

Tensors and Tensor Products for Physicists

Nadir Jeevanjee

NOTE: This is a rough draft and I would appreciate feedback. Please send any and all questions, comments, criticisms, or suggestions to jeevanje@berkeley.edu. Last updated 1/10/2007.

Contents

Preface	4
1 Vector Spaces	6
1.1 Definition and Examples	6
1.2 Span, Linear Independence and Bases	10
1.3 Components	13
1.4 Linear Operators	15
1.5 Dual Spaces	19
1.6 Non-Degenerate Hermitian Forms	21
1.7 Non-Degenerate Hermitian Forms and Dual Spaces	26
Chapter 1 Problems	29
2 Tensors	33
2.1 Definition and Examples	33
2.2 Change of Basis	36
2.3 The Tensor Product - Definition and Properties	43
2.4 Tensor Products of V and V^*	45
2.5 Applications of the Tensor Product in Classical Physics	48
2.6 Applications of the Tensor Product in Quantum Physics	50
2.7 Symmetric and Antisymmetric Tensors	58
Chapter 2 Problems	68
3 Groups, Lie Groups and Lie Algebras	71
3.1 Groups - Definition and Examples	72
3.2 Homomorphism and Isomorphism	82
Chapter 3 Problems	84
Dirac Dictionary	86

Preface

This text is written at the beginning graduate level, and aims to demystify tensors and provide a unified framework for understanding them in all the different contexts in which they arise in physics. The word tensor is ubiquitous in physics (stress tensor, moment of inertia tensor, field tensor, metric tensor, tensor product, etc. etc.) and yet tensors are rarely defined carefully (if at all), and the definition usually has to do with transformation properties, making it difficult to get a feel for these objects. Furthermore, physics texts at the beginning graduate level usually only deal with tensors in their component form, so students wonder what the difference is between a second rank tensor and a matrix, and why new, enigmatic terminology is introduced for something they've already seen. The irony of this situation is that a proper understanding of tensors doesn't require much more mathematics than what most students encounter as undergraduates, and the clarity gained pays dividends far outweighing the modest investment. This text introduces just enough linear algebra to set the stage for tensors, with plenty of examples to keep the discussion grounded. After laying the necessary linear algebraic foundations, we give the modern (component-free) definition of tensors, followed by applications. Exercises and problems are included, and the exercises in particular should be done as they arise, or at least carefully considered, as they often fill out the text and provide good practice in using the definitions.

It should be said that this text aims to be simultaneously intuitive and rigorous. Thus, although much of the language (especially in the examples) is informal, almost all the definitions given are precise and are the same as one would find in a pure math text. This may put off the less mathematically inclined reader; I hope, however, that such a reader will work through his or her discomfort and develop the necessary mathematical sophistication, as the results will be well worth it. As for prerequisites, it is assumed the reader has been through the usual undergraduate physics curriculum, including a "mathematical methods for physicists" course (with at least a cursory treatment of vectors and matrices), as well as the standard upper division courses in classical mechanics, quantum mechanics, and relativity. Any

undergraduate versed in those topics, as well as any graduate student in physics, should be able to read this text. To undergraduates who are eager to learn about tensors but haven't yet completed the standard curriculum, I apologize; many of the examples and practically all of the motivation for the text come from those courses, and to assume no knowledge of those topics would preclude discussion of the many applications that motivated me to write this text. Hopefully, such students will return to this text once they have completed their upper-division coursework, and find it useful then.

Besides the aforementioned prerequisites I've also indulged in the use of some very basic mathematical shorthand for brevity's sake; a guide is below. Enjoy!

Some Mathematical Shorthand

\mathbb{R}	The set of real numbers
\mathbb{C}	The set of complex numbers
\mathbb{Z}	the set of positive and negative integers
\in	"is an element of", "an element of", i.e. $2 \in \mathbb{R}$ reads "2 is an element of the real numbers"
\notin	"is not an element of"
\forall	"for all"
\subset	"is a subset of", "a subset of"
\equiv	denotes a definition
$A \times B$	The set $\{(a, b)\}$ of all ordered pairs where $a \in A, b \in B$. Referred to as the <i>cartesian product</i> of sets A and B . Extends in the obvious way to n -fold products $A_1 \times \dots \times A_n$.
\mathbb{R}^n	$\underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{n \text{ times}}$
\mathbb{C}^n	$\underbrace{\mathbb{C} \times \dots \times \mathbb{C}}_{n \text{ times}}$

Chapter 1

Vector Spaces

Since tensors are basically a special class of functions defined on vector spaces, we must have a good foundation in linear algebra before discussing them. In particular, one needs a little bit more linear algebra than is covered in most sophomore or junior level linear algebra/ODE courses. This chapter starts with the familiar material about vectors, bases, linear operators etc. but eventually moves on to slightly more sophisticated topics that are essential for understanding tensors in physics. As we lay this groundwork, hopefully the reader will also find our slightly more abstract viewpoint useful in clarifying the nature of many of the objects in physics he/she has already encountered.

1.1 Definition and Examples

We begin with the definition of an abstract vector space. We're taught as undergraduates to think of vectors as arrows with a head and a tail, or as ordered triples of real numbers, but physics, and especially quantum mechanics, requires a more abstract notion of vectors. Before reading the definition of an abstract vector space, keep in mind that the definition is supposed to distill all the essential features of vectors as we know them (like addition and scalar multiplication) while detaching the notion of a vector space from specific constructs, like ordered n -tuples of real or complex numbers (denoted as \mathbb{R}^n and \mathbb{C}^n respectively). The mathematical utility of this is that much of what we know about vector spaces depends only on the essential properties of addition and scalar multiplication, not on other properties particular to \mathbb{R}^n or \mathbb{C}^n . If we work in the abstract framework and then come across other mathematical objects that don't look like \mathbb{R}^n or \mathbb{C}^n but that are abstract vector spaces, then most everything we know about \mathbb{R}^n and \mathbb{C}^n will apply to these spaces as well.

Physics also forces us to use the abstract definition since many quantum mechanical vector spaces are infinite-dimensional and cannot be viewed as \mathbb{C}^n or \mathbb{R}^n for any n . An added dividend of the abstract approach is that we will learn to think about vector spaces independently of any basis, which will prove very useful.

That said, an *abstract vector space* is a set V (whose elements are called vectors), together with a set of scalars C (for us, C is always \mathbb{R} or \mathbb{C}) and operations of addition and scalar multiplication that satisfy the following axioms:

1. $v + w = w + v$ for all v, w in V (Commutativity)
2. $v + (w + x) = (v + w) + x$ for all v, w, x in V (Associativity)
3. There exists a unique vector 0 in V such that $v + 0 = v \quad \forall v \in V$
4. $\forall v \in V$ there is a unique vector $-v$ such that $v + (-v) = 0$
5. $c(v + w) = cv + cw \quad \forall v, w \in V, c \in C$ (Distributivity)
6. $1v = v$ for all v in V
7. $(c_1 + c_2)v = c_1v + c_2v$ for all scalars c_1, c_2 and vectors v
8. $(c_1c_2)v = c_1(c_2v)$ for all scalars c_1, c_2 and vectors v

Some parts of the definition may seem tedious or trivial, but they are just meant to ensure that the addition and scalar multiplication operations behave the way we expect them too. In determining whether a set is a vector space or not, one is usually most concerned with defining addition in such a way that the set is closed under addition and that axioms 3 and 4 are satisfied; most of the other axioms are so natural and obviously satisfied that one, in practice, rarely bothers to check them. That said, let's look at some examples from physics, most of which will recur throughout the text.

Example 1.1 \mathbb{R}^n

This is the most basic example of a vector space, and the one on which the abstract definition is modeled. Addition and scalar multiplication are defined in the usual way: for $v = (v^1, v^2, \dots, v^n)$, $w = (w^1, w^2, \dots, w^n)$ in \mathbb{R}^n , we have

$$(v^1, v^2, \dots, v^n) + (w^1, w^2, \dots, w^n) = (v^1 + w^1, v^2 + w^2, \dots, v^n + w^n) \quad (1.1)$$

and

$$c(v^1, v^2, \dots, v^n) = (cv^1, cv^2, \dots, cv^n) \quad (1.2)$$

and the reader should check that the axioms are satisfied. These spaces, of course, are basic in physics; \mathbb{R}^3 is the usual 3 dimensional space we live in, \mathbb{R}^4 is spacetime

in special relativity, and \mathbb{R}^n for higher n occurs in classical physics as configuration spaces for multiparticle systems (i.e. \mathbb{R}^6 is the configuration space in the classic two-body problem, as you need six coordinates to specify the position of two particles in three-dimensional space).

Example 1.2 \mathbb{C}^n

This is another basic example - addition and scalar multiplication are defined as for \mathbb{R}^n , and the axioms are again straightforward to verify. Note, however that \mathbb{C}^n is a *complex* vector space, i.e. the set C in the definition is \mathbb{C} so scalar multiplication by complex numbers is defined, whereas \mathbb{R}^n is only a *real* vector space. This seemingly pedantic distinction can often end up being significant. \mathbb{C}^n occurs in physics primarily as the ket space for finite-dimensional quantum mechanical systems, such as particles with spin but without translational degrees of freedom. For instance, a spin 1/2 particle fixed in space has ket space \mathbb{C}^2 , and more generally a fixed particle with spin $n/2$ has ket space \mathbb{C}^{n+1} .

Example 1.3 $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$, $n \times n$ matrices with real or complex entries

The vector space structure of $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ is similar to that of \mathbb{R}^n and \mathbb{C}^n : denoting the entry in the i th row and j th column of a matrix A as A_{ij} , we define addition and (real) scalar multiplication for $A, B \in M_n(\mathbb{R})$ by

$$(A + B)_{ij} = A_{ij} + B_{ij} \tag{1.3}$$

$$(cA)_{ij} = cA_{ij} \tag{1.4}$$

i.e. addition and scalar multiplication are done component-wise. The same definitions are used for $M_n(\mathbb{C})$, which is of course a complex vector space. The reader can again check that the axioms are satisfied. Though these vector spaces don't appear explicitly in physics very often, they have many important subspaces, one of which we consider in the next example.

Example 1.4 $H_n(\mathbb{C})$, $n \times n$ Hermitian matrices with complex entries

$H_n(\mathbb{C})$, the set of all $n \times n$ hermitian matrices¹, is obviously a subset of $M_n(\mathbb{C})$, and in fact it is a *subspace* of $M_n(\mathbb{C})$ in that it forms a vector space itself. To show this it is not necessary to verify all of the axioms, since most of them are satisfied by virtue

¹hermitian matrices being those which satisfy $A^\dagger \equiv (A^T)^* = A$ where superscript T denotes the transpose and superscript $*$ denotes complex conjugation of the entries.

of $H_n(\mathbb{C})$ being a subset of $M_n(\mathbb{C})$; for instance, addition and scalar multiplication in $H_n(\mathbb{C})$ are just given by the restriction of those operations in $M_n(\mathbb{C})$ to $H_n(\mathbb{C})$, so the commutativity of addition and the distributivity of scalar multiplication over addition follow immediately. What does remain to be checked is that $H_n(\mathbb{C})$ is closed under addition and contains the zero “vector” (in this case, the zero matrix), both of which are easily verified. One interesting thing about $H_n(\mathbb{C})$ is that even though the entries of its matrices can be complex, it does *not* form a complex vector space; multiplying a hermitian matrix by i yields an *anti*-hermitian matrix, so $H_n(\mathbb{C})$ is *not* closed under complex scalar multiplication. As far as physical applications go, we know that physical observables in quantum mechanics are represented by hermitian operators, and if we are dealing with a finite dimensional ket space such as those mentioned in Example 2 then observables can be represented as elements of $H_n(\mathbb{C})$. As an example one can take a fixed spin 1/2 particle whose ket space is \mathbb{C}^2 ; the angular momentum operators are then represented as $L_i = \frac{1}{2}\sigma_i$, where the σ_i are the hermitian *Pauli matrices*

$$\sigma_x \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y \equiv \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.5)$$

Example 1.5 $L^2([a, b])$, *Square-integrable complex-valued functions on an interval*

This example is fundamental in quantum mechanics. A complex-valued function f on $[a, b] \subset \mathbb{R}$ is said to be *square-integrable* if

$$\int_a^b |f(x)|^2 dx < \infty. \quad (1.6)$$

Defining addition and scalar multiplication in the obvious way,

$$(f + g)(x) = f(x) + g(x) \quad (1.7)$$

$$(cf)(x) = cf(x), \quad (1.8)$$

and taking the zero element to be the function which is identically zero (i.e. $f(x) = 0$ for all x) yields a complex vector space. (Note that if we considered only real-valued functions then we would only have a real vector space). Verifying the axioms is straightforward though not entirely trivial, as one must show that the sum of two square integrable functions is again square-integrable (Problem 1). This vector space arises in quantum mechanics as the set of *normalizable* wavefunctions for a particle in a one-dimensional infinite potential well. Later on we’ll consider the more general

scenario where the particle may be unbound, in which case $a = -\infty$ and $b = \infty$ and the above definitions are otherwise unchanged. This vector space is denoted as $L^2(\mathbb{R})$.

Example 1.6 $Y_m^l(\theta, \phi)$ *The Spherical Harmonics*

Consider the set $P_l(\mathbb{R}^3)$ of all complex-coefficient polynomial functions on \mathbb{R}^3 of fixed degree l , i.e. all linear combinations of functions of the form $cx^i y^j z^k$ where $i + j + k = l$ and $c \in \mathbb{C}$. Addition and (complex) scalar multiplication are defined in the usual way and the axioms are again easily verified, so $P_l(\mathbb{R}^3)$ is a vector space. Now consider the vector subspace $\mathcal{H}_l(\mathbb{R}^3) \subset P_l(\mathbb{R}^3)$ of *harmonic* degree l polynomials, i.e. degree l polynomials satisfying $\Delta f = 0$ where Δ is the usual three dimensional Laplacian. The reader may be surprised to learn that the spherical harmonics of degree l are essentially elements of $\mathcal{H}_l(\mathbb{R}^3)$! To see the connection, consider the subset $\{x + iy, z, x - iy\} \subset H_1(\mathbb{R}^3)$. These functions are linear, hence clearly harmonic since Δ is a second-order differential operator. Writing them in spherical coordinates with polar angle θ and azimuthal angle ϕ gives $\{r e^{i\phi} \sin \theta, r \cos \theta, r e^{-i\phi} \sin \theta\}$, which up to normalization and the factor of r give the usual spherical harmonics with $l = 1$. The $l = 2$ case is treated in exercise 1.1 below. Spherical harmonics are discussed further throughout this text; for a complete discussion, see [ST].

Exercise 1.1 Consider the functions

$$(x + iy)^2, z(x + iy), x^2 + y^2 - 2z^2, z(x - iy), (x - iy)^2 \in P_2(\mathbb{R}^3). \quad (1.9)$$

Verify that they are in fact harmonic, and then write them in spherical coordinates and divide by r^2 to obtain, up to normalization, the familiar spherical harmonics for $l = 2$.

Non-example $GL(n, \mathbb{R})$, *invertible $n \times n$ matrices*

$GL(n, \mathbb{R})$, the subset of $M_n(\mathbb{R})$ consisting of invertible $n \times n$ matrices, is not a vector space though it seems like it could be. Why not?

1.2 Span, Linear Independence and Bases

The notion of a basis is probably familiar to most readers, at least intuitively: it's a set of vectors out of which we can 'make' all the other vectors in a given vector space V . In this section we'll make this idea precise and describe bases for some of the examples in the previous section.

First, we need the notion of the *span* of a set $S = \{v_1, v_2 \dots v_k\} \subset V$, denoted $\text{Span} \{v_1, v_2 \dots v_k\}$ or $\text{Span } S$: this is just the set of all vectors of the form $c^1 v_1 + c^2 v_2 + \dots + c^k v_k$. Such vectors are known as *linear combinations* of the v_i , so $\text{Span } S$ is just the set of all linear combinations of the vectors in S . For instance, if $S = \{(1, 0, 0), (0, 1, 0)\} \subset \mathbb{R}^3$, then $\text{Span } S$ is just the set of all vectors of the form $(c^1, c^2, 0)$ with $c^1, c^2 \in \mathbb{R}$. If S has infinitely many elements then the span of S is again all the linear combinations of vectors in S , though in this case the linear combinations can have an arbitrarily large (but finite) number of terms.²

Next we need the notion of linear dependence: a (not necessarily finite) set of vectors S is said to be *linearly dependent* if there exists distinct vectors v_1, v_2, \dots, v_k in S and scalars c^1, c^2, \dots, c^k , not all of which are 0, such that

$$c^1 v_1 + c^2 v_2 + \dots + c^k v_k = 0. \quad (1.10)$$

What this definition really means is that at least one vector in S can be written as a linear combination of the others, and in that sense is dependent (the reader should take a second to convince himself of this). If S is not linearly dependent then we say it is *linearly independent*, and in this case no nonzero vector in S can be written as a linear combination of any others. For instance, the set $S = \{(1, 0, 0), (0, 1, 0), (1, 1, 0)\} \subset \mathbb{R}^3$ is linearly dependent whereas the set $S' = \{(1, 0, 0), (0, 1, 0), (0, 1, 1)\}$ is linearly independent, as the reader can check.

With these definitions in place we can now define a *basis* for a vector space V as an *ordered* linearly independent set $\mathcal{B} \subset V$ whose span is all of V . This means, roughly speaking, that a basis has enough vectors to make all the others, but no more than that. When we say that $\mathcal{B} = \{v_1, \dots, v_k\}$ is an *ordered* set we mean that the order of the v_i is part of the definition of \mathcal{B} , so another basis with the same vectors but a different order is considered inequivalent. The reasons for this will become clear as we progress.

One can show³ that all finite bases must have the same number of elements, so we define the *dimension* of a vector space V , denoted $\dim V$, to be the number of elements of any finite basis. If no finite basis exists, then we say that V is *infinite dimensional*.

Exercise 1.2 Given a vector v and a finite basis $\mathcal{B} = \{e_i\}_{i=1 \dots n}$, show that the expression of v as a linear combination of the e_i is unique.

²We don't generally consider infinite linear combinations like $\sum_{i=1}^{\infty} c^i v_i = \lim_{N \rightarrow \infty} \sum_{i=1}^N c^i v_i$ because in that case we would need to consider whether the limit exists, i.e. whether the sum converges in some sense. More on this later.

³See [HK].

Example 1.7 \mathbb{R}^n and \mathbb{C}^n

\mathbb{R}^n has the following natural basis, also known as the *standard basis*:

$$\{1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, \dots, 1)\}. \quad (1.11)$$

The reader should check that this is indeed a basis, and thus that the dimension of \mathbb{R}^n is, unsurprisingly, n . The same set serves as a basis for \mathbb{C}^n , where of course now the linear combination coefficients c^i are allowed to be complex numbers. Note that although this basis is the most natural, there are infinitely many other perfectly respectable bases out there; the reader should check, for instance, that $\{(1, 1, 0, \dots, 0), (0, 1, 1, 0, \dots, 0), \dots, (0, \dots, 1, 1), (1, 0, \dots, 0, 1)\}$ is also a basis.

Example 1.8 $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$

Let E_{ij} be the $n \times n$ matrix with a 1 in the i th row, j th column and zeros everywhere else. Then the reader can check that $\{E_{ij}\}_{i,j=1\dots n}$ is a basis for both $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$, and that both spaces have dimension n^2 . Again, there are other nice bases out there; for instance, the symmetric matrices $S_{ij} \equiv E_{ij} + E_{ji}$, $i \leq j$, and antisymmetric matrices $A_{ij} \equiv E_{ij} - E_{ji}$, $i < j$ taken together also form a basis for both $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$.

Exercise 1.3 Let $S_n(\mathbb{R})$, $A_n(\mathbb{R})$ be the sets of $n \times n$ symmetric and antisymmetric matrices, respectively. Show that both are real vector spaces, compute their dimensions, and check that $\dim S_n(\mathbb{R}) + \dim A_n(\mathbb{R}) = \dim M_n(\mathbb{R})$, as expected.

Example 1.9 $H_2(\mathbb{C})$

Let's find a basis for $H_2(\mathbb{C})$. First, we need to know what a general element of $H_2(\mathbb{C})$ looks like. In terms of complex components, the condition $A = A^\dagger$ reads

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix} \quad (1.12)$$

where the bar denotes complex conjugation. This means that $a, d \in \mathbb{R}$ and $b = \bar{c}$, so in terms of real numbers we can write a general element of $H_2(\mathbb{C})$ as

$$\begin{pmatrix} t + z & x - iy \\ x + iy & t - z \end{pmatrix} = tI + x\sigma_x + y\sigma_y + z\sigma_z \quad (1.13)$$

where I is the identity matrix and $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices defined in (1.5). The reader can easily check that the set $\mathcal{B} = \{I, \sigma_x, \sigma_y, \sigma_z\}$ is linearly independent, and since (1.13) shows that \mathcal{B} spans $H_2(\mathbb{C})$, \mathcal{B} is a basis for $H_2(\mathbb{C})$. We also see that $H_2(\mathbb{C})$ has (real) dimension 4.

Exercise 1.4 Using the matrices E_{ij} and A_{ij} from example 1.8, construct a basis for $H_n(\mathbb{C})$ and compute its dimension.

Example 1.10 $Y_m^l(\theta, \phi)$

We saw in the previous section that the Y_m^l , or more precisely $r^l Y_m^l$, are elements of the vector space $\mathcal{H}_l(\mathbb{R}^3)$. What's more is that the set $\{r^l Y_m^l\}_{-l \leq m \leq l}$ is actually a basis for $\mathcal{H}_l(\mathbb{R}^3)$. In the case $l = 1$ this is clear since $\mathcal{H}_1(\mathbb{R}^3) = P_1(\mathbb{R}^3)$ and clearly $\{x + iy, x - iy, z\}$ is a basis for $P_1(\mathbb{R}^3)$. For $l > 1$ proving our claim requires a little more effort; see Problem 2. Another simpler basis for $P_1(\mathbb{R}^3)$ would be the *cartesian basis* $\{x, y, z\}$; physicists use the spherical harmonic basis because those functions are eigenfunctions of the orbital angular momentum operator L_z , which on $\mathcal{H}_l(\mathbb{R}^3)$ is represented by $L_z = -i(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x})$. We shall discuss the relationship between the two bases in detail later.

Not Quite Example $L^2([-a, a])$

From doing 1-D problems in quantum mechanics one already 'knows' that the set $\{e^{i \frac{n\pi x}{a}}\}_{n \in \mathbb{Z}}$ is a basis for $L^2([-a, a])$. There's a problem, however; we're used to taking infinite linear combinations of these basis vectors, but our definition above only allows for *finite* linear combinations. What's going on here? It turns out that $L^2([-a, a])$ has more structure than your average vector space: it is an infinite dimensional *Hilbert Space*, and for such spaces we have a generalized definition of a basis, one that allows for infinite linear combinations. We will discuss Hilbert spaces in section 1.6

1.3 Components

One of the most useful things about introducing a basis for a vector space is that it allows us to write elements of the vector space as n -tuples, in the form of either column or row vectors, as follows: Given $v \in V$ and a basis $\mathcal{B} = \{e_i\}_{i=1 \dots n}$ for V , we can write

$$v = \sum_{i=1}^n v^i e_i \tag{1.14}$$

for some numbers v^i , called the *components of v with respect to \mathcal{B}* . We can then represent v by the column vector, denoted $[v]_{\mathcal{B}}$, as

$$[v]_{\mathcal{B}} = \begin{pmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ v^n \end{pmatrix} \quad (1.15)$$

or the row vector

$$[v]_{\mathcal{B}}^T = (v^1, v^2, \dots, v^n) \quad (1.16)$$

where the superscript T denotes the usual transpose of a vector and where the subscript \mathcal{B} just reminds us which basis the components are referred to, and will be dropped if there is no ambiguity. With a choice of basis, then, every n -dimensional vector space can be made to ‘look like’ \mathbb{R}^n or \mathbb{C}^n . Writing vectors in this way greatly facilitates computation, as we’ll see. One must keep in mind, however, that vectors exist independently of any chosen basis, and that their expressions as row or column vectors depend very much on a *choice* of basis \mathcal{B} . This will be evident in the examples.

Example 1.11 E^3 3-D Euclidean space

We said in Example 1.1 that \mathbb{R}^3 represented the 3-dimensional space we inhabit, but this is not quite true. After all, given $(x, y, z) \in \mathbb{R}^3$, what point of space is this referring to? The space we live in is not quite \mathbb{R}^3 but rather 3-dimensional Euclidean space E^3 , defined as follows: pick a point in space, call it the *origin*. Then there exists a directed line segment (arrow) from the origin to every other point in space. *We identify the points in space with the arrows pointing to them from the origin.* These arrows can then be added together head to tail and multiplied by real numbers by scaling, yielding a real vector space. If we pick a basis of three vectors, call it $K = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$, we can then, for any point in space, write the corresponding vector $v = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ as $[v]_K^T = (x, y, z)$, hence identifying E^3 with \mathbb{R}^3 as described above.

Example 1.12 Rigid Body Motion

One area of physics where the distinction between a vector and its expression as an ordered triple is crucial is rigid body motion. In this setting our vector space is E^3 and we usually deal with two bases, an arbitrary but fixed *space axes* $K' = \{\hat{\mathbf{x}}', \hat{\mathbf{y}}', \hat{\mathbf{z}}'\}$

and a time-dependent *body axes* $K = \{\hat{\mathbf{x}}(t), \hat{\mathbf{y}}(t), \hat{\mathbf{z}}(t)\}$ which is fixed relative to the rigid body. When we write down vectors in E^3 , like the angular momentum vector L or the angular velocity vector ω , we must keep in mind what basis we are using, as the component expressions will differ drastically depending on the choice of basis. For example, if there is no external torque on a rigid body, $[L]_{K'}$ will be constant whereas $[L]_K$ will in general be time-dependent.

Example 1.13 $L^2([-a, a])$

We know from experience in quantum mechanics that all square integrable functions on an interval $[-a, a]$ have an expansion⁴ $f = \sum_{m=-\infty}^{\infty} c_m e^{i\frac{m\pi x}{a}}$ in terms of the ‘basis’ $\{e^{i\frac{m\pi x}{a}}\}_{m \in \mathbb{Z}}$. This expansion is known as the *Fourier series* of f , and we see that the c_n , commonly known as the *Fourier coefficients*, are nothing but the components of the vector f in the basis $\{e^{i\frac{m\pi x}{a}}\}_{m \in \mathbb{Z}}$.

1.4 Linear Operators

One of the basic notions in linear algebra, fundamental in quantum mechanics, is that of a linear operator. A *linear operator* on a vector space V is a function T from V to itself satisfying the *linearity* condition

$$T(cv + w) = cT(v) + T(w). \quad (1.17)$$

Sometimes we write Tv instead of $T(v)$. The reader should check that the set of all linear operators on V forms a vector space, denoted $\mathcal{T}(V)$. The reader has doubtless seen many examples of linear operators: for instance, we can interpret a real $n \times n$ matrix as a linear operator on \mathbb{R}^n that acts on column vectors by matrix multiplication. Thus $M_n(\mathbb{R})$ (and, similarly, $M_n(\mathbb{C})$) can be viewed as vector spaces whose elements are themselves linear operators. In fact, that was exactly how we interpreted the vector subspace $H_2(\mathbb{C}) \subset M_2(\mathbb{C})$ in example 1.4; in that case, we identified elements of $H_2(\mathbb{C})$ as the quantum-mechanical angular momentum operators. There are numerous other examples of quantum mechanical linear operators - for instance, the familiar position and momentum operators \hat{x} and \hat{p} act on $L^2([-a, a])$ by

$$\hat{x}f(x) = xf(x) \quad (1.18)$$

$$\hat{p}f(x) = \frac{\hbar}{i} \frac{\partial f}{\partial x}. \quad (1.19)$$

Another class of less familiar examples is given below.

⁴This fact is proved in most real analysis books, see [R].

Example 1.14 $\mathcal{T}(V)$ acting on $\mathcal{T}(V)$

We are familiar with linear operators taking vectors into vectors, but they can also be used to take linear operators into linear operators, as follows: Given $A, B \in \mathcal{T}(V)$, we can define a linear operator $ad_A \in \mathcal{T}(\mathcal{T}(V))$ acting on B by

$$ad_A(B) \equiv [A, B] \tag{1.20}$$

where $[\cdot, \cdot]$ indicates commutator. This action of A on $\mathcal{T}(V)$ is called the *adjoint action* or *adjoint representation*. The adjoint representation has important applications in quantum mechanics; for instance, the Heisenberg picture emphasizes $\mathcal{T}(V)$ rather than V and interprets the Hamiltonian as an operator in the adjoint representation. In fact, for any observable A the Heisenberg equation of motion reads⁵

$$\frac{dA}{dt} = i ad_H(A). \tag{1.21}$$

□

One important property of a linear operator T is whether or not it is *invertible*, i.e. whether there exists a linear operator T^{-1} such that $TT^{-1} = T^{-1}T = I$ where I is the identity operator.⁶ The reader may recall that, in general, an inverse for a map T exists if and only if T is both *1-1*, meaning

$$T(v) = T(w) \implies v = w, \tag{1.22}$$

and *onto*, meaning that $\forall w$ there exists v such that $T(v) = w$. If this is unfamiliar the reader should take a moment to convince himself of this. In the case of a linear operator on a vector space, these two conditions are actually equivalent and turn out also (see exercise 1.6 below) to be equivalent to the statement

$$T(v) = 0 \implies v = 0, \tag{1.23}$$

so T is invertible if and only if the only vector it sends to 0 is the zero vector.

Exercise 1.5 Let T be a linear operator on a vector space V . Show that T being 1-1 is equivalent to T being onto. Feel free to introduce a basis to assist you in the proof.

⁵Normally there would be a factor of \hbar in the denominator in the right hand side of (1.21), but here and below we set $\hbar = 1$.

⁶Throughout this text I will denote the identity operator or identity matrix; it will be clear from context which is meant.

Exercise 1.6 Suppose $T(v) = 0 \implies v = 0$. Show that this is equivalent to T being 1-1, which by the previous exercise is equivalent to T being 1-1 *and* onto, which is then equivalent to T being invertible.

An important point to keep in mind is that a linear operator is *not* the same thing as a matrix; just as with vectors, the identification can only be made *once a basis is chosen*. For operators on finite-dimensional spaces this is done as follows: choose a basis $\mathcal{B} = \{e_i\}_{i=1\dots n}$. Then the action of T is determined by its action on the basis vectors,

$$T(v) = T\left(\sum_{i=1}^n v^i e_i\right) = \sum_{i=1}^n v^i T(e_i) = \sum_{i,j=1}^n v^i T_i^j e_j \quad (1.24)$$

where the numbers T_i^j , again called the *components of T with respect to \mathcal{B}* ,⁷ are defined by $T(e_i) = \sum_{j=1}^n T_i^j e_j$. We then have

$$[v]_{\mathcal{B}} = \begin{pmatrix} v^1 \\ v^2 \\ \cdot \\ \cdot \\ \cdot \\ v^n \end{pmatrix} \quad \text{and} \quad [T(v)]_{\mathcal{B}} = \begin{pmatrix} \sum_{i=1}^n v^i T_i^1 \\ \sum_{i=1}^n v^i T_i^2 \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^n v^i T_i^n \end{pmatrix} \quad (1.25)$$

so if we define the *matrix of T in the basis \mathcal{B}* as

$$[T]_{\mathcal{B}} = \begin{pmatrix} T_1^1 & T_2^1 & \dots & T_n^1 \\ T_1^2 & T_2^2 & \dots & T_n^2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ T_n^1 & T_n^2 & \dots & T_n^n \end{pmatrix} \quad (1.26)$$

we then have

$$[T(v)]_{\mathcal{B}} = [T]_{\mathcal{B}} [v]_{\mathcal{B}} \quad (1.27)$$

as a matrix equation. Thus, $[T]_{\mathcal{B}}$ really does represent T in components, where the action of $[T]_{\mathcal{B}}$ on vectors is by the usual matrix multiplication. Furthermore, if we

⁷Nomenclature to be justified in the next chapter.

have two linear operators A and B and we define their *product* (or *composition*) AB as the linear operator

$$(AB)(v) \equiv A(B(v)), \quad (1.28)$$

the reader can then show that $[AB] = [A][B]$. Thus, composition of operators becomes matrix multiplication of the corresponding matrices.

Exercise 1.7 For two linear operators A and B on a vector space V , show that $[AB] = [A][B]$ in any basis.

Example 1.15 $L_z, \mathcal{H}_l(\mathbb{R}^3)$ and Spherical Harmonics

Recall that $\mathcal{H}_1(\mathbb{R}^3)$ is the set of all linear functions on \mathbb{R}^3 and that $\{rY_m^1\}_{-1 \leq m \leq 1} = \{x + iy, z, x - iy\}$ and $\{x, y, z\}$ are both bases for this space. Now consider the familiar angular momentum operator $L_z = -i(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x})$ on this space. The reader can check that

$$L_z(x + iy) = x + iy \implies (L_z)_1^1 = 1, (L_z)_1^2 = (L_z)_1^3 = 0 \quad (1.29)$$

$$L_z(z) = 0 \implies (L_z)_2^i = 0 \forall i \quad (1.30)$$

$$L_z(x - iy) = x - iy \implies (L_z)_3^3 = -1, (L_z)_3^1 = (L_z)_3^2 = 0 \quad (1.31)$$

so in the spherical harmonic basis,

$$[L_z]_{\{rY_m^1\}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (1.32)$$

This of course just says that the wavefunctions $x + iy, z$ and $x - iy$ have L_z eigenvalues of 1, 0, and -1 respectively.

Meanwhile,

$$L_z(x) = iy \quad (1.33)$$

$$L_z(y) = -ix \quad (1.34)$$

$$L_z(z) = 0 \quad (1.35)$$

so in the cartesian basis,

$$[L_z]_{\{x,y,z\}} = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (1.36)$$

a very different looking matrix.

Exercise 1.8 Compute the matrices of $L_x = -i(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y})$ and $L_y = -i(z\frac{\partial}{\partial x} - x\frac{\partial}{\partial z})$ acting on $\mathcal{H}_1(\mathbb{R}^3)$ in both the cartesian and spherical harmonic bases.

Before concluding this section we should remark that there is much more one can say about linear operators, particularly concerning eigenvectors, eigenvalues and diagonalization. Though these topics are relevant for physics, we will not need them in this text and good references for them abound, so we omit them. The interested reader can consult the first chapter of [SA] for a practical introduction, or [HK] for a thorough discussion.

1.5 Dual Spaces

Another basic construction associated with a vector space, essential for understanding tensors and usually left out of the typical ‘mathematical methods for physicists’ courses, is that of a dual space. Given a vector space V with scalars C , its *dual space* V^* is defined to be the set of C -valued linear functions f on V (referred to as *dual vectors* or *linear functionals*), where ‘linear’ again means $f(cv + w) = cf(v) + f(w)$. It’s easily checked that the usual definitions of addition and scalar multiplication and the zero function turn V^* into a vector space over C . Given a (not necessarily finite) basis $\{e_i\}$ for V , any linear function f on V is determined entirely by its values on the e_i , since for any $v = \sum_{i=1}^n v^i e_i$

$$f(v) = f\left(\sum_{i=1}^n v^i e_i\right) \tag{1.37}$$

$$= \sum_{i=1}^n v^i f(e_i) \tag{1.38}$$

$$\equiv \sum_{i=1}^n v^i f_i \tag{1.39}$$

where the $f_i \equiv f(e_i)$ are, unsurprisingly, referred to as the *components* of f in the basis $\{e_i\}$. To justify this nomenclature, consider a set $\{e^i\}$ (note the raised indices) of dual vectors defined by

$$e^i(e_j) = \delta_j^i. \tag{1.40}$$

If V is finite-dimensional with dimension n , it's easy to check (by evaluating both sides on basis vectors) that we can write⁸

$$f = \sum_{i=1}^n f_i e^i \quad (1.41)$$

so that the f_i really are the components of f . Since f was arbitrary, this means that the e^i span V^* . In exercise 1.9 below the reader will show that the e^i are actually linearly independent, so $\{e^i\}_{i=1\dots n}$ is actually a *basis* for V^* . We sometimes say that the e^i are *dual to* the e_i . Note that we have shown that V and V^* always have the same dimension. Note also that for a given $v = \sum_{i=1}^n v^i e_i$,

$$e^i(v) = \sum_{j=1}^n v^j e^i(e_j) = \sum_{j=1}^n v^j \delta_j^i = v^i \quad (1.42)$$

so we can alternatively think of the i th component of a vector as the value of the i th dual vector on that vector.

Exercise 1.9 By carefully working with the definitions, show that the e^i defined in (1.40) are linearly independent.

Example 1.16 *Dual spaces of \mathbb{R}^n , \mathbb{C}^n , $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$*

Consider the basis $\{e_i\}$ of \mathbb{R}^n and \mathbb{C}^n , where e_i is the vector with a 1 in the i th place and 0's everywhere else; this is just the basis described in Example 1.7. Now consider the element f^j of V^* which eats a vector in \mathbb{R}^n or \mathbb{C}^n and spits out the j th component; clearly $f^j(e_i) = \delta_i^j$ so the f^j are just the dual vectors e^j described above. Similarly, for $M_n(\mathbb{R})$ or $M_n(\mathbb{C})$ consider the dual vector f^{ij} defined by $f^{ij}(A) = A_{ij}$; these vectors are clearly dual to the E_{ij} and thus form the corresponding dual basis. While the f^{ij} may seem a little unnatural or artificial, the reader should note that there is one linear functional on $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ which is familiar: the *Trace* functional, denoted Tr and defined by

$$Tr(A) = \sum_{i=1}^n A_{ii}. \quad (1.43)$$

What are the components of Tr with respect to the f^{ij} ?

⁸If V is infinite-dimensional then this may not work as the sum required may be infinite, and as mentioned before care must be taken in defining infinite linear combinations .

Not Quite Example *Dual space of $L^2([-a, a])$*

We haven't yet properly treated $L^2([-a, a])$ so we clearly cannot yet properly treat its dual, but we would like to point out here that in infinite dimensions, dual spaces get much more interesting. In finite dimensions, we saw above that a basis $\{e_i\}$ for V induces a dual basis $\{e^i\}$ for V^* , so in a sense V^* 'looks' very much like V . This is not true in infinite dimensions - in this case we still have linear functionals dual to a given basis, but these may not span the dual space. Consider the case of $L^2([-a, a])$; the reader can check that $\{e^n\}_{n \in \mathbb{Z}}$ defined by

$$e^n(f(x)) \equiv \frac{1}{2a} \int_{-a}^a e^{-i \frac{n\pi x}{a}} f(x) dx \quad (1.44)$$

satisfy $e^n(e^{i \frac{m\pi x}{a}}) = \delta_n^m$ and are hence dual to $\{e^{i \frac{m\pi x}{a}}\}_{m \in \mathbb{Z}}$. In fact, these linear functionals just eat a function and spit out its n th Fourier coefficient. There are linear functionals, however, that can't be written as a linear combination of the e^i ; one such linear functional is the *Dirac Delta Functional* δ , defined by

$$\delta(f(x)) \equiv f(0). \quad (1.45)$$

We will prove this claim and place δ in its proper context in section 1.7.

1.6 Non-Degenerate Hermitian Forms

Non-degenerate Hermitian forms, of which the Euclidean dot product, Minkowski metric and Hermitian scalar product of quantum mechanics are but a few examples, are very familiar to most physicists. We introduce them here not just to formalize their definition but also to make the fundamental but usually unacknowledged connection between these objects and dual spaces.

A *non-degenerate Hermitian form* on a vector space V is a C -valued function (\cdot, \cdot) which assigns to an ordered pair of vectors $v, w \in V$ a scalar, denoted (v, w) , having the following properties:

1. $(v, w_1 + cw_2) = (v, w_1) + c(v, w_2)$ (linearity in the second argument)
2. $(v, w) = \overline{(w, v)}$ (*Hermiticity*; the bar denotes complex conjugation)
3. For each $v \neq 0 \in V$, there exists $w \in V$ such that $(v, w) \neq 0$ (non-degeneracy)

Note that conditions 1 and 2 imply that $(cv, w) = \bar{c}(v, w)$, so (\cdot, \cdot) is *conjugate-linear* in the first argument. Also note that for a real vector space, condition 2

implies that (\cdot, \cdot) is *symmetric*, i.e. $(v, w) = (w, v)$ ⁹; in this case, (\cdot, \cdot) is called a *metric*. Condition 3 is a little nonintuitive but will be essential in the connection with dual spaces. If, in addition to the above 3 conditions, the Hermitian form obeys

4. $(v, v) > 0$ for all $v \in V, v \neq 0$ (positive-definiteness)

then we say that (\cdot, \cdot) is an *inner product*, and a vector space with such a Hermitian form is called an *inner product space*. Note that condition 4 implies 3. Our reason for separating condition 4 from the rest of the definition will become clear when we consider the examples. One very important use of non-degenerate Hermitian forms is to define preferred sets of bases known as *orthonormal bases*. Such bases $\mathcal{B} = \{e_i\}$ by definition satisfy $(e_i, e_j) = \pm\delta_{ij}$ and are extremely useful for computation, and ubiquitous in physics for that reason. If (\cdot, \cdot) is positive-definite (hence an inner product), then orthonormal basis vectors satisfy $(e_i, e_j) = \delta_{ij}$ and may be constructed out of arbitrary bases by the Gram-Schmidt process. If (\cdot, \cdot) is not positive-definite then orthonormal bases may still be constructed out of arbitrary bases, though the process is slightly more involved. See [HK, sec. 8.2, 10.2] for details.

Exercise 1.10 Let (\cdot, \cdot) be an inner product. If a set of vectors e_1, \dots, e_k is *orthogonal*, i.e. $(e_i, e_j) = 0$ when $i \neq j$, show that they are linearly independent. Note that an *orthonormal* set (i.e. $(e_i, e_j) = \pm\delta_{ij}$) is just an orthogonal set in which the vectors have unit length.

Example 1.17 *The dot product (or Euclidean metric) on \mathbb{R}^n*

Let $v = (v^1, \dots, v^k), w = (w^1, \dots, w^k) \in \mathbb{R}^n$. Define (\cdot, \cdot) on \mathbb{R}^n by

$$(v, w) \equiv \sum_{i=1}^n v^i w^i. \quad (1.46)$$

This is sometimes written as $v \cdot w$. The reader can check that (\cdot, \cdot) is an inner product, and that the standard basis given in example 1.7 is an orthonormal basis.

Example 1.18 *The Hermitian scalar product on \mathbb{C}^n*

Let $v = (v^1, \dots, v^k), w = (w^1, \dots, w^k) \in \mathbb{C}^n$. Define (\cdot, \cdot) on \mathbb{C}^n by

$$(v, w) \equiv \sum_{i=1}^n \bar{v}^i w^i. \quad (1.47)$$

⁹In this case, (\cdot, \cdot) is linear in the first argument as well as the second and would be referred to as *bilinear*.

Again, the reader can check that (\cdot, \cdot) is an inner product, and that the standard basis given in example 1.7 is an orthonormal basis. Such inner products on complex vector spaces are sometimes referred to as *Hermitian scalar products* and are present on every quantum mechanical vector space. In this example we see the importance of condition 2, manifested in the conjugation of the v^i in (1.47); if that conjugation wasn't there, a vector like $v = (i, 0, \dots, 0)$ would have $(v, v) = -1$ and (\cdot, \cdot) wouldn't be an inner product.

Exercise 1.11 Let $A, B \in M_n(\mathbb{C})$. Define (\cdot, \cdot) on $M_n(\mathbb{C})$ by

$$(A, B) = \frac{1}{2} \text{Tr}(A^\dagger B). \quad (1.48)$$

Check that this is indeed an inner product. Also check that the basis $\{I, \sigma_x, \sigma_y, \sigma_z\}$ for $M_2(\mathbb{C})$ is orthonormal with respect to this inner product.

Example 1.19 *The Minkowski Metric on 4-D Spacetime*

Consider two vectors (often called “events” in the physics literature) $v_i = (x_i, y_i, z_i, t_i) \in \mathbb{R}^4$, $i = 1, 2$. The Minkowski metric on spacetime,¹⁰ denoted η , is defined to be¹¹

$$\eta(v_1, v_2) \equiv x_1x_2 + y_1y_2 + z_1z_2 - t_1t_2. \quad (1.49)$$

η is clearly linear in both its arguments (i.e. *bilinear*) and symmetric, hence satisfying conditions 1 and 2, and the reader will check condition 3 in exercise 1.12 below. Notice that for $v = (1, 0, 0, 1)$, $\eta(v, v) = 0$ so η is *not* positive-definite, hence not an inner product. This is why we separated condition 4, and considered the more general non-degenerate Hermitian forms instead of just inner products.

Exercise 1.12 Let $v = (x, y, z, t)$. Show that η is non-degenerate by finding another vector w such that $\eta(v, w) \neq 0$.

We should point out here that the Minkowski metric can be written in components as a matrix, just as a linear operator can. Taking the standard basis $\mathcal{B} = \{e_i\}_{i=1,2,3,4}$ in \mathbb{R}^4 , we can define the *components of η* , denoted η_{ij} , as

$$\eta_{ij} \equiv \eta(e_i, e_j). \quad (1.50)$$

¹⁰As in example 1.11, we can identify physical spacetime with \mathbb{R}^4 once we choose a coordinate system.

¹¹We are, of course arbitrarily choosing the $+++-$ signature, we could equally well choose $---+$.

Then, just as was done for linear operators, the reader can check that if we define the *matrix of η in the basis \mathcal{B}* , denoted $[\eta]_{\mathcal{B}}$, as the matrix

$$[\eta]_{\mathcal{B}} = \begin{pmatrix} \eta_{11} & \eta_{21} & \eta_{31} & \eta_{41} \\ \eta_{12} & \eta_{22} & \eta_{32} & \eta_{42} \\ \eta_{13} & \eta_{23} & \eta_{33} & \eta_{43} \\ \eta_{14} & \eta_{24} & \eta_{34} & \eta_{44} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (1.51)$$

we can write

$$\eta(v_1, v_2) = [v_1]^T [\eta] [v_2] = (x_1, y_1, z_1, t_1) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \\ t_2 \end{pmatrix} \quad (1.52)$$

as some readers may be used to from computations in relativity. Note that the symmetry of η implies that $[\eta]_{\mathcal{B}}$ is a symmetric matrix for any basis \mathcal{B} .

Example 1.20 *The Hermitian scalar product on $L^2([-a, a])$*

For $f, g \in L^2([-a, a])$, define

$$(f, g) \equiv \frac{1}{2a} \int_{-a}^a \bar{f}g \, dx. \quad (1.53)$$

The reader can easily check that this defines an inner product on $L^2([-a, a])$, and that $\{e^{i\frac{n\pi x}{a}}\}_{n \in \mathbb{Z}}$ is an orthonormal set. What's more, this inner product turns $L^2([-a, a])$ into a *Hilbert Space*, which is an inner product space that is *complete*. The notion of completeness is a technical one, so we will not give its precise definition, but in the case of $L^2([-a, a])$ one can think of it as meaning roughly that a limit of square-integrable functions is again square-integrable. Making this precise and proving it for $L^2([-a, a])$ is the subject of real analysis textbooks and far outside the scope of this text¹², so we'll content ourselves here with just mentioning completeness and noting that it is responsible for many of the nice features of Hilbert spaces, in particular the generalized notion of a basis which we now describe.

Given a Hilbert space \mathcal{H} and an orthonormal (and possibly infinite) set $\{e_i\} \subset \mathcal{H}$, the set $\{e_i\}$ is said to be an *orthonormal basis for \mathcal{H}* if

$$(e_i, f) = 0 \quad \forall i \implies f = 0. \quad (1.54)$$

¹²See [R], for instance, for this and for proofs of all the claims made in this example.

The reader can check (see exercise 1.13 below) that in the finite-dimensional case this definition is equivalent to our previous definition of an orthonormal basis. In the infinite-dimensional case, however, this definition differs substantially from the old one in that we no longer require $\text{Span}\{e_i\} = \mathcal{H}$ (recall that spans only include *finite* linear combinations). Does this mean, though, that we now allow *arbitrary* infinite combinations of the basis vectors? If not, which ones are allowed? For $L^2([-a, a])$, for which $\{e^{i\frac{n\pi x}{a}}\}_{n \in \mathbb{Z}}$ is an orthonormal basis, we mentioned in example 1.13 that any $f \in L^2([-a, a])$ can be written as

$$f = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{n\pi x}{a}} \quad (1.55)$$

where

$$\frac{1}{2a} \int_{-a}^a |f|^2 dx = \sum_{n=-\infty}^{\infty} |c_n|^2 < \infty \quad (1.56)$$

(The first equality should be familiar from quantum mechanics and follows from exercise 1.14 below). The converse to this is also true, and this is where the completeness of $L^2([-a, a])$ is essential: if a set of numbers c_n satisfy (1.56), then the series

$$g(x) \equiv \sum_{n=-\infty}^{\infty} c_n e^{i\frac{n\pi x}{a}} \quad (1.57)$$

converges, yielding a square-integrable function g . So $L^2([-a, a])$ is the set of all expressions of the form (1.55), *subject to the condition* (1.56). Now we know how to think about infinite-dimensional Hilbert spaces and their bases: a basis for a Hilbert space is an infinite set whose infinite linear combinations, *together with some suitable convergence condition*, form the entire vector space.

Exercise 1.13 Show that the definition (1.54) of a Hilbert space basis is equivalent to our original definition of a basis for a *finite*-dimensional inner product space V .

Exercise 1.14 Show that for $f = \sum_{n=-\infty}^{\infty} c_n e^{i\frac{n\pi x}{a}}$, $g = \sum_{m=-\infty}^{\infty} d_m e^{i\frac{m\pi x}{a}} \in L^2([-a, a])$,

$$(f, g) = \sum_{n=-\infty}^{\infty} \bar{c}_n d_n \quad (1.58)$$

so that (\cdot, \cdot) on $L^2([-a, a])$ can be viewed as the infinite-dimensional version of the standard Hermitian scalar product on \mathbb{C}^n .

1.7 Non-Degenerate Hermitian Forms and Dual Spaces

We are now ready to explore the connection between dual vectors and non-degenerate Hermitian forms. Given a non-degenerate Hermitian form (\cdot, \cdot) on a finite-dimensional vector space V , we can associate to any $v \in V$ a dual vector $\tilde{v} \in V^*$ defined by

$$\tilde{v}(w) \equiv (v, w). \quad (1.59)$$

Sometimes we write $\tilde{v} = (v, \cdot)$. This map we have just defined, call it L , from V to V^* is conjugate-linear since for $v = cx + z$, $v, z, x \in V$,

$$\tilde{v}(w) = (v, w) = (cx + z, w) = \bar{c}(x, w) + (z, w) = \bar{c}\tilde{x}(w) + \tilde{z}(w) \quad (1.60)$$

so

$$L(v) = L(cx + z) \equiv \tilde{v} = \bar{c}\tilde{x} + \tilde{z} = \bar{c}L(x) + L(z). \quad (1.61)$$

In exercise 1.15 below the reader will show that the non-degeneracy of (\cdot, \cdot) implies that L is 1-1 and onto, so L is an invertible map from V to V^* . The reader will see in the examples below that he is already familiar with L in a couple of different contexts.

Exercise 1.15 Use the non-degeneracy of (\cdot, \cdot) to show that L is 1-1, i.e. that $L(v) = L(w) \implies v = w$. Combine this with the argument used in exercise 1.6 to show that L is onto as well.

Exercise 1.16 Given a basis $\{e_i\}_{i=1\dots n}$, under what circumstances do we have $e^i = \tilde{e}_i$ for all i ?

Example 1.21 *Bras and kets in quantum mechanics*

Let \mathcal{H} be a quantum mechanical Hilbert space with inner product (\cdot, \cdot) . In Dirac notation, a vector $\psi \in \mathcal{H}$ is written as a ket $|\psi\rangle$ and the inner product (ψ, ϕ) is written $\langle\psi|\phi\rangle$. What about bras, written as $\langle\psi|$? What, exactly, are they? Most quantum mechanics texts gloss over their definition, just telling us that they are in 1-1 correspondence with kets and can be combined with kets as $\langle\psi|\phi\rangle$ to get a scalar. We are also told that the correspondence between bras and kets is conjugate-linear, i.e. that the bra corresponding to $c|\psi\rangle$ is $\bar{c}\langle\psi|$. From what we have seen in this section, it is now clear that bras really are dual vectors, labeled in the same way as regular vectors, because the map L allows us to identify the two. In short, $\langle\psi|$ is really just $L(\psi)$, or equivalently (ψ, \cdot) .

Example 1.22 *Raising and lowering indices in relativity*

Consider \mathbb{R}^4 with the Minkowski metric, let $\mathcal{B} = \{e_\mu\}_{\mu=1-4}$ and $\mathcal{B}' = \{e^\mu\}_{\mu=1-4}$ be the standard basis and dual basis for \mathbb{R}^4 (we use a greek index to conform with standard physics notation), and let $v = \sum_{\mu=1}^4 v^\mu e_\mu \in \mathbb{R}^4$. What are the components of \tilde{v} in terms of the v^μ ? Well, as we saw in section 1.5, the components of a dual vector are just given by evaluation on the basis vectors, so

$$\tilde{v}_\mu = \tilde{v}(e_\mu) = (v, e_\mu) = \sum_\nu v^\nu (e_\nu, e_\mu) = \sum_\nu v^\nu \eta_{\nu\mu}. \quad (1.62)$$

In matrices, this reads

$$[\tilde{v}]_{\mathcal{B}'} = [\eta]_{\mathcal{B}} [v]_{\mathcal{B}} \quad (1.63)$$

so matrix multiplication of a vector by the metric matrix gives the corresponding dual vector in the dual basis. Thus, the map L is implemented in coordinates by $[\eta]$. Now, we mentioned above that L is invertible; what does L^{-1} look like in coordinates? Well, by the above, L^{-1} should be given by matrix multiplication by $[\eta]^{-1}$, the matrix inverse to $[\eta]$. Denoting the components of this matrix by $\eta^{\mu\nu}$ (so that $\eta^{\tau\mu}\eta_{\nu\tau} = \delta_\nu^\mu$) and writing $\tilde{f} \equiv L^{-1}(f)$ where f is a dual vector, we have

$$[\tilde{f}] = [\eta]^{-1} [f] \quad (1.64)$$

or in components

$$\tilde{f}^\mu = \sum_\nu \eta^{\nu\mu} f_\nu. \quad (1.65)$$

Now, in physics one usually works with components of vectors, and in relativity the numbers v^μ are called the *contravariant components* of v and the numbers $v_\mu \equiv \sum_\nu v^\nu \eta_{\nu\mu}$ are referred to as the *covariant components* of v . We see now that the contravariant components of a vector are just its usual components, while its covariant components are actually the components of the associated dual vector \tilde{v} . For a dual vector f , the situation is reversed - the covariant components f_μ are its actual components, and the contravariant components are the components of \tilde{f} . Since L allows us to turn vectors into dual vectors and vice-versa, we usually don't bother trying to figure out whether something is 'really' a vector or a dual vector; it can be either, depending on which components we use.

The above discussion shows that the familiar process of "raising" and "lowering" indices is just the application of the map L (and its inverse) in components. For an interpretation of $[\eta]^{-1}$ as the matrix of a metric on \mathbb{R}^{4*} , see the Problems.

Exercise 1.17 Consider \mathbb{R}^3 with the Euclidean metric. Show that the covariant and contravariant components of a vector in an orthonormal basis are identical. This explains why we never bother with this terminology, nor the concept of dual spaces, in basic physics where \mathbb{R}^3 is the relevant vector space. Is the same true for \mathbb{R}^4 with the Minkowski metric?

Example 1.23 $L^2([-a, a])$ and its dual

In our above discussion of the map L we stipulated that V should be finite-dimensional. Why? If you examine the discussion closely, you'll see that the only place where we use the finite-dimensionality of V is in showing that L is onto. Does this mean that L is not necessarily onto in infinite dimensions? Consider the Dirac Delta functional $\delta \in L^2([-a, a])^*$. Does

$$\delta(g) = g(0) \stackrel{?}{=} (\delta(x), g) \tag{1.66}$$

for some function $\delta(x)$? If we write g as $g(x) = \sum_{n=-\infty}^{\infty} d_n e^{i\frac{n\pi x}{a}}$, then simply evaluating g at $x = 0$ gives

$$g(0) = \sum_{n=-\infty}^{\infty} d_n \stackrel{?}{=} (\delta(x), g) \tag{1.67}$$

which, when compared with (1.58), tells us that the function $\delta(x)$ must have fourier coefficients $c_n = 1$ for all n . Such c_n , however, do not satisfy (1.56) and hence $\delta(x)$ cannot be a square-integrable function. So the dual vector δ is not in the image of the map L , hence L is not onto in the case of $L^2([-a, a])$.

Chapter 1 Problems

1. Prove that $L^2([-a, a])$ is closed under addition. You'll need the triangle inequality, as well as the following inequality, valid for all $\lambda \in \mathbb{R}$: $0 \leq \int_{-a}^a (|f| + \lambda|g|)^2 dx$.
2. In this problem we show that $\{r^l Y_m^l\}$ is a basis for $\mathcal{H}_l(\mathbb{R}^3)$. We'll gloss over a few subtleties here; for a totally rigorous discussion see [ST].

- a) Let $f \in \mathcal{H}_l(\mathbb{R}^3)$. Argue that f can be written as $f = r^l Y(\theta, \phi)$. Then write $\Delta f = 0$ in spherical coordinates, separating out the angular part of the laplacian (which we denote as Δ_{S^2}) as

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \Delta_{S^2}. \quad (1.68)$$

You should find

$$\Delta_{S^2} Y = -l(l+1)Y. \quad (1.69)$$

- b) If you have never done so, show that

$$\Delta_{S^2} = L_x^2 + L_y^2 + L_z^2 \equiv \mathbf{L}^2 \quad (1.70)$$

so that (1.69) says that Y is an eigenfunction of \mathbf{L}^2 , as expected. The theory of angular momentum¹³ then tells us that $\mathcal{H}_l(\mathbb{R}^3)$ has dimension $2l+1$.

- c) Exhibit a basis for $\mathcal{H}_l(\mathbb{R}^3)$ by considering the function $f_l^l \equiv (x+iy)^l$ and showing that

$$L_z(f_l^l) = l f_l^l, \quad L_+(f_l^l) \equiv (L_x + iL_y)(f_l^l) = 0. \quad (1.71)$$

The theory of angular momentum then tells us that $(L_-)^m f_l^l \equiv f_{l-m}^l$ satisfies $L_z(f_{l-m}^l) = (l-m)f_{l-m}^l$ and that $\{f_m^l\}_{-l \leq m \leq l}$ is a basis for $\mathcal{H}_l(\mathbb{R}^3)$.

- d) Writing $f_m^l = r^l Y_m^l$ we see that Y_m^l satisfies $\mathbf{L}^2 Y_m^l = -l(l+1)Y_m^l$ and $L_z Y_m^l = m Y_m^l$ as expected. Now use this definition of Y_m^l to compute all the spherical harmonics for $l=1, 2$ and show that this agrees, up to normalization, with the spherical harmonics as tabulated in any quantum mechanics textbook. If you read example 1.6 and did exercise 1.1 then all you have to do is compute f_m^1 , $-1 \leq m \leq 1$ and f_m^2 , $-2 \leq m \leq 2$ and show that these functions agree with the ones given there.

¹³see [SA, G]

3. In discussions of quantum mechanics the reader may have heard the phrase “angular momentum generates rotations”. What this means is that if one takes a component of the angular momentum such as L_z and exponentiates it, i.e. if one considers the operator

$$\begin{aligned}\exp(-i\phi L_z) &\equiv \sum_{n=0}^{\infty} \frac{1}{n!} (-i\phi L_z)^n \\ &= I - i\phi L_z + \frac{1}{2!} (-i\phi L_z)^2 + \frac{1}{3!} (-i\phi L_z)^3 + \dots\end{aligned}\tag{1.72}$$

(the usual power series expansion for e^x) then one gets the operator which represents a rotation about the z axis by an angle ϕ . Confirm this in one instance by explicitly summing the power series for the operator $[L_z]_{\{x,y,z\}}$ of example 1.15 to get

$$\exp(-i[L_z]_{\{x,y,z\}}) = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix},\tag{1.73}$$

the usual matrix for a rotation about the z -axis.

4. a) Let A be a linear operator on a finite-dimensional complex vector space V with inner product (\cdot, \cdot) . Define the *adjoint* of A , denoted A^\dagger , by the equation

$$(A^\dagger v, w) = (v, Aw)\tag{1.74}$$

Show that in an orthonormal basis $\{e_i\}_{i=1\dots n}$, $[A^\dagger] = [A]^\dagger$, where the dagger outside the brackets denotes the usual conjugate transpose of a matrix. You may want to prove and use the fact $A_j^i = e^i(Ae_j)$.

- b) If A satisfies $A = A^\dagger$, A is then said to be *self-adjoint* or *hermitian*. Show that any eigenvalue of A must be real. Note then that in part a) you showed that *in an orthonormal basis* the matrix of a hermitian operator is a hermitian matrix. This is not necessarily true in a non-orthonormal basis.

5. Let g be a non-degenerate bilinear form on a vector space V (we have in mind the Euclidean metric on \mathbb{R}^3 or the Minkowski metric on \mathbb{R}^4). Pick an arbitrary (not necessarily orthonormal) basis, let $[g]^{-1}$ be the matrix inverse of $[g]$ in this basis, and write $g^{\mu\nu}$ for the components of $[g]^{-1}$. Also let $f, h \in V^*$. Define a non-degenerate bilinear form \tilde{g} on V^* by

$$\tilde{g}(f, h) \equiv g(\tilde{f}, \tilde{h}) \quad (1.75)$$

where $\tilde{f} = L^{-1}(f)$ as in example 1.22. Show that

$$\tilde{g}^{\mu\nu} \equiv \tilde{g}(e^\mu, e^\nu) = g^{\mu\nu} \quad (1.76)$$

so that $[g]^{-1}$ is truly a matrix representation of a non-degenerate bilinear form on V^* .

6. In this problem we'll acquaint ourselves with $P(\mathbb{R})$, the set of polynomials in one variable x with real coefficients. We'll also meet several bases for this space which the reader should find familiar.

- a) $P(\mathbb{R})$ is the set of all functions of the form

$$f(x) = c_0 + c_1x + c_2x^2 \dots + c_nx^n \quad (1.77)$$

where n is arbitrary. Verify that $P(\mathbb{R})$ is a (real) vector space. Then show that $P(\mathbb{R})$ is infinite-dimensional by showing that, for any finite set $S \subset P(\mathbb{R})$, there is a polynomial that is not in $\text{Span } S$. Exhibit a simple infinite basis for $P(\mathbb{R})$.

- b) Compute the matrix corresponding to the operator $\frac{d}{dx} \in \mathcal{T}(P(\mathbb{R}))$ with respect to the basis you found in part a).
 c) One can turn $P(\mathbb{R})$ into an inner product space by considering inner products of the form

$$(f, g) \equiv \int_a^b f(x)g(x)W(x) dx \quad (1.78)$$

where $W(x)$ is a nonnegative *weight function*. One can then take the basis $\mathcal{B} = \{1, x, x^2, x^3, \dots\}$ and apply the Gram-Schmidt process to get an orthogonal basis. With the proper choice of range of integration $[a, b]$ and weight function $W(x)$, we can obtain (up to normalization) the various orthogonal polynomials one meets in studying the various differential equations that arise in electrostatics and quantum mechanics.

- i) Let $[a, b] = [-1, 1]$ and $W(x) = 1$. Consider the set $S = \{1, x, x^2, x^3\} \subset \mathcal{B}$. Apply the Gram-Schmidt process to this set to get (up to normalization) the first four *Legendre Polynomials*

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x). \end{aligned}$$

The Legendre Polynomials show up in the solutions to the differential equation (1.69), where we make the identification $x = \cos \theta$. Since $-1 \leq \cos \theta \leq 1$, this explains the range of integration in the inner product.

- ii) Now let $[a, b] = (-\infty, \infty)$ and $W(x) = e^{-x^2}$. Gram-Schmidt S to get the first four *Hermite Polynomials*

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x. \end{aligned}$$

These polynomials arise in the solution to the Schrodinger equation for a one-dimensional harmonic oscillator. Note that the range of integration corresponds to the range of the position variable, as expected.

- iii) Finally, let $[a, b] = (0, \infty)$ and $W(x) = e^{-x}$. Again, Gram-Schmidt S to get the first four *Laguerre Polynomials*

$$\begin{aligned} L_0(x) &= 1 \\ L_1(x) &= -x + 1 \\ L_2(x) &= \frac{1}{2}(x^2 - 4x + 2) \\ L_3(x) &= \frac{1}{6}(-x^3 + 9x^2 - 18x + 6). \end{aligned}$$

These polynomials arise as solutions to the radial part of the Schrodinger equation for the Hydrogen atom. In this case x is interpreted as a radial variable, hence the range of integration $(0, \infty)$.

Chapter 2

Tensors

Now that we're familiar with vector spaces we can finally approach our main subject, tensors. We'll give the modern component-free definition, from which will follow the usual transformation laws that used to *be* the definition.

From here on out we will employ the *Einstein summation convention*, which is that whenever an index is repeated in an expression, once as a superscript and once as a subscript, then summation over that index is implied. Thus an expression like $v = \sum_{i=1}^n v^i e_i$ becomes $v = v^i e_i$. We'll comment on this convention in section 2.2.

2.1 Definition and Examples

A *tensor of type* (r, s) on a vector space V is a C -valued function T on

$$\underbrace{V \times \dots \times V}_r \times \underbrace{V^* \times \dots \times V^*}_s \quad (2.1)$$

which is linear in each argument, i.e.

$$T(v_1 + cw, v_2, \dots, v_r, f_1, \dots, f_s) = T(v_1, \dots, v_r, f_1, \dots, f_s) + cT(w, v_2, \dots, v_r, f_1, \dots, f_s) \quad (2.2)$$

and similarly for all the other arguments. This property is called *multilinearity*. Note that dual vectors are $(1, 0)$ tensors, and that vectors can be viewed as $(0, 1)$ tensors as follows:

$$v(f) \equiv f(v) \quad \text{where } v \in V, f \in V^*. \quad (2.3)$$

Similarly, linear operators can be viewed as $(1, 1)$ tensors as

$$A(v, f) \equiv f(Av). \quad (2.4)$$

We take $(0,0)$ tensors to be scalars, as a matter of convention. The reader will show in exercise 2.1 below that the set of all tensors of type (r,s) on a vector space V , denoted $\mathcal{T}_s^r(V)$ or just \mathcal{T}_s^r , form a vector space. This should not come as much of a surprise since we already know that vectors, dual vectors and linear operators all form vector spaces. Also, just as linearity implies that dual vectors and linear operators are determined by their values on the basis vectors, multilinearity implies the same thing for general tensors. To see this, let $\{e_i\}_{i=1\dots n}$ be a basis for V and $\{e^i\}_{i=1\dots n}$ the corresponding dual basis. Then, denoting the i th component of the vector v_p as v_p^i and the j th component of the dual vector f_q as f_{qj} , we have (by repeated application of multilinearity)

$$T(v_1, \dots, v_r, f_1, \dots, f_s) = v_1^{i_1} \dots v_r^{i_r} f_{1j_1} \dots f_{sj_s} T(e_{i_1}, \dots, e_{i_r}, e^{j_1}, \dots, e^{j_s}) \quad (2.5)$$

$$\equiv v_1^{i_1} \dots v_r^{i_r} f_{1j_1} \dots f_{sj_s} T_{i_1, \dots, i_r}^{j_1 \dots j_s} \quad (2.6)$$

where, as before, the numbers

$$T_{i_1, \dots, i_r}^{j_1 \dots j_s} \equiv T(e_{i_1}, \dots, e_{i_r}, e^{j_1}, \dots, e^{j_s}) \quad (2.7)$$

are referred to as the *components of T* in the basis $\{e_i\}_{i=1\dots n}$. The reader should check that this definition of the components of a tensor, when applied to vectors, dual vectors, and linear operators, agrees with the definitions given earlier. Also note that (2.7) gives us a concrete way to think about the components of tensors: they are the *values of the tensor on the basis vectors*.

Exercise 2.1 By choosing suitable definitions of addition and scalar multiplication, show that $\mathcal{T}_s^r(V)$ is a vector space.

If we have a non-degenerate bilinear form on V , then we may change the type of T by precomposing with the map L or L^{-1} . If T is of type $(1,1)$ with components T_i^j , for instance, then we may turn it into a tensor \tilde{T} of type $(2,0)$ by defining $\tilde{T}(v, w) = T(v, L(w))$. This corresponds to lowering the second index, and we write the components of \tilde{T} as T_{ij} , omitting the tilde since the fact that we lowered the second index implies that we precomposed with L . This is in accord with the conventions in relativity, where given a vector $v \in \mathbb{R}^4$ we write v_μ for the components of \tilde{v} when we should really write \tilde{v}_μ . From this point on, if we have a non-degenerate bilinear form on a vector space then we permit ourselves to raise and lower indices at will and without comment. In such a situation we often don't discuss the type of a tensor, speaking instead of its *rank*, equal to $r + s$, which obviously doesn't change as we raise and lower indices.

Example 2.1 *Linear operators in quantum mechanics*

Thinking about linear operators as $(1, 1)$ tensors may seem a bit strange, but in fact this is what one does in quantum mechanics all the time! Given an operator H on a quantum mechanical Hilbert space spanned by orthonormal vectors $\{e_i\}$ (which in Dirac notation we would write as $\{|i\rangle\}$), we usually write $H|i\rangle$ for $H(e_i)$, $\langle j|i\rangle$ for $\tilde{e}_j(e_i) = (e_j, e_i)$, and $\langle j|H|i\rangle$ for (e_j, He_i) . Thus, (2.4) would tell us that (using basis vectors instead of arbitrary vectors)

$$H_i^j = H(e_i, e^j) \tag{2.8}$$

$$= e^j(He_i) \tag{2.9}$$

$$= \langle j|H|i\rangle \tag{2.10}$$

where we converted to Dirac notation in the last equality to obtain the familiar quantum mechanical expression for the components of a linear operator. These components are often referred to as *matrix elements*, since when we write operators as matrices the elements of the matrices are just the components arranged in a particular fashion, as in (1.26).

Example 2.2 *The Moment of Inertia Tensor*

The moment of inertia tensor, denoted \mathcal{I} , is the symmetric $(2,0)$ tensor which, when evaluated on the angular velocity vector, yields the kinetic energy of a rigid body, i.e.

$$\frac{1}{2}\mathcal{I}(\omega, \omega) = KE \tag{2.11}$$

Alternatively we can raise an index on \mathcal{I} and define it to be the linear operator which eats the angular velocity and spits out the angular momentum, i.e.

$$\mathbf{L} = \mathcal{I}\omega. \tag{2.12}$$

(2.11) and (2.12) are most often seen in components (referred to a cartesian basis), where they read

$$KE = [\omega]^T[\mathcal{I}][\omega] \tag{2.13}$$

$$[\mathbf{L}] = [\mathcal{I}][\omega]. \tag{2.14}$$

Note that since we raise and lower indices with an inner product and usually use orthonormal bases, the components of \mathcal{I} when viewed as a $(2,0)$ tensor and when viewed as a $(1,1)$ tensor are the same, cf. exercise 1.17.

Example 2.3 Metric Tensors

We met the Euclidean metric on \mathbb{R}^n in example 1.17 and the Minkowski metric on \mathbb{R}^4 in example 1.19, and it's easy to verify that both are (2,0) tensors (why isn't the Hermitian scalar product of example 1.47 included?). We also have the inverse metrics, defined in Problem 4 of Chapter 1, and the reader can verify that these are (0,2) tensors.

Exercise 2.2 Show that for a metric g on V ,

$$g_i^j = \delta_i^j, \quad (2.15)$$

so the (1,1) tensor associated to g (via $g!$) is just the identity operator.

2.2 Change of Basis

Now we are in a position to derive the usual transformation laws that historically were taken as the *definition* of a tensor. Suppose we have a vector space V and two bases for V , $\mathcal{B} = \{e_i\}_{i=1\dots n}$ and $\mathcal{B}' = \{e_{i'}\}_{i=1\dots n}$. Since \mathcal{B} is a basis, each of the $e_{i'}$ can be expressed as $e_{i'} = A_{i'}^j e_j$ for some numbers $A_{i'}^j$. The same logic dictates that there exist numbers $A_i^{j'}$ (note that here the upper index is primed) such that $e_i = A_i^{j'} e_{j'}$, and since

$$e_i = A_i^{j'} e_{j'} = A_i^{j'} A_{j'}^k e_k \quad (2.16)$$

we must have

$$A_i^{j'} A_{j'}^k = \delta_i^k. \quad (2.17)$$

Considering (2.16) with the primed and unprimed indices switched also yields

$$A_{i'}^j A_j^{k'} = \delta_{i'}^{k'}, \quad (2.18)$$

so, in a way, $A_i^{j'}$ and $A_{i'}^j$ are inverses of each other. Notice that $A_{i'}^j$ and $A_i^{j'}$ are *not* the components of tensors, as their indices refer to different bases. How do the corresponding dual bases transform? Let $\{e^i\}_{i=1\dots n}$ and $\{e^{i'}\}_{i=1\dots n}$ be the bases dual to \mathcal{B} and \mathcal{B}' . Then the components of $e^{i'}$ with respect to $\{e^i\}_{i=1\dots n}$ are

$$e^{i'}(e_j) = e^{i'}(A_j^{k'} e_{k'}) = A_j^{k'} \delta_{k'}^{i'} = A_j^{i'}, \quad (2.19)$$

i.e.

$$e^{i'} = A_j^{i'} e^j. \quad (2.20)$$

Likewise,

$$e^i = A_{j'}^i e^{j'}. \quad (2.21)$$

Notice how well the Einstein Summation convention and our convention for priming indices work together in the transformation laws. Now we are ready to see how the components of a general (r, s) tensor T transform:

$$T_{i'_1, \dots, i'_r}{}^{j'_1 \dots j'_s} = T(e_{i'_1}, \dots, e_{i'_r}, e^{j'_1}, \dots, e^{j'_s}) \quad (2.22)$$

$$= T(A_{i'_1}^{k_1} e_{k_1}, \dots, A_{i'_r}^{k_r} e_{k_r}, A_{l_1}^{j'_1} e^{l_1}, \dots, A_{l_s}^{j'_s} e^{l_s}) \quad (2.23)$$

$$= A_{i'_1}^{k_1} \dots A_{i'_r}^{k_r} A_{l_1}^{j'_1} \dots A_{l_s}^{j'_s} T(e_{k_1}, \dots, e_{k_r}, e^{l_1}, \dots, e^{l_s}) \quad (2.24)$$

$$= A_{i'_1}^{k_1} \dots A_{i'_r}^{k_r} A_{l_1}^{j'_1} \dots A_{l_s}^{j'_s} T_{k_1 \dots k_r}{}^{l_1 \dots l_s}. \quad (2.25)$$

(2.25) is the standard tensor transformation law, which is taken as the *definition* of a tensor in much of the physics literature; here we have *derived* it as a consequence of our definition of a tensor as a multilinear function on V and V^* . The two are equivalent, however, as the reader will check in exercise 2.3 below. With the general transformation law in hand, we'll now look at specific types of tensors and derive their *matrix* transformation laws; to this end, it will be useful to introduce the matrices

$$A = \begin{pmatrix} A_1^{1'} & A_2^{1'} & \dots & A_n^{1'} \\ A_1^{2'} & A_2^{2'} & \dots & A_n^{2'} \\ \vdots & \vdots & \ddots & \vdots \\ A_1^{n'} & A_2^{n'} & \dots & A_n^{n'} \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} A_{1'}^1 & A_{2'}^1 & \dots & A_{n'}^1 \\ A_{1'}^2 & A_{2'}^2 & \dots & A_{n'}^2 \\ \vdots & \vdots & \ddots & \vdots \\ A_{1'}^n & A_{2'}^n & \dots & A_{n'}^n \end{pmatrix}, \quad (2.26)$$

which by virtue of (2.17) and (2.18) satisfy

$$AA^{-1} = A^{-1}A = I \quad (2.27)$$

as our notation suggests.

Exercise 2.3 Consider a function which assigns to a basis $\{e_i\}_{i=1 \dots n}$ the set of numbers $\{T_{k_1 \dots k_r}{}^{l_1 \dots l_s}\}$ which transform according to (2.25) under a change of basis. Show that this defines a multilinear function T of type (r, s) on V , and be sure to check that your definition is basis independent (i.e. that the value of T does not depend on which basis $\{e_i\}_{i=1 \dots n}$ you choose).

Example 2.4 Vectors and Dual Vectors

Given a vector v (considered as a $(1, 0)$ form as per (2.3)), (2.25) tell us that its components transform as

$$v^{i'} = A_j^{i'} v^j \quad (2.28)$$

while the components of a dual vector f transform as

$$f_{i'} = A_{i'}^j f_j. \quad (2.29)$$

Notice that the components of v transform with the $A_j^{i'}$ whereas the basis vectors transform with the $A_{i'}^j$, so the components of a vector obey the law opposite ('contra') to the basis vectors. This is the origin of the term 'contravariant'. Note also that the components of a dual vector transform in the *same* way as the basis vectors, hence the term 'covariant'. It makes sense that the basis vectors and the components of a vector should transform oppositely; v exists independently of any basis for V and shouldn't change under a change of basis, so if the e_i change one way the v^i should change oppositely. Similar remarks apply to dual vectors.

Incidentally, we can now explain a little bit more about the Einstein summation convention. We knew ahead of time that the components of dual vectors would transform like basis vectors, so we gave them both lower indices. We also knew that the components of vectors would transform like dual basis vectors, so we gave them both upper indices. Since the two transformation laws are opposite, we know (see exercise 2.6 below) that a summation over an upper index and lower index will yield an object that does not transform at all, so the summation represents an object or a process that doesn't depend upon a choice of basis. For instance, the expression $v^i e_i$ represents the vector v which is defined without reference to any basis, and the expression $f_i v^i$ is just $f(v)$, the action of the functional f on the vector v , also defined without reference to any basis. Processes such as these are so important and ubiquitous that it becomes very convenient to omit the summation sign for repeated upper and lower indices, and we thus have the summation convention.

In terms of matrices, we can write (2.28) and (2.29) as

$$[v]_{B'} = A[v]_B \quad (2.30)$$

$$[f]_{B'} = A^{-1T}[f]_B \quad (2.31)$$

where the superscript T again denotes the transpose of a matrix. The reader showed in exercise 1.17 that if we have an inner product (\cdot, \cdot) on a real vector space V and an orthonormal basis $\{e_i\}_{i=1..n}$ then the components of vectors and their corresponding dual vectors are identical, which is why we were able to ignore the distinction between

them for so long. (2.30) and (2.31) seem to contradict this, however, since it looks like the components of dual vectors transform very differently from the components of vectors. How do we explain this? Well, if we change from one *orthonormal* basis to another, we have

$$\delta_{i'j'} = (e_{i'}, e_{j'}) = A_{i'}^k A_{j'}^l (e_k, e_l) = \sum_{k=1}^n A_{i'}^k A_{j'}^k \quad (2.32)$$

which in matrices reads

$$A^{-1T} A^{-1} = I \quad (2.33)$$

so we must have

$$A^{-1T} = A \iff A^{-1} = A^T. \quad (2.34)$$

Such matrices are known as *orthogonal* matrices, and we see here that a transformation from one orthonormal basis to another is always implemented by an orthogonal matrix.¹ For such matrices (2.30) and (2.31) are identical, resolving our contradiction.

Incidentally, for a *complex* inner product space the reader will show that orthonormal basis changes are implemented by matrices satisfying $A^{-1} = A^\dagger$. Such matrices are known as *unitary* matrices and should be familiar from quantum mechanics.

Exercise 2.4 Show that for any invertible matrix A , $(A^{-1})^T = (A^T)^{-1}$, justifying the sloppiness of our notation above.

Exercise 2.5 Show that for a *complex* inner product space V , the matrix A implementing an orthonormal change of basis satisfies $A^{-1} = A^\dagger$.

Exercise 2.6 Show that $f(v) = f^i v_i = [f]^T[v]$ is invariant under a change of basis, as it should be. Prove this using the matrix transformation laws, i.e. show that $[f]_{\mathcal{B}}^T[v]_{\mathcal{B}} = [f]_{\mathcal{B}'}^T[v]_{\mathcal{B}'}$ for any $\mathcal{B}, \mathcal{B}'$.

Example 2.5 Linear Operators

We already noted that linear operators can be viewed as (1,1) tensors as per (2.4). (2.25) then tells us that, for a linear operator T on V ,

$$T_{i'}^{j'} = A_{i'}^k A_l^{j'} T_k^l \quad (2.35)$$

¹See the Problems for more on orthogonal matrices.

which in matrix form reads

$$[T]_{\mathcal{B}'} = A[T]_{\mathcal{B}}A^{-1} \quad (2.36)$$

which is the familiar *similarity transformation* of matrices. This, incidentally, allows us to extend the trace functional from $n \times n$ matrices to linear operators as follows: Given $T \in \mathcal{T}(V)$ and a basis \mathcal{B} for V , define the trace of T as

$$\text{Tr}(T) \equiv \text{Tr}([T]_{\mathcal{B}}). \quad (2.37)$$

The reader can then use (2.36) to show (see exercise 2.9) that $\text{Tr}(T)$ does not depend on the choice of basis \mathcal{B} .

Exercise 2.7 Show that for $v \in V, f \in V^*, T$ a linear operator on V , $f(Tv) = [f]^T[T][v]$ is invariant under a change of basis. Use the matrix transformation laws.

Exercise 2.8 Let $\mathcal{B} = \{x, y, z\}$, $\mathcal{B}' = \{x + iy, z, x - iy\}$ be bases for $\mathcal{H}_1(\mathbb{R}^3)$, and consider the operator L_z for which matrix expressions were found with respect to both bases in example 1.15. Find the numbers $A_j^{i'}$ and $A_{i'}^j$ and use these, along with (2.36), to obtain $[L_z]_{\mathcal{B}'}$ from $[L_z]_{\mathcal{B}}$.

Exercise 2.9 Show that (2.36) implies that $\text{Tr}([T]_{\mathcal{B}})$ does not depend on the choice of basis \mathcal{B} , so that $\text{Tr}(T)$ is well-defined.

Example 2.6 $(2,0)$ Tensors

$(2,0)$ tensors g , which include important examples such as the Minkowski metric and the Euclidean metric, transform as follows according to (2.25):

$$g_{i'j'} = A_{i'}^k A_{j'}^l g_{kl} \quad (2.38)$$

or in matrix form

$$[g]_{\mathcal{B}'} = A^{-1T} [g]_{\mathcal{B}} A^{-1}. \quad (2.39)$$

Notice that if g is an inner product and \mathcal{B} and \mathcal{B}' are orthonormal bases then $[g]_{\mathcal{B}'} = [g]_{\mathcal{B}} = I$ and (2.39) becomes

$$I = A^{-1T} A^{-1}, \quad (2.40)$$

again telling us that A must be orthogonal. Also note that if A is orthogonal, (2.39) is identical to (2.36), so we don't have to distinguish between metric tensors and linear operators (as most of us haven't in the past!). In the case of the Minkowski

metric η we aren't dealing with an inner product but we do have orthonormal bases, with respect to which² η takes the form

$$[\eta] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (2.41)$$

so if we are changing from one orthonormal basis to another we have

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = A^{-1T} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} A^{-1} \quad (2.42)$$

or equivalently

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = A^T \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} A. \quad (2.43)$$

Matrices A satisfying (2.43) are known as *Lorentz Transformations*. Notice that these matrices are not quite orthogonal, so the components of vectors will transform slightly differently than those of dual vectors under these transformations. This is in contrast to the case of \mathbb{R}^n with a positive-definite metric, where if we go from one orthonormal basis to another then the components of vectors and dual vectors transform identically as the reader showed in exercise 1.17.

Exercise 2.10 As in previous exercises, show using the matrix transformation laws that $g(v, w) = [w]^T [g][v]$ is invariant under a change of basis.

Before we conclude this section, we should remark that when we said that the $A_{i'}^j$ were not the components of a tensor, we were lying a little; there *is* a tensor lurking around, namely the linear operator U that takes the old basis vectors into the new, i.e. $U(e_i) = e_{i'} \forall i$ (the action of U on an arbitrary vector is then given by expanding that vector in the basis \mathcal{B} and using linearity). What are the components of this tensor? Well, in the old basis \mathcal{B} we have

$$U_i^j = U(e_i, e^j) = e^j(Ue_i) = e^j(e_{i'}) = e^j(A_{i'}^k e_k) = A_{i'}^j \quad (2.44)$$

²We assume here that the basis vector e_t satisfying $\eta(e_t, e_t) = -1$ is the 4th vector in the basis, which isn't necessary but is somewhat conventional in physics.

so the $A_{i'}^j$ actually are the components of a tensor! Why did we lie, then? Well, the approach we have been taking so far is to try and think about things in a basis-independent way, and although U is a well-defined linear operator, its definition depends entirely on the two bases we've chosen, so we may as well work directly with the numbers that relate the bases. Also, using one primed index and one unprimed index makes it easy to remember transformation laws like (2.20) and (2.21) but is not consistent with our notation for the components of tensors.

If we write out the components of U as a matrix, the reader should verify that

$$[e_{i'}]_{\mathcal{B}} = [U]_{\mathcal{B}}[e_i]_{\mathcal{B}} = A^{-1}[e_i]_{\mathcal{B}} \quad (2.45)$$

which should be compared to (2.30). (2.45) is called an *active* transformation, since we use the matrix A (or, rather, its inverse) to change one vector into another, namely e_i into $e_{i'}$. (2.30), on the other hand, is called a *passive* transformation, since we use the matrix A not to change the vector v but rather to change the basis which v is referred to, hence changing its components. The notation in most physics texts is not as explicit as ours; one usually sees matrix equations like

$$\mathbf{r}' = A\mathbf{r} \quad (2.46)$$

for both passive and active transformations, and one must rely on context to figure out how the equation is to be interpreted. In the active case, one considers the coordinate system fixed and interprets the matrix A as taking the physical vector \mathbf{r} into a *new* vector \mathbf{r}' , where the components of both are expressed in the *same* coordinate system. In the passive case, the physical vector \mathbf{r} doesn't change but the basis does, so one interprets the matrix A as taking the components of \mathbf{r} in the old coordinate system and giving back the components of the *same* vector \mathbf{r} in the new (primed) coordinate system. Notice that in the active case the prime refers to a new vector, and in the passive case to a new coordinate system.

Passive transformations are probably the ones encountered most often in classical physics, since a change of cartesian coordinates induces a passive transformation. Active transformations do crop up, though, especially in the case of rigid body motion. In this scenario, one specifies the orientation of a rigid body by the time-dependent orthogonal basis transformation $A(t)$ which relates the space frame K' to the body frame $K(t)$ (we use here the notation of example 1.12). As we saw above, there corresponds to the time-dependent matrix $A(t)$ a time-dependent linear operator $U(t)$ which satisfies $U(t)(e_{i'}) = e_i(t)$. If K and K' were coincident at $t = 0$ and \mathbf{r}_0 is the position vector of a point p of the rigid body at that time, then the position of p at a later time is just $\mathbf{r}(t) = U(t)\mathbf{r}_0$, which as a matrix equation in K'

would read

$$[\mathbf{r}(t)]_{K'} = A(t)[\mathbf{r}_0]_{K'} \quad (2.47)$$

or in more common and less precise notation,

$$\mathbf{r}(t) = A(t)\mathbf{r}_0. \quad (2.48)$$

In other words, the position of a specific point on the rigid body at an arbitrary time t is given by the active transformation corresponding to the matrix $A(t)$.

The duality between passive and active transformations is also present in quantum mechanics. In the Schrodinger picture, one considers observables like the momentum or position operator as acting on the state ket while the basis kets remain fixed. This is the active viewpoint. In the Heisenberg picture, however, one considers the state ket to be fixed and considers the *observables* to be time-dependent (recall that (1.21) is the equation of motion for these operators). Since the operators are time-dependent, their eigenvectors (which form a basis³) are time-dependent as well, so this picture is the passive one in which the vectors don't change but the basis does. Just as an equation like (2.46) can be interpreted in both the active and passive sense, a quantum mechanical equation like

$$\langle \hat{x}(t) \rangle = \langle \psi | (U^\dagger \hat{x} U) | \psi \rangle \quad (2.49)$$

$$= (\langle \psi | U^\dagger) \hat{x} (U | \psi \rangle), \quad (2.50)$$

where U is the time-evolution operator for time t , can also be interpreted in two ways: in the active sense of (2.50), in which the U 's act on the vectors and change them into new vectors, and in the passive sense of (2.49), where the U 's act on the operator \hat{x} by a similarity transformation to turn it into a new operator, $\hat{x}(t)$.

2.3 The Tensor Product - Definition and Properties

One of the most basic operations with tensors, again commonplace in physics but often unacknowledged (or, at best, dealt with in an ad-hoc fashion) is that of the tensor product. Before giving the precise definition, which takes a little getting used to, we give a rough, heuristic description. Given two vector spaces V and W (over the same set of scalars C), we would like to construct a product space, which we denote $V \otimes W$, whose elements are in some sense products of vectors $v \in V$ and

³For details on why the eigenvectors of hermitian operators form a basis, at least in the finite-dimensional case, see [HK].

$w \in W$. We denote these products by $v \otimes w$. This product, like any respectable product, should be bilinear in the sense that

$$\begin{aligned}(v_1 + v_2) \otimes w &= v_1 \otimes w + v_2 \otimes w \\ v \otimes (w_1 + w_2) &= v \otimes w_1 + v \otimes w_2.\end{aligned}\tag{2.51}$$

Given this property, the product of any two arbitrary vectors v and w can then be expanded in terms of bases $\{e_i\}_{i=1\dots n}$ and $\{f_j\}_{j=1\dots m}$ for V and W as

$$v \otimes w = (v^i e_i) \otimes (w^j f_j)\tag{2.52}$$

$$= v^i w^j e_i \otimes f_j\tag{2.53}$$

so $\{e_i \otimes f_j\}, i = 1 \dots n, j = 1 \dots m$ should be a basis for $V \otimes W$, which would then have dimension nm . Thus the basis for the product space would be just the product of the basis vectors, and the dimension of the product space would be just the product of the dimensions.

Now we make this precise. Given two vector spaces V and W , we define their *tensor product* $V \otimes W$ to be the set of all C -valued bilinear functions on $V^* \times W^*$. Such functions do form a vector space, as the reader can easily check. This definition may seem unexpected or counterintuitive at first, but the reader will see that this definition does yield the vector space described above. Now, given vectors $v \in V, w \in W$, we define the *tensor product* of v and w , written $v \otimes w$, to be the element of $V \otimes W$ defined as follows:

$$(v \otimes w)(h, g) \equiv v(h)w(g) \quad \forall h \in V^*, g \in W^*.\tag{2.54}$$

The bilinearity of the tensor product is immediate and the reader can probably verify it without writing anything down: just check that both sides of (2.51) are equal when evaluated on any pair of dual vectors. To prove that $\{e_i \otimes f_j\}, i = 1 \dots n, j = 1 \dots m$ is a basis for $V \otimes W$, let $\{e^i\}_{i=1\dots n}, \{f^j\}_{j=1\dots m}$ be the corresponding dual bases and consider an arbitrary $T \in V \otimes W$. Using bilinearity,

$$T(h, g) = h_i g_j T(e^i, f^j) = h_i g_j T^{ij}\tag{2.55}$$

where $T^{ij} \equiv T(e^i, f^j)$. If we consider the expression $T^{ij} e_i \otimes f_j$, then

$$(T^{ij} e_i \otimes f_j)(e^k, f^l) = T^{ij} e_i(e^k) f_j(f^l)\tag{2.56}$$

$$= T^{ij} \delta_i^k \delta_j^l\tag{2.57}$$

$$= T^{kl}\tag{2.58}$$

so $T^{ij}e_i \otimes f_j$ agrees with T on basis vectors, hence on all vectors by bilinearity, so $T = T^{ij}e_i \otimes f_j$. Since T was an arbitrary element of $V \otimes W$, $V \otimes W = \text{Span}\{e_i \otimes f_j\}$. Furthermore, the $e_i \otimes f_j$ are linearly independent as the reader should check, so $\{e_i \otimes f_j\}$ is actually a basis for $V \otimes W$ and $V \otimes W$ thus has dimension mn .

The tensor product has a couple of important properties besides bilinearity. First, it *commutes with taking duals*, that is

$$(V \otimes W)^* = V^* \otimes W^*. \quad (2.59)$$

Second, the tensor product it is *associative*, i.e. for vector spaces V_i , $i = 1, 2, 3$,

$$(V_1 \otimes V_2) \otimes V_3 = V_1 \otimes (V_2 \otimes V_3). \quad (2.60)$$

This property allows us to drop the parentheses and write expressions like $V_1 \otimes \dots \otimes V_n$ without ambiguity. One can think of $V_1 \otimes \dots \otimes V_n$ as the set of C -valued multilinear functions on $V_1^* \times \dots \times V_n^*$. Verifying these two properties is somewhat tedious and the reader is referred to [W] for proofs.

Exercise 2.11 If $\{e_i\}, \{f_j\}$ and $\{g_k\}$ are bases for V_1, V_2 and V_3 respectively, convince yourself that $\{e_i \otimes f_j \otimes g_k\}$ is a basis for $V_1 \otimes V_2 \otimes V_3$, and hence that $\dim V_1 \otimes V_2 \otimes V_3 = n_1 n_2 n_3$ where $\dim V_i = n_i$. Extend the above to n -fold tensor products.

2.4 Tensor Products of V and V^*

In the previous section we defined the tensor product for two arbitrary vector spaces V and W . Often, though, we'll be interested in just the iterated tensor product of a vector space and its dual, i.e. in tensor products of the form

$$\underbrace{V^* \otimes \dots \otimes V^*}_{r \text{ times}} \otimes \underbrace{V \otimes \dots \otimes V}_{s \text{ times}}. \quad (2.61)$$

This space is of particular interest because it is actually identical to \mathcal{T}_s^r ! From the previous section we know that the vector space in (2.61) can be interpreted as the set of multilinear functions on

$$\underbrace{V \times \dots \times V}_{r \text{ times}} \times \underbrace{V^* \times \dots \times V^*}_{s \text{ times}}, \quad (2.62)$$

but these functions are exactly \mathcal{T}_s^r ! Since the space in (2.61) has basis $\mathcal{B}_s^r = \{e^{i_1} \otimes \dots \otimes e^{i_r} \otimes e_{j_1} \otimes \dots \otimes e_{j_s}\}$, we can conclude that \mathcal{B}_s^r is a basis for \mathcal{T}_s^r . In fact, we claim that if $T \in \mathcal{T}_s^r$ has components $T_{i_1 \dots i_r}^{j_1 \dots j_s}$, then

$$T = T_{i_1 \dots i_r}^{j_1 \dots j_s} e^{i_1} \otimes \dots \otimes e^{i_r} \otimes e_{j_1} \otimes \dots \otimes e_{j_s} \quad (2.63)$$

is the expansion of T in the basis \mathcal{B}_s^r . To prove this, we just need to check that both sides agree when evaluated on an arbitrary set of basis vectors; on the left hand side we get $T(e_{i_1}, \dots, e_{i_r}, e^{j_1}, \dots, e^{j_s}) = T_{i_1, \dots, i_r}^{j_1, \dots, j_s}$ by definition, and on the right hand side we have

$$\begin{aligned}
& (T_{k_1 \dots k_r}^{l_1 \dots l_s} e^{k_1} \otimes \dots \otimes e^{k_r} \otimes e_{l_1} \otimes \dots \otimes e_{l_s})(e_{i_1}, \dots, e_{i_r}, e^{j_1}, \dots, e^{j_s}) \\
&= T_{k_1 \dots k_r}^{l_1 \dots l_s} e^{k_1}(e_{i_1}) \dots e^{k_r}(e_{i_r}) e_{l_1}(e^{j_1}) \dots e_{l_s}(e^{j_s}) \\
&= T_{k_1 \dots k_r}^{l_1 \dots l_s} \delta_{i_1}^{k_1} \dots \delta_{i_r}^{k_r} \delta_{l_1}^{j_1} \dots \delta_{l_s}^{j_s} \\
&= T_{i_1, \dots, i_r}^{j_1, \dots, j_s} \tag{2.64}
\end{aligned}$$

so our claim is true. Thus, for instance, a $(2, 0)$ tensor like the Minkowski metric can be written as $\eta = \eta_{\mu\nu} e^\mu \otimes e^\nu$. Also, a tensor product like $f \otimes g = f_i g_j e^i \otimes e^j \in \mathcal{T}_0^2$ thus has components $(f \otimes g)_{ij} = f_i g_j$. Notice that we now have two ways of thinking about components: either as the values of the tensor on sets of basis vectors (as in (2.7)) *or* as the expansion coefficients in the given basis (as in (2.63)). This duality of perspective was pointed out in the case of vectors under (1.42), and it's essential that the reader be comfortable thinking about components in either way.

Exercise 2.12 Compute the dimension of \mathcal{T}_s^r .

Exercise 2.13 Let T_1 and T_2 be tensors of type (r_1, s_1) and (r_2, s_2) respectively on a vector space V . Show that $T_1 \otimes T_2$ can be viewed as an $(r_1 + r_2, s_1 + s_2)$ tensor, so that the tensor product of two tensors is again a tensor, justifying the nomenclature.

One important operation on tensors which we are now in a position to discuss is that of *contraction*, which is the generalization of the trace functional to tensors of arbitrary rank : Given $T \in \mathcal{T}_s^r(V)$ with expansion

$$T = T_{i_1 \dots i_r}^{j_1 \dots j_s} e^{i_1} \otimes \dots \otimes e^{i_r} \otimes e_{j_1} \otimes \dots \otimes e_{j_s} \tag{2.65}$$

we can define a contraction of T to be any $(r - 1, s - 1)$ tensor resulting from feeding e^i into one of the arguments, e_i into another and then summing over i as implied by the summation convention. For instance, if we feed e_i into the r th slot and e^i into the $(r + s)$ th slot and sum, we get the $(r - 1, s - 1)$ tensor \tilde{T} defined as

$$\tilde{T}(v_1, \dots, v_{r-1}, f_1, \dots, f_{s-1}) \equiv T(v_1, \dots, v_{r-1}, e_i, f_1, \dots, f_{s-1}, e^i). \tag{2.66}$$

The reader may be suspicious that \tilde{T} depends on our choice of basis but exercise 2.14 below will disabuse him of that. Notice that the components of \tilde{T} are

$$\tilde{T}_{i_1 \dots i_{r-1}}^{j_1 \dots j_{s-1}} = T_{i_1 \dots i_{r-1} l}^{j_1 \dots j_{s-1} l}. \tag{2.67}$$

Similar contractions can be performed on any two arguments of T provided one argument eats vectors and the other dual vectors. In terms of components, a contraction can be taken with respect to any pair of indices provided that one is covariant and the other contravariant. Notice that for a linear operator (or $(1, 1)$ tensor) A we have only one option for contraction: $\tilde{A} = A_i^i = \text{Tr}(A)$. Notice also that if we have two linear operators A and B then their tensor product $A \otimes B \in \mathcal{T}_2^2$ has components

$$(A \otimes B)_{ik}^{jl} = A_i^j B_k^l, \quad (2.68)$$

and contracting on the first and last index gives a $(1, 1)$ tensor AB whose components are

$$(AB)_k^j = A_i^j B_k^i. \quad (2.69)$$

The reader should check that this tensor is just the composition of A and B , as our notation suggests. What linear operator do we get if we consider the other contraction $A_i^j B_j^{l'}$?

Exercise 2.14 Show that if $\{e_i\}_{i=1\dots n}$ and $\{e_{i'}\}_{i=1\dots n}$ are two arbitrary bases that

$$T(v_1, \dots, v_{r-1}, e_i, f_1, \dots, f_{s-1}, e^{i'}) = T(v_1, \dots, v_{r-1}, e_{i'}, f_1, \dots, f_{s-1}, e^{i'}) \quad (2.70)$$

so that contraction is well-defined.

Example 2.7 $V^* \otimes V$

One of the most important examples of tensor products of the form (2.61) is $V^* \otimes V$, which as we mentioned is the same as \mathcal{T}_1^1 , the space of linear operators. How does this identification work, explicitly? Well, given $f \otimes v \in V^* \otimes V$, we can define a linear operator by $(f \otimes v)(w) \equiv f(w)v$. More generally, given

$$T_i^j e^i \otimes e_j \in V^* \otimes V, \quad (2.71)$$

we can define a linear operator T by

$$T(v) = T_i^j e^i(v) e_j = v^i T_i^j e_j \quad (2.72)$$

which is identical to (1.24). This identification of $V^* \otimes V$ and linear operators is actually implicit in many quantum mechanical expressions. Let \mathcal{H} be a quantum mechanical Hilbert space and let $\psi, \phi \in \mathcal{H}$ so that $L(\phi) \in \mathcal{H}^*$. The tensor product of $L(\phi)$ and ψ , which we would write as $L(\phi) \otimes \psi$, is written in Dirac notation as $|\psi\rangle\langle\phi|$ (note the transposition of the factors relative to our convention). If we're given an

orthonormal basis $\mathcal{B} = \{|i\rangle\}$, the expansion (2.71) of an arbitrary operator H can be written in Dirac notation as

$$H = \sum_{i,j} H_{ij} |j\rangle\langle i|, \quad (2.73)$$

an expression which may be familiar from advanced quantum mechanics texts.⁴ In particular, the identity operator can be written as

$$I = \sum_i |i\rangle\langle i|, \quad (2.74)$$

which is referred to as the *resolution of the identity* with respect to the basis $\{|i\rangle\}$.

A word about nomenclature: In quantum mechanics and other contexts the tensor product is often referred to as the *direct* or *outer* product. This last term is meant to distinguish it from the inner product, since both the outer and inner products eat a dual vector and a vector (strictly speaking the inner product eats 2 vectors, but remember that with an inner product we may identify vectors and dual vectors) but the outer product yields a linear operator whereas the inner product yields a scalar.

2.5 Applications of the Tensor Product in Classical Physics

Example 2.8 *Moment of inertia tensor revisited*

We took an abstract look at the moment of inertia tensor in example 2.2; now, armed with the tensor product, we can examine the moment of inertia tensor more concretely. Consider a rigid body and assume that its center of mass is fixed at a point O , so that it has only rotational degrees of freedom. Let O be the origin, pick time-dependent body-fixed axes $K = \{\hat{\mathbf{x}}(t), \hat{\mathbf{y}}(t), \hat{\mathbf{z}}(t)\}$ for E^3 , and let g denote the Euclidean metric on E^3 . Recall that g allows us to define a map L from E^3 to E^{3*} . Also, let the i th particle in the rigid body have mass m_i and position vector \mathbf{r}_i with $[\mathbf{r}_i]_K = (x_i, y_i, z_i)$ relative to O , and let $r_i^2 \equiv g(\mathbf{r}_i, \mathbf{r}_i)$. Then the $(2, 0)$ moment of inertia tensor is given by

$$\mathcal{I}_{(2,0)} = \sum_i m_i (r_i^2 g - L(\mathbf{r}_i) \otimes L(\mathbf{r}_i)) \quad (2.75)$$

⁴We don't bother here with index positions since most quantum mechanics texts don't employ Einstein summation convention, preferring instead to explicitly indicate summation.

while the (1, 1) tensor reads

$$\mathcal{I}_{(1,1)} = \sum_i m_i (r_i^2 I - L(\mathbf{r}_i) \otimes \mathbf{r}_i). \quad (2.76)$$

The reader should check that in components (2.75) reads

$$\mathcal{I}_{jk} = \sum_i m_i (r_i^2 \delta_{jk} - (\mathbf{r}_i)_j (\mathbf{r}_i)_k). \quad (2.77)$$

Writing a couple of components explicitly yields

$$\begin{aligned} \mathcal{I}_{xx} &= \sum_i m_i (y_i^2 + z_i^2) \\ \mathcal{I}_{xy} &= -\sum_i m_i x_i y_i, \end{aligned} \quad (2.78)$$

expressions which should be familiar from classical mechanics. So long as the basis is orthonormal, the components \mathcal{I}_j^k of the (1, 1) tensor in (2.76) will be the same as for the (2, 0) tensor, as remarked earlier. Note that if we had not used body-fixed axes, the components of \mathbf{r}_i (and hence the components of \mathcal{I} , by (2.78)) would in general be time-dependent; this is the main reason for using the body-fixed axes in computation.

Example 2.9 *Maxwell Stress Tensor*

In considering the conservation of total momentum (mechanical plus electromagnetic) in classical electrodynamics one encounters the symmetric rank 2 *Maxwell Stress Tensor*, defined in (2, 0) form as ⁵

$$T_{(2,0)} = \mathbf{E} \otimes \mathbf{E} + \mathbf{B} \otimes \mathbf{B} - \frac{1}{2}(\mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B})g \quad (2.79)$$

where \mathbf{E} and \mathbf{B} are the dual vector versions of the electric and magnetic field vectors. T can be interpreted in the following way: $T(v, w)$ gives the rate at which momentum in the v -direction flows in the w -direction. In components we have

$$T_{ij} = E_i E_j + B_i B_j - \frac{1}{2}(\mathbf{E} \cdot \mathbf{E} + \mathbf{B} \cdot \mathbf{B})\delta_{ij}, \quad (2.80)$$

which is the expression found in most classical electrodynamics textbooks.

⁵Here and below we set all physical constants such as c and ϵ_0 equal to 1.

Example 2.10 *The Electromagnetic Field tensor*

As the reader has probably seen in discussions of relativistic electrodynamics, the electric and magnetic field vectors are properly viewed as components of a rank 2 antisymmetric tensor F , the electromagnetic *field tensor*.⁶ To write F in component-free notation requires machinery outside the scope of this text,⁷ so we settle for its expression as a matrix in an orthonormal basis, which in $(2, 0)$ form is

$$[F_{(2,0)}] = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix}. \quad (2.81)$$

The Lorentz force law

$$\frac{dp^\mu}{d\tau} = qF^\mu{}_\nu v^\nu \quad (2.82)$$

where $p = mv$ is the 4-momentum of a particle, v is its proper velocity and q its charge, can be rewritten without components as

$$\frac{dp}{d\tau} = qF_{(1,1)}(v) \quad (2.83)$$

which just says that the proper force on a particle is given by the action of the field tensor on the particle's proper velocity!

2.6 Applications of the Tensor Product in Quantum Physics

In this section we'll discuss further applications of the tensor product in quantum mechanics, in particular the oft-unwritten rule that to add degrees of freedom one should take the tensor product of the corresponding Hilbert spaces. Before we get to this, however, we must set up a little more machinery and address an issue that we've so far swept under the rug. The issue is that when dealing with spatial degrees of freedom, as opposed to 'internal' degrees of freedom like spin, we often encounter

⁶In this example and the one above we are actually not dealing with tensors but with *tensor fields*, i.e. tensor-valued functions on space and spacetime. For the discussion here, however, we will ignore the spatial dependence, focusing instead on the tensorial properties.

⁷One needs the exterior derivative, a generalization of the curl, divergence and gradient operators from vector calculus. See [SC] for a very readable account.

Hilbert spaces like $L^2([-a, a])$ and $L^2(\mathbb{R})$ which are most conveniently described by ‘basis’ vectors which are eigenvectors of either the position operator \hat{x} or the momentum operator \hat{p} . The trouble with these bases is that they are often non-denumerably infinite (i.e. can’t be indexed by the integers, unlike all the bases we’ve worked with so far) and, what’s worse, the ‘basis vectors’ don’t even belong to the Hilbert space! Consider, for example, $L^2(\mathbb{R})$. The position operator \hat{x} acts on functions $\psi(x) \in L^2(\mathbb{R})$ by

$$\hat{x} \psi(x) = x \psi(x). \quad (2.84)$$

If we follow the practice of most quantum mechanics texts and treat the Dirac delta functional δ as $L(\delta(x))$ where $\delta(x)$, the ‘Dirac delta function’, is infinite at 0 and 0 elsewhere, the reader can check (see exercise 2.15) that

$$\hat{x} \delta(x - x_0) = x_0 \delta(x - x_0) \quad (2.85)$$

so that $\delta(x - x_0)$ is an ‘eigenfunction’ of \hat{x} with eigenvalue x_0 (in Dirac notation we write the corresponding ket as $|x_0\rangle$). The trouble is that, as we saw in example 1.23, there is no such $\delta(x) \in L^2(\mathbb{R})$! Furthermore, since the basis $\{\delta(x - x_0)\}_{x_0 \in \mathbb{R}}$ is indexed by \mathbb{R} and not some subset of \mathbb{Z} , we must expand $\psi \in L^2(\mathbb{R})$ by integrating instead of summing. Integration, however, is a limiting procedure and one should really worry about whether a given integral converges and in what sense it converges. Rectifying all this in a rigorous manner is outside the scope of this text, unfortunately, but we do wish to work with these objects, so we content ourselves with the traditional approach: ignore the fact that the delta functions are not elements of $L^2(\mathbb{R})$, work without discomfort with the basis $\{\delta(x - x_0)\}_{x_0 \in \mathbb{R}}$ ⁸, and fearlessly expand arbitrary functions ψ in the basis $\{\delta(x - x_0)\}_{x_0 \in \mathbb{R}}$ as

$$\psi(x) = \int_{-\infty}^{\infty} dx' \psi(x') \delta(x - x'), \quad (2.86)$$

where the above equation can be interpreted both as the expansion of ψ and just the definition of the delta function. In Dirac notation (2.86) reads

$$|\psi\rangle = \int_{-\infty}^{\infty} dx' \psi(x') |x'\rangle. \quad (2.87)$$

Note that we can think of the numbers $\psi(x)$ as the components of $|\psi\rangle$ with respect to the basis $\{|x\rangle\}_{x \in \mathbb{R}}$. Alternatively, if we define the inner product of our basis vectors to be

$$\langle x|x'\rangle \equiv \delta(x - x') \quad (2.88)$$

⁸Working with the momentum eigenfunctions e^{ipx} instead doesn’t help; though these are legitimate functions, they still are not square integrable since $\int_{-\infty}^{\infty} |e^{ipx}|^2 dx = \infty$!

as is usually done, then using (2.87) we have

$$\psi(x) = \langle x | \psi \rangle \quad (2.89)$$

which gives another interpretation of $\psi(x)$. These two interpretations of $\psi(x)$ are just the ones mentioned below (2.64); that is, the components of a vector can be interpreted either as expansion coefficients (as in (2.87)), or as the value of a given dual vector on the vector, as in (2.89).

Exercise 2.15 By considering the integral

$$\int_{-\infty}^{\infty} (\hat{x} \delta(x - x_0)) f(x) dx \quad (2.90)$$

(where f is an arbitrary square-integrable function), show formally that

$$\hat{x} \delta(x - x_0) = x_0 \delta(x - x_0). \quad (2.91)$$

Exercise 2.16 Check that $\{\delta(x - x_0)\}_{x_0 \in \mathbb{R}}$ satisfies (1.54).

Exercise 2.17 Verify (2.89).

We mentioned in a footnote on the previous page that one could use momentum eigenfunctions instead of position eigenfunctions as a basis for $L^2(\mathbb{R})$. What does the corresponding change of basis look like?

Example 2.11 *The Momentum Representation*

As is well known from quantum mechanics, the eigenfunctions of the momentum operator $\hat{p} = \frac{\hbar}{i} \frac{d}{dx}$ are the wavefunctions $\{e^{\frac{i}{\hbar} p x}\}_{p \in \mathbb{R}}$, and these wavefunctions form a basis for $L^2(\mathbb{R})$. In fact, the expansion of an arbitrary function $\psi \in L^2(\mathbb{R})$ in this basis is just the Fourier expansion of ψ , written

$$\psi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dp \phi(p) e^{\frac{i}{\hbar} p x} \quad (2.92)$$

where the component function $\phi(p)$ is known as the *Fourier Transform* of ψ . One could in fact work exclusively with $\phi(p)$ instead of $\psi(x)$, and recast the operators \hat{x} and \hat{p} in terms of their action on $\phi(p)$ (see exercise 2.18 below); such an approach is known as the *momentum representation*. Now, what does it look like when we

switch from the position representation to the momentum representation, i.e. when we change bases from $\{\delta(x - x_0)\}_{x_0 \in \mathbb{R}}$ to $\{e^{\frac{i}{\hbar}px}\}_{p \in \mathbb{R}}$? Since the basis vectors are indexed by real numbers p and x_0 as opposed to integers i and j , our change of basis will not be given by a matrix with components $A_j^{i'}$ but rather a *function* $A(x_0, p)$, which by (2.19) and the fact that both bases are orthonormal is given by the inner product of $\delta(x - x_0)$ and $e^{\frac{i}{\hbar}px}$. In Dirac notation this would be written as $\langle x_0 | p \rangle$, and we have

$$A(x_0, p) = \langle x_0 | p \rangle = \int_{-\infty}^{\infty} dx \delta(x - x_0) e^{\frac{i}{\hbar}px} = e^{\frac{i}{\hbar}px_0}, \quad (2.93)$$

a familiar equation. □

Exercise 2.18 Use (2.92) to show that in the momentum representation, $\hat{p}\phi(p) = p\phi(p)$ and $\hat{x}\phi(p) = \frac{\hbar}{i} \frac{d\phi}{dp}$.

The next issue to address is that of linear operators: having constructed a new Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ out of two Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 , can we construct linear operators on $\mathcal{H}_1 \otimes \mathcal{H}_2$ out of the linear operators on \mathcal{H}_1 and \mathcal{H}_2 ? Well, given linear operators A_i on \mathcal{H}_i , $i = 1, 2$, we can define a linear operator $A_1 \otimes A_2$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$ by

$$(A_1 \otimes A_2)(v \otimes w) \equiv (A_1 v) \otimes (A_2 w). \quad (2.94)$$

The reader can check that with this definition, $(A \otimes B)(C \otimes D) = AC \otimes BD$. In most quantum mechanical applications either A_1 or A_2 is the identity, i.e. one considers operators of the form $A_1 \otimes I$ or $I \otimes A_2$. These are often abbreviated as A_1 and A_2 even though they're acting on $\mathcal{H}_1 \otimes \mathcal{H}_2$.

The last subject we should touch upon is that of *vector operators*, which are defined to be sets of operators that transform as three dimensional vectors under the adjoint action of the total angular momentum operators J_i . That is, a vector operator is a set of operators $\{B_i\}_{i=1-3}$ (often written collectively as \mathbf{B}) that satisfies

$$ad_{J_i}(B_j) = [J_i, B_j] = i\hbar \epsilon_{ijk} B_k \quad (2.95)$$

where ϵ_{ijk} is the familiar Levi-Civita symbol. The three dimensional position operator $\hat{\mathbf{r}} = \{\hat{x}, \hat{y}, \hat{z}\}$, momentum operator $\hat{\mathbf{p}} = \{\hat{p}_x, \hat{p}_y, \hat{p}_z\}$, and orbital angular momentum operator $\mathbf{L} = \{L_x, L_y, L_z\}$ are all vector operators, as the reader can check.

Exercise 2.19 For spinless particles, $\mathbf{J} = \mathbf{L} = \hat{\mathbf{x}} \times \hat{\mathbf{p}}$. Expressions for the components may be obtained by expanding the cross product or referencing example 1.15 and exercise 1.8. Use these expressions and the canonical commutation relations $[x_i, p_j] = -i\hbar \delta_{ij}$ to show that $\hat{\mathbf{x}}, \hat{\mathbf{p}}$ and \mathbf{L} are all vector operators.

Now we are finally ready to consider some examples, in which we'll take as an axiom that adding degrees of freedom is implemented by taking tensor products of the corresponding Hilbert spaces. The reader will see that this process reproduces familiar results.

Example 2.12 *Addition of translational degrees of freedom*

Consider a spinless particle constrained to move in one dimension; the quantum mechanical Hilbert space for this system is $L^2(\mathbb{R})$ with basis $\{|x\rangle\}_{x \in \mathbb{R}}$. If we consider a second dimension, call it the y dimension, then this degree of freedom has its own Hilbert space $L^2(\mathbb{R})$ with basis $\{|y\rangle\}_{y \in \mathbb{R}}$. If we allow the particle both degrees of freedom then the Hilbert space for the system is $L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$, with basis $\{|x\rangle \otimes |y\rangle\}_{x,y \in \mathbb{R}}$. An arbitrary ket $|\psi\rangle \in L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$ has expansion

$$|\psi\rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy \psi(x, y) |x\rangle \otimes |y\rangle \quad (2.96)$$

with expansion coefficients $\psi(x, y)$. If we iterate this logic, we get in 3 dimensions

$$|\psi\rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy dz \psi(x, y, z) |x\rangle \otimes |y\rangle \otimes |z\rangle. \quad (2.97)$$

If we rewrite $\psi(x, y, z)$ as $\psi(\mathbf{r})$ and $|x\rangle \otimes |y\rangle \otimes |z\rangle$ as $|\mathbf{r}\rangle$ where $\mathbf{r} = (x, y, z)$, then we have

$$|\psi\rangle = \int d^3r \psi(\mathbf{r}) |\mathbf{r}\rangle \quad (2.98)$$

which is the familiar 3D expansion of a ket in terms of position eigenkets. Such a ket is an element of $L^2(\mathbb{R}) \otimes L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$, which is also denoted as $L^2(\mathbb{R}^3)$.⁹

Example 2.13 *Two-particle systems*

Now consider two spinless particles in three dimensional space, possibly interacting through some sort of potential. The two-body problem with a $1/r$ potential is a classic example of this. The Hilbert space for such a system is then $L^2(\mathbb{R}^3) \otimes L^2(\mathbb{R}^3)$, with basis $\{|\mathbf{r}_1\rangle \otimes |\mathbf{r}_2\rangle\}_{\mathbf{r}_i \in \mathbb{R}^3}$. In many textbooks the tensor product symbol is omitted

⁹ $L^2(\mathbb{R}^3)$ is actually defined to be the set of all square integrable functions on \mathbb{R}^3 , i.e. functions f satisfying

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy dz |f|^2 < \infty. \quad (2.99)$$

Not too surprisingly, this space turns out to be identical to $L^2(\mathbb{R}) \otimes L^2(\mathbb{R}) \otimes L^2(\mathbb{R})$.

and such basis vectors are written as $|\mathbf{r}_1\rangle |\mathbf{r}_2\rangle$ or even $|\mathbf{r}_1, \mathbf{r}_2\rangle$. A ket $|\psi\rangle$ in this Hilbert space then has expansion

$$|\psi\rangle = \int d^3r_1 \int d^3r_2 \psi(\mathbf{r}_1, \mathbf{r}_2) |\mathbf{r}_1, \mathbf{r}_2\rangle \quad (2.100)$$

which is the familiar expansion of a ket in a two-particle Hilbert space. One can interpret $\psi(\mathbf{r}_1, \mathbf{r}_2)$ as the probability amplitude of finding particle 1 in position \mathbf{r}_1 and particle 2 in position \mathbf{r}_2 simultaneously.

Example 2.14 *Addition of orbital and spin angular momentum*

Now consider a spin s particle in three dimensions. As remarked in example 1.2, the ket space corresponding to the spin degree of freedom is \mathbb{C}^{2s+1} , and one usually takes a basis $\{|m\rangle\}_{-s \leq m \leq s}$ of S_z eigenvectors with eigenvalue $\hbar m$. The total Hilbert space for this system is $L^2(\mathbb{R}^3) \otimes \mathbb{C}^{2s+1}$, and we can take as a basis $\{|\mathbf{r}\rangle \otimes |m\rangle\}$ where $\mathbf{r} \in \mathbb{R}^3$ and $-s \leq m \leq s$. Again, the basis vectors are often written as $|\mathbf{r}\rangle |m\rangle$ or even $|\mathbf{r}, m\rangle$. An arbitrary ket $|\psi\rangle$ then has expansion

$$|\psi\rangle = \sum_{m=-s}^s \int d^3r \psi_m(\mathbf{r}) |\mathbf{r}, m\rangle \quad (2.101)$$

where $\psi_m(\mathbf{r})$ is the probability of finding the particle at position \mathbf{r} and with $m\hbar$ units of spin angular momentum in the z -direction. These wavefunctions are sometimes written in column vector form

$$\begin{pmatrix} \psi_s \\ \psi_{s-1} \\ \vdots \\ \psi_{-s+1} \\ \psi_{-s} \end{pmatrix}. \quad (2.102)$$

The total angular momentum operator \mathbf{J} is given by $\mathbf{L} \otimes I + I \otimes \mathbf{S}$ where \mathbf{L} is the orbital angular momentum operator. One might wonder why \mathbf{J} isn't given by $\mathbf{L} \otimes \mathbf{S}$; there is a good answer to this question, but it requires delving into the (fascinating) subject of Lie groups and Lie algebras (see [Ha]), which we won't do here. In the meantime, the reader can get a partial answer by checking (exercise 2.20 below) that the operators $L_i \otimes S_i$ don't satisfy the angular momentum commutation relations whereas the $L_i \otimes I + I \otimes S_i$ do.

Exercise 2.20 Check that

$$[L_i \otimes I + I \otimes S_i, L_j \otimes I + I \otimes S_j] = \sum_{k=1}^3 \epsilon_{ijk} (L_k \otimes I + I \otimes S_k). \quad (2.103)$$

Also show that

$$[L_i \otimes S_i, L_j \otimes S_j] \neq \sum_{k=1}^3 \epsilon_{ijk} L_k \otimes S_k. \quad (2.104)$$

Be sure to use the bilinearity of the tensor product carefully.

Example 2.15 *Addition of spin angular momentum*

Next consider two particles of spin s_1 and s_2 respectively, fixed in space so that they have no translational degrees of freedom. The Hilbert space for this system is $\mathbb{C}^{2s_1+1} \otimes \mathbb{C}^{2s_2+1}$, with basis $\{|m_1\rangle \otimes |m_2\rangle\}$ where $-s_i \leq m_i \leq s_i$, $i = 1, 2$. Again, such tensor product kets are usually abbreviated as $|m_1\rangle|m_2\rangle$ or $|m_1, m_2\rangle$. There are several important linear operators on $\mathbb{C}^{2s_1+1} \otimes \mathbb{C}^{2s_2+1}$:

$\mathbf{S}_1 \otimes I$	Vector spin operator on 1st particle
$I \otimes \mathbf{S}_2$	Vector spin operator on 2nd particle
$\mathbf{S} \equiv \mathbf{S}_1 \otimes I + I \otimes \mathbf{S}_2$	Total vector spin operator
$\mathbf{S}^2 \equiv \sum_i \mathbf{S}_i \mathbf{S}_i$	Total spin squared operator

(Why aren't \mathbf{S}_1^2 and \mathbf{S}_2^2 in our list above?) The vectors $|m_1, m_2\rangle$ are clearly eigenvectors of S_{1z} and S_{2z} and hence S_z (we abuse notation as mentioned below (2.95)) but, as the reader will show in exercise 2.21, they are *not* necessarily eigenvectors of \mathbf{S}^2 . However, since the S_i obey the angular momentum commutation relations (as the reader can check), the general theory of angular momentum tells us that we *can* find a basis for $\mathbb{C}^{2s_1+1} \otimes \mathbb{C}^{2s_2+1}$ consisting of eigenvectors of S_z and \mathbf{S}^2 . Furthermore, it can be shown that the \mathbf{S}^2 eigenvalues that occur are $\hbar^2 s(s+1)$ where

$$|s_1 - s_2| \leq s \leq s_1 + s_2 \quad (2.105)$$

and for a given s the possible S_z eigenvalues are $m\hbar$ where $-s \leq m \leq s$ as usual (See [SA] for details). We will write these basis kets as $\{|s, m\rangle\}$ where the above restrictions on s and m are understood, and where we interpret $\{|s, m\rangle\}$ as a state with total angular momentum equal to $\hbar\sqrt{s(s+1)}$ and with $m\hbar$ units of angular momentum pointing along the z -axis. Thus we have two natural and useful bases for $\mathbb{C}^{2s_1+1} \otimes \mathbb{C}^{2s_2+1}$: $\mathcal{B} = \{|m_1, m_2\rangle\}$ and $\mathcal{B}' = \{|s, m\rangle\}$. What does the transformation

between these two bases look like? Well, by their definition, the A_i^j relating the two bases are given by $e^j(e_{i'})$; using s, m collectively in lieu of the primed index and m_1, m_2 collectively in lieu of the unprimed index, we have, in Dirac notation,

$$A_{s,m}^{m_1,m_2} = \langle m_1, m_2 | s, m \rangle. \quad (2.106)$$

These numbers, the notation for which varies widely throughout the literature, are known as *Clebsch-Gordon Coefficients* and methods for computing them can be found in any standard quantum mechanics textbook.

Let us illustrate the foregoing with an example:

Exercise 2.21 Show that

$$\mathbf{S}^2 = \mathbf{S}_1^2 \otimes I + I \otimes \mathbf{S}_2^2 + 2 \sum_i \mathbf{S}_{1i} \otimes \mathbf{S}_{2i}. \quad (2.107)$$

The right hand side of the above equation is usually abbreviated as $\mathbf{S}_1^2 + \mathbf{S}_2^2 + 2\mathbf{S}_1 \cdot \mathbf{S}_2$. Use this to show that $|m_1, m_2\rangle$ is *not* generally an eigenvector of \mathbf{S}^2 .

Example 2.16 *Entanglement*

Consider two Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 and their tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$. Only some of the vectors in $\mathcal{H}_1 \otimes \mathcal{H}_2$ can be written as $\psi \otimes \phi$; such vectors are referred to as *separable states* or *product states*. All other vectors must be written as linear combinations of the form $\sum_i \psi_i \otimes \phi_i$, and these vectors are said to be *entangled*, since in this case the measurement of the degrees of freedom represented by \mathcal{H}_1 will influence the measurement of the degrees of freedom represented by \mathcal{H}_2 . The classic example of an entangled state comes from the previous example of two fixed particles with spin; taking $s_1 = s_2 = 1/2$, we write the standard basis for \mathbb{C}^2 as $\{|+\rangle, |-\rangle\}$ and consider the state

$$|+\rangle|-\rangle - |-\rangle|+\rangle. \quad (2.108)$$

If an observer measure the first particle to be spin up then a measurement of the second particle's spin is guaranteed to be spin-down, and vice-versa, so measuring one part of the system affects what one will measure for the other part. This is the sense in which the system is entangled. Note that for a product state $\psi \otimes \phi$ such a statement cannot be made: a particular measurement of the first particle cannot affect what one measures for the second, since the second particle's state will be ϕ no matter what. The reader will check below that (2.108) is not a product state.

Exercise 2.22 Prove that (2.108) cannot be written as $\psi \otimes \phi$ for any $\psi, \phi \in \mathbb{C}^2$. Do this by expanding ψ and ϕ in the given basis and showing that no choice of expansion coefficients for ψ and ϕ will yield (2.108).

2.7 Symmetric and Antisymmetric Tensors

Given a vector space V there are certain subspaces of $\mathcal{T}_0^r(V)$ and $\mathcal{T}_r^0(V)$ which are of particular interest: the *symmetric* and *antisymmetric* tensors. A symmetric $(r, 0)$ tensor is an $(r, 0)$ tensor whose value is unaffected by the interchange (or *transposition*) of any two of its arguments, i.e.

$$T(v_1, \dots, v_i, \dots, v_j, \dots, v_r) = T(v_1, \dots, v_j, \dots, v_i, \dots, v_r) \quad (2.109)$$

for any i and j . Symmetric $(0, r)$ tensors are defined similarly. The reader can easily check that the symmetric $(r, 0)$ and $(0, r)$ tensors each form vector spaces, denoted $S^r(V^*)$ and $S^r(V)$ respectively. For $T \in S^r(V^*)$, the symmetry condition implies that the components $T_{i_1 \dots i_r}$ are invariant under the transposition of any two indices, hence invariant under *any* rearrangement of the indices (since any rearrangement can be obtained via successive transpositions), so we may consider the expansion

$$T = T_{i_1 \dots i_r} e^{i_1} \otimes \dots \otimes e^{i_r} \quad (2.110)$$

as a totally symmetric product of dual vectors, hence the notation $S^r(V^*)$. Similar remarks apply, of course, to $S^r(V)$. Notice that for rank 2 tensors, the symmetry condition implies $T_{ij} = T_{ji}$ so that $[T]_{\mathcal{B}}$ for any \mathcal{B} is a symmetric matrix. Also note that it doesn't mean anything to say that a linear operator is symmetric, since a linear operator is a $(1, 1)$ tensor and there is no way of transposing the arguments. One might find that the matrix of a linear operator is symmetric in a certain basis, but this won't necessarily be true in other bases. If we have a non-degenerate Hermitian form to raise and lower indices then we can, of course, speak of symmetry by turning our linear operator into a $(2, 0)$ or $(0, 2)$ tensor.

Example 2.17 $S^2(\mathbb{R}^{2*})$

Consider the set $\{e^1 \otimes e^1, e^2 \otimes e^2, e^1 \otimes e^2 + e^2 \otimes e^1\} \subset S^2(\mathbb{R}^{2*})$ where $\{e^i\}_{i=1,2}$ is the standard dual basis. The reader can check that this set is linearly independent, and that any symmetric tensor can be written as

$$T = T_{11}e^1 \otimes e^1 + T_{22}e^2 \otimes e^2 + T_{12}(e^1 \otimes e^2 + e^2 \otimes e^1) \quad (2.111)$$

so this set is a basis for $S^2(\mathbb{R}^{2*})$. In particular, the Euclidean metric g on \mathbb{R}^2 can be written as

$$g = e^1 \otimes e^1 + e^2 \otimes e^2. \quad (2.112)$$

Note that g would *not* take this simple form in a non-orthonormal basis. \square

There are many symmetric tensors in physics, almost all of them of rank 2. Many of them we've met already: the Euclidean metric on \mathbb{R}^3 , the Minkowski metric on \mathbb{R}^4 , the moment of inertia tensor, and the Maxwell stress tensor. The reader should refer to the examples and check that these are all symmetric tensors.

Exercise 2.23 Let $V = \mathbb{R}^n$ with the standard basis \mathcal{B} . Convince yourself that

$$[e_i \otimes e_j + e_j \otimes e_i]_{\mathcal{B}} = S_{ij} \quad (2.113)$$

where S_{ij} is the symmetric matrix defined in example 1.8.

Next we turn to antisymmetric tensors. An antisymmetric (or *alternating*) $(r, 0)$ tensor is one whose value *changes sign* under transposition of any two of its arguments, i.e.

$$T(v_1, \dots, v_i, \dots, v_j, \dots, v_r) = -T(v_1, \dots, v_j, \dots, v_i, \dots, v_r). \quad (2.114)$$

Again, antisymmetric $(0, r)$ tensors are defined similarly and both sets form vector spaces, denoted $\Lambda^r V^*$ and $\Lambda^r V$ (for $r = 1$ we define $\Lambda^1 V^* = V^*$ and $\Lambda^1 V = V$). Note that (2.114) implies that if any of the v_i are equal to each other, then $T(v_1, \dots, v_r) = 0$. In fact, the reader will show in exercise 2.24 below that if $\{v_1, \dots, v_r\}$ is just linearly dependent, then $T(v_1, \dots, v_r) = 0$. The reader will also show that if $\dim V = n$, then the only tensor in $\Lambda^r V^*$ and $\Lambda^r V$ for $r > n$ is the 0 tensor.

An important operation on antisymmetric tensors is the *wedge product*: Given $f, g \in V^*$ we define the wedge product of f and g , denoted $f \wedge g$, to be the antisymmetric $(2, 0)$ tensor defined by

$$f \wedge g \equiv f \otimes g - g \otimes f. \quad (2.115)$$

Note that $f \wedge g = -g \wedge f$, and that $f \wedge f = 0$. Expanding (2.115) in terms of the e^i gives

$$f \wedge g = f_i g_j (e^i \otimes e^j - e^j \otimes e^i) = f_i g_j e^i \wedge e^j \quad (2.116)$$

so that $\{e^i \wedge e^j\}_{i < j}$ spans all wedge products of dual vectors (note the “ $i < j$ ” stipulation, since $e^i \wedge e^j$ and $e^j \wedge e^i$ are not linearly independent). In fact, the reader can check that $\{e^i \wedge e^j\}_{i < j}$ is linearly independent and spans $\Lambda^2 V^*$, hence is a basis for $\Lambda^2 V^*$. The wedge product can be extended to r -fold products of dual vectors as follows: given r dual vectors f_1, \dots, f_r , we define their wedge product $f_1 \wedge \dots \wedge f_r$ to be the sum of all tensor products of the form $f_{i_1} \otimes \dots \otimes f_{i_r}$ where each term gets a + or a - sign depending on whether an odd or an even number of transpositions

of the factors are necessary to obtain it from $f_1 \otimes \dots \otimes f_r$; if the number is odd the term is assigned -1 , if even a $+1$. Thus,

$$f_1 \wedge f_2 = f_1 \otimes f_2 - f_2 \otimes f_1 \quad (2.117)$$

$$\begin{aligned} f_1 \wedge f_2 \wedge f_3 &= f_1 \otimes f_2 \otimes f_3 + f_2 \otimes f_3 \otimes f_1 + f_3 \otimes f_2 \otimes f_1 \\ &\quad - f_3 \otimes f_2 \otimes f_1 - f_2 \otimes f_1 \otimes f_3 - f_1 \otimes f_3 \otimes f_2 \end{aligned} \quad (2.118)$$

and so on. The reader should convince himself that $\{e^{i_1} \wedge \dots \wedge e^{i_r}\}_{i_1 < \dots < i_r}$ is a basis for $\Lambda^r V^*$. Note that this entire construction can be carried out for vectors as well as dual vectors. Also note that all the comments about symmetry above example 2.17 apply here as well.

Exercise 2.24 Let $T \in \Lambda^r V^*$. Show that if $\{v_1, \dots, v_r\}$ is a linearly dependent set then $T(v_1, \dots, v_r) = 0$. Use the same logic to show that if $\{f_1, \dots, f_r\} \subset V^*$ is linearly dependent, then $f_1 \wedge \dots \wedge f_r = 0$. If $\dim V = n$, show that any set of more than n vectors must be linearly dependent, so that $\Lambda^r V = \Lambda^r V^* = 0$ for $r > n$.

Exercise 2.25 Expand the (2,0) Electromagnetic field tensor of (2.81) in the basis $\{e^i \wedge e^j\}$ where $i < j$ and $i, j = 1, 2, 3, 4$.

Exercise 2.26 Let $\dim V = n$. Show that the dimension of $\Lambda^r V^*$ and $\Lambda^r V$ is $\binom{n}{r} = \frac{n!}{(n-r)!r!}$.

Example 2.18 *Identical particles*

In quantum mechanics we often consider systems which contain *identical particles*, i.e. particles of the same mass, charge and spin. For instance, we might consider n non-interacting hydrogen atoms moving in a potential well, or the two electrons of the helium atom orbiting around the nucleus. In such cases we would assume that the total Hilbert space \mathcal{H}_{tot} would be just the n -fold tensor product of the single particle Hilbert space \mathcal{H} . It turns out, however, that nature doesn't work that way; for certain particles (known as *bosons*) only states in $S^n(\mathcal{H})$ are observed, while for other particles (known as *fermions*) only states in $\Lambda^n \mathcal{H}$ are observed. All known particles are either fermions or bosons. This restriction of the total Hilbert space to either $S^n(\mathcal{H})$ or $\Lambda^n \mathcal{H}$ is known as the *symmetrization postulate* and has far-reaching consequences. For instance, if we have two fermions, we cannot measure the same values for a complete set of quantum numbers for both particles, since then the state would have to include a term of the form $|\psi\rangle |\psi\rangle$ and thus couldn't belong to $\Lambda^2 \mathcal{H}$. This fact that two fermions can't be in the same state is known as the *Pauli*

Exclusion Principle. As another example, consider two identical spin 1/2 fermions fixed in space, so that $\mathcal{H}_{tot} = \Lambda^2 \mathbb{C}^2$. $\Lambda^2 \mathbb{C}^2$ is one-dimensional with basis vector

$$|0, 0\rangle = \left| \frac{1}{2} \right\rangle \left| -\frac{1}{2} \right\rangle - \left| -\frac{1}{2} \right\rangle \left| \frac{1}{2} \right\rangle \quad (2.119)$$

where we have used the notation of example 2.15. If we measure \mathbf{S}^2 or S_z for this system we will get 0. This is in marked contrast to the case of two *distinguishable* spin 1/2 fermions; in this case the Hilbert space is $\mathbb{C}^2 \otimes \mathbb{C}^2$ and we have additional possible state kets

$$|1, 1\rangle = \left| \frac{1}{2} \right\rangle \left| \frac{1}{2} \right\rangle \quad (2.120)$$

$$|1, 0\rangle = \left| \frac{1}{2} \right\rangle \left| -\frac{1}{2} \right\rangle + \left| -\frac{1}{2} \right\rangle \left| \frac{1}{2} \right\rangle \quad (2.121)$$

$$|1, -1\rangle = \left| -\frac{1}{2} \right\rangle \left| -\frac{1}{2} \right\rangle \quad (2.122)$$

which yield nonzero values for \mathbf{S}^2 and S_z . □

The next two examples are a little more mathematical than physical but they are necessary for the discussion of pseudovectors below. Hopefully the reader will also find them of interest in their own right.

Example 2.19 *The Levi-Civita tensor*

Consider \mathbb{R}^n with the standard inner product. Let $\{e_i\}_{i=1\dots n}$ be an orthonormal basis for \mathbb{R}^n and consider the tensor

$$\epsilon \equiv e^1 \wedge \dots \wedge e^n \in \Lambda^n \mathbb{R}^{n*}. \quad (2.123)$$

The reader can easily check that

$$\epsilon_{i_1 \dots i_n} = \begin{cases} 0 & \text{if } \{i_1, \dots, i_n\} \text{ contains a repeated index} \\ -1 & \text{if } \{i_1, \dots, i_n\} \text{ is an odd rearrangement of } \{1, \dots, n\} \\ +1 & \text{if } \{i_1, \dots, i_n\} \text{ is an even rearrangement of } \{1, \dots, n\} \end{cases} \quad (2.124)$$

For $n = 3$, the reader can also check that ϵ_{ijk} has the same values as the Levi-Civita symbol (also written as ϵ_{ijk}), so the Levi-Civita symbol can really be thought of as the components (in the standard basis) of an honest tensor, the Levi-Civita tensor!

Note that $\Lambda^n \mathbb{R}^{n*}$ is one dimensional, and that ϵ is the basis for it described under (2.118).

The reader may object that our construction of ϵ seems to depend on a choice of metric and orthonormal basis. The former is true: ϵ *does* depend on the metric, and we make no apologies for that. As to whether it depends on a particular choice of orthonormal basis, we must do a little bit of investigating; this will require a brief detour into the subject of determinants.

Exercise 2.27 Check that the ϵ tensor on \mathbb{R}^3 satisfies

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if } \{ijk\} \text{ is a cyclic permutation}^{10} \text{ of } \{1, 2, 3\} \\ -1 & \text{if } \{ijk\} \text{ is an anticyclic permutation of } \{1, 2, 3\} \\ 0 & \text{otherwise.} \end{cases} \quad (2.125)$$

Is it true for ϵ on \mathbb{R}^4 that $\epsilon_{ijkl} = 1$ if $\{ijkl\}$ is a cyclic permutation of $\{1, 2, 3, 4\}$?

Example 2.20 *The determinant*

The reader has doubtless encountered determinants before, and has probably seen them defined iteratively; that is, the determinant of a 2×2 square matrix A , denoted $|A|$, is defined to be

$$|A| \equiv A_{11}A_{22} - A_{12}A_{21} \quad (2.126)$$

and then the determinant of a 3×3 matrix B is defined in terms of this, i.e.

$$|B| \equiv B_{11} \begin{vmatrix} B_{22} & B_{23} \\ B_{32} & B_{33} \end{vmatrix} - B_{12} \begin{vmatrix} B_{21} & B_{23} \\ B_{31} & B_{33} \end{vmatrix} + B_{13} \begin{vmatrix} B_{21} & B_{22} \\ B_{31} & B_{32} \end{vmatrix}. \quad (2.127)$$

This expression is known as the *cofactor expansion* of the determinant, and is *not* unique; one can expand about any row (or column), not necessarily (B_{11}, B_{12}, B_{13}) .

In our treatment of the determinant we will take a somewhat more sophisticated approach.¹¹ Take an $n \times n$ matrix A and consider its n columns as n column vectors in \mathbb{R}^n , so that the 1st column vector A_1 has i th component A_{i1} and so on. Then, constructing the ϵ tensor using the standard basis and inner product on \mathbb{R}^n , we define the *determinant of A* , denoted $|A|$, to be

$$|A| \equiv \epsilon(A_1, \dots, A_n) \quad (2.128)$$

¹⁰A *cyclic permutation* of $\{1, \dots, n\}$ is any rearrangement of $\{1, \dots, n\}$ obtained by successively moving numbers from the beginning of the sequence to the end. That is, $\{2, \dots, n, 1\}$, $\{3, \dots, n, 1, 2\}$, and so on are the cyclic permutations of $\{1, \dots, n\}$. *Anti-cyclic* permutations are cyclic permutations of $\{n, n-1, \dots, 1\}$.

¹¹For a complete treatment, however, the reader should consult [HK, Ch. 5].

or in components

$$|A| = \sum_{i_1, \dots, i_n} \epsilon_{i_1 \dots i_n} A_{1i_1} \dots A_{ni_n}. \quad (2.129)$$

The reader should check explicitly that this definition reproduces (2.126) and (2.127) for $n = 2, 3$. The reader can also check in the Problems that many of the familiar properties of determinants (sign change under interchange of columns, invariance under addition of rows, factoring of scalars) follow quite naturally from the definition and the multilinearity and antisymmetry of ϵ . With the determinant in hand we may now explore to what extent the definition of ϵ depends on our choice of orthonormal basis. Consider another orthonormal basis $\{e_{i'} = A_{i'}^j e_j\}$. If we define an ϵ' in terms of this basis, we find

$$\begin{aligned} \epsilon' &= e^{1'} \wedge \dots \wedge e^{n'} \\ &= A_{i_1}^{1'} \dots A_{i_n}^{n'} e^{i_1} \wedge \dots \wedge e^{i_n} \\ &= A_{i_1}^{1'} \dots A_{i_n}^{n'} \epsilon^{i_1 \dots i_n} e^1 \wedge \dots \wedge e^n \\ &= |A| \epsilon \end{aligned} \quad (2.130)$$

where in the 3rd equality we used the fact that if $e^{i_1} \wedge \dots \wedge e^{i_n}$ doesn't vanish it can always be rearranged to give $e^1 \wedge \dots \wedge e^n$, and the resulting sign change if any is accounted for by the Levi-Civita symbol. Now since both $\{e_i\}$ and $\{e_{i'}\}$ are orthonormal bases, A must be an orthonormal matrix, so we can use the product rule for determinants $|AB| = |A||B|$ (see the Problems for a simple proof) and the fact that $|A^T| = |A|$ to get

$$1 = |I| = |AA^T| = |A||A^T| = |A|^2 \quad (2.131)$$

which implies $|A| = \pm 1$. Thus $\epsilon' = \epsilon$ if the two orthonormal bases used in their construction are related by an orthogonal transformation A with $|A| = 1$; such a transformation is called a *rotation*¹², and two bases related by a rotation, or by any transformation with $|A| > 0$, are said to have the *same orientation*. If two bases are related by a basis transformation with $|A| < 0$ then the two bases are said to have the *opposite orientation*. We can then define an *orientation* as a maximal¹³ set of bases all having the same orientation, and the reader can show (see Problems)

¹²No doubt the reader is used to thinking about a rotation as a transformation that preserves distances and fixes a line in space (the *axis of rotation*). This definition of a rotation is particular to \mathbb{R}^3 , since even in \mathbb{R}^2 a rotation can't be considered to be "about an axis" since $\hat{\mathbf{z}} \notin \mathbb{R}^2$. For the equivalence of our general definition and the more intuitive definition in \mathbb{R}^3 , see [Go].

¹³i.e. could not be made bigger.

that \mathbb{R}^n has exactly two orientations. In \mathbb{R}^3 these two orientations are the right handed bases and the left handed bases. Thus we can say that ϵ doesn't depend on a particular choice of orthonormal basis, but it does depend on a metric and a choice of orientation, where the orientation chosen is the one determined by the standard basis.

For orientation changing transformations on \mathbb{R}^3 one can show that A can be written as $A = A_0(-I)$, where A_0 is a rotation and $-I$ is referred to as the *inversion* transformation. The inversion transformation plays a key role in the next example.

Example 2.21 *Pseudovectors in \mathbb{R}^3*

NOTE: All indices in this example refer to *orthonormal* bases. The calculations and results below do not apply to non-orthonormal bases.

A *pseudovector* (or *axial* vector) is a tensor on \mathbb{R}^3 whose components transform like vectors under rotations but don't change signs under inversion. Common examples of pseudovectors are the angular velocity vector ω , as well as all cross products, such as the angular momentum vector $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. It turns out that pseudovectors like these are actually elements of $\Lambda^2\mathbb{R}^3$. To show this, we'll first demonstrate that elements of $\Lambda^2\mathbb{R}^3$ (known as *bivectors*) transform like pseudovectors, and then we'll show why \mathbf{L} and ω are naturally interpreted as bivectors.

First of all, what does it mean to say that bivectors in \mathbb{R}^3 transform 'like' vectors? It means that there exists a 1-1 and onto linear map J from $\Lambda^2\mathbb{R}^3$ to \mathbb{R}^3 such that if we transform the components of $\alpha \in \Lambda^2\mathbb{R}^3$ by a rotation *first* and then apply J , or apply J and *then* rotate the components, we get the same thing. Thus J allows us to identify $\Lambda^2\mathbb{R}^3$ with \mathbb{R}^3 , and the fact that J commutes with rotations means that both those spaces behave 'the same' under rotations. The map J is given by

$$J(\alpha^{23}e_2 \wedge e_3 + \alpha^{31}e_3 \wedge e_1 + \alpha^{12}e_1 \wedge e_2) = \alpha^{23}e_1 + \alpha^{31}e_2 + \alpha^{12}e_3 \quad (2.132)$$

or in components as

$$(J(\alpha))^i = \frac{1}{2}\epsilon^i_{jk}\alpha^{jk} \quad (2.133)$$

so $J(\alpha)$ is a contraction of the ϵ tensor and α . The reader can check that this map is 1-1 and onto; that it is onto actually follows from the fact that it is 1-1, as in exercise 1.5, using the fact that $\dim \Lambda^2\mathbb{R}^3 = \dim \mathbb{R}^3 = 3$. To show that J commutes with rotations, we need to show that (letting $\tilde{\alpha} \equiv J(\alpha)$ to clean up the notation)

$$A^i_j \tilde{\alpha}^j = \frac{1}{2}\epsilon^i_{k'l'} A_m^{k'} A_n^{l'} \alpha^{mn}. \quad (2.134)$$

where A is a rotation. On the left hand side J is applied first followed by a rotation, and on the right hand side the rotation is done first, followed by J .

We now compute for A orthogonal:

$$\begin{aligned}
\frac{1}{2}\epsilon^{i'k'l'}A_m^{k'}A_n^{l'}\alpha^{mn} &= \frac{1}{2}\epsilon_{p'k'l'}\delta^{i'p'}A_m^{k'}A_n^{l'}\alpha^{mn} \\
&= \frac{1}{2}\sum_q\epsilon_{p'k'l'}A_q^{i'}A_q^{p'}A_m^{k'}A_n^{l'}\alpha^{mn} \\
&= \frac{1}{2}\sum_q\epsilon_{qmn}|A|A_q^{i'}\alpha^{mn} \\
&= \frac{1}{2}|A|\epsilon_{mn}^qA_q^{i'}\alpha^{mn} \\
&= |A|A_q^{i'}\tilde{\alpha}^q
\end{aligned} \tag{2.135}$$

where in the second equality we used a variant of (2.32) which comes from writing out $AA^T = I$ in components, in the 3rd equality we used the easily verified fact that $\epsilon_{p'k'l'}A_q^{p'}A_m^{k'}A_n^{l'} = |A|\epsilon_{qmn}$, and in the 4th equality we raised an index to resume use of Einstein summation convention and were able to do so because covariant and contravariant components are equal in orthonormal bases. Now, for rotations $|A| = 1$ so in this case (2.135) and (2.134) are identical and $\tilde{\alpha}$ *does* transform like a vector. For inversion, however, $|A| = |-I| = -1$ so (2.135) tells us that the components of $\tilde{\alpha}$ do *not* change sign under inversion, as those of an ordinary vector would. Another way to see this is given in the exercise below.

Exercise 2.28 Use the matrix transformation law for rank 2 tensors (why don't we have to specify type?) to show that the components of α don't change sign under inversion.

Thus pseudovectors are really just bivectors. The next question is, why are cross products and the angular velocity vector naturally interpreted as bivectors? In the case of cross products, the answer is simple: given two vectors $v, w \in \mathbb{R}^3$, their wedge product $v \wedge w$ looks like (using (2.116))

$$\begin{aligned}
v \wedge w &= (v^2w^3 - v^3w^2)e_2 \wedge e_3 + (v^3w^1 - v^1w^3)e_3 \wedge e_1 \\
&\quad + (v^1w^2 - v^2w^1)e_1 \wedge e_2
\end{aligned} \tag{2.136}$$

so comparison with (2.132) shows that $v \times w = J(v \wedge w)$! So you've actually met the wedge product a long time ago, you just didn't know it.

What about the angular velocity vector ω ?

Example 2.22 *The Angular Velocity Vector*

Let K and K' be two orthonormal bases for \mathbb{R}^3 as in example 1.12, with K time-dependent. One can think of K as representing an accelerated reference frame, or a frame attached to a rotating rigid body; either way we'll refer to it as the *body* frame, and K' as the *space* frame. Now, given a time-dependent vector $v(t) \in \mathbb{R}^3 \ \forall t$, we'd like to compare the time-derivatives of v in the two frames. In physics texts this is usually accomplished via a decomposition

$$\left(\frac{dv}{dt}\right)_{total} = \left(\frac{dv}{dt}\right)_{body} + \left(\frac{dv}{dt}\right)_{rotation} \quad (2.137)$$

where $\left(\frac{dv}{dt}\right)_{total}$ represents the time-derivative of v as 'seen in' the fixed inertial frame, $\left(\frac{dv}{dt}\right)_{body}$ represents the time-derivative of v as 'seen in' the rotating or accelerating frame, and $\left(\frac{dv}{dt}\right)_{rotation}$ represents the difference between the two which is attributable solely to the acceleration of K , and which involves the angular velocity vector ω . We will obtain such a decomposition which will make precise the nature of these terms, as well as providing a proper definition of ω where its bivector nature will be manifest.

Let A be the (time-dependent) matrix of the basis transformation taking K' to K , so that

$$[v]_{K'} = A[v]_K \quad (2.138)$$

Then

$$\begin{aligned} \frac{d}{dt}[v]_{K'} &= \frac{dA}{dt}[v]_K + A\frac{d}{dt}[v]_K \\ &= \frac{dA}{dt}A^{-1}[v]_{K'} + A\frac{d}{dt}[v]_K. \end{aligned} \quad (2.139)$$

Now, $\frac{dA}{dt}A^{-1}$ is actually an antisymmetric matrix:

$$\begin{aligned} 0 &= \frac{d}{dt}(I) \\ &= \frac{d}{dt}(AA^T) \\ &= \frac{dA}{dt}A^T + A\frac{dA^T}{dt} \\ &= \frac{dA}{dt}A^T + \left(\frac{dA}{dt}A^T\right)^T \end{aligned} \quad (2.140)$$

so we can define a bivector $\tilde{\omega}$ (note that here the ‘ \sim ’ denotes the bivector, not the vector) whose components in the space frame are given by $[\tilde{\omega}]_{K'} = \frac{dA}{dt} A^{-1}$. Then we define the *angular velocity vector* ω to be

$$\omega \equiv J(\tilde{\omega}). \quad (2.141)$$

It follows from this definition that

$$\frac{dA}{dt} A^{-1} = \begin{pmatrix} 0 & -\omega^{3'} & \omega^{2'} \\ \omega^{3'} & 0 & -\omega^{1'} \\ -\omega^{2'} & \omega^{1'} & 0 \end{pmatrix} \quad (2.142)$$

(we use primed indices since we’re working with the K' components) so then

$$\frac{dA}{dt} A^{-1} [v]_{K'} = \begin{pmatrix} 0 & -\omega^{3'} & \omega^{2'} \\ \omega^{3'} & 0 & -\omega^{1'} \\ -\omega^{2'} & \omega^{1'} & 0 \end{pmatrix} \begin{pmatrix} v^{1'} \\ v^{2'} \\ v^{3'} \end{pmatrix} \quad (2.143)$$

$$= \begin{pmatrix} \omega^{2'} v^{3'} - \omega^{3'} v^{2'} \\ \omega^{3'} v^{1'} - \omega^{1'} v^{3'} \\ \omega^{1'} v^{2'} - \omega^{2'} v^{1'} \end{pmatrix} \quad (2.144)$$

$$= [\omega \times v]_{K'}. \quad (2.145)$$

If we identify

$$\left[\left(\frac{dv}{dt} \right)_{total} \right]_{K'} \equiv \frac{d}{dt} [v]_{K'} \quad (2.146)$$

$$\left[\left(\frac{dv}{dt} \right)_{body} \right]_{K'} \equiv A \frac{d}{dt} [v]_K \quad (2.147)$$

then (2.139) becomes

$$\left(\frac{dv}{dt} \right)_{total} = \omega \times v + \left(\frac{dv}{dt} \right)_{body} \quad (2.148)$$

written in components referred to the space frame. This equation should be familiar from upper-division mechanics texts, and in our treatment we see that ω arises naturally as a bivector. Note that we could also write (2.139) in the body frame, which (as the reader can verify) would then tell us that

$$[\tilde{\omega}]_K = A^T \frac{dA}{dt}. \quad (2.149)$$

Regardless of the frame in which we write (2.137), the reader should spend some time convincing himself that the identifications (2.146) and (2.147) above really do embody what we mean by the time derivatives of v as ‘seen in’ the different frames.

Chapter 2 Problems

1. In this problem we explore the properties of $n \times n$ orthogonal matrices. This is the set of real invertible matrices A satisfying $A^T = A^{-1}$, and is denoted $O(n)$.

- a) Is $O(n)$ a subspace of $M_n(\mathbb{R})$?
- b) Show that the product of two orthogonal matrices is again orthogonal, that the inverse of an orthogonal matrix is again orthogonal, and that the identity matrix is orthogonal. These properties show that $O(n)$ is a *group*, i.e. a set with an associative multiplication operation and identity element such that the set is closed under multiplication and every element has a multiplicative inverse.
- c) Show that the columns of an orthogonal matrix A , viewed as vectors in \mathbb{R}^n , are mutually orthogonal under the usual inner product. Show the same for the rows. Show that for an active transformation, i.e.

$$[e_{i'}]_{\mathcal{B}} = A[e_i]_{\mathcal{B}} \quad (2.150)$$

where $\mathcal{B} = \{e_i\}_{i=1\dots n}$ so that $[e_i]_{\mathcal{B}}^T = (0, \dots, \underbrace{1}_{ith \ slot}, \dots, 0)$, the columns of A are the $[e_{i'}]_{\mathcal{B}}$. In other words, the components of the new basis vectors in the old basis are just the columns of A .

- d) Show that the orthogonal matrices A with $|A| = 1$, the rotations, form a subgroup unto themselves, denoted $SO(n)$. Do the matrices with $|A| = -1$ also form a subgroup?
2. Prove the following basic properties of the determinant directly from the definition (2.128). We will restrict our discussion to operations with columns, though it can be shown that all the corresponding statements for rows are true as well.

- a) Any matrix with a column of zeros has $|A| = 0$.
- b) Multiplying a column by a scalar c multiplies the whole determinant by c .
- c) The determinant changes sign under interchange of any two columns.
- d) Adding two columns together, i.e. sending $A_i \rightarrow A_i + A_j$ for any i and j , doesn't change the value of the determinant.

3. One can extend the definition of determinants from matrices to more general linear operators as follows: We know that a linear operator T on a vector space V (equipped with an inner product and orthonormal basis $\{e_i\}_{i=1\dots n}$) can be extended to arbitrary tensor products of V by

$$T(v_1 \otimes \dots \otimes v_p) = (Tv_1) \otimes \dots \otimes (Tv_p) \quad (2.151)$$

and thus, since $\Lambda^n \mathbb{R}^n \subset \underbrace{\mathbb{R}^n \otimes \dots \otimes \mathbb{R}^n}_{n \text{ times}}$, the action of T extends to $\Lambda^n V$ similarly by

$$T(v_1 \wedge \dots \wedge v_n) = (Tv_1) \wedge \dots \wedge (Tv_n), \quad v_i \in V \quad \forall i. \quad (2.152)$$

Consider then the action of T on the contravariant version of ϵ , the tensor $\tilde{\epsilon} \equiv e_1 \wedge \dots \wedge e_n$; we know from exercise 2.26 that $\Lambda^n V$ is one dimensional so that $T(\tilde{\epsilon}) = (Te_1) \wedge \dots \wedge (Te_n)$ is proportional to $\tilde{\epsilon}$, and we define the *determinant* of T to be this proportionality constant, so that

$$(Te_1) \wedge \dots \wedge (Te_n) \equiv |T| e_1 \wedge \dots \wedge e_n. \quad (2.153)$$

- a) Show by expanding the left hand side of (2.153) in components that this more general definition reduces to the old one of (2.129) in the case of $V = \mathbb{R}^n$.
- b) Use this definition of the determinant to show that for two linear operators B and C on V ,

$$|BC| = |B||C|. \quad (2.154)$$

In particular, this result holds when B and C are square matrices.

- c) Use b) to show that the determinant of a matrix is invariant under similarity transformations (see example 2.5). Conclude that we could have defined the determinant of a linear operator T as the determinant of its matrix in *any* basis.

4. Let V be a vector space V with an inner product and orthonormal basis $\{e_i\}_{i=1\dots n}$. Prove that a linear operator T is invertible if and only if $|T| \neq 0$, as follows:
- Show that T is invertible if and *only if* $\{T(e_i)\}_{i=1\dots n}$ is a linearly independent set (see exercise 1.6 for the ‘if’ part of the statement).
 - Show that $|T| \neq 0$ if and only if $\{T(e_i)\}_{i=1\dots n}$ is a linearly independent set.
5. The determinant of a matrix A in \mathbb{R}^n can actually be interpreted as follows: given the standard basis $\{e_i\}_{i=1\dots n}$, we can imagine the standard n -dimensional cube determined by the points $(0, \dots, \underbrace{1}_{\text{ith slot}}, \dots, 0)$ given by the e_i and whose n -dimensional volume is 1. We can then imagine the cube determined by the Ae_i ; the n -dimensional volume of this parallelepiped is then $|A|$, where the sign of the determinant signifies whether or not the orientation of $\{Ae_i\}$ is the same as $\{e_i\}$. Thus the determinant tells us the *oriented volume* of the n -cube determined by $\{Ae_i\}$. Verify this in the cases $n = 2, 3$, i.e. show that for $n = 2$ the parallelogram spanned by $\{Ae_1, Ae_2\}$ has oriented area $|A|$, and that for $n = 3$ the three dimensional parallelepiped spanned by $\{Ae_1, Ae_2, Ae_3\}$ has oriented volume $|A|$. Feel free to refer to example 2.21 and to use the cross product formulas you learned in vector calculus. Note that if A is not invertible then the Ae_i are linearly dependent, hence span a space of dimension less than n and thus yield n -dimensional volume 0, so the geometrical picture is consistent with the results of the previous problem.
6. Let \mathcal{B} be the standard basis for \mathbb{R}^n , O the set of all bases related to \mathcal{B} by a basis transformation with $|A| > 0$, and O' the set of all bases related to \mathcal{B} by a transformation with $|A| < 0$.
- Using what we’ve learned in the preceding problems, show that a basis transformation matrix A cannot have $|A| = 0$.
 - O is by definition an orientation. Show that O' is also an orientation, and conclude that \mathbb{R}^n has exactly two orientations. Note that both O and O' contain orthonormal and non-orthonormal bases.
 - For what n is $A = -I$ an orientation-changing transformation?

Chapter 3

Groups, Lie Groups and Lie Algebras

In physics we are often interested in how a particular object behaves under a particular set of transformations; for instance, in the classical physics literature one reads that a dipole moment transforms like a vector under rotations and a quadropole moment like a (2nd rank) tensor, and that the electric and magnetic fields transform like vectors under rotations but like a 2nd rank antisymmetric tensor under Lorentz transformations. Similarly, in quantum mechanics one is often interested in the “spin” of a ket (which specifies how it transforms under rotations), or its behavior under the time-reversal or space inversion (parity) transformations. This knowledge is particularly useful as it leads to the many famous “selection rules” which greatly simplify evaluation of matrix elements. Transformations are also crucial in quantum mechanics because all physical observables can be considered as “infinitesimal generators” of particular transformations; e.g., the angular momentum operators “generate” rotations (as we discussed briefly in Problem 3 of Chapter 1) and the momentum operator “generates” translations.

Like tensors, this material is usually treated in a somewhat ad-hoc way which facilitates computation but obscures the underlying mathematical structures. These underlying structures are known to mathematicians as *group theory*, *Lie theory* and *representation theory*, and are known collectively to physicists as just “group theory”. Our aim in this second half of the book is to present the basic facts of this theory as well as its manifold applications to physics, both to clarify and unify the diverse phenomena in physics in which it is involved and also to provide a nice application of what we’ve learned about tensors.

Before we discuss how particular objects transform, however, we must discuss the

transformations themselves. All the transformations we'll be interested in share a few common properties: First, the performance of two successive transformations is always equivalent to the performance of a single, third transformation (just think of rotations and how any two successive rotations about two axes can be considered as a single rotation about a third axis). Second, every transformation has an inverse which undoes it (in the case of rotations, the inverse to any given rotation is a rotation about the same axis in the opposite direction). Sets of transformations like these occur so often in mathematics and physics that they are given a name: groups.

3.1 Groups - Definition and Examples

The definition of a group that we're about to give may appear somewhat abstract, but is meant to embody the most important properties of sets of transformations. After giving the definition and establishing some basic properties, we'll proceed shortly to some concrete examples.

A *group* is a set G together with a multiplication operation, denoted \cdot , that satisfies the following axioms:

1. (Closure) $g, h \in G$ implies $g \cdot h \in G$.
2. (Associativity) For $g, h, k \in G$, $g \cdot (h \cdot k) = (g \cdot h) \cdot k$.
3. (Existence of the Identity) There exists an element $e \in G$ such that $g \cdot e = e \cdot g = g \quad \forall g \in G$.
4. (Existence of Inverses) $\forall g \in G$ there exists an element h such that $g \cdot h = h \cdot g = e$.

In the case of transformations, the multiplication operation is obviously just composition; that is, if R and S are 3-d rotations, for instance, then $R \cdot S$ is just S followed by R . Note that we don't necessarily have $R \cdot S = S \cdot R$ for all rotations R and S ; if we did, we would say that the group is *commutative* (or *abelian*). If we do not, as is the case for rotations and most other groups of physical interest, we say the group is *non-commutative* (or *non-abelian*).

There are several important properties of groups that follow almost immediately from the definition. Firstly, the *identity is unique*, for if e and f are both elements satisfying axiom 3 then we have

$$e = e \cdot f \quad \text{since } f \text{ is an identity} \tag{3.1}$$

$$= f \quad \text{since } e \text{ is an identity.} \tag{3.2}$$

Secondly, *inverses are unique*: Let $g \in G$ and let h and k both be inverses of g . Then

$$g \cdot h = e \tag{3.3}$$

so multiplying both sides on the left by k gives

$$\begin{aligned} k \cdot (g \cdot h) &= k, \\ (k \cdot g) \cdot h &= k \text{ by associativity,} \\ e \cdot h &= k \text{ since } k \text{ is an inverse of } g, \\ h &= k. \end{aligned}$$

We henceforth denote the unique inverse of an element g as g^{-1} .

Thirdly, if $g \in G$ and h is merely a *right* inverse for g , i.e.

$$g \cdot h = e, \tag{3.4}$$

then h is also a *left* inverse for g and is hence the unique inverse g^{-1} . This is seen as follows:

$$\begin{aligned} h \cdot g &= (g^{-1} \cdot g) \cdot (h \cdot g) \\ &= (g^{-1} \cdot (g \cdot h)) \cdot g \text{ by associativity} \\ &= (g^{-1} \cdot e) \cdot g \text{ by (3.4)} \\ &= g^{-1} \cdot g \\ &= e, \end{aligned}$$

so $h = g^{-1}$.

The last few properties concern inverses and can be verified immediately by the reader:

$$(g^{-1})^{-1} = g \tag{3.5}$$

$$(g \cdot h)^{-1} = h^{-1} \cdot g^{-1} \tag{3.6}$$

$$e^{-1} = e. \tag{3.7}$$

Exercise 3.1 Prove the *cancellation laws* for groups, i.e. that

$$g_1 \cdot h = g_2 \cdot h \Rightarrow g_1 = g_2 \tag{3.8}$$

$$h \cdot g_1 = h \cdot g_2 \Rightarrow g_1 = g_2. \tag{3.9}$$

Before we get to some examples, we should note that Properties 2 and 3 in the definition above are usually obviously satisfied and rarely does one bother to check them explicitly. The important thing in showing that a set is a group is verifying that it is closed under multiplication and contains all its inverses. Also, as a matter of notation, from now on we will omit the \cdot when writing a product, and simply write gh for $g \cdot h$.

Example 3.1 *SO(2) Special orthogonal group in two dimensions*

As the reader may recall from Problem 1 of Chapter 2, $SO(2)$ is defined to be the set of all real orthogonal 2×2 matrices with determinant equal to 1 (we remind the reader that A “orthogonal” means $A^{-1} = A^T$). The reader will check in exercise 3.2 that $SO(2)$ is an abelian group and that the general form of an element of $SO(2)$ is

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (3.10)$$

The reader will recognize that such a matrix represents a counterclockwise rotation of θ radians in the $x - y$ plane.

Exercise 3.2 Verify that $SO(2)$ is a group. If the reader has done Problem 1, part d) of Chapter 2 then he has already done this. Also check that $SO(2)$ is abelian. Next, consider an arbitrary matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (3.11)$$

and impose the orthogonality condition, as well as $|A| = 1$. Show that (3.10) is the most general solution to these constraints.

Example 3.2 *SO(3) Special orthogonal group in three dimensions*

This group is defined, just as for $SO(2)$, as the set of all orthogonal 3×3 matrices with determinant equal to one. The reader’s proof that $SO(2)$ is a group should carry over verbatim to show that $SO(3)$ is a group. As we mentioned in example 2.20, orthogonal matrices with determinant equal to 1 actually represent rotations, and though this is obvious from (3.10) in the two dimensional case it is not as obvious in the three dimensional case. We tend to think of three-dimensional rotations as transformations which preserve distances and fix a line in space (called the *axis of rotation*). Do elements of $SO(3)$ do this? To find out, let g be the Euclidean metric

on \mathbb{R}^3 . The distance between two points $x, y \in \mathbb{R}^3$ is then $|x - y| \equiv \sqrt{g(x - y, x - y)}$, and the statement that $R \in SO(3)$ preserves distances means that

$$|Rx - Ry| = |x - y|. \quad (3.12)$$

The reader will verify this in exercise 3.3 below. In fact, the reader will prove the stronger statement that

$$g(Rx, Ry) = g(x, y) \quad \forall x, y \in \mathbb{R}^3 \quad (3.13)$$

which can actually be taken as the *definition* of an orthogonal matrix. These two definitions correspond to the active and passive viewpoints of transformations: our first definition of an orthogonal matrix as one which implements the basis change from one orthonormal basis to another (and hence satisfies $R^{-1} = R^T$) gives the passive viewpoint, while the second definition, given by (3.13), of orthogonal matrices as linear operators which preserve an inner product, gives the active viewpoint.

Now what about the axis of rotation? Does each $R \in SO(3)$ have one? An axis of rotation is a line of points that are unaffected by R , so this line actually comprises a one dimensional eigenspace of R with eigenvalue 1. The question then becomes, does every $R \in SO(3)$ have 1 as an eigenvalue? The answer is yes, as the reader will show in Problem 1. This fact is known as *Euler's Theorem*.

Now that we have convinced ourselves that every $R \in SO(3)$ is a rotation (and these are, in fact, *all* rotations), can we find a general form for R ? As the reader may know from classical mechanics courses, an arbitrary rotation can be described in terms of the *Euler angles*, which tell us how to rotate a given orthonormal basis into another of the same orientation (or handedness). In classical mechanics texts¹, it is shown that this can be achieved by rotating the given axes by an angle ϕ around the original z -axis, then by an angle θ around the *new* y -axis, and finally by an angle ψ around the *new* z -axis. If we take the passive point of view, these three rotations take the form

$$\begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix}, \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.14)$$

so multiplying them together gives a general form for $R \in SO(3)$:

$$\begin{pmatrix} \cos \psi \cos \phi - \cos \theta \sin \phi \sin \psi & -\sin \psi \cos \phi - \cos \theta \sin \phi \cos \psi & \sin \theta \sin \phi \\ \cos \psi \sin \phi + \cos \theta \cos \phi \sin \psi & -\sin \phi \sin \psi + \cos \theta \cos \phi \cos \psi & -\sin \theta \cos \phi \\ \sin \theta \sin \psi & \sin \theta \cos \psi & \cos \theta \end{pmatrix}. \quad (3.15)$$

¹such as [Go]

Exercise 3.3 Assume the standard basis in \mathbb{R}^3 so that $g(x, y) = [x]^T[y]$. Then use orthogonality to show that $g(Rx, Ry) = g(x, y)$. Also, turn the definition around and show that any linear operator R which satisfies $g(Rx, Ry) = g(x, y) \forall x, y \in \mathbb{R}^3$ must satisfy $R^{-1} = R^T$.

Example 3.3 $O(3)$ Orthogonal group in 3 dimensions

Another transformation of our three dimensional space that is familiar from physics and that we've mentioned before is the *inversion* transformation, $-I$, which sends every vector to minus itself. If we try to add this to the rotations to get a group, we get the *orthogonal group* $O(3)$ which is defined to be the set of all 3×3 orthogonal matrices R , this time *without* the condition that the $|R| = 1$. Of course, as we pointed out in example 2.20, the orthogonality condition implies that $|R| = \pm 1^2$, so in going from $SO(3)$ to $O(3)$ we are just adding all the orthogonal matrices with $|R| = -1$ (which of course includes $-I$). The elements with $|R| = 1$ are sometimes referred to as *proper rotations*, and the elements with $|R| = -1$ as the *improper rotations*. The proper and improper rotations are *disconnected*, in the sense that one cannot continuously go from matrices with $|R| = 1$ to matrices with $|R| = -1$. This is represented schematically above. One can obtain any of the improper rotations, however, by multiplying a proper rotation by $-I$; this is easy to see, for if R is an improper rotation then $-R$ is a proper rotation which yields R when multiplied by $-I$.

Example 3.4 $SU(2)$ Special unitary group in two complex dimensions

This group is defined to be the set of all 2×2 complex matrices A which satisfy $|A| = 1$ and

$$A^\dagger = A^{-1}. \tag{3.16}$$

The reader will verify below that $SU(2)$ is a group. The reader will also verify that a generic element of $SU(2)$ looks like

$$\begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \quad \alpha, \beta \in \mathbb{C}, \quad |\alpha|^2 + |\beta|^2 = 1. \tag{3.17}$$

²This fact can be understood geometrically: since orthogonal matrices preserve distances and angles, they should preserve volumes as well, and as we learned in Problem 5 of Chapter 2 the determinant measure how volume changes under the action of a linear operator, so any volume preserving operator should have determinant ± 1 , where the sign is determined by whether or not the orientation is reversed.

We could also use three real parameters with no conditions rather than two complex parameters with a constraint; one such parametrization is

$$\begin{pmatrix} e^{i(\psi+\phi)/2} \cos \frac{\theta}{2} & ie^{i(\psi-\phi)/2} \sin \frac{\theta}{2} \\ ie^{-i(\psi-\phi)/2} \sin \frac{\theta}{2} & e^{-i(\psi+\phi)/2} \cos \frac{\theta}{2} \end{pmatrix} \quad (3.18)$$

where we have used the same symbols for our parameters as we did for the Euler angles. This is no accident, as there is a close relationship between $SU(2)$ and $SO(3)$, which we will discuss in detail later. This relationship underlies the appearance of $SU(2)$ in quantum mechanics, where rotations are implemented on spin 1/2 particles by elements of $SU(2)$.

The reader has already checked in exercise 2.5 that matrices satisfying (3.16) represent transformations between orthonormal bases of a complex vector space. As with orthogonal groups, however, there is another way to interpret the condition (3.16); we could define a unitary matrix as one which, when interpreted as a linear operator on \mathbb{C}^n , *preserves the hermitian scalar product on \mathbb{C}^n* , in the sense of (3.20) below. The reader will check that this condition is equivalent to (3.16), and so we have a passive and active view of unitary transformations, just as we did for orthogonal transformations.

Exercise 3.4 Verify that $SU(2)$ satisfies the group axioms. Also, consider an arbitrary complex matrix

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad (3.19)$$

and impose the unit determinant and unitary conditions. Show that (3.17) is the most general solution to these constraints.

Exercise 3.5 Let $U \in \mathcal{T}(\mathbb{C}^n)$ satisfy

$$(Uv, Uw) = (v, w) \quad \forall v, w \in \mathbb{C}^n. \quad (3.20)$$

Assume the standard basis on \mathbb{C}^n , and show that this condition is equivalent to

$$[U]^\dagger = [U]^{-1}. \quad (3.21)$$

Example 3.5 $SO(3, 1)_o$ *The restricted Lorentz group*

The *restricted Lorentz group* $SO(3, 1)_o$ is defined to be the set of all 4×4 real matrices A satisfying $|A| = 1$, $A_{44} > 1$, and

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = A^T \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} A. \quad (3.22)$$

The reader will verify in problem 2 below that $SO(3, 1)_o$ is a group. Where does its definition come from? Consider \mathbb{R}^4 with the Minkowski metric. As noted in example 2.6, any matrix satisfying (3.22) represents a transformation from one orthonormal basis to another, so the component form of the Minkowski metric is unchanged. As Einstein showed, this condition is precisely the one that guarantees the invariance of the speed of light under the change of inertial reference frame described by A . What about the other conditions? $A_{44} > 0$ just says that the transformation doesn't reverse the direction of time (so that clocks in the new coordinates aren't running backwards) and this, together with $|A| = 1$, implies that A does not reverse the orientation of the space axes. Thus, $SO(3, 1)_o$ represents all the changes from one inertial reference frame to another that don't reverse the orientation of time or space. These transformations are known as *restricted Lorentz transformations*. A transformation between two reference frames that have parallel axes but may be moving relative to each other is called a *pure Lorentz transformation* or *boost*.

What does a generic restricted Lorentz transformation look like? As with the previous examples, there is some arbitrariness in our choice of description, but we can proceed as follows: any restricted Lorentz transformation can be written as the product of a rotation and a boost³, where a rotation in this context looks like

$$R' = \begin{pmatrix} 1 & \vec{0} \\ \vec{0} & R \end{pmatrix} \quad (3.23)$$

where $\vec{0}$ stands for a column (on the left) or row (on the top) of three zeros and $R \in SO(3)$. The reader can check that such a matrix is a Lorentz transformation, so we see that $SO(3, 1)_o$ subsumes $SO(3)$. As for the form of an arbitrary boost, the reader will show later in this book that

$$L = \begin{pmatrix} \frac{u_x^2(\cosh u - 1)}{u^2} + 1 & \frac{u_x u_y(\cosh u - 1)}{u^2} & \frac{u_x u_z(\cosh u - 1)}{u^2} & \frac{u_x}{u} \sinh u \\ \frac{u_y u_x(\cosh u - 1)}{u^2} & \frac{u_y^2(\cosh u - 1)}{u^2} + 1 & \frac{u_y u_z(\cosh u - 1)}{u^2} & \frac{u_y}{u} \sinh u \\ \frac{u_z u_x(\cosh u - 1)}{u^2} & \frac{u_z u_y(\cosh u - 1)}{u^2} & \frac{u_z^2(\cosh u - 1)}{u^2} + 1 & \frac{u_z}{u} \sinh u \\ \frac{u_x}{u} \sinh u & \frac{u_y}{u} \sinh u & \frac{u_z}{u} \sinh u & \cosh u \end{pmatrix} \quad (3.24)$$

is an arbitrary boost, where \vec{u} (with $u \equiv |\vec{u}|$) is a quantity known as the *rapidity*. The reader will verify below that \vec{u} is related to the relative velocity $\vec{\beta}$ between the frames by

$$\vec{\beta} = \frac{\tanh u}{u} \vec{u}. \quad (3.25)$$

³this should be intuitively clear; for a discussion see [Go].

and that in terms of $\vec{\beta}$ and $\gamma \equiv (1 - \beta^2)^{-1/2} = \cosh u$, $\beta \equiv |\vec{\beta}|$, L takes the form

$$L = \begin{pmatrix} \frac{\beta_x^2(\gamma-1)}{\beta^2} + 1 & \frac{\beta_x\beta_y(\gamma-1)}{\beta^2} & \frac{\beta_x\beta_z(\gamma-1)}{\beta^2} & \beta_x\gamma \\ \frac{\beta_y\beta_x(\gamma-1)}{\beta^2} & \frac{\beta_y^2(\gamma-1)}{\beta^2} + 1 & \frac{\beta_y\beta_z(\gamma-1)}{\beta^2} & \beta_y\gamma \\ \frac{\beta_z\beta_x(\gamma-1)}{\beta^2} & \frac{\beta_z\beta_y(\gamma-1)}{\beta^2} & \frac{\beta_z^2(\gamma-1)}{\beta^2} + 1 & \beta_z\gamma \\ \beta_x\gamma & \beta_y\gamma & \beta_z\gamma & \gamma \end{pmatrix}. \quad (3.26)$$

Note that L has three arbitrary parameters so that our arbitrary restricted Lorentz transformation $R'L$ has six parameters total.

Exercise 3.6 Use L as an passive transformation to obtain new coordinates (x', y', z', t') from old by

$$\begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix} = L \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}. \quad (3.27)$$

Show that the origin of the primed frame, defined by $x' = y' = z' = 0$, moves with velocity (3.25) relative to the old frame, and substitute this into (3.24) to get (3.26).

Example 3.6 $O(3, 1)$ *The extended Lorentz group*

In the previous example we restricted our changes of inertial reference frame to those which preserved the orientation of space and time. This is sufficient in classical mechanics, but in quantum mechanics we are often interested in the effects of space and time inversion on the various Hilbert spaces we're working with. If we add spatial inversion, also called *parity* and represented by the matrix

$$P = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.28)$$

as well as time-reversal, represented by

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (3.29)$$

to the restricted Lorentz group, we get the *improper* or *extended* Lorentz group $O(3, 1)$, which is defined by (3.22) with no restrictions on the determinant or A_{44} .

The reader should verify that $P, T \in O(3, 1)$, but $P, T \notin SO(3, 1)_o$. In fact, $|P| = |T| = -1$, which is no accident; as in the case of the orthogonal group, the defining equation (3.22) restricts the determinant, and the reader can check that (3.22) implies that $|A| = \pm 1$. In this case, however, we have four disconnected components instead of two! Obviously those matrices with $|A| = 1$ must be disconnected from those with $|A| = -1$, but those which reverse the orientation of the space axes must also be disconnected from those which do not, and those which reverse the orientation of time must be disconnected from those which do not. This is represented schematically above. Note that, as in the case of $O(3)$, multiplication by the transformations P and T take us to and from the various different components.

Example 3.7 $SL(2, \mathbb{C})$ *Special linear group in two complex dimensions*

$SL(2, \mathbb{C})$ is defined to be the set of all 2×2 complex matrices A with $|A| = 1$. By now it should be obvious that this set is a group. The general form of $A \in SL(2, \mathbb{C})$ is

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad a, b, c, d \in \mathbb{C}, \quad ad - bc = 1. \quad (3.30)$$

The unit determinant constraint means that A is determined by three complex parameters or six real parameters, just as for $SO(3, 1)_o$. This is no coincidence; in fact, $SL(2, \mathbb{C})$ bears the same relationship to $SO(3, 1)_o$ as $SU(2)$ bears to $SO(3)$, in that $SL(2, \mathbb{C})$ implements restricted Lorentz transformations on spin 1/2 particles! We will discuss this in detail later, where we will also show that an arbitrary boost is implemented on a spin 1/2 particle by $\tilde{P} \in SL(2, \mathbb{C})$ of the form

$$\begin{pmatrix} \cosh \frac{u}{2} - \frac{u_z}{u} \sinh \frac{u}{2} & -\frac{1}{u}(u_x - iu_y) \sinh \frac{u}{2} \\ -\frac{1}{u}(u_x + iu_y) \sinh \frac{u}{2} & \cosh \frac{u}{2} + \frac{u_z}{u} \sinh \frac{u}{2} \end{pmatrix}. \quad (3.31)$$

This, together with the facts that an arbitrary rotation is implemented by an $SU(2)$ matrix \tilde{R} of the form (3.18) (note that $SU(2) \subset SL(2, \mathbb{C})$, which makes sense since $SO(3, 1)_o$ subsumes $SO(3)$) and that any Lorentz transformation can be written as a product of a boost and a rotation, yields the general form $\tilde{R}\tilde{P}$ for an element of $SL(2, \mathbb{C})$ in terms of the same parameters we used for $SO(3, 1)_o$.

Example 3.8 \mathbb{Z}_2 *The group with two elements*

Consider the set $\mathbb{Z}_2 \equiv \{e, g\}$ with associative multiplication law given by $g^2 = e$ (note that the rest of the possible products are determined by the stipulation that e is the identity). The reader can easily check that this is a group, in fact an abelian

group. Though this group is defined in a much more abstract way than the groups encountered above, we'll see in the next section that concrete representations of this group pop up in a few places in physics.

Example 3.9 S_n *The symmetric group on n letters*

This group does not usually occur explicitly in physics but is intimately tied to permutation symmetry, the physics of identical particles, and much of the mathematics we discussed in section 2.7. The *symmetric group on n letters* (also known as the *permutation group*), denoted S_n , is defined to be the set of all 1-1 and onto maps of the set $\{1, 2, \dots, n\}$ to itself, where the product is just the composition of maps. The maps are known as *permutations*. The reader should check that any composition of permutations is again a permutation and that permutations are invertible, so that S_n is a group. This verification is simple, and just relies on the fact that permutations are, by definition, 1-1 and onto.

Any permutation σ is specified by the n numbers $\sigma(i)$, $i = 1 \dots n$, and can conveniently be notated as

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}. \quad (3.32)$$

In such a scheme, the identity in S_3 would just look like

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad (3.33)$$

while the cyclic permutation σ_1 given by $1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1$ would look like

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad (3.34)$$

and the transposition σ_2 which switches 1 and 2 and leaves 3 alone would look like

$$\sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}. \quad (3.35)$$

How do we take products of permutations? Well, the product $\sigma_1 \cdot \sigma_2$ would take on the following values:

$$(\sigma_1 \cdot \sigma_2)(1) = \sigma_1(\sigma_2(1)) = \sigma_1(2) = 3 \quad (3.36)$$

$$(\sigma_1 \cdot \sigma_2)(2) = \sigma_1(1) = 2 \quad (3.37)$$

$$(\sigma_1 \cdot \sigma_2)(3) = \sigma_1(3) = 1 \quad (3.38)$$

so we have

$$\sigma_1 \cdot \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}. \quad (3.39)$$

The reader should take the time to inspect (3.39) and understand how to take such a product of permutations without having to write out (3.36)-(3.38).

Though a proper discussion of the applications of S_n to physics must wait until the next section, we can point out here that if we have an n -fold tensor product $V \otimes V \otimes \dots \otimes V$ of a vector space V , then S_n acts on this by

$$\sigma(v_1 \otimes v_2 \otimes \dots \otimes v_n) = v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \dots \otimes v_{\sigma(n)}. \quad (3.40)$$

In the case of n identical particles in quantum mechanics, where the total Hilbert space is the n -fold tensor product of the single-particle Hilbert space \mathcal{H} , this action effectively *interchanges particles*, and we will later restate the symmetrization postulate from example 2.18 in terms of this action of S_n on $\mathcal{H} \otimes \mathcal{H} \otimes \dots \otimes \mathcal{H}$.

3.2 Homomorphism and Isomorphism

In the last section we claimed that there is a close relationship between $SU(2)$ and $SO(3)$, as well as between $SL(2, \mathbb{C})$ and $SO(3, 1)_o$. We now make this relationship precise, and show that a similar relationship exists between S_n and \mathbb{Z}_2 . We will also define what it means for two groups to be ‘the same’, which will then tie into our abstract discussion of \mathbb{Z}_2 in the last section.

Given two groups G and H , a *homomorphism* from G to H is a map $\phi : G \rightarrow H$ such that

$$\phi(g_1 g_2) = \phi(g_1) \phi(g_2) \quad \forall g_1, g_2 \in G. \quad (3.41)$$

Note that the product in the left hand side of (3.41) takes place in G , whereas the product on the right hand side takes place in H . A homomorphism should be thought of as a map from one group to another which preserves the multiplicative structure. Note that ϕ need not be 1-1 or onto; if it is onto then ϕ is said to be a *homomorphism onto* H , and if in addition it is 1-1 then we say ϕ is an *isomorphism*. If ϕ is an isomorphism then it is invertible and thus sets up a one-to-one correspondence which preserves the group structure, so we regard G and H as ‘the same’ group, just with different labels for the elements, with ϕ providing the dictionary between the different labeling schemes.

Exercise 3.7 Let $\phi : G \rightarrow H$ be a homomorphism, and let e be the identity in G and e' the identity in H . Show that

$$\phi(e) = e' \tag{3.42}$$

$$\phi(g^{-1}) = \phi(g)^{-1} \quad \forall g \in G. \tag{3.43}$$

Suppose ϕ is a homomorphism but not an isomorphism. Is there a way to quantify how far it is from being an isomorphism? Define the *kernel* of ϕ to be the set $K \equiv \{g \in G \mid \phi(g) = e'\}$ where e' is the identity in H . In other words, K is the set of all elements of G that get sent to e' under ϕ . If ϕ is an isomorphism, then $K = \{e\}$. Also, if we have $\phi(g_1) = \phi(g_2) = h$, then

$$\phi(g_1 g_2^{-1}) = \phi(g_1) \phi(g_2)^{-1} = h h^{-1} = e \tag{3.44}$$

so $g_1 g_2^{-1}$ is in the kernel of ϕ , i.e. $g_1 g_2^{-1} = k$ for some $k \in K$. Multiplying on the right by g_2 then gives $g_1 = k g_2$ so we see that any two elements of G that give the same element of H under ϕ are related by left multiplication by an element of K . Conversely, if we are given $g \in G$ and $\phi(g) = h$, then for all $k \in K$, $\phi(kg) = \phi(k)\phi(g) = e'\phi(g) = \phi(g) = h$ so if we define $Kg \equiv \{kg \mid k \in K\}$, then Kg are *precisely* those elements (no more, and no less) of G which get sent to h . Thus the size of K tells us how far ϕ is from being an isomorphism, and the elements of K tell us exactly which elements of G will map to a specific element $h \in H$.

Homomorphisms and isomorphisms are ubiquitous in mathematics and occur frequently in physics, as we'll see in the examples below.

Example 3.10 $SU(2)$ and $SO(3)$

Chapter 3 Problems

1. In this problem we prove Euler's theorem that any $R \in SO(3)$ has an eigenvector with eigenvalue 1. Recall that if λ is an eigenvalue of R then there exists a nonzero vector v such that $Rv = \lambda v$, or $(R - \lambda I)v = 0$. This means that $R - \lambda I$ is not invertible, which by Problem 4 of Chapter 2 means that $|R - \lambda I| = 0$. The same problem shows that the converse is true, i.e. that if $|R - \lambda I| = 0$ then $R - \lambda I$ is not invertible, hence there must be some vector v such that $(R - \lambda I)v = 0$, or $Rv = \lambda v$. So, we can prove Euler's theorem if we can show that

$$|R - I| = 0. \quad (3.45)$$

Do this using the orthogonality condition and properties of the determinant. You should not have to work in components.

2. Show that $SO(3, 1)_o$ is a group. Remember that $SO(3, 1)_o$ is defined by 3 conditions: $|A| = 1$, $A_{44} > 1$, and (3.22). Proceed as follows:

- a) Show that $I \in SO(3, 1)_o$.
- b) Show that if $A \in SO(3, 1)_o$, then $A^{-1} \in SO(3, 1)_o$. Do this as follows:
 - i) Verify that $|A^{-1}| = 1$
 - ii) Show that A^{-1} satisfies (3.22). Use this to deduce that A^T does also.
 - iii) Write out the 44 component of (3.22) for both A and A^{-1} . You should get equations of the form

$$a_0^2 = 1 + \vec{a}^2 \quad (3.46)$$

$$b_0^2 = 1 + \vec{b}^2. \quad (3.47)$$

where $b_0 = (A^{-1})_{44}$. Clearly this implies $b_0 < -1$ or $b_0 > 1$. Now, write out the 44 component of the equation $AA^{-1} = I$. You should find

$$a_0 b_0 = 1 - \vec{a} \cdot \vec{b}. \quad (3.48)$$

If we let $a \equiv |\vec{a}|$, $b \equiv |\vec{b}|$ then the last equation implies

$$1 - ab < a_0 b_0 < 1 + ab. \quad (3.49)$$

Assume $b_0 < -1$ and use (3.47) to derive a contradiction to (3.49), hence showing that $b_0 = (A^{-1})_{44} > 1$, and that $A^{-1} \in SO(3, 1)_o$.

d) Show that if $A, B \in SO(3, 1)_o$, then $AB \in SO(3, 1)_o$. You may have to do some inequality manipulating to show that $(AB)_{44} > 0$.

Dirac Dictionary

We summarize here all of the translations given in the text between Dirac notation and standard mathematical notation.

Standard Notation	Dirac Notation
-------------------	----------------

Vector $\psi \in \mathcal{H}$	$ \psi\rangle$
Dual Vector $L(\psi)$	$\langle\psi $
Inner Product (ψ, ϕ)	$\langle\psi \phi\rangle$
$A(\psi)$, $A \in \mathcal{T}(\mathcal{H})$	$A \psi\rangle$
$(\psi, A\phi)$,	$\langle\psi A \phi\rangle$
$T_i^j e^i \otimes e_j$	$\sum_{i,j} T_{ij} j\rangle\langle i $
$e_i \otimes e_j$	$ i\rangle j\rangle$ or $ i, j\rangle$

Bibliography

- [HK] K. Hoffman and D. Kunze, *Linear Algebra*, 2nd ed., Prentice Hall, 1971
- [ST] S. Sternberg, *Group Theory and Physics*, Princeton University Press, 1994
- [R] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw Hill, 1976
- [SA] J.J. Sakurai, *Modern Quantum Mechanics*, 2nd ed., Addison Wesley Longman, 1994
- [G] S. Gasiorowicz, *Quantum Physics*, 2nd ed., John Wiley and Sons, 1996
- [F] G. Folland, *Real Analysis, Modern Techniques and Their Applications*, 2nd ed., Wiley Interscience, 1999
- [SC] B. Schutz, *Geometrical Methods of Mathematical Physics*, Cambridge 1980
- [W] F. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Springer 1979
- [Ha] B. Hall, *Lie Groups, Lie Algebras and Representations: An Elementary Introduction*, Springer-Verlag 2003
- [Go] H. Goldstein, *Classical Mechanics*, 2nd ed., Addison-Wesley 1980
- [He] I. Herstein, *Topics in Algebra*, 2nd ed., Wiley 1975