Scott Corry

Symmetry and Quantum Mechanics

To Sebastian

Contents

Author Biography

Scott Corry obtained his Ph.D. in mathematics from The University of Pennsylvania in 2007 after earning a B.A. in mathematics from Reed College in 2001. He joined the faculty of Lawrence University in 2007 and was promoted to Associated Professor of Mathematics in 2013. Originally working in the areas of Galois theory and p-adic algebraic geometry, his interests have expanded to encompass a broad range of topics, including combinatorics and mathematical physics. The unifying theme in all of his work is a focus on symmetry.

Preface

This book began in summer 2011 as an attempt to understand the mathematical framework underlying quantum mechanics. Although I have long been interested in physics, the immediate impetus for writing came from a reading project with my friend Doug Martin in the Physics Department at Lawrence University. In effect, this text is a first course in quantum mechanics from the mathematical point of view, emphasizing the role of symmetry, and inspired by J.S. Townsend's *A Modern Approach to Quantum Mechanics* [22].

The history of mathematics is deeply entwined with the history of physics, and the two subjects continue to influence each other in dramatic and inspiring ways. Nevertheless, it is an unfortunate fact that physicists and mathematicians often speak past each other and sometimes fail to appreciate the value of each others' concerns.[1] A physicist colleague once sent me a small poster for my office inscribed with the pithy phrase: "Mathematics: Physics without Purpose." I can imagine some of my mathematical colleagues retorting with "Physics: Mathematics without Rigor." While these taunts can be great fun, they do not help bridge the gap between two powerful worldviews. Mathematicians decry physicists' desire to choose coordinates, and we tell them about the beauty of abstract objects, generally defined as sets with further structure. Physicists often don't see the point of these abstractions, and in any case have little intuition for working with them. Instead, they are delighted with a "debauch of indices" (E. Cartan) and are quite skilled at computing, which is their real aim. After all, a physical theory is only justified qua physical theory by its agreement with experiment.

This text takes a middle road, and is loosely structured as a conversation between M(athematician) and P(hysicist). Starting with some basic physical intuitions and experimental results, M and P set out to make a model of the physical world. M introduces abstract mathematical objects, but she always motivates them with reference to experiment and appeals to simplicity. In this way, I hope that physicists already comfortable with the computations of quantum mechanics will gain an appreciation for the natural way in which these abstract objects arise. In response to these abstractions, P tends to

[1] Of course, there are mathematicians who work in the area of mathematical physics, and physicists who identify primarily as mathematical physicists. These two communities presumably understand each other reasonably well, and I am not thinking of them. Instead, I refer to the majority of working mathematicians (who may not know much about modern physics) and the bulk of experimental physicists (who may not know much about modern mathematics).

choose coordinates, but M is careful to make him account for all the *other* choices he could have made. P's instinct is to say: "of course I could have chosen differently, but then I would just need to do some bookkeeping to translate between the resulting computations." But M insists that they study the particular *structure* of the collection of possible choices at each stage. In general, this is a group structure, and M and P are naturally led to build a model of the physical world based on group representations. Remarkably, much of the mathematical structure of quantum mechanics falls out from this procedure, giving it an aura of inevitability and extreme beauty. In this way, I hope that mathematicians already comfortable with Lie groups and their representations will gain an appreciation for quantum mechanics and its myriad connections to pure mathematics. Of course, the main audience for this book is the advanced undergraduate or beginning graduate student whose understanding of both physics and mathematics is just beginning to grow.

Indeed, the student I have in mind will have taken courses in multivariable calculus, linear algebra, abstract algebra, real analysis, and perhaps topology. But she may not have seen any truly rich connections between these various subjects, and in any case would benefit from an opportunity to review them in a new context, where she will gain exposure to some graduate-level topics: smooth manifolds, group representations, Lie algebras, and Hilbert spaces. My aim is to introduce these new topics in a natural way, as an outgrowth of a compelling physical and mathematical exploration. Moreover, an introduction is all that I intend, leaving a deeper and fuller account to other texts and future courses. Especially with regard to the analytic subtleties that arise in the context of self-adjoint operators on Hilbert spaces, I am content to raise awareness of the difficulties while avoiding getting bogged down in the details. Students will be better able to comprehend a graduate course in real analysis if they have some prior understanding of why one should bother with those technical details in the first place. A great place to learn about the details in the context of quantum mechanics is B.C. Hall's excellent text *Quantum Theory for Mathematicians* [10].

This book is *not* intended as a replacement for introductory physics texts such as [9, 22]: the reader will not learn perturbation theory nor gain proficiency at computing the energy levels or eigenstates of any but the simplest quantum mechanical systems. Nevertheless, M and P *do* try out their model on systems such as the infinite spherical well, the harmonic oscillator, and the hydrogen atom. But the point is always to illustrate the underlying mathematical structure, not the explicit form of the solutions or their physical consequences. A highly recommended undergraduate text for students of mathematics that does present peturbation theory with applications to scattering problems is [3], which has a more analytic focus than our text, and also provides a fuller discussion of the relationship between quantum and classical mechanics.

Likewise, this text is not meant as a replacement for more advanced mathematical treatments of quantum mechanics such as [10, 20]. In particular, M introduces a piece of mathematics only if she feels it is demanded by physical considerations. And even then, she resists the temptation to develop the ideas in even modest generality, preferring to stay close to the physical model under development. Overall, one might read this book as a motivating introduction to Lie groups and their representations, with focus on the quantum mechanically relevant Heisenberg group H_3 and special unitary group $SU(2)$.

While the results described herein are well-known, the presentation is somewhat novel. In any case, my aim is to whet the appetite for further study, and I hope this text serves to reveal the simplicity and beauty of a subject that is often perceived as complicated and intimidating. Exercises occur throughout, and I have provided solutions to those tagged by the symbol ♣ in appendix A.4. In an effort to bridge the gap between the physics and mathematics literature, I have adopted some notation that may be more familiar to physicists than mathematicians. In particular, time-derivatives are denoted by $\dot{c}(t)$ rather than $c'(t)$, and primes instead decorate objects viewed from M's point of view as compared with P's. In addition, I denote complex conjugation by α^* rather than $\overline{\alpha}$, and use $L^\dagger$ to denote the adjoint/hermitian conjugate of an operator rather than L^*. A brief review of key material from linear algebra, multivariable calculus, and analysis is provided in appendices A.1–A.3.

This book is dedicated to my son, Sebastian, and I certainly could not have written it without the love, support, and patience of my wife Madera. I would also like to thank my students Karl Mayer, Sanfer D'souza, and Daniel Martinez Zambrano for working through early drafts of this material as part of their Senior Experiences at Lawrence University—their questions and comments have been extremely helpful. My colleague Allison Fleshman from the Chemistry Department provided enthusiastic conversations and insightful comments about the periodic table, and Doug Martin in Physics has given me continual encouragement and inspiration. Finally, many thanks go to an anonymous reviewer for excellent suggestions that have substantially improved the exposition.

Scott Corry
Appleton, WI
May, 2016

Plan of the Book

Part I: Spin

Chapter 1: Physical Space *. . . in which M and P discover the group of rotations, $SO(3)$.*

The book begins with a description of physical space as isomorphic to $\mathbb{R}^3$, but care is taken to note that the *choice* of isomorphism is arbitrary. The argument by which this freedom of choice leads to an action of $SO(3)$ is carefully rehearsed so that it may serve as a template for later discussions in less familiar contexts.

Chapter 2: Spinor Space *. . . in which M and P discover the special unitary group $SU(2)$ and its relation to the group of rotations $SO(3)$.*

On the basis of the Stern-Gerlach experiment, spinor space is described as isomorphic to $\mathbb{C}^2$, but once again the choice of isomorphism is arbitrary. Following closely the pattern of Chapter 1, this freedom of choice leads to an action of $SU(2)$. The relationship between physical space and spinor space is established by showing that $SU(2)$ is the universal cover of $SO(3)$.

Chapter 3: Observables and Uncertainty *. . . in which M and P discover the Lie algebra $\mathfrak{su}(2)$ and its complexification $\mathfrak{sl}_2(\mathbb{C})$.*

A discussion of quantum observables leads to the definition of several Lie algebras and an exploration of their relationship to the corresponding Lie groups. The Lie bracket is shown to have a physical interpretation in terms of uncertainty.

Chapter 4: Dynamics *. . . in which M and P discover the Schrödinger equation.*

The time-evolution of spin-states is modeled as a curve in the unitary group $U(2)$, and this curve is shown to be determined by the Hamiltonian of the physical system, obtained by quantizing the classical expression for the energy.

Chapter 5: Higher Spin *. . . in which M and P classify the representations of $SU(2)$.*

The complex irreducible representations of $SU(2)$ are classified by working with the corresponding representations of the Lie algebra $\mathfrak{sl}_2(\mathbb{C})$. These representations are described explicitly and a physical interpretation is provided in terms of higher spin particles measured by a Stern-Gerlach apparatus. The final section studies the irreducible representations of $SO(3)$ as preparation for the theory of orbital angular momentum in the second part of the book.

Chapter 6: Multiple Particles *. . . in which M and P learn about the tensor product.*

The tensor product of spinor spaces provides a model for the spin-states of a system of two particles. This leads to the Clebsch-Gordan problem for $SU(2)$, whose solution describes how the tensor product of irreducible representations decomposes as a direct sum of irreducibles.

Part II: Position & Momentum

Chapter 7: A One-dimensional World *. . . in which M and P discover the Heisenberg group, H_1.*

Position space $L^2(\mathbb{R})$ is introduced in order to model the position of a particle in one dimension. The freedom of choice of an origin in physical space leads to an action of the group $(\mathbb{R}, +)$. This translation action extends to an action of the Heisenberg group H_1, and the corresponding Lie algebra action provides the position and momentum operators. The resulting framework is applied to several physical systems: the free particle, the infinite square well, and the harmonic oscillator.

Chapter 8: A Three-dimensional World *. . . in which M and P combine their studies of the Heisenberg group H_3 and the rotation group $SO(3)$.*

Following the pattern developed for the one-dimensional world, position space $L^2(\mathbb{R}^3)$ is introduced for three dimensions. This space carries a translation action of the group $(\mathbb{R}^3, +)$ which extends to an action of the Heisenberg group H_3. Hearkening back to Chapter 1, the choice of a basis for physical space leads to an action of $SO(3)$ on position space, which combines with the Heisenberg action to yield an action of the group $G = H_3 \rtimes SO(3)$. The corresponding Lie algebra action provides the position, linear momentum, and orbital angular momentum operators. Several physical systems are studied: the free particle, the infinite spherical well, the harmonic oscillator, and the Coulomb potential. In order to incorporate spin, the G-action is extended to an action of the group $\widetilde{G} = H_3 \rtimes SU(2)$ on spinor-valued wave functions. This framework is applied to the hydrogen atom, and a discussion of identical particles leads to the Pauli exclusion principle and some insight into the structure of the periodic table of the elements.

Chapter 9: Towards a Relativistic Theory *. . . in which M and P discover the central extension of the Galilean group, the restricted Lorentz group $SO^+(1,3)$, and the Dirac equation.*

The final chapter considers the relationship between the time-evolutions of wave functions for observers in uniform relative motion. This leads to an action of the central extension of the Galilean group in the non-relativistic context, and to actions of the restricted Lorentz and Poincaré groups in special relativity. The text ends with a discussion of the Dirac equation describing a relativistic spin-$\frac{1}{2}$ particle.

List of Figures

Part I

Spin

Chapter 1

Physical Space

In which M and P discover the group of rotations, $SO(3)$.

1.1 Modeling space

Let us imagine two observers, M(athematician) and P(hysicist). They are located together, looking at empty space, armed with meter sticks and protractors. Their sense impressions will probably lead them to agree about the following statements:

1. Space is three-dimensional and flat (i.e. not curved);
2. Their meter sticks are identical;
3. Their protractors are identical.

Of course, statement number 1 is imprecise. What does three-dimensional really mean? What about flat? Nonetheless, these shared intuitions lead M to propose the following model of the empty physical space surrounding them:

Definition 1.1. Physical space *is a three-dimensional real inner product space* $(V, \langle,\rangle)$.

P politely asks M to motivate her choice of model. M responds by unpacking her concise definition[1]: V is a vector space over the field of real numbers $\mathbb{R}$. To say that V is three-dimensional means that there exists an ordered set of three vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \subset V$ such that every $\mathbf{v}$ in V can be written uniquely as

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$$

for some particular real numbers c_1, c_2, c_3. The uniqueness means that two distinct 3-tuples of real numbers will yield distinct linear combinations of the basis vectors $\mathbf{v}_i$. Now P understands that M is using a vector space to capture the intuition that space is "flat", and in this context "three-dimensional" acquires a precise meaning that meshes with physical intuition: there are exactly three independent directions in space, no more and no less.

An *inner product* on V is a function $\langle,\rangle\colon V \times V \to \mathbb{R}$ such that if $c \in \mathbb{R}$ and $\mathbf{v}, \mathbf{v}', \mathbf{w} \in V$, then:

[1] See appendix A.1 for a summary of basic concepts in linear algebra.

i) $\langle c\mathbf{v} + \mathbf{v}', \mathbf{w}\rangle = c\langle \mathbf{v}, \mathbf{w}\rangle + \langle \mathbf{v}', \mathbf{w}\rangle$ (*linearity in first component*);

ii) $\langle \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle$ (*symmetry*);

iii) $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$ (*positive definite*).

Exercise 1.2. *Show that conditions i) and ii) for an inner product imply linearity in the second component:* $\langle \mathbf{w}, c\mathbf{v} + \mathbf{v}'\rangle = c\langle \mathbf{w}, \mathbf{v}\rangle + \langle \mathbf{w}, \mathbf{v}'\rangle$.

M summarizes by saying that an inner product is a bilinear, symmetric, positive definite function on $V \times V$. So far, P isn't very impressed with this as motivation for M's definition of space. But M continues: define the *length* of a vector $\mathbf{v}$ in V to be $|\mathbf{v}| := \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$, which is a non-negative real number by condition iii). Furthermore, define the *angle*[2] between two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ to be $\theta(\mathbf{v}, \mathbf{w}) := \arccos\left(\frac{\langle \mathbf{v}, \mathbf{w}\rangle}{|\mathbf{v}||\mathbf{w}|}\right)$. Observer P now understands: an inner product on V is an algebraic gadget that captures the notion of length and angle. Since M and P have identical meter sticks and protractors, they have the same notion of length and angle, hence they agree about the inner product on V.

Exercise 1.3. *Suppose that* $(X, \langle , \rangle)$ *is a real inner product space. Show that the inner product* $\langle , \rangle$ *is uniquely determined by the corresponding length function. That is, show that if two inner products define the same length function, then they are the same. (Hint: Compute* $|\mathbf{x}_1 + \mathbf{x}_2|^2$*).*

Example 1.4. *Fix an integer* $n \geq 1$*, and consider the set* $\mathbb{R}^n$ *of all* n*-tuples of real numbers:*

$$\mathbb{R}^n := \{(x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R}\}.$$

Then $\mathbb{R}^n$ *is an* n*-dimensional real vector space under the operations of component-wise addition and scalar multiplication. The* standard basis *for* $\mathbb{R}^n$ *is given by* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$*, where the vector* $\mathbf{e}_i$ *has a* 1 *in the* i*th slot and zeros elsewhere. Define the* dot product *of two vectors in* $\mathbb{R}^n$ *by the formula*

$$(x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) := \sum_{i=1}^{n} x_i y_i.$$

The reader should check that $\cdot \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *defines an inner product on* $\mathbb{R}^n$*. The resulting real inner product space* $(\mathbb{R}^n, \cdot)$ *is called* real Euclidean n-space.

P is tired of all this formalism, and wants to start doing experiments. So he begins to set up his lab. He is going to want to make measurements, so his first order of business is to set up a coordinate system. In terms of the model, this requires a specific choice of ordered basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for the vector space V.

[2] Here $\arccos : [-1, 1] \to [0, \pi]$ is the inverse of the cosine function. This definition makes sense because of the *Cauchy-Schwarz inequality* which states that $|\langle \mathbf{v}, \mathbf{w}\rangle| \leq |\mathbf{v}||\mathbf{w}|$ for all vectors $\mathbf{v}, \mathbf{w} \in V$.

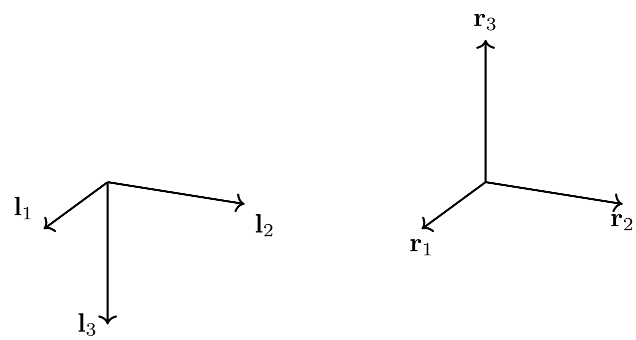

FIGURE 1.1: The basis on the left is left-handed; the basis on the right is right-handed.

The fact that V is three-dimensional guarantees that at least one such basis exists, but P quickly realizes that there are in fact infinitely many distinct bases to choose from, each of which corresponds to a different coordinate system for his lab. However, P doesn't like all coordinate systems equally. He prefers one in which the coordinate axes are orthogonal. Moreover, it strikes P as convenient to normalize his basis vectors to have length 1 (i.e. the same length as his meter stick). Thus he decides to choose an *orthonormal* basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for the inner product space $(V, \langle , \rangle)$, which means that the basis vectors are pairwise-orthogonal and have unit length[3]. Observer M agrees that this is a reasonable condition to impose on a choice of basis: after all, since M and P agree about the inner product on V, they will also agree about whether a given basis is orthonormal.

P has one more preference about coordinate systems: because he is right-handed, he wants to use a right-handed coordinate system. By a right-handed system he means the following (see figure 1.1): if he points the fingers of his right hand along the direction of his first coordinate axis, and curls them toward his second axis, then his thumb will point in the direction of his third axis (rather than in the opposite direction). Using physical intuition, P observes that any orthogonal coordinate system is either right-handed or left-handed, and that any two right-handed systems are related by a rotation, and similarly for any two left-handed systems. However, to move a right-handed system onto a left-handed system requires a reflection. At first, M is reluctant to build these considerations into their model of space, but she finally relents after specifying the following implication: the distinction between right- and left-handed coordinate systems divides the collection of all bases for V into two disjoint subsets. P's preference for right-handed systems corresponds to the selection of one of these two subsets as the *positively oriented* bases. If you flip the sign of the third basis vector in a positively oriented basis, you get a negatively

[3]The *Gram-Schmidt orthogonalization algorithm* guarantees that such a basis exists.

oriented basis and vice-versa (this flipping corresponds to the reflection across the plane spanned by the first two basis vectors). The specification of which subset is positive is called an *orientation* on V, so their model of space is now an *oriented* three-dimensional inner product space. The next exercise makes the notion of orientation rigorous and generalizes it to n dimensions.

Exercise 1.5. *Suppose that X is an n-dimensional real vector space, and that $\gamma := \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and $\gamma' := \{\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_n\}$ are two ordered bases for X. Then there is a unique isomorphism $\Phi : X \to X$ with the property that $\Phi(\mathbf{x}_i) = \mathbf{x}'_i$ for all i. Since Φ is invertible, the determinant of Φ is a nonzero real number. Define a relation $\sim$ on the set of all bases of X by saying that $\gamma \sim \gamma'$ if and only if* $\det(\Phi) > 0$. *Show that $\sim$ is an equivalence relation that partitions the set of all bases for X into two subsets. The choice of one of these subsets as positive is called an* orientation *on X.*

After all of this discussion, P finally chooses a positively oriented orthonormal basis $\beta := \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for the oriented inner product space $(V, \langle,\rangle)$. Having made this choice, he can think about space in more concrete terms as follows. Given an arbitrary vector $\mathbf{v} \in V$, there exist unique real numbers a, b, c such that $\mathbf{v} = a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3$. This correspondence defines an isomorphism of vector spaces $\varphi \colon V \to \mathbb{R}^3$ defined by $\varphi(\mathbf{v}) = (a, b, c)$. Moreover, since P's basis is orthonormal, φ actually yields an isomorphism of inner product spaces $\varphi \colon (V, \langle,\rangle) \to (\mathbb{R}^3, \cdot)$. That is, φ not only preserves the vector space structure, it also preserves the inner products.

Exercise 1.6. *Suppose that $(X, \langle,\rangle)$ is an n-dimensional real inner product space, and that $\gamma := \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for X. For any vector $\mathbf{x} \in X$, the fact that γ is a basis means that there exist unique real numbers $a_1, a_2, \ldots, a_n$ such that $\mathbf{x} = \sum_{i=1}^n a_i \mathbf{u}_i$. Show that the function $\varphi \colon X \to \mathbb{R}^n$ defined by $\varphi(\mathbf{x}) := (a_1, a_2, \ldots, a_n)$ is an isomorphism of vector spaces. Further, show that φ preserves the inner products, hence is an isomorphism of inner product spaces: for all $\mathbf{x}_1, \mathbf{x}_2 \in X$, we have $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \phi(\mathbf{x}_1) \cdot \phi(\mathbf{x}_2)$.*

Thus, *once P chooses a positively oriented orthonormal basis*, he has identified space with $(\mathbb{R}^3, \cdot)$ endowed with the familiar right-handed orientation in which the standard basis is positive. With this, P feels like he is back on firm ground and starts thinking about some experiments he wants to perform. But M interrupts with an annoying question: how does this more concrete description of space depend on P's choice of orthonormal basis?

To make the question more precise, M chooses a different positively oriented orthonormal basis for $(V, \langle,\rangle)$, which she denotes by $\beta' := \{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\}$. As above, this choice of basis defines an isomorphism $\varphi' \colon (V, \langle,\rangle) \to (\mathbb{R}^3, \cdot)$. But note that φ' is *not* the same isomorphism as φ, so when P and M each decide to describe space as $(\mathbb{R}^3, \cdot)$, their descriptions do not agree. Nevertheless, they are both working with $(V, \langle,\rangle)$, so there must be a way of translating between the two descriptions. To discover the translation, consider the composed function $\varphi' \circ \varphi^{-1} \colon (\mathbb{R}^3, \cdot) \to (\mathbb{R}^3, \cdot)$. This function is an isomorphism, since it is the composition of two isomorphisms. Since it maps $\mathbb{R}^3$ to itself, we say that it is an *automorphism* of the inner product space $(\mathbb{R}^3, \cdot)$. We can picture the situation via the commutative diagram below:

$$\begin{array}{ccc} V & = & V \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\varphi'} \\ \mathbb{R}^3 & \xrightarrow{\varphi' \circ \varphi^{-1}} & \mathbb{R}^3. \end{array}$$

In order to better understand the automorphism $\varphi' \circ \varphi^{-1}$, we pause to provide a brief review of certain classes of linear operators on real Euclidean n-space.

1.2 Real linear operators and matrix groups

Suppose that $L \colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator, and let $\varepsilon := \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ denote the standard basis of $\mathbb{R}^n$ (see example 1.4) . Then L is represented (with respect to ε) by an $n \times n$ matrix of real numbers, which we also denote by L:

$$L = [L_{ij}] \qquad \text{where} \qquad L(\mathbf{e}_j) = \sum_{i=1}^{n} L_{ij}\mathbf{e}_i.$$

This means that the jth column of the matrix L is the vector $L(\mathbf{e}_j)$. We may now express the effect of the linear operator L as left-multiplication on column vectors: if $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, then viewing $\mathbf{v}$ as a column yields $L(\mathbf{v}) = L\mathbf{v} \in \mathbb{R}^n$, where the ith component of $L\mathbf{v}$ is

$$(L\mathbf{v})_i = \sum_{j=1}^{n} L_{ij} v_j,$$

the dot product of the ith row of L with $\mathbf{v}$.

The next few propositions express properties of the linear operator L in terms of the corresponding matrix.

Proposition 1.7. *The linear operator L is an automorphism of the vector space $\mathbb{R}^n$ if and only if the matrix L is invertible.*

Proof. The operator L is an automorphism if and only if it possesses an inverse: a linear operator $M \colon \mathbb{R}^n \to \mathbb{R}^n$ such that $L \circ M = M \circ L = \mathrm{id}$, the identity transformation. But composition of linear operators corresponds to multiplication of the corresponding matrices, so the existence of an inverse operator is equivalent to the existence of a matrix M such that $LM = ML = I_n$, the $n \times n$ identity matrix. $\square$

Denote by $GL(n,\mathbb{R})$ the set of all invertible $n \times n$ matrices with real entries. Note that this set has the following properties with respect to the operation of matrix multiplication:

i) (*closure*) If A and B are in $GL(n,\mathbb{R})$, then so is their matrix product AB, since the inverse of AB is equal to $B^{-1}A^{-1}$.

ii) (*identity*) The identity matrix $I_n \in GL(n,\mathbb{R})$, and it satisfies $I_nA = AI_n = A$ for all $A \in GL(n,\mathbb{R})$.

iii) (*inverses*) If A is in $GL(n,\mathbb{R})$, then so is A^{-1}, since $(A^{-1})^{-1} = A$.

iv) (*associativity*) Matrix multiplication is associative: $(AB)C = A(BC)$ for all matrices A, B, C of compatible sizes. This follows from the fact that matrix multiplication corresponds to the composition of linear operators, and composition of functions is associative.

These statements mean that $GL(n,\mathbb{R})$ forms a *group* under matrix multiplication; it is non-abelian since $AB \neq BA$ for matrices in general.

Definition 1.8. *The group of all invertible $n \times n$ real matrices, denoted $GL(n,\mathbb{R})$, is called the* real general linear group. *It is the symmetry group of the vector space $\mathbb{R}^n$.*

Proposition 1.9. *Suppose that $L \in GL(n,\mathbb{R})$. Then L preserves the dot product on $\mathbb{R}^n$ if and only if $L^{-1} = L^T$, the transpose of the matrix L. (By "preserves the dot product" we mean that $L\mathbf{v} \cdot L\mathbf{w} = \mathbf{v} \cdot \mathbf{w}$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.)*

Proof. Recall that the transpose of a matrix is obtained by reflecting across the main diagonal: $(L^T)_{ij} := L_{ji}$. We first show that for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $L\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot L^T\mathbf{w}$. Indeed, note that if we think of $\mathbf{v}$ and $\mathbf{w}$ as column vectors, then $\mathbf{w}^T$ is a row vector, and we may express the dot product as matrix multiplication: $\mathbf{v} \cdot \mathbf{w} = \mathbf{w}^T\mathbf{v}$. Replacing $\mathbf{v}$ by $L\mathbf{v}$ yields

$$L\mathbf{v} \cdot \mathbf{w} = \mathbf{w}^T(L\mathbf{v}) = (\mathbf{w}^TL)\mathbf{v} = (L^T\mathbf{w})^T\mathbf{v} = \mathbf{v} \cdot L^T\mathbf{w}.$$

(Here we have used the fact that $(AB)^T = B^TA^T$ for any two matrices of compatible sizes.) We now apply this identity to the dot product of $L\mathbf{v}$ and $L\mathbf{w}$:

$$L\mathbf{v} \cdot L\mathbf{w} = \mathbf{v} \cdot L^TL\mathbf{w}.$$

Clearly, if $L^{-1} = L^T$, then $L^TL = I_n$ and L preserves the dot product as claimed. Going the other direction, if L preserves the dot product, then we find that $\mathbf{v} \cdot \mathbf{w} = \mathbf{v} \cdot L^TL\mathbf{w}$ for all $\mathbf{v}, \mathbf{w}$. Subtraction yields

$$0 = \mathbf{v} \cdot \mathbf{w} - \mathbf{v} \cdot L^TL\mathbf{w} = \mathbf{v} \cdot (I_n - L^TL)\mathbf{w}.$$

Since this equation holds for all $\mathbf{v}$ and $\mathbf{w}$, we may take $\mathbf{v} = (I_n - L^TL)\mathbf{w}$ to find that

$$(I_n - L^TL)\mathbf{w} \cdot (I_n - L^TL)\mathbf{w} = 0.$$

Since the dot product is positive definite, it follows that $(I_n - L^TL)\mathbf{w} = 0$ for all $\mathbf{w}$. Thus, $I_n - L^TL$ is the zero operator, so that $L^TL = I_n$ and $L^{-1} = L^T$. $\square$

Denote by $O(n) \subset GL(n, \mathbb{R})$ the subset of matrices satisfying the condition $L^{-1} = L^T$. Such matrices are called *orthogonal*. The reader should check that $O(n)$ is a *subgroup* of $GL(n, \mathbb{R})$: a subset containing the identity matrix and closed under matrix multiplication and inverses.

Definition 1.10. *The group of all invertible $n \times n$ real matrices satisfying $L^{-1} = L^T$, denoted $O(n)$, is called the* orthogonal group*. It is the symmetry group of real Euclidean n-space $(\mathbb{R}^n, \cdot)$.*

Exercise 1.11. *Suppose that $L \colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator. Show that $L \in O(n)$ if and only if L preserves the lengths of vectors: $|L\mathbf{v}| = |\mathbf{v}|$ for all $\mathbf{v} \in \mathbb{R}^n$. (Hint: use proposition 1.9 and compute $|L(\mathbf{v} + \mathbf{w})|^2$.)*

Proposition 1.12. *The determinant of an orthogonal matrix is ± 1.*

Proof. If L is orthogonal, then $I_n = L^T L$. Taking the determinant of both sides yields $1 = \det(I_n) = \det(L^T L) = \det(L^T)\det(L) = \det(L)^2$. It follows that $\det(L) = \pm 1$. □

Now endow $(\mathbb{R}^n, \cdot)$ with the orientation for which the standard basis ε is positive. If $L \in GL(n, \mathbb{R})$ is an invertible matrix, we say that L is *orientation preserving* if L sends positively oriented bases to positively oriented bases. By exercise 1.5, we see that L is orientation preserving if and only if $\det L > 0$. It follows from the previous proposition that an orthogonal matrix L is orientation preserving if and only if $\det L = 1$; such matrices are called *special orthogonal*, and they form a subgroup of the orthogonal group.

Definition 1.13. *The* special orthogonal group *is the subgroup $SO(n) \subset O(n)$ of orthogonal matrices with determinant 1. It is the symmetry group of oriented real Euclidean n-space.*

Before returning to our observers M and P in three dimensions, we take a close look at all of these groups in the cases $n = 1$ and $n = 2$.

Example 1.14. *For $n = 1$ we have the following groups:*

- *$GL(1, \mathbb{R}) = \mathbb{R}^\times$, the group of nonzero real numbers;*
- *$O(1) = \{\pm 1\}$, the sign group;*
- *$SO(1) = \{1\}$, the trivial group.*

Example 1.15. *The real general linear group $GL(2, \mathbb{R})$ consists of 2×2 real matrices with nonzero determinant. Hence, we have*

$$GL(2, \mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid ad - bc \neq 0 \right\}.$$

Equivalently, a 2×2 real matrix is an element of $GL(2, \mathbb{R})$ if and only if its *columns form a basis for $\mathbb{R}^2$. Concretely, this means that neither column is a scalar multiple of the other.*

Writing out the othogonality condition $I_2 = L^T L$ we find

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix}.$$

Equating entries, we obtain the conditions $a^2+c^2 = b^2+d^2 = 1$ and $ab+cd = 0$. Interpreting these relations as dot products, we see that L is orthogonal exactly when the columns of L have unit length and are orthogonal to each other. Thus, a 2×2 real matrix is an element of $O(2)$ if and only if its columns form an orthonormal basis for $(\mathbb{R}^2, \cdot)$.

Note that for any choice of a, c such that $a^2 + c^2 = 1$, the point (a, c) lies on the unit circle in $\mathbb{R}^2$. Hence, there exists a unique angle $\theta \in [0, 2\pi)$ such that $a = \cos(\theta)$ and $c = \sin(\theta)$ (see figure 1.2). The remaining two conditions on b and d then imply that $(b, d) = \pm(-\sin(\theta), \cos(\theta))$. Thus, we have the following description of the orthogonal group $O(2)$:

$$O(2) = \left\{ \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix} \mid 0 \le \theta < 2\pi \right\}.$$

Finally, if we demand that the determinant is $+1$, we find that only half of the matrices in $O(2)$ remain as special orthogonal matrices:

$$SO(2) = \left\{ L_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \mid 0 \le \theta < 2\pi \right\}.$$

Hence, each element of $SO(2)$ is uniquely determined by an angle $\theta \in [0, 2\pi)$. In fact, the special orthogonal matrix L_θ describes a counter-clockwise rotation through the angle θ (see figure 1.2). The first standard basis vector $\mathbf{e}_1 = (1, 0)$ is sent to $(\cos(\theta), \sin(\theta))$, while the second standard basis vector $\mathbf{e}_2 = (0, 1)$ is sent to $(-\sin(\theta), \cos(\theta)$.

Exercise 1.16. *Generalize the analysis in the preceding example to show that for all $n \ge 1$:*

a) an $n \times n$ real matrix L is in $GL(n, \mathbb{R})$ if and only if the columns of L form a basis for $\mathbb{R}^n$;

b) an $n \times n$ real matrix L is in $O(n)$ if and only if the columns of L form an orthonormal basis for $(\mathbb{R}^n, \cdot)$.

Part b) of the previous exercise has the following important consequence. We have defined a linear operator L on real Euclidean n-space to be orthogonal if its matrix with respect to the standard basis satisfies $L^T = L^{-1}$. But in fact, the next proposition shows that this relation between the transpose and inverse of an orthogonal operator holds for the matrix of L with respect to *any* orthonormal basis. We will make use of this fact in proposition 1.19 in the next section.

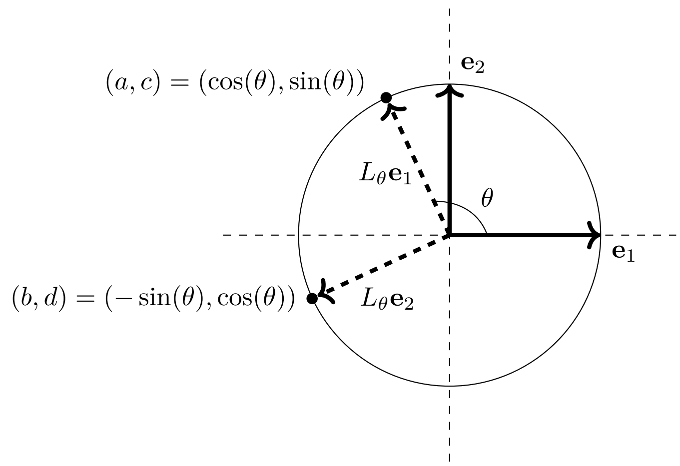

FIGURE 1.2: The 2×2 special orthogonal matrix L_θ defines a counterclockwise rotation through the angle θ.

Proposition 1.17. *Let $L\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator, and $\gamma := \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be any orthonormal basis for $(\mathbb{R}^n, \cdot)$. Denote by $[L]_\gamma$ the matrix representing L in the basis γ. Then L is an orthogonal transformation if and only if $[L]_\gamma^T = [L]_\gamma^{-1}$.*

Proof. Let Q be the change of basis matrix from γ to ε, the standard basis. Recall that the jth column of Q is simply the basis vector $\mathbf{u}_j$. Hence, by part b) of the previous exercise, Q is orthogonal and $Q^{-1} = Q^T$. Then writing L for $[L]_\varepsilon$ as usual, we have $[L]_\gamma = QLQ^{-1} = QLQ^T$. We may then compute

$$[L]_\gamma^T [L]_\gamma = (QLQ^T)^T QLQ^T = QL^TQ^TQLQ^T = Q(L^TL)Q^T.$$

From this equation we see that $[L]_\gamma^T [L]_\gamma = I_n$ if and only if $L^TL = I_n$ as claimed. □

1.3 SO(3) is the group of rotations

Recall the situation of M and P, illustrated by the following diagram:

$$\begin{array}{ccc} V & = & V \\ {\scriptstyle \varphi}\downarrow & & \downarrow{\scriptstyle \varphi'} \\ \mathbb{R}^3 & \xrightarrow{\varphi' \circ \varphi^{-1}} & \mathbb{R}^3. \end{array}$$

P has chosen an orthonormal basis $\beta = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for the inner product space $(V, \langle , \rangle)$, thereby obtaining an isomorphism $\varphi \colon (V, \langle , \rangle) \to (\mathbb{R}^3, \cdot)$. Meanwhile, M has chosen a different orthonormal basis $\beta' = \{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\}$, thereby obtaining a different isomorphism $\varphi' \colon (V, \langle , \rangle) \to (\mathbb{R}^3, \cdot)$. To study the difference in their descriptions of physical space, we are considering the composition $\varphi' \circ \varphi^{-1}$.

Since this composition of isomorphisms is an automorphism of $\mathbb{R}^3$, we may identify it with an invertible matrix of real numbers, i.e. an element of $GL(3, \mathbb{R})$. Moreover, because the automorphism $\varphi' \circ \varphi^{-1}$ preserves the dot product, the corresponding matrix is actually an element of the orthogonal group $O(3)$. But because M and P both chose positively oriented bases, the automorphism $\varphi' \circ \varphi^{-1}$ preserves the orientation on $(\mathbb{R}^3, \cdot)$, which means that its determinant is 1. As a concise summary, M says that the difference between the two descriptions of space is the automorphism $\varphi' \circ \varphi^{-1}$, which may be identified with an element of the special orthogonal group $SO(3)$.

Exercise 1.18. *Show that the matrix representing $\varphi' \circ \varphi^{-1}$ with respect to the standard basis on $\mathbb{R}^3$ is simply the change of basis matrix from the basis β to the basis β'. Recall that this is the matrix representing the identity transformation* $\mathrm{id} \colon V \to V$ *with respect to the bases β and β':*

$$[\mathrm{id}]_{\beta}^{\beta'} := [b_{ij}] \qquad \textit{where} \qquad \mathbf{u}_j = \sum_{i=1}^{3} b_{ij} \mathbf{u}'_i.$$

In example 1.15, we saw that the group $SO(2)$ consists entirely of rotations. In the next proposition, we establish the corresponding result for three dimensions.

Proposition 1.19. *The special orthogonal group $SO(3)$ consists of rotations in real Euclidean 3-space. More precisely, for each non-identity element $A \in SO(3)$, there is a unique line of $\mathbb{R}^3$ that is fixed pointwise by A. Moreover, A acts as a rotation through some angle θ around this fixed axis.*

Proof. Let $A \in SO(3)$ be an arbitrary special orthogonal matrix. We begin by showing that 1 is an eigenvalue for A, so that A fixes the line spanned by a corresponding eigenvector in $\mathbb{R}^3$; this line will turn out to be the axis of rotation.

Note that the characteristic polynomial $p_A(\lambda) := \det(A - \lambda I_3)$ is a degree 3 polynomial with real coefficients and hence has at least one real root, say $\lambda_0 \in \mathbb{R}$. Thus, λ_0 is an eigenvalue for A, and we may choose a unit length eigenvector $\mathbf{u} \in \mathbb{R}^3$ such that $A\mathbf{u} = \lambda_0 \mathbf{u}$. Now use the fact that A preserves the dot product:

$$1 = \mathbf{u} \cdot \mathbf{u} = A\mathbf{u} \cdot A\mathbf{u} = \lambda_0 \mathbf{u} \cdot \lambda_0 \mathbf{u} = \lambda_0^2 \mathbf{u} \cdot \mathbf{u} = \lambda_0^2.$$

Thus, we see that any real eigenvalue of A satisfies $\lambda_0 = \pm 1$.

There are two cases depending on the factorization of the characteristic polynomial over $\mathbb{R}$:

i) the polynomial $p_A(\lambda)$ factors into linear factors over $\mathbb{R}$, so that A has 3 real eigenvalues $\lambda_0, \lambda_1, \lambda_2$, each of which is ± 1 by the previous argument. Then $1 = \det(A) = \lambda_0\lambda_1\lambda_2$, which implies that at least one of the eigenvalues is $+1$. If all three of them are $+1$ then A is the identity. Otherwise, only one of the eigenvalues is $+1$, and the corresponding 1-dimensional eigenspace is the unique line that is fixed pointwise by A.

ii) the polynomial $p_A(\lambda)$ factors into a linear factor $(x - \lambda_0)$ and an irreducible quadratic. By the quadratic formula, the complex roots of the irreducible quadratic factor are complex conjugates, say $\lambda_1 = \lambda_2^* \in \mathbb{C}$. It follows that $1 = \det(A) = \lambda_0\lambda_1\lambda_2 = \lambda_0|\lambda_1|^2$. Since $|\lambda_1|^2 > 0$, we see that $\lambda_0 = -1$ is impossible in this case, so that $\lambda_0 = 1$. Again, the corresponding 1-dimensional eigenspace is the unique line that is fixed pointwise by A.

Thus, we have established that 1 is an eigenvalue for A, and we have chosen a unit length eigenvector $\mathbf{u}$ such that $A\mathbf{u} = \mathbf{u}$. Expand $\{\mathbf{u}\}$ to an orthonormal basis $\gamma := \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ for $(\mathbb{R}^3, \cdot)$. In the basis γ, the linear operator A is represented by a matrix of the following form:

$$[A]_\gamma = \begin{bmatrix} 1 & v & w \\ 0 & a & b \\ 0 & c & d \end{bmatrix}.$$

In fact, $v = w = 0$. Indeed, we simply need to use the orthonormality of the basis together with the orthogonality of the matrix A to compute:

$$v = A\mathbf{v} \cdot \mathbf{u} = \mathbf{v} \cdot A^T\mathbf{u} = \mathbf{v} \cdot A^{-1}\mathbf{u} = \mathbf{v} \cdot \mathbf{u} = 0.$$

The proof that $w = 0$ is the same, with $\mathbf{w}$ in place of $\mathbf{v}$.

Therefore, in the basis γ, our special orthogonal transformation A is represented by a matrix

$$[A]_\gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & B \end{bmatrix},$$

where we have written B for the 2×2 matrix in the lower right. Now by proposition 1.17, the matrix $[A]_\gamma$ satisfies $[A]_\gamma^T [A]_\gamma = I_3$. But explicit computation then shows that

$$I_3 = [A]_\gamma^T [A]_\gamma = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & B^T B \end{bmatrix},$$

so that $B^T B = I_2$ and $B \in O(2)$. But we also have that $1 = \det(A) = \det(B)$, so that in fact $B \in SO(2)$. As revealed in example 1.15, the elements of $SO(2)$ describe rotations, and we may write $[A]_\gamma$ explicitly as

$$[A]_\gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}$$

for some angle $0 \leq \theta < 2\pi$. The matrix clearly reveals the effect of the linear operator A: it fixes the axis spanned by the eigenvector $\mathbf{u}$ while rotating the plane perpendicular to that axis through the angle θ. □

Once P recognizes $SO(3)$ as the group of rotations, he adds to the previous discussion as follows: his choice of the basis β leads to the identification $\varphi\colon (V, \langle,\rangle) \to (\mathbb{R}^3, \cdot)$, sending β to the standard basis ε of $\mathbb{R}^3$. Then M's choice of basis β' determines an automorphism $\mathcal{A}$ of $(V, \langle,\rangle)$ defined by $\mathcal{A}(\mathbf{u}_i) = \mathbf{u}'_i$. In physical terms, $\mathcal{A}$ rotates P's coordinate axes onto M's. In terms of P's description, the automorphism $\mathcal{A}$ is represented with respect to the basis β by the matrix $A := [\mathcal{A}]_\beta \in SO(3)$. Explicitly, we have

$$A := [a_{ij}] \qquad \text{where} \qquad \mathcal{A}(\mathbf{u}_j) = \mathbf{u}'_j = \sum_{i=1}^{3} a_{ij}\mathbf{u}_i.$$

It follows from exercise 1.18 that A is the inverse of the matrix representing $\varphi' \circ \varphi^{-1}$. Thus, the matrix $A \in SO(3)$ that describes (with respect to β) the rotation that moves P's coordinate axes onto M's is the inverse of the change of basis matrix describing the translation from P's description to M's. Conversely, an arbitrary element of $SO(3)$ will describe for P a way of rotating his coordinate axes to obtain a new right-handed coordinate system, hence a new positively oriented orthonormal basis for $(V, \langle,\rangle)$. Thus, the group $SO(3)$ acts as the group of rotations on P's copy of space $(\mathbb{R}^3, \cdot)$, serving to connect P's basis with all other possible choices of positive orthonormal basis. The next definition specifies exactly what it means for a group to *act* on a set.

Definition 1.20. *Let G be a group, and X a set. Then a G-*action *on X is a function $G \times X \to X$ (denoted by $(g, x) \mapsto g \star x$) with the following properties:*

- *$(g_1 g_2) \star x = g_1 \star (g_2 \star x)$ for all $g_1, g_2 \in G$ and $x \in X$;*
- *$e \star x = x$ for all $x \in X$, where $e \in G$ is the identity element.*

In the case where X is a vector space, we can make the additional requirement that each $g \in G$ acts linearly on X:

$$g \star (cx + y) = c(g \star x) + (g \star y) \qquad \text{for all } x, y \in X \text{ and scalars } c.$$

Such a linear G-action is called a representation *of the group G. Almost all of the group actions considered in this text will be linear representations, and we will have much more to say about them in Chapter 5.*

Exercise 1.21. *Show that left-multiplication $(A, \mathbf{x}) \mapsto A \star \mathbf{x} := A\mathbf{x}$ defines an action of $SO(3)$ on $\mathbb{R}^3$. This is clearly a linear action, and we refer to it as the* defining representation *of $SO(3)$. We will study the other representations of $SO(3)$ in section 5.6.*

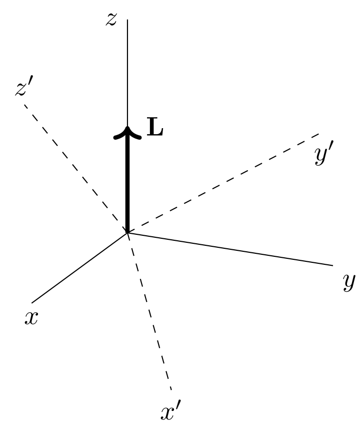

FIGURE 1.3: Observer M's rotated coordinate system (dashed) drawn on top of P's coordinate system (solid). The vector **L** represents a spinning top's angular momentum.

Observer P wants to try all this out to make sense of it. So having laid out his coordinate system, he starts a top spinning at the origin, with its axis of rotation along the third axis. When he looks down on it from the positive third axis, it is spinning counter-clockwise.The angular momentum[4] of the top is represented by a vector $\mathbf{L} = (0, 0, c)$, where $c > 0$ is the magnitude (see figure 1.3). This entire description derives from P's initial choice of the basis β, which yielded the identification with $\mathbb{R}^3$. As above, suppose that M's basis β' is obtained from β via the rotation $\mathcal{A}$. What is the column vector[5] representing the top's angular momentum under M's identification of space with $\mathbb{R}^3$? As explained above, since the rotation sending P's basis to M's is represented by the orthogonal matrix $A := [\mathcal{A}]_\beta$, the observed coordinates transform according to $A^{-1} = A^T$. Hence, M will measure $A^T[0, 0, c]^T$ for the angular momentum of the top. Take a simple example: suppose that M's coordinate system is obtained by rotating P's through an angle of $\frac{\pi}{2}$ counter-clockwise around P's second axis (see Figure 1.4). Then $\mathbf{u}'_1 = -\mathbf{u}_3, \mathbf{u}'_2 = \mathbf{u}_2$, and $\mathbf{u}'_3 = \mathbf{u}_1$. Thus, the rotation $\mathcal{A}$ is defined by

$$\mathcal{A}(\mathbf{u}_1) = -\mathbf{u}_3, \qquad \mathcal{A}(\mathbf{u}_2) = \mathbf{u}_2, \qquad \mathcal{A}(\mathbf{u}_3) = \mathbf{u}_1.$$

The matrix of $\mathcal{A}$ with respect to P's basis β is

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

[4]See section 2.1 for a brief discussion of angular momentum.

[5]For reasons of typographical economy, we have been writing elements of $\mathbb{R}^3$ as row vectors (a, b, c), although they are really columns $[a, b, c]^T$ for the purposes of matrix multiplication.

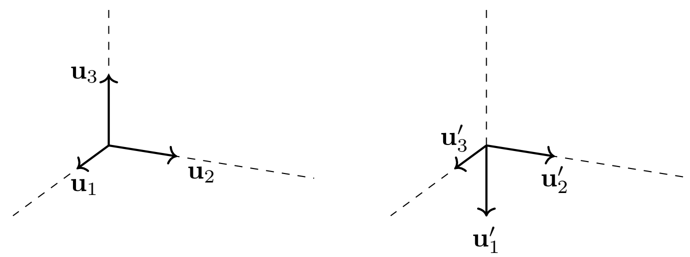

FIGURE 1.4: A rotation of $\pi/2$ radians about the $\mathbf{u}_2$-axis. The dashed lines in both pictures show P's coordinate axes.

Hence, M will measure $A^T[0, 0, c]^T = [-c, 0, 0]^T$ for the angular momentum of the top. This is what we should expect, because P's positive third axis points along M's negative first axis.

P thinks of all this as an elaborate bookkeeping device. M understands this point of view, but advocates for a richer viewpoint. Namely, because P's choice of positively oriented orthonormal basis is arbitrary, the only way to make his identification of space with $(\mathbb{R}^3, \cdot)$ independent of this choice is to remember that $SO(3)$ acts on this inner product space. Moreover, any physically meaningful mathematical object connected with this model of space should also be independent of the choice of basis, hence must support an action of $SO(3)$. That is, we expect to find that the mathematical gadgets that serve as models for physical systems will support a natural action of the group $SO(3)$. Observer P nods politely and changes the subject ... he wants to tell M about something called the Stern-Gerlach experiment.

Chapter 2

Spinor Space

In which M and P discover the special unitary group $SU(2)$ and its relation to the group of rotations $SO(3)$.

In 1922, Otto Stern and Walter Gerlach sent a beam of silver atoms through an inhomogeneous magnetic field and measured the resulting deflection of the atoms. Before we can understand the surprising results of their experiment, we need just a bit of information about the classical theory of angular momentum.

2.1 Angular momentum in classical mechanics

Suppose that the function $\mathbf{r}\colon \mathbb{R} \to \mathbb{R}^3$ describes the position of a particle with mass m, so that at time t, the particle is at the location $\mathbf{r}(t) = (x(t), y(t), z(t))$. Then the velocity of the particle is given by the time-derivative $\dot{\mathbf{r}}$, and its *linear momentum* is defined to be $\mathbf{p} := m\dot{\mathbf{r}}$. Thus, the linear momentum is a measure of the particle's linear motion, taking into account both its velocity and mass. For a similar measure of rotational motion, we define the *angular momentum* of the particle with respect to the origin as $\mathbf{L} := \mathbf{r} \times \mathbf{p}$, the cross product of the particle's position and linear momentum (see figure 2.1).

Now let's revisit the top, T, that observer P started spinning at the end of the previous chapter. The top has its axis of symmetry aligned with the z-axis, and is spinning counter-clockwise at the rate of ω radians per second—so T

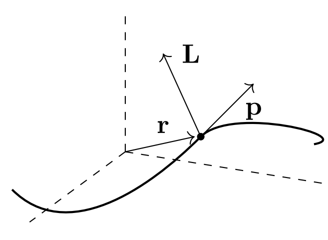

FIGURE 2.1: The position, linear momentum, and angular momentum of a particle moving in three-dimensional space.

makes a full revolution every $2\pi/\omega$ seconds. Consider a point of T at height z and distance $r \geq 0$ from the z-axis. As T spins, the point follows a circular trajectory:

$$\mathbf{r}(t) = (r\cos(\omega t), r\sin(\omega t), z). \tag{2.1}$$

If we assume that T has a uniform mass density, ρ, then a small volume ΔV centered at our point will have mass $\rho\Delta V$. Treating this small volume as a single particle, it has angular momentum $\mathbf{r} \times \dot{\mathbf{r}}\rho\Delta V$. Integrating over the spatial extent of $T \subset \mathbb{R}^3$ at a given instant yields the angular momentum of the spinning object T:

$$\mathbf{L} := \int_T \mathbf{r} \times \dot{\mathbf{r}}\rho \mathrm{d}V.$$

Let's work this out explicitly in the case where $T = B_R$ is the ball of radius $R > 0$ centered at the origin. Note that the derivative of the circular trajectory (2.1) is

$$\dot{\mathbf{r}}(t) = (-r\omega\sin(\omega t), r\omega\cos(\omega t), 0),$$

so the cross product is $\mathbf{r} \times \dot{\mathbf{r}} = r\omega(-\cos(\omega t)z, -\sin(\omega t)z, r)$. Using cylindrical coordinates, we compute that at any instant of time we have:

$$\begin{aligned}\mathbf{L} &:= \int_{B_R} \mathbf{r} \times \dot{\mathbf{r}}\rho \mathrm{d}V \\ &= \omega\rho \int_{z=-R}^{R} \int_{r=0}^{\sqrt{R^2-z^2}} \int_{\theta=0}^{2\pi} (-r\cos(\theta)z, -r\sin(\theta)z, r^2) r \mathrm{d}\theta \mathrm{d}r \mathrm{d}z.\end{aligned}$$

The first two components of this integral are zero due to the inner integration over θ, while for the z-component we have

$$\begin{aligned}L_z &= \omega\rho \int_{z=-R}^{R} \int_{r=0}^{\sqrt{R^2-z^2}} \int_{\theta=0}^{2\pi} r^3 \mathrm{d}\theta \mathrm{d}r \mathrm{d}z \\ &= 2\pi\omega\rho \int_{z=-R}^{R} \int_{r=0}^{\sqrt{R^2-z^2}} r^3 \mathrm{d}r \mathrm{d}z \\ &= \frac{\pi}{2}\omega\rho \int_{z=-R}^{R} (R^2 - z^2)^2 \mathrm{d}z \\ &= \frac{8}{15}\pi\omega\rho R^5.\end{aligned}$$

Note that the total mass of the ball is $M = \frac{4}{3}\pi R^3\rho$, so that the z-component of the angular momentum may be rewritten as $L_z = I\omega$, where $I := \frac{2}{5}MR^2$ is the *moment of inertia* of the spinning ball. Finally, defining the *angular velocity vector* as $\boldsymbol{\omega} := (0, 0, \omega)$, we may write the angular momentum as $\mathbf{L} = I\boldsymbol{\omega}$. We see that the angular momentum is a vector quantity in $\mathbb{R}^3$ that incorporates both the rate of rotation and the distribution of mass around the axis of rotation. Observe that spinning the ball clockwise rather than counter-clockwise corresponds to replacing $\omega > 0$ by $-\omega < 0$, which has the effect of reversing the direction of the angular momentum $\mathbf{L}$. In particular, for any angular speed ω, the vector $\mathbf{L}$ points along the z-axis, with magnitude determined by the absolute value of ω, and direction ("up" or "down") determined by the sign of ω.

Of course, there is nothing special about the z-axis here. If $\boldsymbol{\omega}$ is an arbitrary vector in $\mathbb{R}^3$, then it spans a line of rotational symmetry for the ball B_R, and $\mathbf{L} = I\boldsymbol{\omega}$ is the angular momentum of B_R when it spins counter-clockwise around $\boldsymbol{\omega}$ at the angular speed of $|\boldsymbol{\omega}|$ radians per second. Of course, a counter-clockwise rotation around $\boldsymbol{\omega}$ is a clockwise rotation around $-\boldsymbol{\omega}$, so changing the rotational sense (without changing the angular speed) simply replaces $\mathbf{L}$ by $-\mathbf{L}$.

We would like to find a way of measuring the angular momentum of the ball in a laboratory. It turns out that if the ball is small and electrically charged, then there is an ingenious way of measuring its angular momentum that relies on the classical theory of magnetic fields. We will describe the details below, but the upshot is that with the right experimental setup, the observed deflection of the ball in the z-direction will be directly proportional to the z-component, L_z, of its angular momentum, so that we can measure the ball's angular momentum by instead measuring the magnitude of its spatial deflection.

So suppose that our spinning ball is small, and that it carries a distribution of electric charge. The rotating charge turns the ball into a little magnet, characterized by its *magnetic dipole moment*, $\boldsymbol{\mu}$, a vector quantity proportional to the angular momentum, $\mathbf{L}$:

$$\boldsymbol{\mu} = \gamma \mathbf{L}.$$

Here, the constant of proportionality, γ, depends on the charge distribution on the ball. The important point for us is that this magnetic dipole moment determines the force that the ball will experience if placed within an external magnetic field.

In particular, if $\mathbf{B}$ is an inhomogeneous magnetic field with dominant direction z, and strength increasing linearly in the positive z-direction, then the ball will experience a force in the z-direction, proportional to the z-component of $\boldsymbol{\mu}$ (i.e., proportional to the z-component of $\mathbf{L}$). Hence, if we were to send the ball down the x-axis through the field $\mathbf{B}$, then the ball would be deflected in the z-direction, traveling up if $L_z > 0$, down if $L_z < 0$ (if $L_z = 0$, then the ball would experience no deflection). Moreover, if we confine the field $\mathbf{B}$ to a region of fixed length along the x-axis, and if we know the constant velocity of the ball as it travels down the x-axis, then the magnitude of the z-deflection will be directly proportional to the magnitude of L_z. This is the fact that

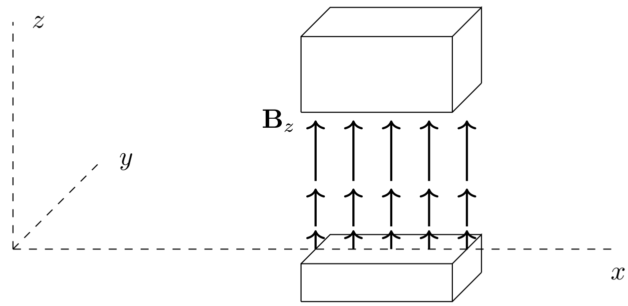

FIGURE 2.2: A Stern-Gerlach device oriented in the positive z-direction. The vertical arrows indicate the z-component of the resulting inhomogeneous magnetic field.

Stern and Gerlach used as the basis of their experiment with silver atoms[1] in 1922. By a *Stern-Gerlach device* (see figure 2.2), we will mean a device of fixed length that produces such an inhomogeneous field $\mathbf{B}$.

Now imagine sending a beam of these spinning balls down the x-axis, through a Stern-Gerlach device oriented in the positive z-direction as described in the previous paragraph. We assume that all of the balls have the same linear velocity, but that their angular momenta are distributed among all directions and a large range of angular speeds. That is, we have carefully prepared the translational motion, but have made no special preparation of the rotational motions. In particular, the z-components of the angular momenta, L_z, will form a continuous range of values, positive and negative. Since the z-deflection of an individual ball is proportional to the z-component of its angular momentum, we should find a continuous spread of the beam in the positive and negative z-directions.

Being small and charged, we expect our spinning ball to provide a crude classical model of the electron, considered as a charged point particle. So, applying the previous thought experiment to a beam of electrons, we record our conclusion as a

> **Classical Expectation:** *We should observe a continuous spread of the electron beam in the positive and negative z-directions, reflecting a continuous range of values for L_z among the individual electrons (see figure 2.3).*

But the experimental results are strikingly at odds with this expectation:

[1] The experiment described actually requires a neutral particle in order to avoid the Lorentz force that a moving charge would experience. However, the magnetic moment of a silver atom is due almost entirely to the magnetic moment of its outermost electron, so in effect, Stern and Gerlach were detecting the angular momentum of the electron (see [22, pp. 1-4]). Hence, for the idealized Stern-Gerlach thought experiments discussed here, we speak in terms of the negatively charged electron, even though the actual historical experiment requires a neutral particle.

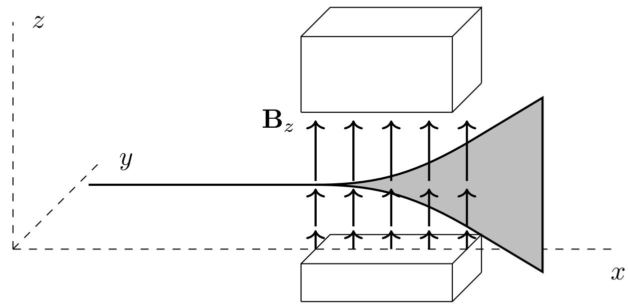

FIGURE 2.3: The classical expectation for the behavior of an electron-beam in a Stern-Gerlach device.

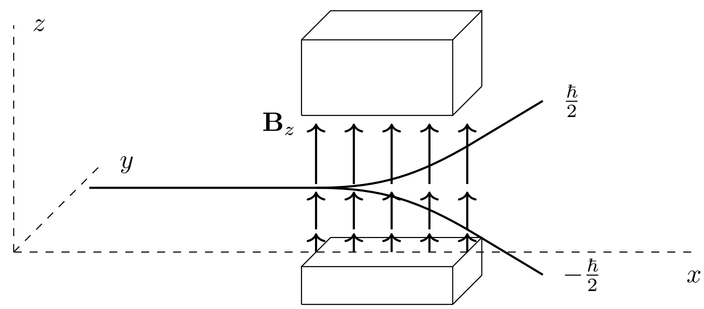

FIGURE 2.4: The actual behavior of an electron-beam in a Stern-Gerlach device.

> **Experimental Result:** *The electron beam splits into two discrete pieces, with half the electrons deflecting upward as if they have* $L_z = \frac{\hbar}{2}$*, while the other half deflects downward by the same amount, as if they have* $L_z = -\frac{\hbar}{2}$ *(see figure 2.4). Here,* $\hbar = 1.054573 \times 10^{-34}\ kg \cdot m^2/s$ *is the* reduced Planck constant.

Faced with this experimental fact, the only conclusion we can draw is that the crude classical model of the electron as a spinning charged ball is wrong. Instead of displaying a continuous range of angular momenta, the beam behaves as if it contains a 50-50 mix of two types of electrons: those that are "spin up" and those that are "spin down" along the z-direction. While the *sign* for each electron appears to be random, the *magnitude* of the z-component of angular momentum is fixed at $\frac{\hbar}{2}$. But the phenomenon is actually stranger still, because there is nothing special about the z-direction!

Indeed, suppose that we turn on the electron beam before establishing the magnetic field $\mathbf{B}$. Then we could choose any unit vector, $\mathbf{u}$, orthogonal to the beam's direction, and set up a Stern-Gerlach device with orientation $\mathbf{u}$, inducing an inhomogeneous magnetic field as described above, but now with dominant direction $\mathbf{u}$. From the rotational symmetry of physical space, the electron beam will behave just as before: half of the electrons will be deflected in the $+\mathbf{u}$-direction, half in the $-\mathbf{u}$-direction, but the *amount* of deflection will always be the same, corresponding to a $\mathbf{u}$-component of angular momentum of magnitude $\frac{\hbar}{2}$. This is quite bizarre, because (thinking classically) it seems to suggest that each electron is spinning at the same rate around every axis, with half of them spinning clockwise, half counter-clockwise! As described below (belief 2), one way out of this difficulty is to give up on the idea that individual electrons possess a definite angular momentum, and instead think in terms of definite probabilities for measurement outcomes.

2.2 Modeling spin

As P finishes his description of various Stern-Gerlach experiments[2], M is stunned. Nevertheless, these things have been revealed through careful experiment, and there is no denying them. The question is: how to model this phenomenon? Just as P and M made a short list of shared intuitions that led to their model of physical space in Chapter 1, they now make a list of shared beliefs about the electron, coming from their knowledge of the Stern-Gerlach experiments. If $\mathbf{u}$ is a unit vector in physical space $(V, \langle , \rangle)$, then $SG\mathbf{u}$ denotes a Stern-Gerlach device producing an inhomogeneous magnetic field with dominant direction $\mathbf{u}$.

1. An electron passing through an $SG\mathbf{u}$ will return an angular momentum measurement of $\pm\frac{\hbar}{2}$, which we think of as “spin up” and “spin down” along the direction $\mathbf{u}$;

2. Until we make a measurement with an $SG\mathbf{u}$, a particular electron may have no definite spin along $\mathbf{u}$, but it *does* have a definite probability of returning each of the values $\pm\frac{\hbar}{2}$ when measured by an $SG\mathbf{u}$;

3. If an electron exits an $SG\mathbf{u}$ spin up, then it will measure spin up if measured immediately by a successive $SG\mathbf{u}$. Likewise, if an electron exits an $SG\mathbf{u}$ spin down, then it will measure spin down if measured immediately by a successive $SG\mathbf{u}$;

4. More generally, if the angle between $\mathbf{u}$ and $\mathbf{u}'$ is α, then an electron that exits an $SG\mathbf{u}$ spin up will measure spin up with probability $\cos^2\left(\frac{\alpha}{2}\right)$ if measured immediately by an $SG\mathbf{u}'$. Similarly, an electron that exits an $SG\mathbf{u}$ spin down will measure spin down with probability $\cos^2\left(\frac{\alpha}{2}\right)$ if measured immediately by an $SG\mathbf{u}'$.

[2]One can imagine a number of different experiments involving multiple Stern-Gerlach devices arranged in sequence, with different orientations (see [22, pp. 5-9]). Beliefs 3 and 4 come from the results of such experiments

P thinks about this list for a while, and observes that together these beliefs imply that measurement with a Stern-Gerlach device *does something* to the electron. Indeed, by beliefs 1 and 3 he can produce electrons that will measure spin up along $\mathbf{u}$ with probability 1, by blocking the electrons that emerge spin down from the $SG\mathbf{u}$. Given such a beam of electrons, belief 4 says that if $\mathbf{u}'$ is orthogonal to $\mathbf{u}$, then half of the electrons will measure spin up when measured by an $SG\mathbf{u}'$. But again by 4, the electrons in this spin up along $\mathbf{u}'$ stream will each measure spin up along $\mathbf{u}$ with probability $\frac{1}{2}$. Thus, the measurement with $SG\mathbf{u}'$ has changed the definite probabilities announced in belief 2. M agrees with P and concludes that any successful model of spin will have to include a concept of measurements as "operating" on spin-states.

Since belief 1 says that there are two distinct measurement outcomes (spin up and spin down), while belief 2 suggests that general states are some kind of combination of these possibilities, it seems reasonable to look for a model based on a 2-dimensional vector space. Roughly speaking, a basis should correspond to states of definite spin up and spin down (announced by belief 3), while a general state should be a linear combination of the basis. Moreover, belief 4 (which arises from experimental data) reminds P of Malus' law about the intensity of polarized light transmitted through a linear polarizer. M and P discuss this for a while, and after trying and failing with real vector spaces, they instead propose the following model involving the complex numbers.[3]

Definition 2.1. Spinor space *is a two-dimensional complex inner product space* $(W, \langle | \rangle)$*. The* spin-states *of an electron are represented by unit vectors in* W*, and two unit vectors represent the same spin-state if one is a scalar multiple of the other. That is, the unit vectors* $\mathbf{w}$ *and* $\mathbf{w}'$ *represent the same spin-state if and only if* $\mathbf{w} = e^{i\theta}\mathbf{w}'$ *for some* $\theta \in \mathbb{R}$*.*

Recall that a *complex inner product space* is a complex vector space W together with a function $\langle | \rangle \colon W \times W \to \mathbb{C}$ such that if $\alpha \in \mathbb{C}$ and $\mathbf{a}, \mathbf{a}', \mathbf{b} \in W$, then[4]

i) $\langle \alpha\mathbf{a} + \mathbf{a}' | \mathbf{b} \rangle = \alpha^* \langle \mathbf{a} | \mathbf{b} \rangle + \langle \mathbf{a}' | \mathbf{b} \rangle$ (*conjugate-linear in first component*)[5];

[3]The reader may well wonder why it is necessary to employ complex numbers. While there are a variety of reasons for the use of complex numbers in quantum mechanics, the immediate reason in terms of our current story is that if we were to use real numbers, then we would need to find a natural way of translating rotations of physical space $\mathbb{R}^3$ into rotations of $\mathbb{R}^2$. But there is no relationship between the rotation groups $SO(3)$ and $SO(2)$ that is suitable for this purpose. As we will see in the course of this chapter, there is an extremely elegant relationship between the groups $SO(3)$ and $SU(2)$, the analogue of the rotation group for $\mathbb{C}^2$, and this relationship plays a central role in the theory of quantum mechanical spin.

[4]Here, α^* denotes the complex conjugate of a complex number α.

[5]We follow the convention common in the physics literature of defining a complex inner product to be conjugate-linear in the first component and linear in the second. Mathematicians usually do the opposite, and take inner products to be linear in the first component and conjugate-linear in the second. This difference in conventions can lead to confusion, so beware.

ii) $\langle \mathbf{a}|\mathbf{b}\rangle = \langle \mathbf{b}|\mathbf{a}\rangle^*$ (*conjugate symmetry*);

iii) $\langle \mathbf{a}|\mathbf{a}\rangle \geq 0$ with equality if and only if $\mathbf{a} = 0$ (*positive definite*).

Exercise 2.2. *Show that conditions i) and ii) for an inner product imply linearity in the second component:* $\langle \mathbf{a}|\alpha\mathbf{b} + \mathbf{b}'\rangle = \alpha \langle \mathbf{a}|\mathbf{b}\rangle + \langle \mathbf{a}|\mathbf{b}'\rangle$.

As in the real case, an inner product on W determines a *norm* $\|\mathbf{a}\| := \sqrt{\langle \mathbf{a}|\mathbf{a}\rangle}$. Moreover, the Cauchy-Schwarz inequality holds: $|\langle \mathbf{a}|\mathbf{b}\rangle| \leq \|\mathbf{a}\|\|\mathbf{b}\|$ for all $\mathbf{a}, \mathbf{b} \in W$. Two vectors $\mathbf{a}$ and $\mathbf{b}$ are defined to be *orthogonal* in $(W, \langle|\rangle)$ if and only if $\langle \mathbf{a}|\mathbf{b}\rangle = 0$.

Exercise 2.3. *Suppose that* $(X, \langle|\rangle)$ *is a complex inner product space. Show that the inner product* $\langle|\rangle$ *is uniquely determined by the corresponding norm. (Hint: compute* $\|\mathbf{x}_1 + \mathbf{x}_2\|^2$ *and* $\|\mathbf{x}_1 + i\mathbf{x}_2\|^2$*. Compare exercise 1.3).*

Example 2.4. *Fix an integer* $n \geq 1$*, and consider the set* $\mathbb{C}^n$ *of all* n*-tuples of complex numbers:*

$$\mathbb{C}^n := \{(\alpha_1, \alpha_2, \ldots, \alpha_n) \mid \alpha_i \in \mathbb{C}\}.$$

The set $\mathbb{C}^n$ *is an* n*-dimensional complex vector space under the operations of component-wise addition and scalar multiplication. Moreover, define the* dot product *of two vectors in* $\mathbb{C}^n$ *by the formula*

$$(\alpha_1, \alpha_2, \ldots, \alpha_n) \cdot (\beta_1, \beta_2, \ldots, \beta_n) := \sum_{i=1}^{n} \alpha_i^* \beta_i.$$

Then $\cdot\colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ *defines an inner product on* $\mathbb{C}^n$*. The complex inner product space* $(\mathbb{C}^n, \cdot)$ *is called* complex Euclidean n-space.

After unpacking all of this terminology, M reiterates the meaning of definition 2.1: the possible spin-states of an electron are given by the unit vectors in $(W, \langle|\rangle)$, where two unit vectors correspond to the same spin-state if and only if they differ by a complex number of modulus 1, called a *phase*. This phase ambiguity is somewhat mysterious at this point, and the first order of business is to find a way of producing some quantities associated to spin-states in a phase-independent way.

To this end, we now present some convenient and powerful notation, introduced by P.A.M. Dirac in 1939 and popularized in his classic textbook [2]. If we use the symbol ψ to denote a spin-state, then ψ is actually an equivalence class of unit vectors in $(W, \langle|\rangle)$. Nevertheless, we will generally think of ψ as an actual unit vector, always remembering that the vector is only well defined up to a phase $e^{i\theta}$. We will often write $|\psi\rangle$ instead of ψ to emphasize that we have chosen a unit vector in W to represent the spin-state. The unit vector $|\psi\rangle$ is called a *ket*, being the latter half of a bracket $\langle \mathbf{a}|\mathbf{b}\rangle$. Hence, every ket determines a unique spin-state, but each spin-state is represented by infinitely many kets, any two of which differ by a phase $e^{i\theta}$. For each ket $|\psi\rangle$, there is a corresponding *bra* $\langle\psi|$, which is the linear mapping from W to $\mathbb{C}$ given by the inner product[6]:

$$\langle\psi|(\mathbf{a}) := \langle\psi|\mathbf{a}\rangle = \langle\mathbf{a}|\psi\rangle^*.$$

We must remember that the inner product $\langle\psi|\mathbf{a}\rangle$ depends on a choice of a representing ket $|\psi\rangle$, and not only on the spin-state.

Exercise 2.5 (♣[7]). *Show that the bra corresponding to the ket* $e^{i\theta}|\psi\rangle$ *is* $e^{-i\theta}\langle\psi|$. *More generally, if* $|\psi\rangle = c_1|\phi_1\rangle + c_2|\phi_2\rangle$ *is a complex linear combination, then* $\langle\psi| = c_1^*\langle\phi_1| + c_2^*\langle\phi_2|$.

The next proposition eliminates the phase ambiguity in our description by showing that spin-states are in one-to-one correspondence with orthogonal projections onto lines in W.

Proposition 2.6. *There is a one-to-one correspondence between spin-states and rank one orthogonal projections on* W*:*

$$\psi \longleftrightarrow P_\psi = |\psi\rangle\langle\psi|.$$

Here, $P_\psi\colon W \to W$ *is given by the formula* $P_\psi(\mathbf{a}) = \langle\psi|\mathbf{a}\rangle|\psi\rangle$.

Proof. Suppose that ψ is a spin-state, and choose a representing ket $|\psi\rangle$. Any other ket representing ψ is of the form $e^{i\theta}|\psi\rangle$, and hence spans the same complex line $\mathbb{C}|\psi\rangle$ contained in W. Any $\mathbf{a} \in W$ can be written uniquely as $\mathbf{a} = \mathbf{b} + \mathbf{b}^\perp$ for $\mathbf{b} \in \mathbb{C}|\psi\rangle$ and $\mathbf{b}^\perp \in (\mathbb{C}|\psi\rangle)^\perp$, the orthogonal complement to $\mathbb{C}|\psi\rangle$. In terms of this decomposition, the orthogonal projection onto the line $\mathbb{C}|\psi\rangle$ is defined as $P_\psi(\mathbf{a}) = P_\psi(\mathbf{b} + \mathbf{b}^\perp) := \mathbf{b}$ (see figure 2.5). But note that $\mathbf{b} = \langle\psi|\mathbf{a}\rangle|\psi\rangle =: |\psi\rangle\langle\psi|(\mathbf{a})$, so that $P_\psi = |\psi\rangle\langle\psi|$ as claimed.

Conversely, suppose that $P\colon W \to W$ is any rank one orthogonal projection on W. Choose a basis ket $|\psi\rangle$ for the 1-dimensional range of P, and note that any other choice differs from $|\psi\rangle$ by a phase $e^{i\theta}$. It follows that P determines a unique spin-state ψ such that $P = P_\psi$. □

Note that if two kets differ by a phase, then so will their projections onto the line spanned by ψ:

$$P_\psi(e^{i\theta}|\phi\rangle) = e^{i\theta}P_\psi(|\phi\rangle).$$

We can eliminate this phase dependence by taking the squared norm of the projections:

$$\|P_\psi(e^{i\theta}|\phi\rangle)\|^2 = \|P_\psi(|\phi\rangle)\|^2 = |\langle\psi|\phi\rangle|^2|\langle\psi|\psi\rangle|^2 = |\langle\psi|\phi\rangle|^2,$$

where we have used the fact that $e^{i\theta}$ has modulus 1 and $|\psi\rangle$ has unit norm. Thus, we have suceeded in producing a quantity that depends only on the

[6] That is, $\langle\psi| := \langle\psi|-\rangle\colon W \to \mathbb{C}$ is an element of the *dual space* of W.

[7] ♣ indicates an exercise with a solution in appendix A.4.

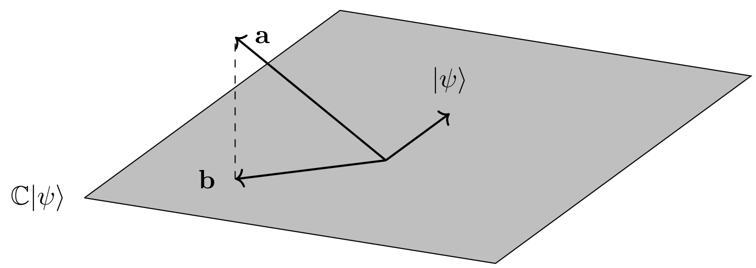

FIGURE 2.5: Orthogonal projection onto the complex line spanned by $|\psi\rangle$, with one real dimension suppressed.

spin-states ψ and ϕ, and not on the choice of representing kets. Note that the real number $|\langle\psi|\phi\rangle|^2$ is between 0 and 1, since it is the squared length of the orthogonal projection of a unit vector (alternatively, this follows from the Cauchy-Schwarz inequality). Hence, we may interpret this number as a probability, as recorded in the following interpretation, fundamental to all that follows.

> **Probability Interpretation:** *Given two spin-states ψ and ϕ, the probability that ϕ, when measured with a "ψ-device", will be found in the state ψ is given by the squared modulus of the inner product $|\langle\psi|\phi\rangle|^2$.*

But what kind of thing is a "ψ-device"? To answer this and to make a connection with their list of beliefs coming from the Stern-Gerlach experiments, M and P need to establish a connection between spinor space $(W, \langle|\rangle)$ and physical space $(V, \langle,\rangle)$. Observer P is eager to help, so as in the previous chapter, he chooses a right-handed orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for V which yields the identification of physical space with $(\mathbb{R}^3, \cdot)$ together with the rotation action of $SO(3)$. Henceforth, we will denote a Stern-Gerlach device $SG\mathbf{u}_3$ by SGz since such a device is oriented along P's positive z-axis if P labels his three coordinate axes by x, y, z as usual.

Having thus installed his SGz, every electron that P measures comes out either spin up or spin down along the z-direction (beliefs 1 and 3). These two outcomes correspond to two distinct spin-states, represented by kets $|+z\rangle$ and $|-z\rangle$ in W, uniquely determined up to individual multiplication by phases. Moreover, these kets form an orthonormal basis of W. Indeed, belief 3 says that the probability of $|+z\rangle$ being found in the state $|+z\rangle$ upon measurement is 1, while the probability of it being found in the state $|-z\rangle$ is 0. Since these probabilities are given by the squared absolute values of the inner products, the kets are orthonormal as claimed.

Thus, an arbitrary ket $|\phi\rangle$ in W can be written uniquely as $|\phi\rangle = c_+|+z\rangle + c_-|-z\rangle$, where c_+, c_- are complex numbers satisfying $|c_+|^2 + |c_-|^2 = 1$. Moreover, the probability that an electron with spin-state

ϕ will be spin up when measured with P's SGz is given by

$$|\langle +z|\phi\rangle|^2 = |\langle +z|c_+|+z\rangle + \langle +z|c_-|-z\rangle|^2 = |c_+|^2.$$

Similarly, $|c_-|^2$ is the probability that ϕ will be spin down when measured by an SGz. Thus, the model nicely captures belief 2: general spin-states are *superpositions* of the spin up and spin down states, with coefficients that determine the definite probabilities for measurements. The complex coefficients c_+, c_- are called *probability amplitudes*.

The preceding discussion shows that every unit vector in physical space yields an ordered pair of orthogonal spin-states. In particular, P's choice of basis for physical space (together with his installation of an SGz) has determined an orthonormal basis for spinor-space W, up to phases. Continuing with her investigation from Chapter 1, observer M wonders how a different choice of basis for V would change the basis for W? Before we take up her question, we pause to remind the reader of some different types of linear operators on complex Euclidean n-space.

2.3 Complex linear operators and matrix groups

This section follows the pattern of section 1.2, extending the results obtained there for real linear operators to the complex case. Most of the proofs extend easily to the complex situation, so we only briefly mention the necessary changes.

Suppose that $L\colon \mathbb{C}^n \to \mathbb{C}^n$ is a linear operator, and let $\varepsilon := \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ denote the standard basis of $\mathbb{C}^n$ (see example 2.4) . Then L is represented (with respect to ε) by an $n \times n$ matrix of complex numbers, which we also denote by L:

$$L = [L_{ij}] \qquad \text{where} \qquad L(\mathbf{e}_j) = \sum_{i=1}^{n} L_{ij}\mathbf{e}_i.$$

The proof of proposition 1.7 works just as well in the complex case to show that L is an automorphism of $\mathbb{C}^n$ if and only if the matrix L is invertible.

Definition 2.7. *The group of all invertible $n \times n$ complex matrices, denoted $GL(n, \mathbb{C})$, is called the* complex general linear group*. It is the symmetry group of the vector space $\mathbb{C}^n$.*

Definition 2.8. *If B is a complex $m \times n$ matrix, then its* conjugate transpose *$B^\dagger$ is the $n \times m$ matrix obtained by taking the ordinary transpose of B and then replacing each entry with its complex conjugate:*

$$(B^\dagger)_{ij} := B^*_{ji}.$$

Proposition 2.9. *Suppose that* $L \in GL(n, \mathbb{C})$. *Then* L *preserves the dot product on* $\mathbb{C}^n$ *if and only if* $L^{-1} = L^\dagger$, *the conjugate transpose of the matrix* L.

Proof. Adapt the proof of proposition 1.9 by replacing the ordinary transpose with the conjugate transpose. □

Definition 2.10. *The group of all invertible* $n \times n$ *complex matrices satisfying* $L^{-1} = L^\dagger$, *denoted* $U(n)$, *is called the* unitary group. *It is the symmetry group of complex Euclidean* n*-space* $(\mathbb{C}^n, \cdot)$.

Exercise 2.11. *Suppose that* $L \colon \mathbb{C}^n \to \mathbb{C}^n$ *is a linear operator. Show that* $L \in U(n)$ *if and only if* L *preserves the norm of vectors:* $\|L\mathbf{w}\| = \|\mathbf{w}\|$ *for all* $\mathbf{w} \in \mathbb{C}^n$. *(Hint: use proposition 2.9 and compute* $\|L(\mathbf{v} + \mathbf{w})\|^2$ *and* $\|L(\mathbf{v} + i\mathbf{w})\|^2$. *Compare exercise 1.11.)*

Proposition 2.12. *The determinant of any unitary matrix is a complex number of modulus 1.*

Proof. If L is unitary, then $I_n = L^\dagger L$. Taking the determinant of both sides yields $1 = \det(I_n) = \det(L^\dagger L) = \det(L^\dagger)\det(L) = \det(L)^* \det(L) = |\det(L)|^2$. It follows that $|\det(L)| = 1$ as claimed. □

Definition 2.13. *The* special unitary group *is the subgroup* $SU(n) \subset U(n)$ *of unitary matrices with determinant 1.*

Example 2.14. *The case* $n = 1$ *is somewhat more interesting in the complex case than in the real case:*

- $GL(1, \mathbb{C}) = \mathbb{C}^\times$, *the group of nonzero complex numbers;*
- $U(1) = \{e^{i\theta} \mid \theta \in \mathbb{R}\}$, *the phase group;*
- $SU(1) = \{1\}$, *the trivial group.*

Exercise 2.15. *Show that the unitary group* $U(1)$ *is isomorphic to the special orthogonal group* $SO(2)$.

Example 2.16. *The complex general linear group* $GL(2, \mathbb{C})$ *consists of* 2×2 *complex matrices with nonzero determinant:*

$$GL(2, \mathbb{C}) = \left\{ \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \mid \alpha\delta - \beta\gamma \neq 0 \right\}.$$

As in the real case, a 2×2 *complex matrix is an element of* $GL(2, \mathbb{C})$ *if and only if its columns form a basis for* $\mathbb{C}^2$.

Writing out the unitarity condition $I_2 = L^\dagger L$ *we find*

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} |\alpha|^2 + |\gamma|^2 & \alpha^*\beta + \gamma^*\delta \\ \beta^*\alpha + \delta^*\gamma & |\beta|^2 + |\delta|^2 \end{bmatrix}.$$

Thus, the conditions for L to be unitary are $|\alpha|^2 + |\gamma|^2 = |\beta|^2 + |\delta|^2 = 1$ and $\alpha^*\beta + \gamma^*\delta = 0$. Interpreting these relations as dot products, we see that L is unitary exactly when the columns of L have unit norm and are orthogonal to each other. Thus, a 2×2 complex matrix is an element of $U(2)$ if and only if its columns form an orthonormal basis for $(\mathbb{C}^2, \cdot)$.

If in addition we require that $\det(L) = 1$, then using the formula for the inverse of a 2×2 matrix, we find that the unitarity condition $L^\dagger = L^{-1}$ becomes

$$\begin{bmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{bmatrix} = \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix} \iff \alpha = \delta^* \quad \text{and} \quad \beta = -\gamma^*.$$

Thus, we find that the special unitary group is

$$SU(2) = \left\{ \begin{bmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{bmatrix} \mid |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

This group will play a major role in the remainder of our story.

Exercise 2.17. *Generalize the analysis in the preceding example to show that for all $n \geq 1$:*

a) an $n \times n$ complex matrix L is in $GL(n, \mathbb{C})$ if and only if the columns of L form a basis for $\mathbb{C}^n$;

b) an $n \times n$ complex matrix L is in $U(n)$ if and only if the columns of L form an orthonormal basis for $(\mathbb{C}^n, \cdot)$.

c) Let $L\colon \mathbb{C}^n \to \mathbb{C}^n$ be a linear operator, and $\gamma := \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be any orthonormal basis for $(\mathbb{C}^n, \cdot)$. Denote by $[L]_\gamma$ the matrix representing L in the basis γ. Then L is a unitary transformation if and only if $[L]_\gamma^\dagger = [L]_\gamma^{-1}$. (Hint: adapt the proof of proposition 1.17.)

Now we rejoin observers M and P, who are still puzzling over spinor space. Recall M's question: how would a different choice of basis for physical space V affect the basis for spinor space W obtained by the installation of a Stern-Gerlach device along the third axis? To study this question, P continues to consider the basis $\gamma := \{|{+z}\rangle, |{-z}\rangle\}$ for W coming from his third basis vector $\mathbf{u}_3$ in physical space. But M considers the different orthonormal basis $\gamma' := \{|{+z'}\rangle, |{-z'}\rangle\}$ for W that arises from her basis β' for V, together with her installation of an SGz' oriented along her third basis vector $\mathbf{u}_3'$. Just as in Chapter 1, these two orthonormal bases for W determine distinct isomorphisms $\Phi, \Phi'\colon (W, \langle|\rangle) \to (\mathbb{C}^2, \cdot)$, defined by sending γ, γ' respectively to the standard basis of $\mathbb{C}^2$. As before, in order to determine the translation between their two descriptions of spinor space, we consider the automorphism of $(\mathbb{C}^2, \cdot)$ defined by the composition $\Phi' \circ \Phi^{-1}$. The situation is pictured in the following diagram:

$$\begin{array}{ccc} W & = & W \\ \Phi \downarrow & & \downarrow \Phi' \\ \mathbb{C}^2 & \xrightarrow{\Phi' \circ \Phi^{-1}} & \mathbb{C}^2. \end{array}$$

By the discussion above, the automorphism $\Phi' \circ \Phi^{-1}$ may be identified with an invertible matrix of complex numbers, i.e. an element of $GL(2, \mathbb{C})$. Moreover, because the automorphism $\Phi' \circ \Phi^{-1}$ preserves the dot product, the corresponding matrix is actually an element of the unitary group $U(2)$.

Exercise 2.18. *Show that the matrix of $\Phi' \circ \Phi^{-1}$ with respect to the standard basis on $\mathbb{C}^2$ is the* change of basis matrix *from γ to γ'. (Compare exercise 1.18.)*

Going further, M observes that by exploiting phases, she can change her basis kets (without changing the corresponding spin states) to arrange for the matrix to be an element of the special unitary group $SU(2)$. Indeed, setting $\delta := \det(\Phi' \circ \Phi^{-1})$, proposition 2.12 shows that $|\delta| = 1$.

Exercise 2.19. *Choose a square root of δ, and denote it by $\sqrt{\delta}$. Show that $\sqrt{\delta} \in U(1)$, and hence may be used as a phase. Then consider the orthonormal basis of W given by $\gamma'' := \left\{\sqrt{\delta}\,|{+z'}\rangle, \sqrt{\delta}\,|{-z'}\rangle\right\}$. This basis, while distinct from γ', corresponds to the same pair of orthogonal spin-states. As usual, sending the basis γ'' to the standard basis of $\mathbb{C}^2$ determines an isomorphism $\Phi''\colon (W, \langle|\rangle) \to (\mathbb{C}^2, \cdot)$. Use exercise 2.18 to show that*

$$\Phi'' \circ \Phi^{-1} = \frac{1}{\sqrt{\delta}} \Phi' \circ \Phi,$$

and conclude that $\Phi'' \circ \Phi^{-1}$ is an element of $SU(2)$.

We now assume (after adjustment by a phase as above) that M's basis kets $|{\pm z'}\rangle$ yield an automorphism $\Phi' \circ \Phi^{-1}$ which is an element of the special unitary group $SU(2)$. Following in the pattern of his comments about $SO(3)$ in Chapter 1, observer P summarizes the situation as follows: his choice of the z-basis γ leads to the identification $\Phi\colon (W, \langle|\rangle) \to (\mathbb{C}^2, \cdot)$, sending γ to the standard basis of $\mathbb{C}^2$. Then M's choice of basis γ' determines an automorphism $\mathcal{B}$ of $(W, \langle|\rangle)$ defined by $\mathcal{B}(|{+z}\rangle) = |{+z'}\rangle$ and $\mathcal{B}(|{-z}\rangle) = |{-z'}\rangle$. In terms of P's description, the automorphism $\mathcal{B}$ is represented with respect to the basis γ by a matrix $B := [\mathcal{B}]_\gamma$. Explicitly, we have $B := [\beta_{ij}]$ where

$$|{+z'}\rangle = \beta_{11}|{+z}\rangle + \beta_{21}|{-z}\rangle \quad \text{and} \quad |{-z'}\rangle = \beta_{12}|{+z}\rangle + \beta_{22}|{-z}\rangle.$$

It follows from exercise 2.18 that B is the inverse of the special unitary matrix $\Phi' \circ \Phi^{-1}$. Thus, the matrix $B \in SU(2)$ that describes (with respect to γ) the automorphism that moves the z-basis onto the z'-basis is the inverse of the change of basis matrix describing the translation from P's description to M's. Conversely, an arbitrary element of $SU(2)$ will describe for P a way of superposing the z-basis to obtain a new orthonormal basis for W, hence a new pair of orthogonal spin-states.

Thus, $SU(2)$ acts on spinor space $(\mathbb{C}^2, \cdot)$ similarly to the way $SO(3)$ acts on physical space $(\mathbb{R}^3, \cdot)$. But M wants an answer to the following question:

if the rotation $A \in SO(3)$ connects P's right-handed orthonormal basis for V to M's, how can we determine the corresponding matrix $B \in SU(2)$ that connects P's z-basis for W to M's z'-basis? This is really a question about the relationship between the action of $SO(3)$ on $\mathbb{R}^3$ and the action of $SU(2)$ on $\mathbb{C}^2$. Before considering the actions, we should ask a more basic question: what is the relationship between the groups themselves?

2.4 The geometry of $SU(2)$

The group $SU(2)$ consists of unitary 2×2 matrices with determinant 1. Thus, the 2×2 complex matrix B is in $SU(2)$ if and only if $B^\dagger = B^{-1}$ and $\det(B) = 1$. In example 2.16, we discovered the explicit form of these matrices:

$$B = \begin{bmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{bmatrix} \qquad \alpha, \beta \in \mathbb{C} \text{ and } |\alpha|^2 + |\beta|^2 = 1.$$

If we write $\alpha = a_1 + ia_2$ and $\beta = b_1 + ib_2$ for $a_j, b_j \in \mathbb{R}$, the condition on α and β becomes

$$a_1^2 + a_2^2 + b_1^2 + b_2^2 = 1,$$

which defines the unit sphere $S^3 \subset \mathbb{R}^4$. Thus, as a topological space, $SU(2)$ is the three-dimensional unit sphere. In particular, it is path connected and simply connected[8].

In an effort to establish a connection between the groups $SU(2)$ and $SO(3)$, we would like to discover a natural way in which $SU(2)$ acts on $\mathbb{R}^3$ via rotations. We begin with the observation that $SU(2)$ acts on itself by conjugation: $(B, M) \mapsto B \star M := BMB^{-1}$ defines an $SU(2)$-action on $SU(2)$ in the sense of definition 1.20. Thinking of the copy of $SU(2)$ being acted upon as the 3-sphere S^3, we have an action of $SU(2)$ on S^3. This is close to what we want, because the *tangent space* to S^3 at any point is a copy of $\mathbb{R}^3$.

2.4.1 The tangent space to the circle $U(1) = S^1$

In order to motivate and clarify the notion of the tangent space to $SU(2) = S^3 \subset \mathbb{R}^4$, consider the simpler case of the group $U(1)$ consisting of complex numbers of modulus 1. Note that a complex number $z = x + iy \in U(1)$ if and only if $x^2 + y^2 = 1$, so $U(1)$ is the unit circle $S^1 \subset \mathbb{R}^2$.

From figure 2.6, it is clear that the tangent space to the circle S^1 at the point $(1, 0)$ is the vertical line $x = 1$. Since we want our tangent spaces to

[8]Path connected means that any two elements of S^3 may be joined by a continuous path in S^3. Simply connected means that all loops in S^3 may be continuously contracted to a point, which formalizes the idea that S^3 has "no holes".

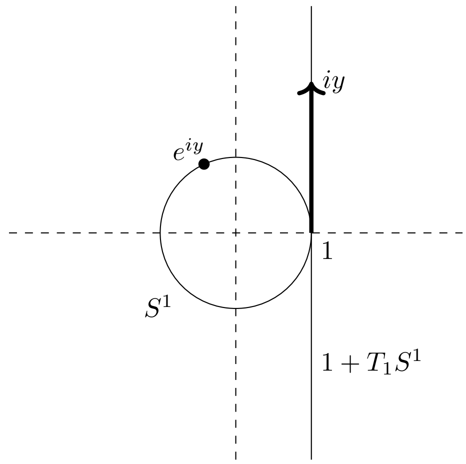

FIGURE 2.6: The tangent line to the circle S^1 at the identity.

be closed under vector addition and scalar multiplication, we will refer to the vertical line $x = 1$ as the *translated tangent space*, and use the term *tangent space* for the vertical line $x = 0$, which is a vector subspace of $\mathbb{R}^2$. Since the point $(1, 0)$ corresponds to the identity element $z = 1$ of the group $U(1)$, we denote this tangent space by T_1S^1 and write

$$T_1S^1 = \{iy \mid y \in \mathbb{R}\} = i\mathbb{R} \subset \mathbb{C}.$$

The translated tangent space is then given by

$$1 + T_1S^1 = 1 + i\mathbb{R} \subset \mathbb{C}.$$

But how could we determine this tangent space without relying on the picture? Well, suppose that $c : (-\epsilon, \epsilon) \to \mathbb{C} = \mathbb{R}^2$ is a one-to-one differentiable curve with the property that $c(t) \in U(1) = S^1$ for all t and $c(0) = 1 \in U(1)$. That is, c is a parametrization of the curve S^1 near the identity. Then the derivative $\dot{c}(0) \in \mathbb{R}^2$ is a tangent vector to S^1 at the identity. The totality of all such tangent vectors forms the tangent line T_1S^1. But note that we have $1 = c(t)c(t)^*$ for all t, since each $c(t) \in U(1)$ is a complex number of modulus 1. Taking the derivative with respect to t and evaluating at $t = 0$ yields:

$$\begin{aligned} 0 &= \dot{c}(0)c(0)^* + c(0)\dot{c}(0)^* \\ &= \dot{c}(0) + \dot{c}(0)^*. \end{aligned}$$

We see that the derivative $\dot{c}(0)$ must be a purely imaginary complex number: $\dot{c}(0) = iy$ for some $y \in \mathbb{R}$. Thus, this computation reproduces the description of the tangent space T_1S^1 provided above.

Note that every element of $i\mathbb{R}$ does indeed arise from a curve c. Indeed, for any $y \in \mathbb{R}$, consider the curve $c(t) := e^{ity} \in U(1)$. Then $c(0) = 1$ and $\dot{c}(0) = iy$.

2.4.2 The tangent space to the sphere $SU(2) = S^3$

We wish to determine the tangent space at the identity of $SU(2)$ by following the strategy described for the circle group $U(1)$ in the previous section. Recall that $SU(2) = S^3 \subset \mathbb{R}^4$. It will be convenient to make the explicit identification of the element $(x_0, x_1, x_2, x_3) \in \mathbb{R}^4$ with the matrix

$$\begin{bmatrix} x_0 + ix_3 & x_2 + ix_1 \\ -x_2 + ix_1 & x_0 - ix_3 \end{bmatrix}.$$

Such a matrix is in $SU(2)$ if and only if $x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1$, which defines the three-dimensional sphere $S^3 \subset \mathbb{R}^4$. The identity matrix I corresponds to the point $(1, 0, 0, 0)$, and we denote the tangent space to S^3 at this point by $T_I S^3$.

So suppose that $c : (-\epsilon, \epsilon) \to \mathbb{R}^4$ is a one-to-one differentiable curve satisfying $c(t) \in SU(2) = S^3$ for all t and $c(0) = I \in SU(2)$. Then as in the case of the circle, the derivative $\dot{c}(0) \in \mathbb{R}^4$ is a tangent vector to S^3 at the identity, and the totality of all such tangent vectors forms the tangent space $T_I S^3$. But since the curve c lies entirely within $SU(2)$, we have $I = c(t)c(t)^\dagger$ for all t. Taking the derivative and evaluating at $t = 0$ yields

$$\begin{aligned} 0 &= \dot{c}(0)c(0)^\dagger + c(0)\dot{c}(0)^\dagger \\ &= \dot{c}(0) + \dot{c}(0)^\dagger. \end{aligned}$$

From this, we see that the derivative must be a skew-hermitian matrix: $\dot{c}(0)^\dagger = -\dot{c}(0)$. But $\dot{c}(0)$ corresponds to a vector $(x_0, x_1, x_2, x_3) \in \mathbb{R}^4$, and this vector yields a skew-hermitian matrix if and only if $x_0 = 0$. It follows that

$$\dot{c}(0) = \begin{bmatrix} ix_3 & x_2 + ix_1 \\ -x_2 + ix_1 & -ix_3 \end{bmatrix} \qquad \text{for some } x_1, x_2, x_3 \in \mathbb{R}.$$

Note that this matrix has trace zero in addition to being skew-hermitian. We will show below that every such matrix is the tangent vector of some curve c, so that the tangent space is

$$T_I S^3 = \left\{ \begin{bmatrix} ix_3 & x_2 + ix_1 \\ -x_2 + ix_1 & -ix_3 \end{bmatrix} \mid x_1, x_2, x_3 \in \mathbb{R} \right\}.$$

Recall that in the case of $U(1) = S^1$, we found that the tangent space was i times the vector space $\mathbb{R}$. Following in this pattern, consider the real vector space $H_0(2)$ of 2×2 hermitian matrices with trace zero. Thus, $X \in H_0(2)$ if and only if $X^\dagger = X$ and $\text{tr}(X) = 0$.

Exercise 2.20. *Show that a general element of $H_0(2)$ looks like*

$$X = \begin{bmatrix} x_3 & x_1 - ix_2 \\ x_1 + ix_2 & -x_3 \end{bmatrix} \qquad x_1, x_2, x_3 \in \mathbb{R}.$$

Check that the tangent space to S^3 at the identity may be described as $T_I S^3 = iH_0(2)$.

As promised, we now wish show that every element of $iH_0(2)$ does arise from a curve c in the manner described above. Recall how we did this for the circle at the end of the previous section: given an element $iy \in i\mathbb{R}$, we wrote down the curve $c(t) := e^{ity}$ with tangent vector iy at $t = 0$. We can make an analogous argument for the sphere S^3 provided we have a suitable exponential function for matrices. In the next section we show that such a function exists. For now we will simply assume the existence of the matrix exponential function together with the properties listed below in proposition 2.21. So suppose that $X \in H_0(2)$ is an arbitrary 2×2 hermitian matrix with trace zero. Then define $c(t) := \exp(itX)$, which defines a differentiable curve in the space of 2×2 matrices with complex entries. In fact, in proposition 2.31 we will see that $c(t) \in SU(2)$ for all t, so that we have a curve in $SU(2) = S^3$ as desired. Moreover, $c(0) = I$ and $\dot{c}(0) = iX \in iH_0(2)$, thus showing that $iX \in T_I S^3$.

2.4.3 The exponential of a matrix

If A is an arbitrary $n\times n$ complex matrix, we want to define a matrix $\exp(A)$ with properties that generalize the usual exponential for complex numbers. We will do so by making use of the power series for the ordinary exponential function, replacing the scalar argument by a matrix. In order to justify this construction, we will need to extend some familiar analytic results to the setting of matrices, which is the purpose of this section. For the convenience of readers wishing to skip the analytic justification, we begin with a proposition that lists the basic properties of the matrix exponential; after reading proposition 2.21, the reader can safely jump to proposition 2.31.

Proposition 2.21. *There exists a function* $\exp\colon M(n,\mathbb{C}) \to GL(n,\mathbb{C})$ *that assigns to each* $n \times n$ *matrix* A *an invertible matrix* $\exp(A)$ *defined by the following absolutely convergent power series:*

$$\exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!} A^k.$$

This matrix exponential function *satisfies the following properties:*

a) $\exp(0) = I_n$*;*

b) If $AB = BA$ *then* $\exp(A + B) = \exp(A)\exp(B) = \exp(B)\exp(A)$*;*

c) $\exp(A)^{-1} = \exp(-A)$*;*

d) If B *is invertible, then* $\exp(BAB^{-1}) = B\exp(A)B^{-1}$*;*

e) For a fixed matrix A*, the function* $c\colon \mathbb{R} \to GL(n,\mathbb{C})$ *defined by* $c(t) := \exp(tA)$ *is differentiable, and* $\dot{c}(t) = Ac(t)$ *for all* $t \in \mathbb{R}$*. In particular,* $\dot{c}(0) = A$*.*

Bibliography

[1] M.L. Boas. *Mathematical Methods in the Physical Sciences*. John Wiley & Sons, New Jersey, 3rd edition, 2006.

[2] P.A.M. Dirac. *The Principles of Quantum Mechanics*. Oxford UP, London, 4th edition, 1958.

[3] L.D. Faddeev and O.A. Yakubovskiĭ. *Lectures on Quantum Mechanics for Mathematics Students*, volume 47 of *Student Mathematical Library*. American Mathematical Society, Providence, 2009.

[4] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics. Wiley-Interscience, New York, 2nd edition, 1999.

[5] G.B Folland. *Quantum Field Theory: A Tourist Guide for Mathematicians*, volume 149 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, 2008.

[6] S. Friedberg, A. Insel, and L. Spence. *Linear Algebra*. Prentice Hall, New Jersey, 4th edition, 2003.

[7] W. Fulton and J. Harris. *Representation Theory: A First Course*. Springer, New York, 1991.

[8] J. Gallier and J. Quaintance. Notes on differential geometry and lie groups. Available at `http://www.cis.upenn.edu/~jean/gbooks/manif.html`. Book in progress.

[9] D.J. Griffiths. *Introduction to Quantum Mechanics*. Prentice Hall, New Jersey, 1995.

[10] B.C. Hall. *Quantum Theory for Mathematicians*, volume 267 of *Graduate Texts in Mathematics*. Springer, New York, 2013.

[11] B.C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, volume 222 of *Graduate Texts in Mathematics*. Springer, New York, 2nd edition, 2015.

[12] P.R. Halmos. *Finite-Dimensional Vector Spaces*. Undergraduate Texts in Mathematics. Springer, New York, 1987.

[13] J.H. Hubbard and B.H. West. *Differential Equations: A Dynamical Systems Approach (Higher-Dimensional Systems)*, volume 18 of *Texts in Applied Mathematics*. Springer, New York, 1995.

[14] T. Lancaster and S.J. Blundell. *Quantum Field Theory for the Gifted Amateur*. Oxford UP, Oxford, 2014.

[15] S.F. Singer. *Linearity, Symmetry, and Prediction in the Hydrogen Atom*. Undergraduate Texts in Mathematics. Springer, New York, 2005.

[16] M. Spivak. *Calculus on Manifolds*. Addison-Wesley, New York, 1965.

[17] M. Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 1. Publish or Perish, Houston, 3rd edition, 1999.

[18] J. Stillwell. *Naive Lie Theory*. Undergraduate Texts in Mathematics. Springer, New York, 2008.

[19] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, 5th edition, 2016.

[20] L. Takhtajan. *Quantum Mechanics for Mathematicians*, volume 95 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2008.

[21] B. Tekin. Stern-gerlach experiment with higher spins. *arXiv preprint physics.atom-ph*, 2015. http://arxiv.org/abs/1506.04632.

[22] J.S. Townsend. *A Modern Approach to Quantum Mechanics*. McGraw-Hill, New York, 1992.

[23] S. Weinberg. *The Quantum Theory of Fields*, volume I. Cambridge University Press, Cambridge, 1995.

[24] E. Zeidler. *Quantum Field Theory I: Basics in Mathematics and Physics*. Springer, Berlin, 2006.

[25] E. Zeidler. *Quantum Field Theory II: Quantum Electrodynamics*. Springer, Berlin, 2009.

[26] E. Zeidler. *Quantum Field Theory III: Gauge Theory*. Springer, Berlin, 2011.