interscience (1979).