Eur. Phys. J. C 55, 145–157 (2008)
DOI 10.1140/epjc/s10052-008-0563-0

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article – Theoretical Physics

Spinors in the Lorentz group and their implications for quantum mechanics

G. Coddens[a]

Laboratoire des Solides Irradiés, Ecole Polytechnique, CEA/DSM/IRAMIS, CNRS, Route de Saclay, 91128 Palaiseau Cedex, France

Received: 5 December 2007 /
Published online: 18 March 2008 –

Abstract. We investigate what the precise meaning is of a spinor in the rotation and Lorentz groups. We find that spinors correspond to a special coding of a group element. This is achieved by coding the whole reference frame into a special isotropic or "zero-length" vector. The precise form of that special vector in the Lorentz group is lacking in the literature, and this leads to some confusion, as the point that the coding can be complete has been missed. We then apply these ideas to quantum mechanics and find that the Dirac equation can be derived by just trying to describe a rotating electron.

PACS. 02.20.-a; 03.65.-w; 03.65.Fd

1 Introduction

The present paper is the second one in a personal attempt to get an understanding of what spinors are. The reader is referred to [1]; but the other references listed below are of course much more important. The motivation of this work was not to derive new results on the mathematics of spinors. In this respect, the existing literature [2–6] – in particular, there is a detailed treatment of spinors in [4] – contains very complete accounts that we cannot aim to surpass. Our goal was rather to obtain a better understanding of the underlying mathematical ideas. We thus did not expect to find new results. Despite this, we have been able to derive an important remark about the standard interpretation of the information contents of a spinor in the Lorentz group. In fact, we are able to show that these information contents are richer than transpires in Penrose's interpretation in terms flags and flagpoles, as it is possible to code unambiguously the whole transformed reference frame into a spinor. The formulation in terms of flags and flagpoles is precisely based on the belief that this would not be possible.

The motivation of our work was our profound conviction that a thorough understanding of the ideas behind the mathematics is absolutely necessary if one wants to sort out the conceptual difficulties inherent to quantum mechanics. As both the mathematics and the physics involved are difficult, one might at a certain stage even not be able to figure out which difficulties are purely mathematical (and therefore are not mysterious) and which ones belong to physics. It will be futile to try to search for a physical explanation of some difficult point if the difficulty in question is in reality purely mathematical. If one wants to avoid such situations, one has to know exactly where the demarcation line lies between the physics and the mathematics, i.e. one must master all the underlying ideas of the mathematics. In an axiomatic presentation, some of these ideas can remain totally hidden, and it may take quite an effort to figure them out from the presentation [1]. It is this effort that we wanted to make. So far as the initial motivations of this paper are concerned, they are entirely personal, and as the reader will notice, the presentation is also entirely personal and non-standard, because I was not concerned with an axiomatic presentation or with deriving theorems. However, in the final paragraphs of the paper we have attempted to apply what we have understood to quantum mechanics. We will see that here the difficulties in the mathematics and in the physics truly interfere as anticipated. A number of points that are felt as very counter-intuitive physics are mathematical. We are able to derive the Dirac equation with very few additional assumptions. What comes out is an extremely simple and visual picture, well eclipsed behind the subtleties of the mathematical machinery, which offers interesting new view points. This justifies a posteriori the initial motivations of our quest.

2 Preamble

In this section we start from a small remark about group representation theory. We will use the group of three-

[a] e-mail: Gerrit.Coddens@polytechnique.fr

[1] The following is an incomplete list of text books that give detailed but conventional accounts of group representation theory in physics: [7–11].

dimensional rotations as a case in point. Some parts of the theory of spinors in the rotation group will be further developed in the next section. It is not our aim to reproduce here the full theory of spinors in SO(3): an ample description is given in [1–6]. We just want to focus onto a number of important points in the development that will clarify the spinor idea and by analogy will also make it possible to understand the meaning of spinors in the Lorentz group.

A representation of a group G is an isomorphism between group elements $g \in G$ and matrices $D(g)$:

$$\begin{aligned} &\text{if} && g_1 \to D(g_1) \\ &&& g_2 \to D(g_2) \\ &\text{then} && g_2 \circ g_1 \to D(g_2)D(g_1)\,. \end{aligned} \tag{1}$$

A well-known example are the 3×3 rotation matrices in $\mathbb{R}^3$. In this case the matrices operate on 3×1 column vectors, which are images of vectors $\mathbf{v} \in \mathbb{R}^3$. With two representations $D^{(1)}$ and $D^{(2)}$ of dimensions d_1 and d_2 respectively, one may construct a new one D of dimension $d_1 d_2$, namely

$$g \to D(g) = D^{(2)}(g) \otimes D^{(1)}(g)\,, \tag{2}$$

which works on vectors $\mathbf{w} \otimes \mathbf{v}$. This is a compact form to write that starting from quantities and transformation laws $w'_j = \sum_m D^{(2)}_{jm} w_m$ and $v'_k = \sum_n D^{(1)}_{kn} v_n$, one can define quantities $a'_{jk} = w_j v_k$ and $a_{mn} = w_m v_n$, with a transformation law $a'_{jk} = \sum_{mn} D_{jk,mn} a_{mn}$, where $D_{jk,mn} = D^{(2)}_{jm} D^{(1)}_{kn}$, as is easily checked. If the eigenvalues corresponding to the eigenvectors $\mathbf{w}_j$ of $D^{(2)}$ are λ_j and the eigenvalues corresponding to the eigenvectors $\mathbf{v}_k$ of $D^{(1)}$ are μ_k, then the eigenvectors of $D^{(2)}(g) \otimes D^{(1)}(g)$ will be $\mathbf{w}_j \otimes \mathbf{v}_k$ with eigenvalues $\lambda_j \mu_k$. When the vectors $\mathbf{v}$ and $\mathbf{w}$ are identical, the indices (jk) and (kj) must be grouped together, which will reduce the dimension to less than $d_1 d_2$. In the rotation group, this point summarizes the derivation of the whole set of harmonic polynomials.

3 The group of three-dimensional rotations

When we apply this idea the other way around, to the rotation group in $\mathbb{R}^3$, we see that the eigenvalues for a rotation over an angle φ are $1, \mathrm{e}^{\mathrm{i}\varphi}, \mathrm{e}^{-\mathrm{i}\varphi}$, which is of the type $\lambda_1\mu_2, \lambda_1\mu_1, \lambda_2\mu_2$, with $\lambda_1 = \mathrm{e}^{\mathrm{i}\varphi/2}$, $\lambda_2 = \mathrm{e}^{-\mathrm{i}\varphi/2}$ and $\mu_1 = \mathrm{e}^{\mathrm{i}\varphi/2}$, $\mu_2 = \mathrm{e}^{-\mathrm{i}\varphi/2}$. This suggests that there might exist a two-dimensional representation D wherein the eigenvalues are $\lambda_1 = \mathrm{e}^{\mathrm{i}\varphi/2}$ and $\lambda_2 = \mathrm{e}^{-\mathrm{i}\varphi/2}$, for which $D \otimes D$ corresponds to the three-dimensional representation. The three-dimensional vectors would then in reality be composed quantities of the type $\mathbf{v} \otimes \mathbf{v}$ in terms of more basic quantities $\mathbf{v}$. (Of course, the two-dimensional representation we are talking about is well known. The reader is supposed to have some knowledge about it in order to be able to follow the argument.) This surprising possibility has profound consequences: there must exist representations of the rotation group whose matrices do not work on images of vectors, as the two eigenvalues mentioned do not fit into a scheme for vector images. The question arises then on what kind of images these representations might be working. The answer is simple: they work on images of rotations. In fact, the group structure of the rotation group exists without any reference to a vector of $\mathbb{R}^3$. All that the group structure really defines is the multiplication table of group elements. By considering the rotations as elements of a group, we are shifting viewpoints: we consider the rotations as acting on other rotations, as expressed by the composition law of an abstract group. We no longer consider them as acting on coordinates. There is no longer any mention of any vector $\mathbf{r}$ in such an abstract description in terms of rotations acting on other rotations, nor of any length R of such a vector. One focuses all the attention onto the abstract structure of the group. One will thus end up with a minimal description, from which everything that is not essential has been stripped away and thus e.g. any reference to vectors has been removed, because we might need to consider vectors for carrying out the task of writing down the group multiplication table, but once it is written and we have its structure, these vectors are not part of the formulation. Perhaps we could then reason as follows: the most fundamental representation of a group should be devoid of any reference to quantities like vectors. In this minimal approach, all a rotation can work on is another rotation. The most basic representation *must* therefore be one that works on images of group elements: therefore, the two components of the 2×1 column vectors that the 2×2 matrices of the representation work on *must* code for a rotation. We will show that this reasoning holds for the rotation group. By analogy we will develop it as a guiding principle for the Lorentz group.

This raises the question of the coding: how can one turn the image of a rotation into a 2×1 column vector? A rotation is a linear mapping, and a linear mapping is entirely defined by its restriction to a basis. Hence, to know a rotation completely we must know how it works on the triad ("*Dreibein*" in German) of three basis vectors of an orthonormal reference frame. The three vectors have in total nine components, but they are not all independent: the vectors are normalized (three conditions) and mutually orthogonal (three more conditions). There remain only three independent variables: the direction of a first unit vector (e.g. $\mathbf{e}_1 = (x_1, y_1, z_1)$) can be coded with two independent variables. The direction of a unit vector that is orthogonal to it (e.g. $\mathbf{e}_2 = (x_2, y_2, z_2)$) can then be coded with just one more independent variable. This fixes at once the value of the third vector $\mathbf{e}_3 = \mathbf{e}_1 \wedge \mathbf{e}_2$. Hence, a triad could serve as an image of a rotation.

A scheme for coding the whole information contents of a triad has been developed by Cartan. It starts with vectors, but it works in such a way that we can clearly see at a given stage how we leap from vectors to rotations. We combine two unit vectors of the triad into a single vector by blending them into a single quantity $\mathbf{e}_1 + \mathrm{i}\mathbf{e}_2 = (x, y, z)$. That is, $x = x_1 + \mathrm{i}x_2, y = y_1 + \mathrm{i}y_2, z = z_1 + \mathrm{i}z_2$. By doing so the whole information about the triad is coded unambiguously. We can always decode it back again. Whatever rotations we perform on this quantity, we will always be

able to identify the rotated images of the two basis vectors $\mathbf{e}'_1$ and $\mathbf{e}'_2$ afterwards by separating the real and imaginary parts.

The extrapolation of the algebra involved in the Euclidean distance function of $\mathbb{R}^3$ to $\mathbb{C}^3$ leads to the finding that the quantity $\mathbf{e}_1 + \mathrm{i}\mathbf{e}_2$ obtained in this way by combining two orthogonal unit vectors is a so-called isotropic vector, i.e. a vector of "zero length", but the latter formulation is an abuse of language, since after the extrapolation of the Euclidean distance function towards $\mathbb{C}^3$ it no longer defines a distance function. In fact, $x^2 + y^2 + z^2 = (x_1 + \mathrm{i}x_2)^2 + (y_1 + \mathrm{i}y_2)^2 + (z_1 + \mathrm{i}z_2)^2 = 0$, which is completely at variance with a basic axiom for a distance function. Hence the transition from vectors to rotations can be made through the use of isotropic vectors. Note that in reality we do not need to work on the unit vectors, but only on their directions: a vector remains isotropic when we multiply it with a constant. This is why we can use homogeneous coordinates and why we come across representations in terms of homogeneous coordinates (i.e. the harmonic polynomials) in the example of the three-dimensional rotation group.

The generalization of coding the whole "*N-bein*" to higher dimensions is a leading principle for the representation theory, although it is a priori not obvious how one can e.g. code the four unit vectors of the tetrad ("*Vierbein*") in $\mathbb{R}^4$ into one quantity: we run out of quantities like i. There is no commutative number field beyond $\mathbb{C}$ and therefore it looks at first sight as though we are in for a party of non-commutative algebra. We will explain how we overcome these difficulties for $\mathbb{R}^4$, and especially for the Lorentz group of special relativity, later on. It will contain a number of surprises, but eventually we will be able to appreciate that the idea of coding the "*N-bein*" as an image of the group element is rigorously respected and the guiding principle for the development. Spinors are nothing else than the appropriate coding in the form of a column matrix of the elements of the rotation group in $\mathbb{R}^n$ in the form of an "*N-bein*". This is precise and clear, and it is noteworthy that such a clear statement is hard to find in the specialized literature, with the immediate consequence that to many people the spinor concept looks impenetrable and shrouded in mystery. But if we combine the argument about what is the minimal information that defines a group and the idea that a group element can be pictured through an "*N-bein*", the further development looks quite cogent.

For the two-dimensional representation of the rotation group we must thus combine two three-dimensional vectors into an isotropic vector. But we are confronted with a problem: how can we code three-dimensional vectors into a two-dimensional formalism in the first place? This is where Dirac's trick comes in. We consider a unit vector $\mathbf{a} = (a_x, a_y, a_z)$. We will use this unit vector to code a reflection (the generators of the rotation group are the reflections!) with respect to the plane that is normal to this vector. We thus first attempt to code the reflections into 2×2 matrices. This leap from vectors to reflections is easier than the one from isotropic vectors to rotations. Rotations will follow by combining reflections. The components of the vector $\mathbf{a}$ that defines the reflection A will somewhere pop up in the matrix we are looking for as parameters, but we do not know how or where. Therefore we decompose the matrix $\mathbf{A}$ that codes the reflection A defined by $\mathbf{a}$ linearly as $a_x\sigma_x + a_y\sigma_y + a_z\sigma_z$, where σ_x, σ_y and σ_z are unknown matrices. If we know the matrix σ_x it will tell us where a_x pops up in $\mathbf{A}$. The matrices σ_x, σ_y and σ_z can be found by expressing isomorphically through $\mathbf{AA} = \mathbb{1}$ what defines a reflection, viz. that the reflection operator A is idempotent. We find out that this can be done provided we take the Pauli matrices for σ_x, σ_y and σ_z. If we take the idea that vectors identify reflections a step further, we can use σ_x, σ_y and σ_z to code any vector $\mathbf{v}$ of $\mathbb{R}^3$ as a 2×2 matrix $\mathbf{V}$. We then find that $\mathbf{V}^2 = \mathbf{v} \cdot \mathbf{v}\mathbb{1}$ codes $\mathbf{v} \cdot \mathbf{v}$. We can learn from this that, more generally, $2\mathbf{w} \cdot \mathbf{v} = \mathbf{VW} + \mathbf{WV}$. As a reflection defined by a unit vector $\mathbf{a}$ maps a vector $\mathbf{v}$ onto $\mathbf{v} - 2(\mathbf{a} \cdot \mathbf{v})\mathbf{a}$, we find that the outcome of operating the reflection A on the vector represented by the matrix $\mathbf{V}$ is given by $-\mathbf{AVA}$. We stumble here onto a genuine, major problem, because the result is quadratic in the matrix $\mathbf{A}$: in other words, the representation is not linear.

Similarly, a rotation resulting from two successive reflections A and B, respectively defined by unit vectors $\mathbf{a}$ and $\mathbf{b}$, will be given by $\mathbf{BAVAB}$, which is of the form $\mathbf{SVS}^{-1}$, with $\mathbf{S} = \mathbf{BA}$. To render the formalism linear, we will try to split it by writing $\mathbf{V}$ as $\mathbf{S}_1\mathbf{L}\mathbf{S}_1^{-1}$, such that in the end we can represent the group linearly by matrices $\mathbf{A}$ working on the left on $\mathbf{S}_1$ or by matrices $\mathbf{A}^{-1} = \mathbf{A}$ working on the right on $\mathbf{S}_1^{-1}$. To get more clarity about this problem we can diagonalize $\mathbf{V}$ as $\mathbf{S}_1\mathbf{L}\mathbf{S}_1^{-1}$. By doing so, we see that the structure of $\mathbf{S}_1\mathbf{L}\mathbf{S}_1^{-1}$ perfectly matches the structure of $\mathbf{SVS}^{-1}$ for a rotation.

In fact, the representation is not linear because it is written in terms of vectors instead of rotations: remember that the eigenvalues in a representation based on vectors were of the type λ_1^2, $\lambda_1\lambda_2$ and λ_2^2, and that we are searching for the representation that turns out eigenvalues λ_1 and λ_2. We should thus not be surprised about the idea of halving the formalism, and we can expect a vector to be of the second degree in the more basic quantities. This reminds us of the fact that $\mathbf{V}$ is a vector. We should in some way turn it into a rotation, by taking a kind of square root. As it is the isotropic vectors that must code a rotation, we must replace $\mathbf{V}$ by a matrix $\mathbf{M}$ that codes an isotropic vector. Suppose as a matter of heuristics that we can also diagonalize the matrix $\mathbf{M}$ as $\mathbf{T}\mathbf{\Lambda}\mathbf{T}^{-1}$. Because the structures are the same we will then no longer be able to tell isotropic vectors and rotations apart if we can identify the two diagonalizations, provided $\mathbf{\Lambda}$ can be identified with the value of $\mathbf{L}$ for some element of the group. In this way, we will have found a way to jump logically from a representation in the form of vectors towards a representation in the form of rotations.

But when we try to put this idea to practice we strike disappointment. The idea is totally thwarted by the fact that an isotropic vector has "zero length". This implies that $\mathbf{M}^2 = 0$ and therefore that $\mathbf{M}$ cannot be diagonalized because both its eigenvalues are zero! What saves us now is a *renormalization procedure*. We first diagonalize the matrix $\mathbf{V}$ corresponding to a non-isotropic vector. The eigenvalues turn out to be R and $-R$, where R is the length

of the vector. But the amplitude R can be factorized into two terms $\sqrt{R}$, one of which we relegate to the left to combine it with $\mathbf{S}_1$, and one which we move to the right to combine it with $\mathbf{S}_1^{-1}$. The magic is that when after doing this we let R tend to zero, we find that all quantities remain finite and non-trivial, and this renders our idea of identifying the diagonalization procedures viable again, be it in a modified form. Simultaneously we recover a *principle of homogeneity* that becomes enabled by the prior renormalization: all remaining expressions become homogeneous in the coordinates (x, y, z). This principle is a kind of surprising mental leap. It embodies the leap from vectors to rotations. Without it, all we write is still in terms of vectors. It is therefore crucial. The homogeneity principle and the concern to code the triad become simultaneously enabled by taking the limit $R \to 0$. As we mentioned before, logically the precise length R of a vector has a priori not its place in a formulation of the group. Everything should be independent of the precise value of R. This transpires in the final formalism through the obvious property that the formalism is scale invariant. This is exactly what the homogeneity principle is about: we end up with scale invariant homogeneous coordinates. But this is here achieved in the most radical fashion: by kicking out the quantity R all together by putting it equal to zero. We can understand this as follows: we remove R from the formalism because R does not appear in the group multiplication table. It has no rôle in a minimal set of parameters that should allow one to discuss the group table. That is why we can think of kicking it out. What is left of $\mathbf{\Lambda}$ after taking R out can be identified with a reflection matrix, such that we really satisfy our initial goals. There exists a derivation of the two-dimensional representation of the rotation group starting from the stereographic projection of a vector [5]. A similar derivation for the Lorentz group has been worked out by Penrose [4]. At a certain stage homogeneous coordinates are laconically introduced to turn a homeographic transformation into a matrix transformation, and the group structure appears. When one comes across it for the first time, one wonders what is behind this surprising mental leap. As discussed in the preceding lines and in more detail in [1], the solution to this question is that it embodies the leap from vectors to group elements. All this is exactly what we had in mind when we thought of coding a rotation as an isotropic vector, viz. coding rotations rather than vectors. The ideas applied to the diagonalization of an isotropic vector do indeed reproduce exactly the definition of the spinors for the rotation group as given by Cartan. It also shows that vectors are second-degree tensor products of spinors. The details of the calculations described run as follows: the matrix $\mathbf{R}$ that codes the vector $\mathbf{r} = (x, y, z)$ of length R is given by

$$\mathbf{R} = x\sigma_x + y\sigma_y + z\sigma_z = \begin{pmatrix} z & x - \mathrm{i}y \\ x + \mathrm{i}y & -z \end{pmatrix} . \tag{3}$$

We have introduced here the notation (x, y, z) for a vector that is not of "zero length", anticipating that eventually it will be replaced by the isotropic vector (x, y, z) defined above. The eigenvalues of this matrix are R and $-R$. Diagonalization leads to

$$\mathbf{R} = \frac{1}{\sqrt{-2R(x - \mathrm{i}y)}} \begin{pmatrix} x - \mathrm{i}y & x - \mathrm{i}y \\ -z + R & -z - R \end{pmatrix} \begin{pmatrix} R & 0 \\ 0 & -R \end{pmatrix} \times \begin{pmatrix} -z - R & -(x - \mathrm{i}y) \\ z - R & x - \mathrm{i}y \end{pmatrix} \frac{1}{\sqrt{-2R(x - \mathrm{i}y)}} , \tag{4}$$

and after applying the renormalization procedure

$$\mathbf{R} = \frac{1}{\sqrt{-2(x - \mathrm{i}y)}} \begin{pmatrix} x - \mathrm{i}y & x - \mathrm{i}y \\ -z + R & -z - R \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \begin{pmatrix} -z - R & -(x - \mathrm{i}y) \\ z - R & x - \mathrm{i}y \end{pmatrix} \frac{1}{\sqrt{-2(x - \mathrm{i}y)}} . \tag{5}$$

After taking the limit $R \to 0$ we obtain after some algebra for both columns in the left-hand matrix:

$$\begin{pmatrix} \xi_0 \\ \xi_1 \end{pmatrix} = \pm \begin{pmatrix} \sqrt{\frac{x - \mathrm{i}y}{2}} \\ -\sqrt{\frac{-x - \mathrm{i}y}{2}} \end{pmatrix} , \tag{6}$$

in agreement with the result given by Cartan. Note that $(x, y, z) = (x_1 + \mathrm{i}x_2, y_1 + \mathrm{i}y_2, z_1 + \mathrm{i}z_2)$ as discussed above and therefore this codes the whole triad of a rotated reference frame. We have

$$x = \xi_0^2 - \xi_1^2 , \tag{7}$$

$$y = \mathrm{i}\left(\xi_0^2 + \xi_1^2\right) , \tag{8}$$

$$z = -2\xi_0\xi_1 , \tag{9}$$

from which we can appreciate that (ξ_0, ξ_1) code the whole triad of the rotated reference frame. We must note that in the previous equations x, y and z are complex numbers. For this reason we should be cautious not to use complex conjugation too routinely on ξ_0, ξ_1, and other quantities like $x + \mathrm{i}y$: e.g. the complex conjugate of $x + \mathrm{i}y$ is not $x - \mathrm{i}y$ but $x^* - \mathrm{i}y^*, \ldots$

From the quadratic expressions, we already see transpiring here an essential feature that we find back in quantum mechanics. Probabilities behave like charges; they are conserved quantities. In relativity, they become therefore parts of a more general probability "charge-current" four-vector, subject to a continuity equation. As also in the Lorentz group four-vectors will be second-degree tensors of spinors, this will explain why probability densities are expressed through a quadratic expression $\psi^\dagger\psi$ subject to a continuity equation, where ψ is a spinor. We will have to render this more precise later on.

It may be noted that rotations operate on reflections in a different way than they operate on vectors: there is a factor of 2 in the respective angles involved. This becomes only paradoxical if we confuse these different representations based on reflections and based on vectors. We want to draw attention to a particularity, an important remark, that will allow us to discover also something quite beautiful about the formalism. The three-dimensional representation with the homogeneous (harmonic) polynomials x, y, z can be derived from the two-dimensional one using the ten-

sor product:

$$\begin{pmatrix} z & x-\mathrm{i}y \\ x+\mathrm{i}y & -z \end{pmatrix} = 2\begin{pmatrix} -\xi_0\xi_1 & \xi_0\xi_0 \\ -\xi_1\xi_1 & \xi_0\xi_1 \end{pmatrix} = \sqrt{2}(-\xi_1,\xi_0)\otimes\begin{pmatrix}\xi_0\\ \xi_1\end{pmatrix}\sqrt{2}\,. \quad (10)$$

There is some consistency checking to be done here. We can check that the determinant of the 2×2 matrix with the quantities ξ_1 and ξ_2 is indeed zero, and also that $z/(x+\mathrm{i}y) = -(x-\mathrm{i}y)/z = \xi_0/\xi_1$, etc. This is all fine if we think of the polynomials as abstract quantities. But we are used to think of (x,y,z) as the coordinates of a particle in $\mathbb{R}^3$, while on the other hand, in the spinor formalism, (x,y,z) are not the coordinates of a particle, but of an isotropic vector that codes a rotation, and the numbers (x,y,z) belong to the isotropic cone of $\mathbb{C}^3$. It is the power of analyticity in $\mathbb{C}$ that here accomplishes a miracle. We can extend the algebra from the isotropic cone to the whole of $\mathbb{C}^3$, and then restrict it again to $\mathbb{R}^3$. The original and final domains only have the trivial $\mathbf{0}$ vector in common. But the whole algebraic group structure remains the same and carries through. It is quite beautiful that due to this analytic extension the algebra plays us two movies at the same time. On the isotropic cone the algebra plays us the movie of the rotations, while on $\mathbb{R}^3$ the same algebra plays us the movie of particle motion. But we have to recall that the meaning of the symbols x,y,z is totally different in the two different domains. In a perfectly analogous way, in the Lorentz group we will find that the algebra on the light cone in $\mathbb{C}^4$ will describe the orientation of the axes of a reference frame, while on $\mathbb{R}^4$ the very same algebra will describe its translational motion, and vectors like the probability current-density four-vector $(\psi^\dagger\psi, c\psi^\dagger\boldsymbol{\alpha}\psi)$. But we will again have to remind the reader that the variables (x,y,z,t) in $\psi(x,y,z,t)$ used to describe the probability density will get a completely different meaning and range of values when we use $\psi(x,y,z,t)$ to describe the orientation of the basic tetrad. This is a point that has been largely missed in physics text books.

4 The group of Lorentz transformations

We will now try to build a representation of the group of Lorentz transformations in $\mathbb{R}^4$ based on the same principle that the column matrices that the representation matrices work upon must be images of Lorentz transformations through the coding of a tetrad. There is no major difficulty in repeating the Dirac trick on a vector (that will represent a reflection in $\mathbb{R}^4$) as for the three-dimensional case: this is just analogous to what Dirac did for the four-vector $(E, c\mathbf{p})$ in $\mathbb{R}^4$. That means that we can then code the four-dimensional vectors with the aid of 4×4 matrices. We can even start applying the renormalization and the principle of homogeneity on such a vector. We will then find an expression that codes the tetrad, but we will have to find out how this coding can be built and written down at the level of the coordinates of all the unit vectors. The difficulty resides in the question: how do we code the whole tetrad into one column vector in a constructive step-by-step fashion? How do we code the information contents of four vectors into a single vector, such that it can also be decoded again?

We can appreciate that if the tetrad consists of the vectors $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$ and $\mathbf{e}_{ct}$, then we can build in a first stage the 2×2 matrices $\mathbf{D}$ corresponding to $\mathbf{d} = \mathbf{e}_x + \mathrm{i}\mathbf{e}_y$ and $\mathbf{F}$ corresponding to $\mathbf{f} = \mathbf{e}_z + \mathrm{i}\mathbf{e}_{ct}$. (We anticipate here on something we will develop below, viz. that there exists a two-dimensional representation of the Lorentz group.) But it is hard to see what can be done next. By going to four dimensions we increase the number of available mutually anticommuting matrices. We do not only have

$$\tilde{\sigma}_x = \begin{pmatrix}\sigma_x & \mathbf{0}\\ \mathbf{0} & \sigma_x\end{pmatrix}, \quad \tilde{\sigma}_y = \begin{pmatrix}\sigma_y & \mathbf{0}\\ \mathbf{0} & \sigma_y\end{pmatrix}, \quad \tilde{\sigma}_z = \begin{pmatrix}\sigma_z & \mathbf{0}\\ \mathbf{0} & \sigma_z\end{pmatrix}, \quad (11)$$

but also new matrices, like

$$\begin{pmatrix}\mathbf{0} & \mathbb{I}\\ \mathbb{I} & \mathbf{0}\end{pmatrix}, \quad (12)$$

etc. But putting this superalgebra to work to code the tetrad does not work. If we use e.g. 4×4 super Pauli matrices $\tilde{\sigma}_x$, $\tilde{\sigma}_y$ and $\tilde{\sigma}_z$, to make $\mathbf{D}\tilde{\sigma}_D + \mathbf{F}\tilde{\sigma}_F = \mathbf{D}\otimes\sigma_D + \mathbf{F}\otimes\sigma_F$, it does not work because $\mathbf{D}$ and $\mathbf{F}$ do not commute, as they are not scalar quantities. In fact, taking e.g. $(\mathbf{D}\tilde{\sigma}_z + \mathbf{F}\tilde{\sigma}_x)^2$ we obtain

$$\begin{pmatrix}\mathbf{D} & \mathbf{F}\\ \mathbf{F} & -\mathbf{D}\end{pmatrix}^2 = \begin{pmatrix}\mathbf{D}^2+\mathbf{F}^2 & \mathbf{DF}-\mathbf{FD}\\ \mathbf{FD}-\mathbf{DF} & \mathbf{D}^2+\mathbf{F}^2\end{pmatrix} \neq (\mathbf{D}^2+\mathbf{F}^2)\otimes\mathbb{I}_{2\times 2}\,. \quad (13)$$

Hence we cannot use the supermatrices as a kind of numbers that would play a rôle similar to the one played by i in the rotation group. The whole problem is tied up with the fact that there is no commutative field of numbers beyond $\mathbb{C}$. Coding the information contents of four vectors into a single vector, such that after applying Lorentz transformations on it, this information can be retrieved again unambiguously, looks like an insuperable task. Nevertheless, the analysis of the meaning of spinors presented above holds us in the firm belief that this should be possible. What is the point we have missed? Here the special nature of the Minkowski pseudo-metric used in the Lorentz group brings an unexpected solution. Let us assume that the signature of the pseudo-metric is $(1,1,1,-1)$ such that the pseudo-norm of (x,y,z,ct) is $x^2+y^2+z^2-c^2t^2$. There are then two very distinct ways to obtain a vector of "zero length". In fact, we can take e.g. $\mathbf{e}_x + \mathrm{i}\mathbf{e}_y$, which is an isotropic vector built according to the same recipe as is being used in the rotation group. But we can also take $\mathbf{e}_x + \mathbf{e}_{ct}$, which is a "zero-length" vector or light ray in Minkowski space-time. One can imagine an analogy of the latter quantity with an isotropic vector, by assuming that we do not combine the quantities as $\mathbf{e}_x + \mathrm{i}\mathbf{e}_{ct}$, due to the fact that the signature is not $(1,1,1,1)$. With an isotropic vector like $\mathbf{e}_x + \mathrm{i}\mathbf{e}_y$, the two parts $\mathbf{e}_x$ and $\mathbf{e}_y$ can be reconstructed by separating the real and imaginary parts. Could we perhaps also reconstruct $\mathbf{e}_x$ and $\mathbf{e}_{ct}$ from the "zero-length" vector $\mathbf{e}_x + \mathbf{e}_{ct}$?

The answer is that we can reconstruct the three vectors $\mathbf{r}_1 = \mathcal{L}(\mathbf{e}_x)$, $\mathbf{r}_2 = \mathcal{L}(\mathbf{e}_y)$ and $\mathbf{r}_4 = \mathcal{L}(\mathbf{e}_{ct})$ (where $\mathcal{L}$ denotes a Lorentz transformation) from a combination of the type $\mathbf{r}_1 + \mathrm{i}\mathbf{r}_2 + \mathbf{r}_4$. (From the knowledge of these three vectors, the whole tetrad $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, $\mathbf{r}_4$ can then be reconstructed.) In fact, we can first separate out $\mathbf{r}_2$ by taking the imaginary part of the vector. The real part, $\mathbf{u} = \mathbf{r}_1 + \mathbf{r}_4$, codes a "zero-length" vector of the three-dimensional vector space that is complementary to $\mathbf{r}_2$. This vector space is spanned by the vectors $\mathbf{r}_1$, $\mathbf{r}_3$ and $\mathbf{r}_4$. In other words (u_x, u_y, u_z, u_t) codes a "zero-length" vector, whereby $u_x = x_1 + x_4$, $u_y = y_1 + y_4$, $u_z = z_1 + z_4$, $u_t = ct_1 + ct_4$. A given vector $\mathbf{u}$ on the light cone in the space spanned by $\mathbf{r}_1$, $\mathbf{r}_3$ and $\mathbf{r}_4$, can be split in many ways into two parts $\mathbf{s}_1$, $\mathbf{s}_4$, such that $\mathbf{u} = \mathbf{s}_1 + \mathbf{s}_4$. But there is only one solution for which $\mathbf{s}_1 - \mathbf{s}_4$ also belongs to the light cone and simultaneously also $(\mathbf{s}_1 - \mathbf{s}_4)\cdot(\mathbf{s}_1 + \mathbf{s}_4) = 2$ remains satisfied. This is the solution $\mathbf{s}_1 - \mathbf{s}_4 = \mathbf{r}_1 - \mathbf{r}_4$, $\mathbf{s}_1 + \mathbf{s}_4 = \mathbf{r}_1 + \mathbf{r}_4$. From the knowledge of $\mathbf{s}_1 - \mathbf{s}_4$ and $\mathbf{s}_1 + \mathbf{s}_4$, the vectors $\mathbf{r}_1$ and $\mathbf{r}_4$ can then be reconstructed unambiguously. The proof of the uniqueness of this solution is relegated to Appendix A.

We have thus found an unambiguous way to code the whole information content of a tetrad of Minkowski spacetime into a single vector. While this confirms our general argument that the representation must work on Lorentz transformations rather than on vectors and that therefore the whole tetrad must be in some way coded into the formalism, finding the key to code and to decode the information is less obvious. We have not studied how this idea can be generalized to SO(4) or to higher-dimension rotation groups SO(n). We suspect that the more general approach would rather consist in making tensor products of isotropic vectors, but we have not explored this possibility. We are also not aware of any mention of this possibility of coding in the literature. We suspect Cartan to have worked it out, but the laconic style of his presentation renders it almost impenetrable. It has certainly escaped the attention of many a scientist, as can be appreciated from the discussion in terms of flags and flag poles by Penrose, as reported by Misner et al. in paragraph 41.9 of their book [6]. If one codes only two vectors into a linear combination $\mathbf{r}_1 + \mathbf{r}_4$, the task of making the leap from a description in terms of vectors towards a description in terms of Lorentz transformations cannot be accomplished: the ensuing description is incomplete and the tetrad cannot be entirely reconstructed. This transpires in Penrose's notion of a flag, which just acknowledges in a particularly pictorial way the fact that the tetrad is not described completely. In such a presentation, which could be misinterpreted as an impossibility or incompleteness proof, the true meaning of a spinor, in that it codes the whole tetrad unambiguously, cannot shine through in total clarity.

At this stage we may note that it is possible to code a four-dimensional vector $\mathbf{v} = (x, y, z, ct)$ in a two-dimensional formalism, by taking

$$\begin{aligned}\mathbf{v} = (x, y, z, ct) \rightarrow \mathbf{V} &= ct\mathbb{1} + (\sigma_x x + \sigma_y y + \sigma_z z) \\ &= \begin{pmatrix} ct + z & x + \mathrm{i}y \\ x - \mathrm{i}y & ct - z \end{pmatrix} . \end{aligned} \tag{14}$$

There is an alternative representation, given by

$$\begin{aligned}\mathbf{v} = (x, y, z, ct) \rightarrow \mathbf{V}' &= ct\mathbb{1} - (\sigma_x x + \sigma_y y + \sigma_z z) \\ &= \begin{pmatrix} ct - z & -(x + \mathrm{i}y) \\ -(x - \mathrm{i}y) & ct + z \end{pmatrix} . \end{aligned} \tag{15}$$

We see that the elements of $\mathbf{V}'$ are the minors of the matrix $\mathbf{V}$. Therefore, in these representations $\mathbf{V}'' = \mathbf{V}$ and $\mathbf{V}'\mathbf{V} = (\det \mathbf{V})\mathbb{1} = (c^2t^2 - x^2 - y^2 - z^2)\mathbb{1}$. For unit vectors $\mathbf{a}$, we thus have $\mathbf{A}' = \mathbf{A}^{-1}$. The scalar product of two vectors $\mathbf{v}\cdot\mathbf{w}$ is now given by $\frac{1}{2}(\mathbf{V}'\mathbf{W} + \mathbf{W}'\mathbf{V})$. In this formalism we see that the reflection defined by a unit vector $\mathbf{a}$ whose matrix is $\mathbf{A}$ operates on a vector $\mathbf{v}$ whose matrix is $\mathbf{V}$ as $\mathbf{V} \rightarrow -\mathbf{A}\mathbf{V}'\mathbf{A}$. A rotation leads to $\mathbf{V} \rightarrow \mathbf{B}\mathbf{A}'\mathbf{V}\mathbf{A}'\mathbf{B}$. That is, we must leap backwards and forwards between two different representations in order to obtain a complete formalism. We can return to a complete and pure reflection formalism like we had in the rotation group by combining the two representations into a single one as follows:

$$\mathbf{v} = (x, y, z, ct) \rightarrow \begin{pmatrix} \mathbf{0} & \mathbf{V} \\ \mathbf{V}' & \mathbf{0} \end{pmatrix} , \tag{16}$$

since the reflection then works like

$$\begin{aligned}\begin{pmatrix} \mathbf{0} & \mathbf{V} \\ \mathbf{V}' & \mathbf{0} \end{pmatrix} &\rightarrow - \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{V} \\ \mathbf{V}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{pmatrix} \\ &= - \begin{pmatrix} \mathbf{0} & \mathbf{A}\mathbf{V}'\mathbf{A} \\ \mathbf{A}'\mathbf{V}\mathbf{A}' & \mathbf{0} \end{pmatrix} , \end{aligned} \tag{17}$$

with

$$\begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbf{0} \\ \mathbf{0} & \mathbb{1} \end{pmatrix} . \tag{18}$$

This then becomes equivalent to the four-dimensional description we can derive from the Dirac trick. The form of (18) is a standard form used by Cartan. It may not be the standard form a physicist is used to, but it can be argued that this is not important as Pauli has shown that all representations are equivalent up to a similarity transformation. The two-dimensional matrices can only yield a representation of the proper Lorentz transformations. The Lorentz reflections cannot be accounted for in a single two-dimensional representation, due to the fact that there is not a fourth 2×2 matrix that anticommutes with $\sigma_x, \sigma_y, \sigma_z$. In the two-dimensional representation, we will obtain a Lorentz transformation if $\det(\mathcal{L}(\mathbf{V})) = \det(\mathbf{V})$. One can show (see Appendix B) that this will be true if we make $\mathcal{L}$ correspond to a transformation $\mathbf{V} \rightarrow \mathbf{L}\mathbf{V}$, whereby $\det(\mathbf{L}) = 1$, in other words $\mathbf{L} \in SL(2, \mathbb{C})$. Therefore, the two-dimensional representation of the group of proper Lorentz transformations is linear right from the start. It does not need any further linearizing as in the four-dimensional representation, given by (17), which allows also for reflections but is quadratic in the matrix

$$\begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}' & \mathbf{0} \end{pmatrix} . \tag{19}$$

Some more remarks about the relationships between the two-dimensional representations and the four-dimen-

sional one we want to examine here are given in Appendix C.

We are now getting to the point of deriving the expressions for the spinors along the methods that we outlined for the rotation group. We will work it out for the case of the energy-momentum four-vector:

$$\begin{pmatrix} \mathbf{0} & E\mathbb{1}+c\mathbf{p}\cdot\sigma \\ E\mathbb{1}-c\mathbf{p}\cdot\sigma & \mathbf{0} \end{pmatrix}, \tag{20}$$

where we continue to use the Cartan form. The eigenvalues of this matrix are twice m_0c^2 and twice $-m_0c^2$. The corresponding eigenvectors (including a determinant-based normalization factor) can be calculated and make up the matrix $\mathbf{S}$:

$$\frac{1}{2(m_0c^2)^2}\times \begin{pmatrix} E+cp_z & cp_x+\mathrm{i}cp_y & -m_0c^2 & 0 \\ cp_x-\mathrm{i}cp_y & E-cp_z & 0 & -m_0c^2 \\ m_0c^2 & 0 & E-cp_z & -(cp_x+\mathrm{i}cp_y) \\ 0 & m_0c^2 & -(cp_x-\mathrm{i}cp_y) & E+cp_z \end{pmatrix}. \tag{21}$$

Here the first and second columns correspond to the positive eigenvalue m_0c^2, and the third and fourth column to the negative eigenvalue $-m_0c^2$. At this stage we can already move out the square roots $\sqrt{m_0c^2}$ from the diagonal matrix $\mathbf{\Lambda}$ to the left and to the right in the full expression $\mathbf{S\Lambda S}^{-1}$. This leads to a change of the determinant-based pre-factor from $\frac{1}{2(m_0c^2)^2}$ to $\frac{1}{2}$ in (21). The same applies to the matrix $\mathbf{S}^{-1}$ on the right-hand side. By carrying out this move we have renormalized the matrix $\mathbf{S}$, just like in the rotation group.

The idea behind the diagonalization procedure is to make the leap from vectors to Lorentz transformations, by taking a special vector that can code the tetrad, and then making the algebraic structure of that special vector and that of a Lorentz transformation coincide to render the isomorphism between the special vector and the Lorentz transformation conspicuous. More precisely, it is the structure of the eigenvector matrix $\mathbf{S}$ of the reflection matrix, rather than the structure of the reflection matrix itself that must coincide with the structure of the Lorentz transformation. We can e.g. appreciate from the Cartan representation that the special vector and the Lorentz transformation can never be cast into the same algebraic structure and never coincide, because the block structure of a vector is represented by (16), while the block structure of a Lorentz transformation is along the main diagonal. In this respect the Cartan representation permits us to see this point very neatly. It would be less clear in another equivalent representation where the information is scattered over all blocks. The main point is however that we have to be careful about the order of the eigenvectors in the matrix $\mathbf{S}$ if we want to have the two structures coincide, and this is what we have done. We see from (21) that we recover exactly the structure of a Lorentz transformation if we apply the same expedient as in the rotation group, viz. kicking every reference to m_0c^2 out by taking the limit $m_0c^2 \to 0$. We end up then with two block matrices $\mathbf{V}$ and $\mathbf{V}'$ whose determinant is now zero.

An important point, however, is that in the Lorentz group the vector $\mathbf{e}_1+\mathrm{i}\mathbf{e}_2+\mathbf{e}_4$ that is able to code the tetrad *is not a "zero-length" vector!* Separately taken, $\mathbf{e}_1+\mathrm{i}\mathbf{e}_2$ and $\mathbf{e}_1+\mathbf{e}_4$ are both "zero-length" vectors. But not the linear combination $\mathbf{e}_1+\mathrm{i}\mathbf{e}_2+\mathbf{e}_4$. Hence the non-"zero-length" vector $\mathbf{e}_1+\mathrm{i}\mathbf{e}_2+\mathbf{e}_4$ should suffice to code the information. But this seems to contradict the idea that we should let $m_0c^2 \to 0$ to get rid of any scale factor. Hence, in order to obtain a "zero-length" vector, we should rather take $\sqrt{2}\mathbf{e}_1+\mathrm{i}\mathbf{e}_2+\mathbf{e}_4$. Hence in the Lorentz group the concerns of kicking out the parameter m_0c^2 and of coding the tetrad do not exactly coincide as in the rotation group. For the rest of the development we can take the approach given by (41.57) in [6], which shows how we factorize the two block matrices whose determinant has become zero. We then end up with two proportional 2×1 column matrices $\eta_0[\xi_0,\xi_1]^\top$, $\eta_1[\xi_0,\xi_1]^\top$, that are also called "spinors" in the literature, and which (with the aid of their complex ratio $\eta_0:\eta_1$) code the total information of the tetrad.

In [6] there is a problem with the fact that it has not been realized that we should take $\sqrt{2}\mathbf{e}_1+\mathrm{i}\mathbf{e}_2+\mathbf{e}_4$, and that $\mathbf{e}_1+\mathrm{i}\mathbf{e}_2$ and $\mathbf{e}_1+\mathbf{e}_4$ are not orthogonal. As the $\mathrm{SL}(2,\mathbb{C})$ representation of $\mathbf{V}$ is already linear, there is no further need of linearization in the two-dimensional representations. The need to consider 2×2 matrices whose determinant becomes zero becomes only really clear by considering the four-dimensional representation.

In [6] that need for linearization is sort of *reintroduced* by asking the matrices to be Hermitian (thereby restricting the formalism to vectors; see Appendix C). Hence this is putting a constraint on $\mathrm{SL}(2,\mathbb{C})$ in order to preserve vectors, just for the whim of taking the reverse road to the complete group $\mathrm{SL}(2,\mathbb{C})$ later on when we want to recover $\mathrm{SL}(2,\mathbb{C})$ by making the leap from vectors to Lorentz transformations. Making it back to Lorentz transformations means that we must make the formalism scale invariant. We are then faced with a matrix whose determinant is zero. The way to treat such determinants proposed by [6] is based on an algebraic approach. It takes inspiration from the structure of (10) as a method to factorize a matrix that has determinant zero. One can also argue that it transposes the quadratic structure of vectors in terms of spinors that we found in the rotation group to the context of the Lorentz group, such that it can be preserved and embedded. But in a further stage it has not been realized what kind of linear combination of vectors has to be taken in order to preserve also the coding of the tetrad, and one is then forced to go into the notion of flags and flag poles to acknowledge the loss of information that follows from taking orthogonality for granted. The method to keep in line with all the ideas that emerged from the treatment in the rotation group is thus to adopt the "spinor" factorization of a zero determinant as used by [6], and to code the "zero-length" vector $\sqrt{2}\mathbf{e}_1+\mathrm{i}\mathbf{e}_2+\mathbf{e}_4$ into it. It would then remain to prove that the three vectors $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_4$ can still be decoded from it. It must then be very carefully noted that the two "zero-length" vectors that code the tetrad, viz. the isotropic vector $\mathbf{e}_1+\mathrm{i}\mathbf{e}_2$ and the light ray $\mathbf{e}_1+\mathbf{e}_4$ *are not orthogonal*. If we miss that point and treat them as orthogonal, the description becomes incomplete and we loose

information. We may note that after all this we recover the notion that vectors are rank-2 tensors in terms of "spinors" within the Lorentz group. The final treatment of the matrices whose determinant has become zero permits one to see how the set of rotational spinors is imbedded within the set of the Lorentz "spinors": it suffices to take space-like vectors with $ct = 0$ to recover the rotations. There is another, perhaps more natural, way to code the tetrad: that is to code it as the tensor product of two "zero-length" vectors. We must then consider the four-dimensional representation as a tensor product of two two-dimensional representations. We have not explored this possibility.

It must finally be noted that in using the Lorentz group in physics, e.g. when we calculate the eigenvectors of a Dirac matrix, the spinor structure is not introduced. The eigenvectors remain vectors. They keep containing the quantity m_0c^2. Hence this does not fully respect the representation theory. This is perhaps a problem and it would be interesting to investigate if it has a relation with renormalization problems.

5 Quantum mechanics

We have now understood and demonstrated one thing: a spinor can code a Lorentz transformation by coding the complete orientation of the basic tetrad. Let us turn back to rotations. They are a subgroup of the Lorentz group. We will try to check if the spin as described in the Dirac equation corresponds to a rotating particle. Let us first work out an example to show how the spinor formalism works. The Rodrigues formula,

$$\cos\frac{\theta}{2}\mathbb{I} - \mathrm{i}\mathbf{e}\cdot\sigma\sin\frac{\theta}{2}\,, \tag{22}$$

can be derived by just writing the product of two reflections and it describes a rotation over an angle θ around the axis with unit vector $\mathbf{e}$. Let us check this for a rotation around the z-axis. The reference frame is given by $\mathbf{e}_x = (1,0,0)$, $\mathbf{e}_y = (0,1,0)$ and $\mathbf{e} = \mathbf{e}_z = (0,0,1)$, such that $\mathbf{e}_x + \mathrm{i}\mathbf{e}_y = (1,\mathrm{i},0)$. Hence we have $x = 1$ and $y = \mathrm{i}$. The spinor is given by

$$\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = \begin{pmatrix}\sqrt{(x-\mathrm{i}y)/2}\\ \sqrt{(-x-\mathrm{i}y)/2}\end{pmatrix} = \begin{pmatrix}1\\ 0\end{pmatrix}. \tag{23}$$

Now $\cos\frac{\theta}{2}\mathbb{I} - \mathrm{i}\sigma_z\sin\frac{\theta}{2}$ becomes

$$\begin{pmatrix}\mathrm{e}^{-\mathrm{i}\theta/2} & 0\\ 0 & \mathrm{e}^{+\mathrm{i}\theta/2}\end{pmatrix}. \tag{24}$$

Operating on the spinor, this yields

$$\begin{pmatrix}\sqrt{(x'-\mathrm{i}y')/2}\\ \sqrt{(-x'-\mathrm{i}y')/2}\end{pmatrix} = \begin{pmatrix}\psi_1'\\ \psi_2'\end{pmatrix} = \begin{pmatrix}\mathrm{e}^{-\mathrm{i}\theta/2}\\ 0\end{pmatrix}, \tag{25}$$

such that $(x'-\mathrm{i}y')/2 = \mathrm{e}^{-\mathrm{i}\theta}$ and $(-x'-\mathrm{i}y')/2 = 0$. From this we derive $x' = \mathrm{e}^{-\mathrm{i}\theta}$ and $y' = \mathrm{i}\mathrm{e}^{-\mathrm{i}\theta}$ and $x_1' = \cos\theta$, $x_2' = -\sin\theta$, $y_1' = \sin\theta$, $y_2' = \cos\theta$, such that $\mathbf{e}_x' = (\cos\theta, \sin\theta)$ and $\mathbf{e}_y' = (-\sin\theta, \cos\theta)$, which corresponds indeed to a rotation over an angle θ around the z-axis.

Let us now code a rotation with angular velocity ω around the axis with unit vector $\mathbf{e}$. According to the Rodrigues formula this must be

$$\cos\frac{\omega t}{2}\mathbb{I} - \mathrm{i}\mathbf{e}\cdot\sigma\sin\frac{\omega t}{2}\,, \tag{26}$$

such that for the corresponding spinors we must have

$$\Psi(t) = \left[\cos\frac{\omega t}{2}\mathbb{I} - \mathrm{i}\mathbf{e}\cdot\sigma\sin\frac{\omega t}{2}\right]\Psi(0)\,. \tag{27}$$

Of course, the Rodrigues formula is not directly intuitive, but the reader can replace θ by ωt in the example worked out to convince himself of the fact that this represents a rotating frame and to "see" the frame turning "with his own eyes". This leads to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Psi(t) &= \frac{\omega}{2}\left[-\sin\frac{\omega t}{2}\mathbb{I} - \mathrm{i}\mathbf{e}\cdot\sigma\cos\frac{\omega t}{2}\right]\Psi(0)\\ &= \frac{\omega}{2}\left[\cos\frac{(\omega t+\pi)}{2}\mathbb{I} - \mathrm{i}\mathbf{e}\cdot\sigma\sin\frac{(\omega t+\pi)}{2}\right]\Psi(0)\,,\end{aligned} \tag{28}$$

such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Psi(t) = -\mathrm{i}\mathbf{e}\cdot\sigma\frac{\omega}{2}\Psi(t)\,. \tag{29}$$

One can also derive

$$\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\Psi = -\frac{\omega^2}{4}\Psi\,. \tag{30}$$

By putting $\hbar\omega/2 = m_0c^2$ (which we could interpret loosely by stating that the whole rest energy of the electron is due to its rotation with angular momentum $h/2$) we obtain

$$\frac{1}{c^2}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\Psi \equiv \left[\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}\right]\Psi = -\frac{m_0^2c^2}{\hbar^2}\Psi\,, \tag{31}$$

which is the Klein–Gordon equation. From this it becomes obvious that the Dirac operator

$$\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}\tau} \to \sum\gamma_\mu\frac{\partial}{\partial x_\mu} \tag{32}$$

expresses the derivation with respect to proper time in the representation of the Lorentz group spanned by the Dirac matrices. When we use this and $\hbar\omega/2 = m_0c^2$ on (29) we obtain

$$\sum\gamma_\mu\frac{\partial}{\partial x_\mu}\Psi = -\mathrm{i}\mathbf{e}\cdot\sigma\frac{m_0c}{\hbar}\Psi\,, \tag{33}$$

which is, except for the factor $\mathbf{e}\cdot\sigma$, the Dirac equation for a free electron. On our example of the rotation around the z-axis we can check that $\mathbf{e}\cdot\sigma\Psi(t) = \Psi(t)$. But this is not general. Without it, it is not obvious how to derive the

Dirac equation, and unfortunately it is not generally valid. Let us however decompose the *spinor* $\Psi(t)$ in the *vector basis* of the eigenvectors of $\mathbf{e}\cdot\sigma$. We must emphasize that this is a strange and hybrid construction, which has no directly obvious geometrical meaning. It is only a calculation expedient. As the matrix $\mathbf{e}\cdot\sigma$ is Hermitian, its eigenvalues are real (viz. 1 and -1) and its eigenvectors are orthogonal. (To get some intuition for what happens, the reader can assume without loss of generality that $\mathbf{e}=\mathbf{e}_z$.) In other words, Ψ splits up into two parts Ψ_+ and Ψ_-. These two quantities are not spinors, they are vector projections of spinors. But Ψ_+ is a solution of $\frac{d}{dt}\Psi_+(t)=-\mathrm{i}\frac{\omega}{2}\Psi_+(t)$ and therefore of the Dirac equation:

$$\sum\gamma_\mu\frac{\partial}{\partial x_\mu}\Psi_+=-\mathrm{i}\frac{m_0c}{\hbar}\Psi_+\,, \tag{34}$$

while Ψ_- is a solution of $\frac{\mathrm{d}}{\mathrm{d}t}\Psi_-(t)=+\mathrm{i}\frac{\omega}{2}\Psi_-(t)$ and therefore of another Dirac-like equation with a reversed sign:

$$\sum\gamma_\mu\frac{\partial}{\partial x_\mu}\Psi_-=+\mathrm{i}\frac{m_0c}{\hbar}\Psi_-\,. \tag{35}$$

As, however, (34) and (35) have the same solutions, it follows ultimately that Ψ as a whole is a solution of the Dirac equation. We thus obtain the Dirac equation and we have found that quantum mechanics describes a rotating particle by attaching a reference frame to it and by treating this co-rotating frame with the aid of group representation theory. The equation just expresses (26) using the derivative $\frac{1}{c}\frac{\mathrm{d}}{\mathrm{d}\tau}\rightarrow\sum_\mu\gamma_\mu\frac{\partial}{\partial x_\mu}$ with respect to the proper time of the co-moving frame. As the derivative is vector-like it is entered as a reflection operator. We can thus see the Dirac equation as treating a spinning particle subjected to Lorentz transformations. We first treat the spinning particle in its rest frame, and then use Lorentz invariance to express this in any other moving frame. The pure state solutions of the Dirac equation and the original Rodrigues equation are not equivalent. As the Dirac equation has just been guessed, we may have to rely on the Rodrigues equation for matters of deciding what a pure state is. As the Schrödinger equation can be derived from the Dirac equation, we have derived a lot about quantum mechanics from the simple *Ansatz* of a rotating particle. The only point that we have not justified is the gimmick of replacing ω according to $\hbar\omega/2=m_0c^2$. In some way the mystery of quantum mechanics must lie hidden behind this formula, which is analogous to the equation $E=\hbar\omega$ of Planck and Einstein. It must be noted that the Dirac equation does a fine job but does not reproduce the experimental results exactly. It must therefore be wrong in some small details.

What changes the rate ω at which the electron spins for another observer is the translational velocity of the electron. The velocity four-vector can be obtained from the kinetic energy. This in turn can be obtained by subtracting the potential energy from the total energy. Hence the velocity four-vector is given by $\frac{1}{m_0c}(c\mathbf{p}-e\mathbf{A},E-eV)$, and this is what explains the use of $\mathbf{p}\rightarrow\mathbf{p}-e\mathbf{A}/c$, $E\rightarrow E-eV$. The corresponding operators are just covariant derivatives. But there is a problem here: the quantities $\sum\gamma_\mu\frac{\partial}{\partial x_\mu}$ and m_0 are quantities from the rest frame of the electron, while $(eV,e\mathbf{A}/c)$ are expressed in the frame of the observer. In reality, in the rest frame of the electron one sees a fluctuating potential due to its motion in what is a stationary $1/r$ potential in the *observer's frame*. Hence the Dirac equation with a potential is not exact and only an approximation. This is a first problem with the Dirac equation.

Note that in a non-flat potential the clock will not only change its rate, also the orientation of the rotation axis will change and exhibit Thomas precession. The time dilatation and the orientation of the rotation axis do not exhaust the complete definition of the tetrad, but to know the orientation and the total accumulated time dilatation we have to know the whole traveling history of the spinning electron. If we do not know the whole history we may run into ambiguity due to the twin paradox: two twins that come together after having made different journeys have aged differently. If their clock readings (and more generally their tetrads) do not coincide that would then mean that the spinor (or "wave function") is not single-valued! It may well be that we will only be able to avoid this if the electron follows a well-defined orbit.

The Dirac equation does not address why the accelerated motion of the electron does not cause radiation losses. Just like in the early Bohr and Bohr–Sommerfeld models, that problem is just ignored.

It must be noted that $(x,y,z,ct,\tau)\in\mathbb{R}^5$ in (31) and (33), and which describe the trajectory of the electron, have nothing to do with the spinor quantities $(x,y,z,ct)\in\mathbb{C}^4$ in (14), (15), ..., which describe the orientation of the tetrad. In (23)–(25) the numbers $(x,y,z)\in\mathbb{C}^3$ describe the internal rotation of the electron with the aid of complex numbers coding the triad, and using $c\tau\in\mathbb{R}$ as a real evolution parameter of the triad that is conceptually not subject to translational motion. Let us therefore temporarily rename these three complex numbers as $(\xi,\eta,\zeta)\in\mathbb{C}^3$ in order to distinguish them clearly from $(x,y,z)\in\mathbb{R}^3$ and to avoid ambiguity. The numbers $(\xi,\eta,\zeta)\in\mathbb{C}^3$ describe the tetrad, the numbers $(x,y,z,ct,\tau)\in\mathbb{R}^5$ describe the trajectory. In (23)–(25) only $c\tau$ and ct have the same meaning as in (31) and (33). As such they should affect, in principle, the numbers $(\xi,\eta,\zeta)\in\mathbb{C}^3$ in the triad making them subject to time dilatation and Thomas precession. But the translation–tetrad coupling between the particle coordinates $(x,y,z)\in\mathbb{R}^3$ and the tetrad coordinates $(\xi,\eta,\zeta)\in\mathbb{C}^3$ that should allow one to describe the Thomas precession is not correctly introduced into the formulation. The reason for this is as follows: most unfortunately the algebraic structure of the calculations on $(\xi,\eta,\zeta)\in\mathbb{C}^3$ can be carried over to $\mathbb{R}^3$ where (ξ,η,ζ) also become noted as (x,y,z) and interpreted as real particle coordinates. It is thereby not noted that this algebraic structure can only get a meaningful interpretation when (x,y,z) are understood as truncated extrapolations to $\mathbb{R}^3$ of $(\xi,\eta,\zeta)\in\mathbb{C}^3$. When (x,y,z) are taken to be forced real (because they are interpreted as particle coordinates) it becomes impossible to code correctly into the algebraic structure all the changes of the tetrad, for which one needs six real coding parameters ($(\xi,\eta,\zeta)\in\mathbb{C}^3$), rather than three ($(x,y,z)\in\mathbb{R}^3$)! This is a second problem with the Dirac equation.

The rotation described by the Rodrigues formula treats the rotation non-relativistically, in the sense that it con-

siders a rotation purely as a transformation in three-dimensional space, while this is no longer true when we go to space-time. The rotational spinors are embedded into the more general formalism of Lorentz spinors, and for a problem in a rest frame (like we started from) they must coincide. Hence that part of our derivation is correct. However, in space-time, rotational states can be changed due to boosts, because the rotations are a subgroup of the Lorentz transformations, and because the rotational motion around an axis is only a part or aspect of a complete Lorentz transformation from which it cannot be separated, if we want to carry through the calculations correctly. As long as we stick to a single boost the description is still correct, but when we go to compositions of boosts, the Rodrigues description will fail. When the motion is not uniform, the quantity $[\mathbf{e}\cdot\sigma]$ will become time-dependent in the observer's frame, and our way of proceeding, by decomposing Ψ into Ψ_+ and Ψ_-, will no longer be correct. In this respect, the Dirac equation is an approximation. If we follow Einstein's principle of equivalence, in the case of gravitation or in the case of accelerated motion without potential, $[\mathbf{e}\cdot\sigma]$ remains a constant for an observer within an accelerated frame, because he is on a geodesic line and feels no forces. In the co-moving frame $\frac{\mathrm{d}}{\mathrm{d}\tau}[\mathbf{e}\cdot\sigma]=0$. The resulting correction for $[\mathbf{e}\cdot\sigma]$ would still remain a simple expression. But this cannot be claimed when there is a fluctuating electromagnetic potential in the co-moving frame. There will thus not only exist a spin–orbit coupling, but also a spin–potential coupling. Due to the fluctuating electromagnetic potential, $\frac{\mathrm{d}}{\mathrm{d}\tau}[\mathbf{e}\cdot\sigma]$ will no longer be zero (*even in the co-moving frame!*), and it must enter into the calculations at the level of the time derivative taken in (28) and (29). Even the mass of the electron in the co-moving frame will no longer be m_0.

The correct wave function is by definition $\psi(\xi,\eta,\zeta,\sigma)$ (where σ plays the rôle of the complex parameter ct in (14), (15), ...). This codes the orientation of the tetrad. But this is no longer only a function of the instant particle coordinates x,y,z,ct but also of the particle history due to the twin paradox. And to perform the calculations on ξ,η,ζ,σ correctly, one will need the instant velocity parameters β_x, β_y, β_z and γ. In principle, one does not need to calculate *all possible* histories. The actual history should suffice. But if one does not know the actual history, one ends up in being forced to calculate them all. It is necessary to follow the whole tetrad since otherwise one cannot calculate the time dilatation correctly. Getting the time dilatation wrong means getting ω wrong, and this will show up directly in the calculated energies $\hbar\omega$. In the Dirac equation the time dilatation is only treated correctly up to a certain approximation.

For a non-rotating particle, the total energy splits up into kinetic energy and potential energy. From the theory of relativity we learned that the mass of a particle changes as it moves. But there was no mechanism given for it. It just comes out of the formulas for the Lorentz transformation of the energy-momentum four-vector. Quantum mechanics might propose us an explanation for this phenomenon. With a rotating particle, the total energy is split into three parts: rotational energy, kinetic energy and potential energy. When a particle has no translational motion its rotational energy has a fixed value, which contributes to the total energy m_0c^2 of the electron at rest. When the particle moves, energy is stored into its rotational motion and we see therefore its mass increase. When there is more energy in the rotational motion, it means that the angular frequency ω should be higher. We can evaluate how much energy will go into a rotational motion with frequency ω and see how the energy and the frequency are related. In fact, a Lorentz transformation leaves the quantity $Et-\mathbf{r}\cdot\mathbf{p}$ invariant. It has the dimension of angular momentum. On the other hand it also leaves the quantity $\omega t-\mathbf{k}\cdot\mathbf{r}$ invariant. This is a dimensionless quantity. If $Et-\mathbf{r}\cdot\mathbf{p}$ is related to the angular momentum of the rotation of a particle, the ratio $(Et-\mathbf{r}\cdot\mathbf{p})/(\omega t-\mathbf{k}\cdot\mathbf{r})$ must remain constant. From this we infer that the energy mc^2 of the rotating particle with angular momentum $h/2$ must be $\hbar\omega/2$. It could be argued that we must be cautious and consider that the equation $m_0c^2=\hbar\omega/2$ is an extrapolation to the whole energy scale of something that in reality is only valid for changes of rotational energy: that is, it would be more correct to write $\Delta E=(m-m_0)c^2=\hbar\Delta\omega/2$. In fact, we can imagine that m_0c^2 could contain other energies than just rotational energy, e.g. energies related to its charge. But this caveat does not seem to be pertinent as in the derivation of the Klein–Gordon and Dirac equations the total substitutions $m_0c^2=\hbar\omega/2$ do occur. This seems to imply that the whole rest energy of the electron is rotational energy. It should be noted, however, that this whole reasoning is purely phenomenological. A true explanation requires that one develops a dynamical model for the electron that allows one to calculate its total energy.

We pointed out the problem of the distinction between (x,y,z,ct) and (ξ,η,ζ). If one does not realize all these fundamental distinctions one can become quite bewildered with the "unphysical values" (x,y,z,ct) in (14), (15), ... (where conceptually $(x,y,z,ct)\in\mathbb{C}^4$) might take in $\mathbb{R}^4$ when one believes that they are defined as $(x,y,z,ct)\in\mathbb{R}^4$ and thereby unknowingly extrapolates the formalism from the light cone to $\mathbb{R}^4$. Especially for the wave function $\psi(x,y,z,ct)$ it is not realized that (x,y,z,ct) can have a meaning totally different from the coordinates of a particle (viz. when it codes the tetrad), and that there are two movies rather than one. Similarly, in his treatment of the spinors in the rotation group Cartan [2] jumps backwards and forwards between $(x,y,z)\in\mathbb{C}^3$ when they are on the isotropic cone and $(x,y,z)\in\mathbb{R}^3$ when they describe the coordinates of a real vector without any warning or mention. This is a real headache for the non-adverted reader.

There is another very delicate point we want to mention. It is often stated that the linearity of the Dirac equation warrants the superposition principle. This is however treating spinors like vectors. Two spinors do not necessarily add up to a spinor, because they belong to a manifold (the isotropic cone or the light cone) rather than to a vector space. The sum of two spinors must belong to the group ring rather than to the group. However, if we extrapolate the formalism to $(x,y,z,ct)\in\mathbb{R}^4$, we drop the constraint that (x,y,z,ct) are points of the isotropic cone. Two spinor "remnants" $\psi_1(x,y,z,ct)$ and $\psi_2(x,y,z,ct)$ can then be combined to $a_1\psi_1+a_2\psi_2$ like in a vector space. Moreover,

we know that in $\mathbb{R}^4$ we get into the "second movie" and into the realm of probability calculus. Hence the linearity of the Dirac equation is only valid for problems of probability calculus. The linear combinations correspond then to so-called mixed states, which follow the well-known rule that $|a_i|^2$ represents the probability to measure ψ_i. But such linear combinations do not have any meaning on the isotropic cone for the "first movie". Spinors cannot be added like vectors. However, in the derivation of the Dirac equation we have shown how in certain cases this can be circumvented, by restricting the superposition principle to mutually orthogonal states. Perhaps, we are touching here the essence of quantum mechanics. The states of nature are spinor states. But we are used to think about nature in terms of vectors or four-vectors. In an experiment we try to measure vector components, while it is only the spinor state that exists. We get then some value back for the presumed vector component with some probability, which can be calculated from the state. Quantum mechanics is about the problem that in our experiments we treat spinors like vectors. This is a conclusion that can only be reached by having been extremely meticulous about the meaning of the mathematics, i.e. by sticking to the program we announced in Sect. 1.

It may shock that we obtain the Klein–Gordon equation and the Dirac equation starting from the same *Ansatz* for the rotating frame, as the Klein–Gordon equation is known to apply to particles of spin 0, while the Dirac equation applies to particles of spin $\frac{1}{2}$. Note, however, that the Klein–Gordon equation can be obtained by applying the Dirac equation to itself. But from the calculations it looks more as though the "spin" would be a compound quantity. In our derivation the particle always has the same true angular momentum related to the internal rotation, with a value of $h/2$. But this is multiplied with some weighting number, which is rather related to some kind of algebra with half the rank of the tensor that intervenes (a spinor being of rank 1, while it corresponds to a "polynomial" of degree $\frac{1}{2}$) than the property of a particle. We may note in this respect something funny: if ψ_1 and ψ_2 are both solutions of $\frac{\mathrm{d}\psi}{\mathrm{d}t} \propto \psi$, then $\frac{\mathrm{d}\psi_1\psi_2}{\mathrm{d}t} \propto 2\psi_1 \otimes \psi_2, \ldots$ Also, the matrix $[\mathbf{e}\cdot\sigma]$ has some properties that remind us of the angular momentum operator J_z, but it is $\frac{\mathrm{d}}{\mathrm{d}t}$ that projects out the number $\frac{1}{2}$, not the matrix $\mathbf{e}\cdot\sigma$. In the same logic we can get a feeling for the fact that $\frac{\mathrm{d}^2}{\mathrm{d}t^2}$ would yield "spin zero". The "angular momentum" states in the hydrogen atom also seem to fit into this scheme. It is not then the angular momentum that is quantized but the degree of the harmonic polynomials that occur within the representation that occurs in the calculation. It looks much less mysterious that the degree of a polynomial is a quantized quantity.

Let us now conclude with a final remark. In the very beginning of wave mechanics it was noted that the phase velocity of a de Broglie wave is larger than c. We can understand this now. The electron spin is like a clock. The Dirac equation is thus describing the proper time as read on this clock. When we extrapolate the algebra from the light cone to $\mathbb{R}^4$ this will remain valid, be it in a weird form interpreted as a "matter wave", describing (in the most elementary simplified form) the probability density, which is subject to a continuity equation. This continuity equation expresses that the probability density of an electron (which behaves completely analogously to a charge density) is time invariant in a reference frame wherein the electron is translationally at rest. The wave equation therefore expresses in spinor form translational invariance for the current-density vector with respect to time translations, just like Bloch waves in a crystal express translational invariance with respect to space translations. But that means also that the de Broglie wave is a time wave, rather than a wave in space. The slope of a time axis in Minkowski space-time is c^2/v, which is exactly the speed of the de Broglie wave, which confirms the idea that the de Broglie wave should be understood as a time wave rather than a wave "propagating" in space. There does thus not reside any conflict with the theory of relativity in this superluminal velocity c^2/v of the de Broglie wave.

Appendix A: Unambiguity of the coding of the tetrad

Let us denote $\mathbf{r}_j = (x_j, y_j, z_j, ct_j)$. We must reconstruct all these quantities from knowledge of $\mathbf{r}_2$ and $\mathbf{r}_1 + \mathbf{r}_4$. By construction, we know that there is a solution to our problem, but only the quantities $(u_x, u_y, u_z, u_t) = (x_1 + x_4, y_1 + y_4, z_1 + z_4, ct_1 + ct_4)$, and (x_2, y_2, z_2, ct_2), with $u_x x_2 + u_y y_2 + u_z z_2 - u_t ct_2 = 0$ are given. Let us nevertheless show that there is no second solution that can be derived from these quantities. As the Lorentz transformation conserves the Minkowski pseudo-metric we must have $\mathbf{r}_j \cdot \mathbf{r}_k = \mathbf{e}_j \cdot \mathbf{e}_k = g_{jk}$, where g_{jk} are the elements of the metric tensor. From this the following equations can be derived immediately:

$$\begin{aligned} [1]: &\quad \mathbf{r}_1\cdot\mathbf{r}_1 - \mathbf{r}_4\cdot\mathbf{r}_4 = 2\,, \\ [2]: &\quad \mathbf{r}_1\cdot\mathbf{r}_1 + \mathbf{r}_4\cdot\mathbf{r}_4 = 0\,, \\ [3]: &\quad \mathbf{r}_1\cdot\mathbf{r}_2 = \mathbf{r}_4\cdot\mathbf{r}_2 = 0\,, \end{aligned}$$

where the metric tensor intervenes in the definition of the dot product. Let us search for a second solution. For this purpose imagine that we decompose $\mathbf{u} = (u_x, u_y, u_z, u_t)$ differently into $\mathbf{r}_1 + \mathbf{w} = (x_1 + w_x, y_1 + w_y, z_1 + w_z, ct_1 + w_t)$, and $\mathbf{r}_4 - \mathbf{w} = (x_4 - w_x, y_4 - w_y, z_4 - w_z, ct_4 - w_t)$. Then $\mathbf{r}_1 + \mathbf{w}$ must satisfy all the conditions fulfilled by $\mathbf{r}_1$, and $\mathbf{r}_4 - \mathbf{w}$ all the conditions fulfilled by $\mathbf{r}_4$ expressed in [1]–[3]. From expressing this idea for [1] we obtain

$$\begin{aligned} \mathbf{w}\cdot\mathbf{u} = {} & w_x(x_1 + x_4) + w_y(y_1 + y_4) + w_z(z_1 + z_4) \\ & - w_t(ct_1 + ct_4) = 0\,, \end{aligned} \tag{A.1}$$

while from [2] we obtain

$$\begin{aligned} \mathbf{w}\cdot(\mathbf{r}_1 - \mathbf{r}_4) + \mathbf{w}\cdot\mathbf{w} = {} & w_x(x_1 - x_4) + w_y(y_1 - y_4) \\ & + w_z(z_1 - z_4) - w_t(ct_1 - ct_4) \\ & + w_x^2 + w_y^2 + w_z^2 - w_t^2 = 0\,. \end{aligned} \tag{A.2}$$

From [3] we can derive

$$\mathbf{w}\cdot\mathbf{r}_2 = w_x x_2 + w_y y_2 + w_z z_2 - w_t ct_2 = 0\,. \tag{A.3}$$

Expressing that also $\mathbf{r}_1 - \mathbf{r}_4$ must be a "zero-length" vector, because it is the image of the "zero-length" vector $\mathbf{e}_1 - \mathbf{e}_4$, yields (A.2) again. Expressing $(\mathbf{r}_1 - \mathbf{r}_4)\cdot(\mathbf{r}_1 + \mathbf{r}_4) = 2$, because it is the image of $(\mathbf{e}_1 - \mathbf{e}_4)\cdot(\mathbf{e}_1 + \mathbf{e}_4) = 2$, yields

$$\mathbf{w}\cdot\mathbf{w} = w_x^2 + w_y^2 + w_z^2 - w_t^2 = 0\,, \tag{A.4}$$

which combined with (A.2) yields

$$\mathbf{w}\cdot(\mathbf{r}_1 - \mathbf{r}_4) = w_x(x_1 - x_4) + w_y(y_1 - y_4) + w_z(z_1 - z_4) - w_t(ct_1 - ct_4) = 0\,. \tag{A.5}$$

Call the components of the vector w_x, w_y, w_z, w_t in the basis $\mathbf{r}_j$: a_1, a_2, a_3, a_4. Then we find from (A.1) and (A.3)–(A.5)): $a_1 = 0$, $a_2 = 0$, $a_3 = 0$ and $a_4 = 0$, which implies that there is no second solution.

Let us now briefly indicate the path towards the solution. Call the vector $\mathbf{r}_1 - \mathbf{r}_4$ we have to find $\mathbf{d}$. We know that $\mathbf{r}_2$ is a unit vector. Take an orthonormal basis of space-time $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4$, such that $\mathbf{q}_2 = \mathbf{r}_2$. There are many possibilities to do this but any particular choice will do. As $\mathbf{u}\cdot\mathbf{q}_2 = 0$ and $\mathbf{d}\cdot\mathbf{q}_2 = 0$, we have $\mathbf{u} = u_1\mathbf{q}_1 + u_3\mathbf{q}_3 + u_4\mathbf{q}_4$ and $\mathbf{d} = d_1\mathbf{q}_1 + d_3\mathbf{q}_3 + d_4\mathbf{q}_4$. From $\mathbf{u}\cdot\mathbf{u} = 0$ we get $u_1^2 + u_3^2 - u_4^2 = 0$. From $\mathbf{d}\cdot\mathbf{d} = 0$ we get $d_1^2 + d_3^2 - d_4^2 = 0$. We therefore put $u_1 = u_4\cos\varphi$, $u_3 = u_4\sin\varphi$, $d_1 = d_4\cos\psi$ and $d_3 = d_4\sin\psi$. Note that these formulations are equivalent to the original ones, and that φ and ψ must be real. Then $\mathbf{d}\cdot\mathbf{u} = 2$ leads to $u_4 d_4(\cos\varphi\cos\psi + \sin\varphi\sin\psi - 1) = 2$, or $d_4 u_4 \sin^2\frac{\psi-\varphi}{2} = -1$. For each possible real value of ψ there is a value of d_4 that goes with it. Which combination (d_4, ψ) is the good one depends on the circumstance that $(\mathbf{u}+\mathbf{d})/2$ and $(\mathbf{u}-\mathbf{d})/2$ must be unit vectors. That is, $(u_4\cos\varphi, u_4\sin\varphi, u_4) - \frac{2}{(1-\cos(\psi-\varphi))u_4}(\cos\psi, \sin\psi, 1)$ must square to 4, while $(u_4\cos\varphi, u_4\sin\varphi, u_4) + \frac{2}{(1-\cos(\psi-\varphi))u_4}(\cos\psi, \sin\psi, 1)$ must square to -4. The first of these conditions leads to an equation in ψ or $\psi - \varphi$, with all other parameters (viz. u_4 and φ) known. After solving it for ψ we can determine d_4. That raises the question if the solution that one could derive analogously from the alternative second condition will be consistent with the one derived from the first condition. But as we already know that there is a solution and that it is unique, this must be indeed the case and we do not need to discuss this problem further.

Appendix B: How the group SL(2, ℂ) comes about

Let us denote the matrices $\mathbf{V}$ and $\mathbf{V}'$ from (14) and (15)

$$\mathbf{V} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{V}' = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{B.1}$$

such that $\mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = (ad - bc)\mathbb{1} = \det(\mathbf{V})\mathbb{1}$. Here $\mathbf{V}'$ is derived from $\mathbf{V}$ by taking its minors. Let us now write a Lorentz transformation

$$\mathbf{X} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \tag{B.2}$$

If $\mathbf{X}$ is a Lorentz transformation, then $\mathbf{XV}$ must have the same properties as $\mathbf{V}$. For example, we must have $(\mathbf{XV})'(\mathbf{XV}) = (ad - bc)\mathbb{1} = \det(\mathbf{V})\mathbb{1}$. The calculation of $\mathbf{XV}$, $(\mathbf{XV})'$ and $(\mathbf{XV})'(\mathbf{XV})$ is straightforward. The elements $[(\mathbf{XV})'(\mathbf{XV})]_{11} = [(\mathbf{XV})'(\mathbf{XV})]_{22} = (AD - BC)(ad - bc)$ lead both to the condition $AD - BC = 1$. The elements $[(\mathbf{XV})'(\mathbf{XV})]_{12}$ and $[(\mathbf{XV})'(\mathbf{XV})]_{21}$ are zero, in conformity with $[\mathbf{V}'\mathbf{V}]_{12} = [\mathbf{V}'\mathbf{V}]_{21} = 0$ and therefore do not lead to any supplementary condition. From this it follows that $\mathbf{X} \in \mathrm{SL}(2, \mathbb{C})$. We note also that $(\mathbf{XV})' = \mathbf{V}'\mathbf{X}'$.

Appendix C: Some remarks about the two-dimensional representations of the Lorentz group

The two-dimensional approach based on $\mathrm{SL}(2, \mathbb{C})$ is completely different from the construction: $\mathbf{V} \to \mathbf{L}^\dagger\mathbf{V}\mathbf{L}$, with $\det(\mathbf{L}) = 1$, as postulated by Misner et al. The inconvenience in the two-dimensional would-be linear representation based on $\mathrm{SL}(2, \mathbb{C})$ is, however, that $\mathbf{V}$ still codes a vector, rather than a rotation. Multiplying $\mathbf{V}$ with $\mathbf{L} \in \mathrm{SL}(2, \mathbb{C})$ may therefore very well yield a result that no longer respects the initial structure of $\mathbf{V}$, e.g. it is quite possible that the elements on the main diagonal are no longer real. As we have learned that a Lorentz transformation $\mathcal{L}$ corresponds to a matrix $\mathbf{L} \in \mathrm{SL}(2, \mathbb{C})$, and since $\mathrm{SL}(2, \mathbb{C})$ is a group, we must find a way to replace $\mathbf{V}$ by some element of $\mathrm{SL}(2, \mathbb{C})$ to make the description complete and self-consistent. We may note that a Lorentz transformation, as defined by a tetrad, depends on six independent real variables. Similarly, an element of $\mathrm{SL}(2, \mathbb{C})$ depends on six independent real variables. This is the path towards a two-dimensional representation followed by Hladik (see his Eqs. 6.2.1–6.2.6), who immediately postulates the group $\mathrm{SL}(2, \mathbb{C})$, without explaining how it can be derived from the vector image given by (14) or (15). (It is in turn, more logical to derive (14) or (15) from (17), which one can derive from the Dirac trick.) We may note that this genuine representation is different from what one might derive from the reflection-based pseudo-representation $\mathbf{V} \to \mathbf{B}\mathbf{A}'\mathbf{V}\mathbf{A}'\mathbf{B}$, which serves to construct the four-dimensional representation, and may have inspired the approach $\mathbf{V} \to \mathbf{L}^\dagger\mathbf{V}\mathbf{L}$ proposed by Misner et al. This approach by Misner et al. is rather vector-oriented in the sense that the combination of $\mathbf{L}^\dagger$ and $\mathbf{L}$ is intended to keep the result real, such that it can continue to be interpreted as a four-vector (see their Eqs. 41.23–41.24). The step towards Lorentz transformations is then made in a further stage. These are thus two very different ways to render a formalism self-consistent, so as to move from vector images to group element images. The approach based on $\mathrm{SL}(2, \mathbb{C})$ intends to achieve self-consistency by dropping at once any reference to vectors. The approach by Misner et al. first seeks for self-consistency by preserving the structure of the four-vectors, postponing the leap towards Lorentz transformations to an ulterior stage. This documents the relationships between the various formalisms one may find in the literature.

References

1. G. Coddens, Eur. J. Phys. **23**, 549 (2002), but of course the following references are much more important
2. E. Cartan, The Theory of Spinors (Dover, New York, 1981)
3. J. Hladik, Les Spineurs en Physique (Masson, Paris, 1996)
4. R. Penrose, W. Rindler, Spinors and Space-Time, Vol. I, Two-spinor Calculus and Relativistic Fields (Cambridge University Press, Cambridge, 1984)
5. V. Smirnov, Cours de Mathémathiques Supérieures, Vol. 2 and 3 (Mir, Moscow, 1972)
6. C.W. Misner, K.S. Thorne, J.A. Wheeler, Gravitation (Freeman, San Francisco, 1970)
7. J.F. Cornwell, Group Theory in Physics (Academic Press, Londen, 1984), three volumes
8. S. Sternberg, Group Theory in Physics (Cambridge University Press, Cambridge, 1994)
9. H.F. Jones, Groups, Representations and Physics (Adam Hilger, Bristol, 1990)
10. M. Chaichian, R. Hagedorn, Symmetries in Quantum Mechanics, From Angular Momentum to Supersymmetry (IOP, Bristol, 1998)
11. T. Inui, Y. Tanabe, Y. Onodera, Group Theory and its Applications in Physics (Springer, Heidelberg, 1990)