The Mathematical Physics of Gravity and Gauge Theories

A Self-Study Guide

Physics 6938
General Relativity
Fall 2007

Christopher Beetle
Department of Physics
Florida Atlantic University

Lecture 1

Abstract Vector Spaces

> It is not so easy to free oneself from the idea that coordinates must have an immediate metrical meaning.
>
> —Albert Einstein (1949)

Vector spaces play several important roles in differential geometry. For instance, in a sense made more precise below, the set of tangent vectors at a point of a curved space provides an organized way to investigate the geometry in a suitably small surrounding region. While physicists are certainly accustomed to working with vectors, however, their early training often emphasizes ideas that aren't necessarily appropriate generally. Let us illustrate this point in order to motivate the discussion of vectors below, which is admittedly a bit more abstract and mathematical than some physics students might initially prefer.

First, physicists often think of vectors in terms of their components relative to some particular coordinate system. In contrast, one of the deepest physical principles of general relativity is that the coordinates one uses to describe phenomena in spacetime generally have no definite physical meaning. Any system of coordinates is as good as any other, and we know very well that the same vector, in two different coordinate systems, may be described by two entirely different sets of numerical components.

There are (at least) two natural responses to this problem. One is to work with numerical components nonetheless, but then to make sure that all of the formulae in any physical calculation are ultimately independent of the coordinate system used. This approach has the advantage of relative familiarity since one ends up working once again with vectors as collections of numbers, but the formulas it produces can be daunting. Relativists, including Einstein, nonetheless worked this way for a long time. This was at least partially responsible for the widely-held view that general relativity involves arcane manipulations of excessively complicated formulae. An alternative response is to seek a description of vectors that deemphasizes the role of coordinates from the very beginning. This point of view is generally more popular among modern relativists, and we shall develop it extensively below.

While currently more in vogue, our preferred, coordinate-independent approach to geometry also has deep mathematical and philosophical roots. Indeed, Euclid's original method, technically called *synthetic* geometry, talks about line segments and vectors in space without any reference to numerical coordinates. It was only much more recently, in Descartes' development of *analytic* geometry, that the component approach came to dominate. Presumably, this was because of its relative simplicity for computing numerical results. While the presentation of vectors and vector spaces below is somewhat "synthetic" in the sense that it eschews components as fundamental descriptors of a vector, it also remains close to the "analytic" tradition in the sense that such components, relative to any coordinate system one likes, are easy to compute at any stage of a calculation. This hybrid approach allows us to exploit both the conceptual clarity of the synthetic approach and the computational facility of the analytic, which is why it has become popular.

A second drawback to the traditional view of vectors in freshman physics is that it introduces so many different mathematical structures at once. Students learn vector addition, scalar multiplication, dot products and cross products all at virtually the same moment, conflating several quite distinct mathematical structures. In dimension higher than three, for example, the cross-product of two vectors gives a bi-vector, a special type of tensor, rather than another vector. In general relativity, the situation is even worse. The very definitions of the dot and cross products rely on a mathematical structure, the spacetime metric, that depends on the gravitational field. Thus, the dot product of two vectors cannot be evaluated before the gravitational field equations have been solved. The field is determined by the distribution of energy (in relativistic physics) in the universe, however, and the expression for the energy of matter often depends on dot products or norms of certain vectors. Therefore, to solve the gravitational field equations, we must be able to write down expressions where the metric is not known *a priori*. This means we must be able to work in vector spaces *without a metric*, or at least without a predetermined metric.

Given the discussion above, we clearly must first disentangle the essential idea of a vector from the more geometrical notions of the dot and cross products before we can make much progress in relativity theory. The modern view of vector spaces does this by eliminating everything from the discussion except the key features of vector addition and scalar multiplication. As a result, in the following initial discussion, the only mathematical question we can answer concerning a general pair of vectors is whether they point in the same direction. If they do point in the same direction, we can also establish the ratio of their lengths. The answers to all further questions, such as the absolute length of a single vector, or the angle between vectors pointing in two different directions, must await an additional structure, an inner product, to be introduced below.

1.1 VECTOR SPACES

A vector space, by definition, is a set in which one can form linear combinations. More precisely, a ***vector space*** consists of:

1. a set V whose elements are called ***vectors***,
2. a rule for adding two vectors $\mathbf{v}, \mathbf{v}' \in V$ to yield another vector $\mathbf{v} + \mathbf{v}' \in V$, and
3. a rule for multiplying a vector $\mathbf{v} \in V$ by a ***scalar*** $\alpha \in \mathbb{F}$ to yield a new vector $\alpha\mathbf{v} \in V$.

The scalars are ordinary numbers, either real or complex. The vector space V is itself called ***real*** or ***complex*** depending on whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. The addition and scalar multiplication operations, however they are defined on V, must obey certain axioms. In the following list, $\mathbf{v}$, $\mathbf{v}'$, and $\mathbf{v}''$ represent arbitrary vectors, while α and α' are arbitrary scalars.

1. V is an ***Abelian group*** under addition. That is,

   (a) addition is ***associative***

   $$(\mathbf{v}+\mathbf{v}')+\mathbf{v}''=\mathbf{v}+(\mathbf{v}'+\mathbf{v}''), \tag{1.1}$$ {vsAdAs}

   (b) there is an ***additive identity***, the ***zero vector*** $\mathbf{0} \in V$, such that

   $$\mathbf{0}+\mathbf{v}=\mathbf{v}+\mathbf{0}=\mathbf{v}, \tag{1.2}$$ {vsAdId}

   (c) every vector $\mathbf{v} \in V$ has an ***additive inverse*** $-\mathbf{v} \in V$ such that

   $$\mathbf{v}+(-\mathbf{v})=(-\mathbf{v})+\mathbf{v}=\mathbf{0}, \tag{1.3}$$ {vsAdIn}

   and

   (d) addition is ***commutative***:

   $$\mathbf{v}+\mathbf{v}'=\mathbf{v}'+\mathbf{v}. \tag{1.4}$$ {vsAdCo}

   The first three conditions here define a general ***group***. The fourth specializes to the case of a ***commutative*** or ***Abelian*** group.

2. Scalar multiplication ***distributes*** over both scalar and vector addition:

   $$(\alpha+\alpha')\mathbf{v}=\alpha\mathbf{v}+\alpha'\mathbf{v} \quad \text{and} \quad \alpha(\mathbf{v}+\mathbf{v}')=\alpha\mathbf{v}+\alpha\mathbf{v}'. \tag{1.5}$$ {vsAMDis}

3. Scalar multiplication commutes with multiplication of scalars:

   $$\forall \alpha,\alpha' \in \mathbb{F}; \mathbf{v} \in V: \quad \alpha(\alpha'\mathbf{v})=(\alpha\alpha')\mathbf{v}. \tag{1.6}$$ {vsAMCom}

4. Multiplication by the unit scalar preserves all vectors:

   $$\forall \mathbf{v} \in V: \quad 1\mathbf{v}=\mathbf{v}. \tag{1.7}$$ {vsMuUn}

These abstract axioms mimic the algebra of addition and scalar multiplication for ordinary column vectors in matrix theory. However, there are many other interesting examples of spaces satisfying these conditions. We begin with a simple, familiar example that might help shed some light on the conceptual utility of the abstract approach.

Example 1.8: (*solutions of the Bessel equation*)

Perhaps the earliest examples of *abstract* vector spaces in mathematics arose in the solution of differential equations. To take an example of direct interest in physics, the Bessel equation of order ν reads

$$z^2 B''(z) + zB'(z) + (z^2 - \nu^2)B(z) = 0. \tag{1.8a}$$ {BesEq}

Like any linear, second-order, ordinary differential equation, (1.8a) does not have a unique solution. There are two undetermined constants of integration, and we may write

$$\text{either} \qquad B(z) = B_1 J_\nu(z) + B_2 N_\nu(z) \qquad \text{or} \qquad B(z) = B_+ H^+_\nu(z) + B_- H^-_\nu(z). \tag{1.8b}$$ {BesSol}

Here, $J_\nu(z)$ and $N_\nu(z)$ denote the standard Bessel functions of the first and second kind, while $H^\pm_\nu(z)$ denote the Hankel functions of the first and second kind, and B_1, B_2 and $B_\pm$ denote constants of integration. Recall that the Hankel functions are conventionally defined by

$$H^\pm_\nu(z) := J_\nu(z) \pm \mathrm{i} N_\nu(z) \qquad \text{so that} \qquad B_\pm = \tfrac{1}{2}(B_1 \mp \mathrm{i} B_2). \tag{1.8c}$$ {BesBas}

It is therefore straightforward to move back and forth between the Bessel- and Hankel-function descriptions of a general solution $B(z)$.

As for any linear equation, one can superpose any two solutions of (1.8a) to find another. The *set* of solutions is therefore naturally a vector space, but it is *not* naturally a *column*-vector space. The same solution $B(z)$ in (1.8b) may be described equally well by either of the column vectors

$$\mathbf{B}_{12} := \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \qquad \text{or} \qquad \mathbf{B}_\pm := \begin{pmatrix} B_+ \\ B_- \end{pmatrix}. \tag{1.8d}$$ {BesCol}

Each of these column-vector *representations* of a solution $B(z)$ captures perfectly well the effect of superposing general solutions of (1.8a). That is, ordinary column-vector addition in (1.8d) accurately reflects the usual superposition of solutions (1.8b) as functions of $z \in \mathbb{C}$. The only thing to choose between the two representations is the particular problem being solved, and the taste and insight of the person solving it. Accordingly, it is generally useful to consider the set of *all* Bessel functions of order ν as a (two-dimensional) *abstract* vector space. The particular column-vector representations of this space in (1.8d) may be either more or less useful, depending on the physical context of the particular problem at hand.

Exercise 1.9: In the abstract, axiomatic approach, a number of fairly obvious facts arise only as the result of calculations. For example,

$$0\mathbf{v} = 0\mathbf{v} + \mathbf{0} = 0\mathbf{v} + (0\mathbf{v} - 0\mathbf{v}) = (0\mathbf{v} + 0\mathbf{v}) - 0\mathbf{v} = (0+0)\mathbf{v} - 0\mathbf{v} = 0\mathbf{v} - 0\mathbf{v} = \mathbf{0}. \tag{1.9a}$$ {vs0v0}

Justify each step of the above calculation using individual axioms from the definition. Note that the definition $\alpha\mathbf{v} - \alpha'\mathbf{v}' := \alpha\mathbf{v} + (-\alpha'\mathbf{v}')$, which has been used here implicitly, is merely a matter of notation.

Exercise 1.10: Show axiomatically that $(-\alpha)\mathbf{v} = -(\alpha\mathbf{v})$.

Exercise 1.11: Show that the space $\mathbb{F}^n$ consisting of all ordered n-tuples $(f^1, f^2, \ldots, f^n)$ of scalars f^i is indeed a vector space in the abstract sense. Define addition, scalar multiplication and the zero vector, and check that your definitions satisfy the axioms above.

Exercise 1.12: (*function spaces*)
Show that the set $\mathring{\mathscr{F}}(S)$ of all scalar-valued functions on a given set S is naturally a vector space.
Solution: Vector addition and scalar multiplication in $\mathring{\mathscr{F}}(S)$ can be defined naturally in a point-wise sense:

$$[f + f'](s) := f(s) + f(s') \quad \text{and} \quad [\alpha f](s) := \alpha\, f(s). \tag{1.12a}$$ {fsVec}

That is, the function $f + f'$ on S is defined to take the value $f(s) + f'(s)$ at each point $s \in S$. The zero vector in $\mathring{\mathscr{F}}(S)$ is the ***zero function*** $0(s)$, which of course takes the scalar value 0 at each point $s \in S$.

It is straightforward to confirm that the above definitions make $\mathring{\mathscr{F}}(S)$ a vector space. The process is typified by the proof of the distributive property for scalar multiplication:

$$[\alpha(f+f')](s) := \alpha[f+f'](s) := \alpha[f(s)+f'(s)] = \alpha f(s)+\alpha f'(s) =: [\alpha f](s)+[\alpha f'](s) =: [\alpha f+\alpha f'](s). \quad (1.12b)$$ {fsDis}

Here, we have simply used the definitions of scalar multiplication and addition in $\mathring{\mathscr{F}}(S)$ to evaluate the function $\alpha(f+f')$ at an arbitrary point $s \in S$. Then, after using the distributive property of multiplication of scalars in $\mathbb{F}$, we use the vector-space definitions once again to show that the function $\alpha f + \alpha f'$ takes the same value at each $s \in S$. This proves the result. We call $\mathring{\mathscr{F}}(S)$ the ***function space*** over S.

Exercise 1.13: (*sequence spaces*)
The particular function space $\mathring{\mathscr{S}} := \mathring{\mathscr{F}}(\mathbb{Z}^+)$ over the positive integers $\mathbb{Z}^+$ is called the (countable) ***sequence space***. A conventional notation sets $f^n := f(n)$, and sequences are denoted $(f^1, f^2, \ldots)$. Note that this space contains *all* sequences of scalars, with no criterion of convergence of any sort. Write down definitions of vector addition and scalar multiplication in the sequence notation, and check explicitly that these define a vector space.

1.2 BASES AND DIMENSION

There is clearly a big difference between an ordinary, three-dimensional vector and one in a higher-dimensional vector space such as one of the Hilbert spaces encountered in quantum mechanics. So far in our discussion, the precise dimension of a vector, or rather that of the space containing it, is still an open question. The answer is quite straightforward and intuitive. The dimension of a vector space V is the minimum number of scalars needed to specify an arbitrary vector in that space; it is the number of components any vector in that space has. To make this intuitive idea precise, we must describe carefully how scalar components, relative to some basis, specify a vector. We must not do this, however, using the familiar method of *orthogonal* components, which of course relies on the dot product.

A subset K of a vector space V is said to be ***linearly independent*** if, for any finite collection $\mathbf{k}_1, \ldots, \mathbf{k}_n \in K$, we have

$$\alpha^1 \mathbf{k}_1 + \cdots + \alpha^n \mathbf{k}_n = \mathbf{0} \quad \Rightarrow \quad \alpha^1 = \cdots = \alpha^n = 0. \quad (1.14)$$ {vsLinInd}

In words, the only linear combination of an arbitrary finite subset of K which gives the zero vector is the trivial combination in which each vector is scaled to zero before being summed. A collection $\mathbf{k}_1, \ldots, \mathbf{k}_n$ that admits a *non-trivial* linear combination yielding the zero vector is of course said to be ***linearly dependent***.

A subset $L \subset V$ is said to ***span*** a vector space V if *every* vector $\mathbf{v} \in V$ can be written as a finite linear combination

$$\mathbf{v} = v^1 \boldsymbol{\ell}_1 + \cdots + v^n \boldsymbol{\ell}_n \quad (1.15)$$ {vExp}

of vectors $\boldsymbol{\ell}_i \in L$. The set of *all* vectors, the entire vector space V, obviously spans V, but also is obviously linearly dependent. There are always smaller subsets of V that also span the vector space. A linearly independent subset $B \subset V$ that *also* spans V is called a ***basis***.

Exercise 1.16: Let V be the space of 3-dimensional column vectors from matrix theory. Show that

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \mathbf{b}_3 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \quad (1.16a)$$ {vsExBas1}

is a basis for V. You must show (a) that these three vectors are linearly independent and (b) that any larger set of vectors would be linearly dependent. Expand the vector

$$\mathbf{v} = \begin{pmatrix} 4 \\ 3 \\ -2 \end{pmatrix} \qquad (1.16b) \quad \{\texttt{vsExVec1}\}$$

in this basis.

Exercise 1.17: If B is a basis for V, show that the representation

$$\mathbf{v} = v^1\, \mathbf{b}_1 + \cdots + v^n\, \mathbf{b}_n \qquad (1.17a) \quad \{\texttt{vBas}\}$$

of an arbitrary vector $\mathbf{v} \in V$ as a linear combination of basis elements $\mathbf{b}_i$ must be *unique*. The coefficients v^i appearing in this expansion are called the ***components*** of $\mathbf{v}$ in the basis B.

Exercise 1.18: Show that the "components" v^i in the expansion (1.15) are *not* necessarily unique if the subset L underlying the expansion spans V, but is linearly dependent.
Comment: Strictly speaking, we refer to "components" v^i of a vector *only* when the expansion 1.15 refers to a *basis* B, as in 1.17a.

The ***dimension*** of a vector space V is defined to be the number of vectors in any one of its basis sets B. This number may be finite or infinite. In the case of a finite-dimensional vector space V, which is the only case we consider in these notes (see, however, the optional discussion of the infinite-dimensional case below), we must ensure that this definition makes sense by showing that all bases for V contain the same number of vectors. The intuitive scheme of the proof is to pick a vector in one basis, B, and show that one can replace it with a vector from the other basis, B', to give a new basis, $\tilde{B}$. After iternating this process many times, we can eliminate all of the original vectors from B to find two bases $\tilde{B}$ and B', with $\tilde{B}$ containing exactly as many vectors as B, all drawn from B'. Thus, $\tilde{B}$ is a subset of B'. If it is a *proper* subset, then B' certainly cannot be a basis because it would have to be linearly dependent. Thus, we must have $\tilde{B} = B'$, and both original bases contain the same number of vectors. We include the full proof only to illustrate the power of the definitions above, and to highlight the critical role played by our assumption of finite dimension.

Theorem 1.19: If a vector space V admits a finite basis B, then every other basis B' for V is also finite, and contains exactly the same number of vectors as B.
Proof: Let B and B' be basis sets for a vector space V, and suppose that B' contains more vectors than B, which is finite. Then, pick an arbitrary vector $\mathbf{b}' \in B'$. Since B is a basis, $\mathbf{b}'$ must be expressible as a linear combination of elements of B, so

$$\mathbf{b}' = \alpha^1\, \mathbf{b}_1 + \cdots + \alpha^n\, \mathbf{b}_n. \qquad (1.19a) \quad \{\texttt{vsVexp}\}$$

Now pick any vector $\mathbf{b}_i \in B$ for which $\alpha^i \neq 0$ and remove it from B, replacing it with $\mathbf{b}'$. The resulting set $\tilde{B} \subset V$ must still be a basis for V because we can solve

$$\mathbf{b}_i = \frac{1}{\alpha^i}\, \mathbf{b}' - \frac{\alpha^1}{\alpha^i}\, \mathbf{b}_1 - \cdots - \frac{\alpha^{i-1}}{\alpha^i}\, \mathbf{b}_{i-1} - \frac{\alpha^{i+1}}{\alpha^i}\, \mathbf{b}_{i+1} - \cdots - \frac{\alpha^n}{\alpha^i}\, \mathbf{b}_n. \qquad (1.19b) \quad \{\texttt{vsVrep}\}$$

Substituting this expression for $\mathbf{b}_i$ into the expansion (1.17a) of an arbitrary vector $\mathbf{v}$ in the basis B then shows that $\mathbf{v}$ can be written as a linear combination of elements of the new set $\tilde{B}$. Moreover, $\tilde{B}$ must also be linearly independent since otherwise we would have

$$\beta^1\, \mathbf{b}_1 + \cdots \beta^{i-1}\, \mathbf{b}_{i-1} + \beta^i\, \mathbf{b}' + \beta^{i+1}\, \mathbf{b}_{i+1} + \cdots + \beta^n\, \mathbf{b}_n = \mathbf{0}, \qquad (1.19c) \quad \{\texttt{vsnewBLI}\}$$

with $\beta^i \neq 0$ since the original set B was linearly independent. Solving for $\mathbf{b}'$ then would show that $\mathbf{b}'$ has two different expansions in the basis B, one involving $\mathbf{b}_i$ and the other not. The exercise above shows that this is impossible if B is linearly independent.

Repeat the above process multiple times, always choosing another vector $\mathbf{b}' \in B'$ to replace another vector $\mathbf{b} \in \tilde{B}$. One might worry that at some stage it would be impossible to continue, that the expression for a newly-chosen $\mathbf{b}'$ would have have non-zero components only along vectors in $\tilde{B}$ which had previously been brought in from B'. However, if this were the case, then B' would be linearly dependent, contradicting the assumption that it is a basis for V. Thus, in each iteration, we can always choose a vector from $\tilde{B}$ that originally belonged to B and replace it with a vector from B'. The process only terminates when all of the original vectors in B have been replaced. Moreover, $\tilde{B}$ must still be a basis for V. But if B' is larger than $\tilde{B}$ at this stage, then it must surely be linearly dependent since $\tilde{B}$ is a basis. Thus, we must have $B' = \tilde{B}$. Since the process described here has always replaced one vector of B with one vector from B', $\tilde{B}$ must contain the same number of vectors as B, and so therefore must B'.

Exercise 1.20: Show that the set $\mathscr{P}_3(x)$ of real-valued polynomials in a real variable x with order at most three is a real vector space. Find two distinct bases for this space. What is its dimension?

Exercise 1.21: Suppose that a vector space V has a finite basis B with n elements. Show that every other linearly independent set of n vectors in V must also be a basis.

Hint: We have already proved a closely related fact above, that any two bases on V must have the same number of elements if one is finite. Here, you are going the other direction, showing that any linearly independent set with n elements is in fact a basis. Together, these results characterize completely the set of all bases on a finite-dimensional vector space.

Exercise 1.22: Show that the countable sequence space $\mathring{\mathscr{S}}$ introduced above is infinite-dimensional.

Hint: Proceed in two steps. First, suppose that you have a finite "basis" B for $\mathring{\mathscr{S}}$ containing n elements, and consider the linearly-independent set K containing the sequences

$$\mathbf{k}_i = (0, \ldots, 0, 1, 0, \ldots), \tag{1.22a}$$

where the "1" falls in the i^{th} position in the sequence and i runs from 1 to n. Show that either (a) one of the $\mathbf{k}_i$ cannot be expanded in terms of B, so that B cannot be a basis, or (b) K must be a basis whenever B is. In the latter case, find a vector in $\mathscr{S}^0$ that cannot be expanded in terms of K. (This bit is *really* easy.) Then, K cannot be a basis, and therefore neither can B.

Any two bases on a given finite-dimensional vector space V are linked by a square matrix defined as follows. Let the bases in question be $B = \{\mathbf{b}_i\}$ and $\tilde{B} = \{\tilde{\mathbf{b}}_j\}$. Following the discussion above, the labels i and j on the basis vectors take the same values: $i, j = 1, \ldots, n$. Applying the basis expansion (1.17a) to each of the vectors in the $\tilde{B}$ basis, we find a set of scalars ${\lambda^i}_j$ with

$$\tilde{\mathbf{b}}_j = \sum_i {\lambda^i}_j \, \mathbf{b}_i. \tag{1.23}$$ {vsBasBas}

That is, ${\lambda^i}_j$ is the component of the basis vector $\tilde{\mathbf{b}}_j$ along the vector $\mathbf{b}_i$ from the other basis. This relation between bases can be arranged in a convenient, matrix form as

$$\begin{pmatrix} \tilde{\mathbf{b}}_1 & \tilde{\mathbf{b}}_2 & \cdots & \tilde{\mathbf{b}}_n \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix} \begin{pmatrix} {\lambda^1}_1 & {\lambda^1}_2 & & {\lambda^1}_n \\ {\lambda^2}_1 & {\lambda^2}_2 & & {\lambda^2}_n \\ & & \ddots & \\ {\lambda^n}_1 & {\lambda^n}_2 & & {\lambda^n}_n \end{pmatrix}. \tag{1.24}$$ {vsCOBmat}

We have arranged the basis vectors appearing here to form a row vector. We could have arranged them in a column vector instead, and had the *transpose* of the square matrix appearing here multiply it from the left instead of from the right. We have chosen this representation since vectors in elementary linear algebra are often taken conventionally to be column vectors, and one could think of each element of the row vector as a column containing the components of that basis vector relative to another, arbitrary basis. It is not necessary to think this way, however. The point here is that the the change-of-basis transformation (1.23) is captured perfectly by the familiar *algebra* of matrix multiplication. The matrix λ introduced here is called the ***change-of-basis matrix***.

Exercise 1.25: Show that the change-of-basis matrix λ in (1.24) must be invertible.
Hint: Find an easy way to calculate its inverse, which might be denoted $\tilde{\lambda}$.

Exercise 1.26: Let v^i and $\tilde{v}^j$ denote the scalar components of a given vector $\mathbf{v} \in V$ relative to the bases B and $\tilde{B}$, respectively. Use the scalars $\lambda^i{}_j$ defined above to relate these two sets of components. Arrange your result in a matrix form. Is it better to use a row or a column vector for the components? Compare the form of your result to (1.24). Which set of components is easier to express in terms of the other?

1.3 SUBSPACES, COMPLEMENTS AND QUOTIENTS

A ***subspace*** of a vector space V is a subset $U \subset V$ that contains all finite linear combinations of its elements. A subspace is automatically a vector space in its own right. The operations of vector addition and scalar multiplication in U descend from those already given on V. Indeed, the essential content of the definition of a subspace above is that sums and scalings of vectors in U, which certainly belong to V, also belong to U. Addition and scalar multiplication of vectors in U can therefore be regarded as intrinsic operations on U. In addition, taking any $\mathbf{u} \in U$, we have $\mathbf{0} = 0\mathbf{u} \in U$, so U contains an additive identity, and $-\mathbf{u} = (-1)\mathbf{u} \in U$, so U contains additive inverses for its elements. It is straightforward to verify that U obeys all of the vector-space axioms.

Exercise 1.27: We can describe a subset $U \subset \mathbb{R}^3$ by constraining the components (x, y, z) of a vector $\mathbf{v} \in \mathbb{R}^3$. Which of the following sets of constraints define a subspace $U \subset \mathbb{R}^3$?

$$\begin{matrix} z=0 & x=y & z=1 & x^2=y & x^2=y^2 & x=y=0 & x=y=1 \\ \text{(A)} & \text{(B)} & \text{(C)} & \text{(D)} & \text{(E)} & \text{(F)} & \text{(G)} \end{matrix} \tag{1.27a}$$

Answer: Only A, B and F define subspaces.

Exercise 1.28: Verify that the operations of vector addition and scalar multiplication on a vector space V, when restricted to a subspace $U \subset V$, satisfy all of the axioms needed to make U a vector space.

Certainly, any vector space V contains subsets $K \subset V$ that are *not* subspaces. For example, any subset $U = \{\mathbf{v}\}$ containing just one vector cannot be a subspace (unless $\mathbf{v} = \mathbf{0}$) because it does not contain scalar multiples of $\mathbf{v}$. However, even if a sub*set* $K \subset V$ is not a sub*space*, we can still construct such a subspace from it. Intuitively, it is the subspace formed by the set of all possible finite linear combinations of vectors in K. We denote this subspace span(K), the ***span*** of K, or the subspace of V ***generated*** by K. To avoid subtleties associated especially with infinite-dimensional spaces, the subspace span(K) is rigorously defined as the smallest subspace of V containing every vector in K.

Exercise 1.29: Let U_λ be a collection of subspaces of a vector space V indexed by elements $\lambda \in \Lambda$ of an arbitrary set, which may be finite or infinite. Show that the intersection

$$U := \bigcap_{\lambda \in \Lambda} U_\lambda \tag{1.29a}$$ {vsSubInt}

is also a subspace of V.
Hint: You need to show that U is closed under the formation of linear combinations. This will follow from the closure of each individual U_λ and ordinary properties of the intersection.

Exercise 1.30: Show that the span of a subset $K \subset V$ can be realized as the intersection of all subspaces $U \subset V$ that contain K as subsets. That is, show that

$$\text{span}(K) = \bigcap_{K \subset U} U. \tag{1.30a}$$

Hint: You need to show that this intersection, which is a subspace by the previous exercise, (a) contains every vector in K and (b) is the *smallest* subspace of V to do so. After showing (a), suppose that there is a smaller subspace containing K and argue (b) by contradiction.

Exercise 1.31: Show that a subset $B \subset V$ is a basis for V if and only if (a) the span of B is all of V, and (b) the span of any proper subset of B is not all of V.

The first exercise above shows that the intersection of two subspaces, U and U', of a vector space V is again a subspace of V. This idea is illustrated in Fig. 1.1(a), where a pair of two-dimensional subspaces of a three-dimensional space intersect to give a smaller, one-dimensional subspace. The same cannot be said for the union $U \cup U'$, except in the trivial case that one of U or U' is fully contained within the other. The problem is illustrated in Fig. 1.1(b). If each subspace contains at least one vector not contained in the other, say $\mathbf{u} \in U$ and $\mathbf{u}' \in U'$, then the linear combination $\alpha\mathbf{u} + \alpha'\mathbf{u}'$ cannot lie in either space, unless of course one of the scalars, α or α', vanishes. However, the *span* of the union of subspaces, like the span of *any* subset of V, *is* a subspace. This particular combination of subspaces, called simply their ***sum*** $U + U'$, can be quite useful. Again, the idea is illustrated in Fig. 1.1(b). The *union* of subspaces is the cross formed by the two lines, while the *sum* is the entire plane containing that cross. Note the symmetry in the definitions: $U \cap U'$ is the *largest subspace contained within* both U and U', while $U + U'$ is the *smallest subspace containing* both U and U'. If we replace "subspace" here with "subset," then these are exactly the properties that define the ordinary set-theoretic operations of intersection and union, respectively. However, there are significant differences between the usual set operations and the present subspace ones.

{vsSubDis}

Exercise 1.32: An important feature of the intersection and union of sets is that the two operations are mutually ***distributive***. That is, for any trio of sets A, B and C, we have

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \qquad \text{and} \qquad (A \cap B) \cup C = (A \cup C) \cap (B \cup C). \tag{1.32a}$$ {vsSetDis}

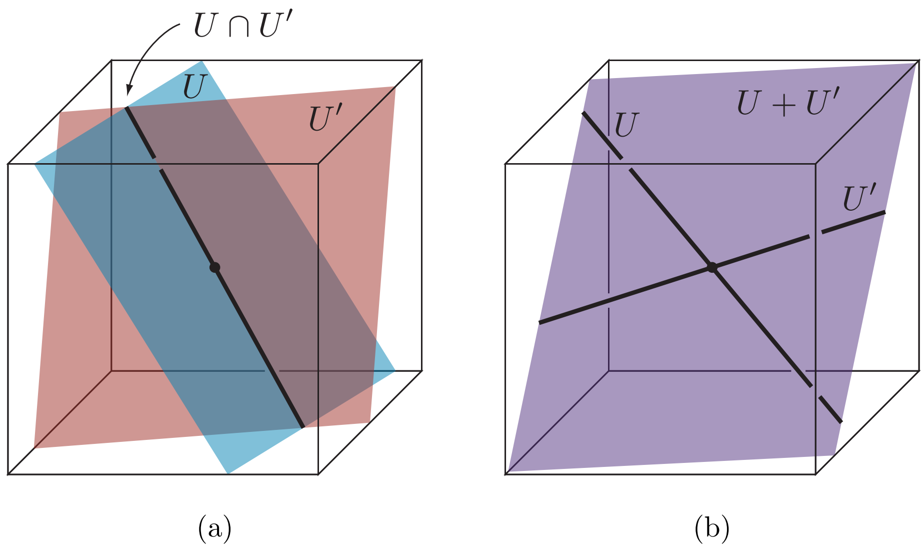


Figure 1.1: Operations with pairs of vector subspaces. (a) Two two-dimensional subspaces in a three-dimensional space intersect in a one-dimensional subspace $U \cap U'$. (b) The set-wise union of two one-dimensional subspaces is the cross formed by the two lines, which is not subspace. The span of the union *is* a subspace, represented the shaded two-dimensional plane $U + U'$. {vssUnion}

The intersection and sum operations on subspaces, in contrast, are not distributive. Consider, for example, the following three subsets of $\mathbb{R}^2$:

$$A := \{(x, y) \mid x = 0\}, \qquad B := \{(x, y) \mid y = 0\} \qquad \text{and} \qquad C := \{(x, y) \mid x = y\}. \tag{1.32b}$$

Show that $(A + B) \cap C \neq (A \cap C) + (B \cap C)$.

Comment: In the quantum theory of a spin-half particle, with x and y allowed to take complex values, the subspace A could describe spin eigenstates polarized along the $+z$ axis, B, along the $-z$-axis, and C, along the $+x$-axis. The subspace $A + B$ describes the space of spin states with spin either up *or* down, which of course is the whole Hilbert space. The subspace $A \cap C$, on the other hand, describes states with spin both up *and* right, which is physically impossible. Indeed the intersection $A \cap C$ contains only the zero vector.

This simple example shows how, in quantum mechanics, the sum and intersection of subspaces are associated with the *logic* of assertions one might make about the state of a system (e.g., "the spin is up *and* the energy is -13.6 eV"). In classical mechanics, such assertions are associated not with subspaces, but with volumes in phase space. The logic of classical assertions corresponds to the union and intersection of those sets. Classical logic is therefore distributive, and quantum logic is not. This fact is at the heart of many of the exotic features of quantum theory, such as the EPR phenomenon and Bell's theorem.

Exercise 1.33: Find an example of subspaces A, B and C of a vector space V such that

$$(A \cap B) + C \neq (A + C) \cap (B + C). \tag{1.33a}$$

That is, show that *both* distributive laws (1.32a) fail when unions of sets are replaced by sums of subspaces.
Hint: This is pretty easy. See the previous example.

The sum and intersection operations allow us to manufacture larger and smaller subspaces from any two given subspaces of a vector space V. There are clearly subspaces of V

that are overall the largest and smallest possible. These are respectively V itself and the ***trivial subspace***, denoted $\{\mathbf{0}\}$ or even simply $\mathbf{0}$ because it contains only the zero vector. Certain pairs of subspaces, U and U', when summed or intersected, may yield these extremal results. When $U \cap U' = \mathbf{0}$, for example, we say that U and U' are ***disjoint***. Note that they are not really disjoint in the set-theoretic sense, but that their intersection in the zero vector is wholly unremarkable. (See exercise 1.34 below.) When $U + U' = V$, on the other hand, we say that U and U' ***span*** V. When U and U' are disjoint *and* span V, we say that they ***complement*** one another. This last case is particularly important since intuitively it means that the subspaces U and U' account for all vectors (they span V) with no redundancy (they are disjoint). Subspaces U and U' can fail to complement one because they are not complete, not disjoint, or neither. As explicit examples, Fig. 1.1(a) shows a complete pair of subspaces that are not disjoint, while Fig. 1.1(b) shows a disjoint pair that is not complete.

{vsTrivGen}

Exercise 1.34: Show that the trivial subspace $\mathbf{0}$ is a subspace of *every* subspace U of a vector space V. This is why it is called trivial. Conclude that *every* intersection of subspaces $U \cap U'$ contains the trivial subspace. The intersection is interesting only when it is strictly larger than the generic, trivial result.

Exercise 1.35: Show that every pair of the subspaces A, B and C from exercise 1.32 complement one another in $\mathbb{R}^2$. That is, show that each pair is both disjoint and complete.

A given vector space may be split into many different pairs of complementary subspaces. This ambiguity is a sort of higher-dimensional generalization of the freedom to choose a basis discussed above. Just as there is no preferred basis B on a vector space V with no additional structure, there also is no preferred decomposition of V into complementary subspaces. In fact, even if one subspace $U \subset V$ is given, there are generally infinitely many complements $U' \subset V$ for it. This situation is depicted in Fig. 1.2(a), which shows two different complements, U' and $\tilde{U}'$, for a single subspace $U \subset V$. Each of these complementary subspaces is one-dimensional. This is because U in this case is a two-dimensional subspace of a three dimensional space V. (See exercise 1.36 below.) However, not every one-dimensional subspace of V complements U. The line $\mathring{U}'$, for example, lies within the plane defined by U in Fig. 1.2(a), and therefore has $U + \mathring{U}' = U$. The problem in this last case, of course, is that U and $\mathring{U}'$ are not disjoint.

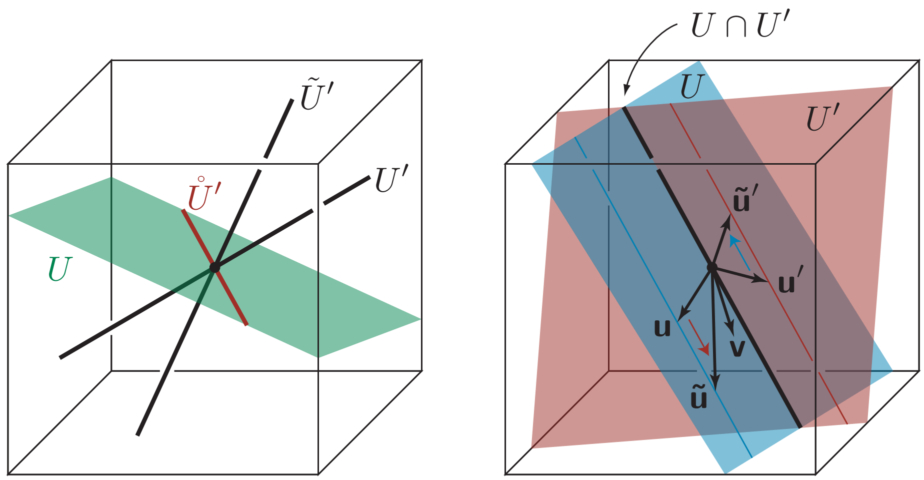
{vsComDim}

Exercise 1.36: If U and U' are complementary subspaces of a finite-dimensional vector space V, show that both must be finite-dimensional and that the sum of their dimensions must equal that of V.
Hint: Do the second part first. Pick arbitrary bases B and B' for U and U', respectively, and prove that the union $B \cup B'$ of basis sets must define a basis for V. That is, show that the union is linearly independent and spans V. Then, count.

Exercise 1.37: Expanding on the previous problem, let U and U' be *arbitrary* subspaces of V. Show that the dimensions of these subspaces and their sum and intersection are linked by

$$\dim U + \dim U' = \dim(U + U') + \dim(U \cap U'). \tag{1.37a}$$

Hint: First, choose a basis $B_\cap$ for the intersection subspace $U \cap U'$. Argue that this set can be extended in two different ways to form bases B and B' for the subspaces U and U'. Show that the (set-wise) union of these bases gives a basis for $U + U'$. How many *distinct* vectors does that union contain?



(a) (b)

Figure 1.2: Sums of vector subspaces. (a) The one-dimensional subspace U' complements the two-dimensional subspace U in three dimensions, but the complement is not unique: $\tilde{U}'$ also complements U. The only one-dimensional subspaces that fail to complement U in this example are those, like $\mathring{U}'$, that lie within the plane of U. (b) When two subspaces span V but intersect, the decomposition of a vector $\mathbf{v} \in V$ into components $\mathbf{u} \in U$ and $\mathbf{u}' \in U'$ always exists, but is not unique. In this example, we could equally well choose $\tilde{\mathbf{u}} \in U$ and $\tilde{\mathbf{u}}' \in U'$ to represent $\mathbf{v}$. {vssComp}

{vssComUn}

Exercise 1.38: If U and U' are complementary subspaces of a finite-dimensional vector space V, show that *any* vector $\mathbf{v} \in V$ can be written *uniquely* as a sum

$$\mathbf{v} = \mathbf{u} + \mathbf{u}' \qquad \text{with} \qquad \mathbf{u} \in U \quad \text{and} \mathbf{u}' \in U'. \tag{1.38a}$$ {vssComps}

We refer to $\mathbf{u}$ and $\mathbf{u}'$ simply as the ***components*** of $\mathbf{v}$ in the decomposition $V = U + U'$.

Hint: To prove uniqueness, you will need to show first that the decomposition of the zero vector is unique, or $\mathbf{u} + \mathbf{u}' = \mathbf{0}$ implies $\mathbf{u} = \mathbf{u}' = \mathbf{0}$ in V. Then, proceed to non-zero vectors $\mathbf{v} \in V$.

Exercise 1.39: Show that if subspaces U and U' span a vector space V, but are not disjoint, then any $\mathbf{v} \in V$ can be written as a sum of vectors $\mathbf{u} \in U$ and $\mathbf{u}' \in U'$, but that these vectors are not unique.

Solution: Showing that $\mathbf{u}$ and $\mathbf{u}'$ exist is easy. It follows immediately from the definition: If U and U' span V, then the space of linear combinations of the form $\alpha\mathbf{u} + \alpha'\mathbf{u}'$, with $\mathbf{u} \in U$ and $\mathbf{u}' \in U'$, comprises all of V. Thus, every $\mathbf{v} \in V$ can be written as such a combination.

The peculiarity when U and U' are not disjoint is illustrated in Fig. 1.2(b). In that Figure, we can shift the vector $\mathbf{u}$ by any certain amount along the line through $\mathbf{u}$ parallel to the intersection $U \cap U'$ to give a new vector $\tilde{\mathbf{u}}$. As long as we shift the vector $\mathbf{u}'$ by an equal amount, in the opposite direction, along the corresponding line through $\mathbf{u}'$, the sum remains unchanged. Analytically, and in broad generality, the point is that we can choose any vector $\mathbf{w} \in U \cap U'$ and, since this vector belongs both to U and to U', we have

$$\tilde{\mathbf{u}} := \mathbf{u} + \mathbf{w} \in U \qquad \text{And} \qquad \tilde{\mathbf{u}}' := \mathbf{u}' - \mathbf{w} \in U'. \tag{1.39a}$$

The new vectors belong to U and U', respectively, precisely because these are subspaces of V. Thus, when two spanning subspaces intersect, the "component" vectors, $\mathbf{u} \in U$ and $\mathbf{u}' \in U'$, of a fixed $\mathbf{v} \in V$ are determined only up to the addition and subtraction of an arbitrary vector in the intersection.

Physicists are sometimes accustomed to thinking not of *a* complement to a given subspace $U \subset V$, but rather of *the* complement. What they have in mind, of course, is the *orthogonal* complement, which we define below. This subspace consists of all vectors in V that are orthogonal to every vector in U. It is perfectly well-defined, but does depend in an obvious way on the inner product one uses to decide whether two vectors are orthogonal. Since a single vector space may support a wide variety of different inner products, there generally is no preferred complement $U' \subset V$ to a given $U \subset V$. This fact becomes especially important in relativity theory because the indefinite metric on Minkowski space admits subspaces, called null subspaces, with *no* orthogonal complement. These can cause needless problems if one does not appreciate the general issues discussed here.

Although a given $U \subset V$ may admit many complementary subspaces, these various complements have certain features in common. For instance, we have seen in Exercise 1.36 that the various complements to a given subspace of a finite-dimensional vector space must have the same dimension. The root of these common features lies in the notion of the ***quotient space*** V/U. This vector space is a sort of *universal* complement to U, but the trick is that it is not naturally identified with *any* particular subspace of V. Rather, V/U is equally well identified with *any* complementary subspace $U' \subset V$, which is why these complementary subspaces share common features.

Given a subspace $U \subset V$, the quotient space is defined as the set of certain sets of vectors in V. These sets are defined by the ***equivalence relation***

$$\mathbf{v}' \sim \mathbf{v} \qquad \text{if and only if} \qquad \mathbf{v}' - \mathbf{v} \in U. \tag{1.40}$$ {vsQuotER}

Certainly, unless U is all of V, some pairs of vectors will not be equivalent in this sense. The space V therefore divides up into various ***equivalence classes*** of vectors denoted $\{\mathbf{v}\}_U$, or simply $\{\mathbf{v}\}$ when the subspace U is unambiguous. Each equivalence class is a subset of V containing *all* vectors differing from one another by a vector in U, and V/U is the set of all such classes. The vector structure on V induces one on V/U when we set

$$\alpha\{\mathbf{v}\} + \alpha'\{\mathbf{v}'\} := \{\alpha\mathbf{v} + \alpha'\mathbf{v}'\}. \tag{1.41}$$ {vsQuotLC}

That is, a linear combination of the equivalence classes containing the vectors $\mathbf{v}$ and $\mathbf{v}'$ is defined as the equivalence class containing the same linear combination of $\mathbf{v}$ and $\mathbf{v}'$ in V. In order for this definition to make sense in V/U, the equivalence class containing $\alpha\mathbf{v} + \alpha'\mathbf{v}'$ must be independent of the particular vectors $\mathbf{v} \in \{\mathbf{v}\}$ and $\mathbf{v}' \in \{\mathbf{v}'\}$ one uses to form the linear combination in V. We check this by replacing the vector $\mathbf{v}$ with another vector $\tilde{\mathbf{v}} = \mathbf{v} + \mathbf{u}$ in the same equivalence class, and likewise $\mathbf{v}'$ with $\tilde{\mathbf{v}}' = \mathbf{v}' + \mathbf{u}'$. We find

$$\alpha\tilde{\mathbf{v}} + \alpha'\tilde{\mathbf{v}}' = \alpha(\mathbf{v} + \mathbf{u}) + \alpha'(\mathbf{v}' + \mathbf{u}') = (\alpha\mathbf{v} + \alpha'\mathbf{v}') + (\alpha\mathbf{u} + \alpha'\mathbf{u}') \sim \alpha\mathbf{v} + \alpha'\mathbf{v}'. \tag{1.42}$$ {vsQuotLCwd}

The last equivalence follows precisely because U is a subspace, and so contains the linear combination of $\mathbf{u}$ and $\mathbf{u}'$ appearing here.

Exercise 1.43: Show that the equivalence relation defined in (1.40) is transitive. That is, show that $\mathbf{v} \sim \mathbf{v}'$ and $\mathbf{v}' \sim \mathbf{v}''$ imply $\mathbf{v} \sim \mathbf{v}''$.

Exercise 1.44: Define the zero vector in V/U and the additive inverse $-\{\mathbf{v}\}$, and show that vector addition and scalar multiplication in V/U satisfy all the required associative, commutative and distributive laws.

Exercise 1.45: Show that the sub*set* $\{\mathbf{v}\} \subset V$ is not generally a sub*space* of V. When is it a subspace? *Answer*: If $\mathbf{v} \in U$, then $\{\mathbf{v}\} = U$ is a subspace. Otherwise, $\{\mathbf{v}\}$ is not a subspace.

The quotient space is illustrated in Fig. 1.3, which we now describe. The large box represents the vector space V, with its origin at the center of the cube. The heavy vertical line passing through the origin represents a subspace $U \subset V$. The remaining heavy vertical lines represent the equivalence classes containing the four vectors $\mathbf{v}$, $\tilde{\mathbf{v}}$, $\mathbf{v}'$ and $\mathbf{v}''$. Note that $\mathbf{v}$ and $\tilde{\mathbf{v}}$ differ by a vertical vector, which lies in U, so $\{\mathbf{v}\} = \{\tilde{\mathbf{v}}\}$. The quotient space V/U is represented by the horizontal plane beneath the cube. The vertical lines representing equivalence classes of vectors in V become simply points in this space. Note that V/U is *not* a subspace of V in this diagram, but an entirely different vector space. We can define a map $\pi : \mathbf{v} \mapsto \{\mathbf{v}\}$, called the ***projection***, taking vectors in V to vectors in V/U. The diagram also shows two shaded planes, representing actual subspaces U' and $\tilde{U}'$ of V. Since neither of these planes contains the vertical direction, both complement U. Once a particular complement is chosen, say U', then one can easily define a ***lift*** function $\sigma : V/U \to U' \subset V$. Because U' complements U in V, every $\mathbf{v} \in V$ can be written in the component form (1.38a), and we define

$$\sigma(\{\mathbf{v}\}) := \mathbf{u}'. \tag{1.46}$$

We must check, of course, that this definition makes sense by ensuring that the right side is independent of which $\mathbf{v} \in \{\mathbf{v}\}$ is used to represent the given vector in V/U. The details are left as an exercise, but the intuitive picture is clear from Fig. 1.3. A subspace $U' \subset V$ complements $U \subset V$ if and only if it cuts each vertical line parallel to U in exactly one point, and this point defines the vector $\mathbf{u}'$. For each vertical line, it is clearly unique. Finally, if one changes the chosen complementary subspace from U' to $\tilde{U}'$, then the lift function likewise changes to $\tilde{\sigma}$. Given a complementary subspace in V, there is always a lift from V/U into V, but one must be given the complementary subspace.

Exercise 1.47: If U' is a complementary subspace to $U \subset V$, show that *every* equivalence class of vectors $\{\mathbf{v}\}_U$ in V contains *exactly one* vector belonging to U'.

Exercise 1.48: Let U and U' be complementary subspaces in V, and let π and σ be the projection and lift maps discussed above. Given $\mathbf{v} \in V$, describe the vector $\sigma \circ \pi(\mathbf{v}) := \sigma(\pi(\mathbf{v}))$. Likewise, given $\mathbf{z} \in V/U$, what is $\pi \circ \sigma(\mathbf{z})$?

Exercise 1.49: Show that the sum of the dimensions of a subspace $U \subset V$ of a finite-dimensional vector space V and the quotient space V/U is equal to that of V itself.

Comment 1.50: (*On Natural Structures*)
We have encountered here for the first time a clear example of the difference between a ***natural*** and a ***fiducial*** structure. Given a subspace U of a vector space V, there is *only one* quotient space V/U. Its structure does not depend on any choice we make: it is "natural." The subspace $U \subset V$ has many complements $U' \subset V$,

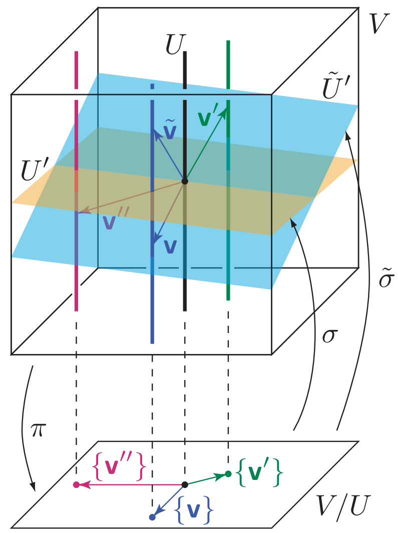


Figure 1.3: Structure of the quotient space V/U. This figure shows a vector space V, with a subspace U, and the projection of several vectors in V to V/U. Also shown are the projection π from V to V/U, and the lift functions σ and $\tilde{\sigma}$ from V/U a pair of subspaces U' and $\tilde{U}'$ complementary to U in V. `{vsquotient}`

and in certain calculations it may be useful to pick one of these in order to establish a result that, in the end, will not depend on which complement was chosen. We call such a choice a "fiducial" complement: it is chosen temporarily and at random to make progress, but ultimately plays no fundamental role in the result.

These terms are actually important in theoretical physics. In gauge theory, for example, the gauge potential corresponds to the difference between the gauge connection and a *fiducial* connection defined by the coordinate system. We explain this terminology in more detail below. For now, note that we use the gauge potential all the time in our calculations, but ultimately physical results must be gauge-invariant; they must be independent of the particular fiducial connection used to define the potential. In fact, it turns out that this example is closely related to the problem of picking a complementary subspace discussed in this section. The definition of the gauge theory automatically (*naturally*) picks out a subspace of so-called vertical vectors in a certain vector space. The gauge connection literally picks out a particular complement of horizontal vectors in that same space. Of course, this gauge connection is the mathematical manifestation of a dynamical, physical field. This, in essence, is why it is important to pay close attention to the difference between natural structures and fiducial ones in modern theoretical physics. Fiducial structures are likely to be associated with physical entities, which have observable, dynamical degrees of freedom all their own.

Exercise 1.51: Let $T \subset U \subset V$ be a nested sequence of subspaces of a vector space V. Prove that the sets $(V/T)/(U/T)$ and V/U are naturally equivalent. Proceed as follows:

a. Define the quotient spaces V/U, V/T and U/T.

b. Show that U/T is *naturally* identified with a subspace of V/T.

c. Show that each vector in the quotient $(V/T)/(U/T)$ *naturally* corresponds to a unique one in V/U, and *vice versa*.

These identifications really apply at the level of vector spaces; they respect the formation of linear combinations. Mathematically, this means that they are *linear* maps, which we introduce only in the next section. For now, focus only on showing that there is a natural one-to-one correspondence between elements of $(V/T)/(U/T)$ on the one hand and elements of V/U on the other.

PROBLEMS FOR LECTURE 1

1. Consider the bases

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \mathbf{b}_3 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

and

$$\tilde{\mathbf{b}}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{\mathbf{b}}_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \tilde{\mathbf{b}}_3 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

on $V = \mathbb{R}^3$. Note that only the third basis vectors differ.

(a) Expand the vector

$$\mathbf{v} = \begin{pmatrix} 2 \\ 3 \\ -2 \end{pmatrix}$$

in components relative to each basis. Explain why only the third components *agree* in your two expansions.

(b) Arrange the basis vectors $\mathbf{b}_i$ to form the columns of a matrix:

$$\boldsymbol{B} := \left(\begin{array}{c|c|c} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{array} \right).$$

Show explicitly that the matrix product $\boldsymbol{B}^{-1}\mathbf{v}$ yields a column vector whose entries are the components of $\mathbf{v}$ in the basis $\mathbf{b}_i$.

(c) Calculate the change-of-basis matrix $\boldsymbol{\lambda}$ from (1.24) that links the two bases in this problem. Also calculate the matrix $\tilde{\boldsymbol{\lambda}}$ from (1.24) with the roles of $\mathbf{b}_i$ and $\tilde{\mathbf{b}}_j$ interchanged. Show that the matrix product $\boldsymbol{\lambda}\tilde{\boldsymbol{\lambda}}$ gives the identity. Explain this result.

(d) Describe the two-dimensional subspaces spanned the pairs $\{\mathbf{b}_1, \mathbf{b}_2\}$, $\{\mathbf{b}_2, \mathbf{b}_3\}$, $\{\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2\}$ and $\{\tilde{\mathbf{b}}_2, \tilde{\mathbf{b}}_3\}$. In each case, write down the generic vector in the appropriate subspace as an explicit function of two independent variables. Describe the six possible intersections of these four subspaces with one another.

(e) Let U denote the subspace of V spanned by $\{\mathbf{b}_1, \mathbf{b}_2\}$, and let $\mathbf{v}$ be the vector given above. Describe the equivalence class $\{\mathbf{v}\} \in V/U$. Write down the generic vector in this class as a function of one independent variable. Is $\{\mathbf{v}\}$ a subspace of V?

(f) Show that the one-dimensional subspaces spanned by $\mathbf{b}_3$ and $\tilde{\mathbf{b}}_3$ both complement the subspace U from the previous problem. What are the lifts $\sigma(\{\mathbf{v}\})$ and $\tilde{\sigma}(\{\mathbf{v}\})$ of the vector $\{\mathbf{v}\} \in V/U$ from the previous problem into these subspaces?

You may solve this problem any way you like, including using a computer-algebra package to handle the detailed arithmetic.

2. Let $T \subset U \subset V$ be a nested sequence of subspaces of a vector space V. Let $T' \subset U$ complement T within U, and let $U' \subset V$ complement U within V. That is, suppose that $T + T' = U$ and $U + U' = V$ with no redundancy in either case. Show that $T' + U'$ complements T within V. Can every complement of T within V be written as such a sum?

The remaining material in this lecture is optional.

1.A EXISTENCE OF BASES

Bases have an obvious utility when analyzing vector spaces. But how do we know that a given vector space actually admits *any* basis? The definition above certainly does not guarantee that any subset $B \subset V$ should exist satisfying *both* criteria of linear independence and spanning V. It turns out, happily, that any worry of this sort is unfounded. Every vector space admits infinitely many bases. To understand why this is so, it is necessary to make a brief, but *very* mathematical, detour. Although outside the main thrust of our discussion, the ideas and methods introduced below to establish that every vector space does have a basis are both interesting and conceptually useful in their own right.

The set $\mathscr{K}$ of all subsets of a vector space V is an example of a ***partially ordered set***. It is endowed with a natural (logical) ***ordering relation*** $\preceq$ defined by simple set-wise inclusion:

$$K \preceq K' \quad \Leftrightarrow \quad K \subseteq K'. \tag{1.52}$$ {vsSubOrd}

That is, the set $K \subset V$ precedes $K' \subset V$ logically if every vector in K is also in K'. The ordering $\preceq$ in this case is only partial since there are certainly pairs of sets $K, K' \subset V$, each of which contains at least one vector not contained in the other. Such sets are not ordered in either direction (neither $K \preceq K'$ nor $K' \preceq K$) by the inclusion relation. We now need three more related definitions:

- A partially ordered set is called ***totally ordered*** if, for *any* pair of elements K and K', either $K \preceq K'$ or $K' \preceq K$.
- A subset $\mathscr{N} \subset \mathscr{K}$ of a partially ordered set is said to be ***bounded above*** if there exists at least one element $L \in \mathscr{K}$ with $N \preceq L$ for *all* $N \in \mathscr{N}$.
- Finally, an element $M \in \mathscr{K}$ of a partially ordered set is ***maximal*** if there is no element $M' \in \mathscr{K}$ which succeeds it logically: $M \preceq M'$ implies $M' = M$.

The four definitions above are linked together by

Proposition 1.53: (*Zorn's lemma*)
Let $\mathscr{K}$ be a partially ordered set with the property that every totally ordered subset of $\mathscr{K}$ is bounded above. Then, $\mathscr{K}$ has at least one maximal element.

This standard result of set theory is actually *unprovable*. It is equivalent to the ***axiom of choice***, which we do not state here, but which has grudgingly been accepted by mathematicians as a necessary assumption underlying set theory, and therefore of all mathematics, or at least all practical mathematics, built upon it.

To apply Zorn's lemma to the problem of vector-space bases, we make the following observations. First, the set $\overline{\mathscr{K}} \subset \mathscr{K}$ of *linearly independent* subsets of a vector space V inherits the partial ordering from the set $\mathscr{K}$ of *all* subsets of V. Second, if $K_\lambda \in \overline{\mathscr{K}}$ is a nested sequence of successively larger linearly independent sets, indexed by a parameter λ taking values in some (possibly infinite, or even uncountable) set Λ, then

$$L := \bigcup_{\lambda \in \Lambda} K_\lambda \tag{1.54}$$ {vsSubLim}

contains every K_λ, and thus bounds the sequence K_λ above. The set L must be linearly independent since, if it were not, there would exist a finite sequence of vectors $\mathbf{k}_i$, each belonging to some K_{λ_i}, such that

$$\alpha^1 \,\mathbf{k}_1 + \cdots + \alpha^n \,\mathbf{k}_n = \mathbf{0}. \tag{1.55}$$ {linDep}

However, because the sequence of all K_λ is nested, whichever of the finite number of K_{λ_i} is largest will contain all the other K_{λ_i}, and therefore all the other $\mathbf{k}_i$. Since this largest K_{λ_i} must be linearly independent, (1.55) is clearly impossible, and L is linearly independent. In the language of Zorn's lemma, K_λ is a generic totally ordered subset of $\overline{\mathscr{K}}$ since one element of any pair $(K_\lambda, K_{\lambda'})$ in our nested sequence must certainly contain the other. Moreover, this sequence is bounded above, within $\overline{\mathscr{K}}$, by the union L. Zorn's lemma then guarantees the existence of a *maximal* linearly independent set B. This must be a basis, spanning the vector space V. To see this, we suppose otherwise, that B does not span V. Then, there would have to exist at least one vector $\mathbf{v} \in V$ which could *not* be expanded in the form (1.17a). If we augmented the set B to the set $B' := B \bigcup \{\mathbf{v}\}$, the result would have to be linearly independent since otherwise there would exist vectors $\mathbf{b}_1, \ldots, \mathbf{b}_n \in B$ and scalars $\alpha^0, \alpha^1, \ldots, \alpha^n \in \mathbb{F}$, not all zero, such that

$$\alpha^0 \,\mathbf{v} + \alpha^1 \,\mathbf{b}_1 + \cdots + \alpha^n \,\mathbf{b}_n = \mathbf{0}. \tag{1.56}$$ {vsMaxSpan}

We cannot have $\alpha^0 = 0$ since B was linearly independent, and thus this relation can be solved for $\mathbf{v}$ in the form (1.17a), with $v^i = -\alpha^i/\alpha^0$. Thus, B' would be linearly independent, and would contain B, so $B \preceq B' \in \overline{\mathscr{K}}$. But this contradicts the maximality of B guaranteed by Zorn's lemma. No such $\mathbf{v}$ can exist, and B must span V. Maximal linearly independent sets are bases.

The logic of subsets of a vector space V is sketched roughly in Fig. 1.4. The ordering by inclusion runs from the bottom to the top in this figure, and the set of bases is indicated by the heavy black line at the border between the sets of linearly independent and spanning subsets of V. Note that, as this diagram suggests, no two different bases can be ordered relative to one another.

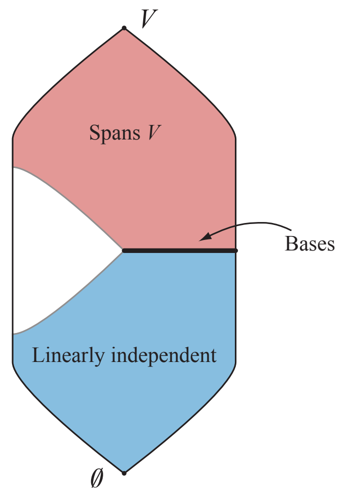


Figure 1.4: Sketch of the logic of the set of all subsets of a vector space V. Subsets are ordered by inclusion from the bottom of the diagram to the top. Linearly independent subsets, the smallest of which is the empty set $\emptyset$, are represented by the blue region at the bottom of the diagram. Subsets which span V, the largest of which is V itself, are represented by the red region at the top. The two intersect in the set of all bases for V. The unshaded region represents linearly dependent sets which fail to span V. {vsLatt}

Exercise 1.57: Show that the set $\mathscr{B}$ of bases for a given vector space V is "totally unordered." That is, if B and B' are both bases for V, then we have *neither* $B \preceq B'$ *nor* $B' \preceq B$ unless $B = B'$.

Exercise 1.58: Let V be a vector space, and let $K \subset V$ be linearly independent and $L \subset V$ span V. Show that there exists a basis B for V with $K \subseteq B \subseteq L$.

Exercise 1.59: Let V be a vector space. Prove that $B \subset V$ is a basis for V if and only if both (a) removing any element $\mathbf{b} \in B$ results in a subset $K \subset V$ which fails to span V and (b) adding any vector $\mathbf{v} \in V$ which is not in B results in a subset $L \subset V$ which is linearly dependent.

Zorn's lemma can also be used to establish the existence of bases with certain additional properties. We give a couple examples here in order to illustrate the flavor of the results.

Exercise 1.60: Use Zorn's lemma to show that there always exists at least one subspace $U' \subset V$ complementary to a given $U \subset V$. Show further that this complement is unique if and only if $U = V$.

Hint: Consider the collection of subspaces of V that intersect U only in the zero vector, ordered by inclusion.

Exercise 1.61: Let B be a basis for a given subspace $U \subset V$. Use Zorn's lemma to prove that there always exists at least one basis C for V that contains B as a subset.

Hint: The intuitive goal here is to "keep adding vectors to B until we have a set C spanning V." You simply need to formalize this picture using Zorn's lemma.

1.B INFINITE-DIMENSIONAL SPACES

It is a common fact of mathematics that finite things are easier to deal with than infinite. Vector spaces are no exception: finite-dimensional ones are *much* simpler than infinite-dimensional ones.

A careful reader may have noticed that the definition of a basis given here refers always to finite linear combinations of vectors. This may seem a bit surprising since examples of infinite-dimensional vector spaces are common today, even in physics. Obviously, any basis for an infinite-dimensional vector space will have to contain infinitely many vectors, so why should the definition of a basis refer only to finite sums? The answer lies in the subtlety inherent in defining an infinite sum of *anything*. In the simplest example of infinite series of scalars, we know quite well that infinite sums are often undefined. A proper mathematical meaning attaches to such a sum only when the corresponding sequence of its partial sums has a well-defined limit. To define an infinite sum, we must have at our disposal a proper notion of *convergence*. A general infinite-dimensional vector space is not equipped with the required structure, a ***topology***, which is needed to distinguish convergent sequences from non-. A vector space with this structure is called, of course, a ***topological vector space***.

The Hilbert space in quantum mechanics is a particular example of a topological vector space. The notion of convergence is defined just as in the usual case of sequences of scalars, with the Hilbert-space norm $\|\psi - \psi_0\|$ replacing the absolute value $|z - z_0|$ of scalars. However, the definition of a basis is slightly, but critically, different in a topological vector space than that given above. Specifically, in a topological vector space V we do *not* require that *every* vector $\mathbf{v} \in V$ can be written as a finite linear combination of basis vectors. Instead, we require that every vector $\mathbf{v} \in V$ can be written as the *limit* of a sequence such vectors. In the language of functional analysis, we require that the set of all finite linear combinations of basis vectors is ***dense*** in V. The difference is profound. For example, it is well-known that the standard Hilbert space $L^2(\mathbb{R}, \mathrm{d}x)$ for quantum mechanics, while not finite-dimensional, is ***separable***. That is, as a Hilbert space, it has a countable basis, such as the Hermite polynomials, the energy eigen-basis of the simple harmonic oscillator. As an ordinary *vector* space, divorced of its topology, this same object has a much larger, uncountable dimension.

Exercise 1.62: Show that the Hilbert space $L^2([0,1], \mathrm{d}x)$ admits a countable *Hilbert space* basis, but no countable *vector space* basis.

Hint: Do the first part explicitly, for example by using well-known properties of the Legendre polynomials. The second part cannot be done explicitly. However, you might note that vectors in the Hilbert space can be identified with equivalence classes of Cauchy sequences of vectors formed by finite linear combinations of the Legendre polynomials. Show that a basis for this set of functions[1] must be uncountable.

When working with Hilbert spaces in practice, for example, one *always* uses bases and dense subspaces in the topological sense. Indeed, most Hilbert spaces of interest in physics are *defined* beginning with a set of vectors whose finite linear combinations define a vector

[1] Technically, this should read "this set of equivalence classes of functions," where "equivalent" means essentially equal in the functional-analytic sense.

space V. Only then does one define an inner product on this space which turns it into a *pre-*Hilbert space. The Hilbert space itself is then defined as the set of all convergent sequences in this inner product space, and generally contains elements quite different from any of the original basis vectors. This happens especially in quantum field theory, for example. To do any calculation, one effectively works always with finite linear combinations of basis elements, and then proves some sort of continuity result needed to establish convergence. It is this second step which leads to the rich mathematical structure of function spaces.

The relative simplicity of finite-dimensional vector spaces lies in the general absence of questions of convergence. Every vector in a finite-dimensional space can be written in terms the same *finite* set of basis elements. While this doesn't quite get rid of the problem of defining infinite sums of vectors, it does reduce that problem to one of finding several infinite sums of *scalar* components relative to a single, fixed basis. This component-based notion of convergence is actually independent of the basis chosen, and mathematically is said to define the *unique* (Hausdorff, vector-space) topology on a given finite-dimensional vector space. It is the uniqueness of this topology which makes finite-dimensional vector spaces so simple. It obviates any need to discuss issues of convergence.

Throughout this course, we will focus exclusively on finite-dimensional vector spaces and, later, their generalizations to finite-dimensional manifolds. However, we will continue to emphasize a few differences between the finite- and infinite-dimensional cases below.

Exercise 1.63: Working in the vector space $\mathbb{R}^2$, define what it means for a sequence of vectors

$$\mathbf{v}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \tag{1.63a}$$ {vs2dVec}

to converge. Use the basis

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.63b}$$ {vs2dBas}

Show that if such a sequence converges using this basis, then it will also converge if one uses any other basis for $\mathbb{R}^2$, and that it converges to the same vector regardless of which basis is chosen.

Hint: This amounts to showing that if $x_n \to x$ and $y_n \to y$, then $ax_n + by_n \to ax + by$.

Exercise 1.64: Can the matrix representation (1.24) of the change-of-basis transformation be used, in principle, when dealing with infinite-dimensional vector spaces? What about when dealing with infinite-dimensional topological vector spaces? What new subtleties arise?

Lecture 2

The Algebra of Tensors and Densities

2.1 LINEAR MAPS

A function $\varphi : V \to W$ mapping one vector space to another is said to be ***linear*** if it commutes with the formation of linear combinations:

$$\forall \alpha, \alpha' \in \mathbb{F}; \mathbf{v}, \mathbf{v}' \in V : \quad \varphi(\alpha \mathbf{v} + \alpha' \mathbf{v}') = \alpha \varphi(\mathbf{v}) + \alpha' \varphi(\mathbf{v}'). \tag{2.1}$$ {vsLinMap}

The linear combination in the argument of φ on the left is taken in the vector space V, while that on the right is taken W. The two vector spaces appearing here are assumed either to be both real or both complex. Often, a linear map is referred to as a ***linear operator***, especially when it maps a vector space to itself.

Example 2.2: If $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, then any real $m \times n$ matrix $\boldsymbol{A}$ defines a linear map from V to W when it acts on n-dimensional column vectors from the left in the usual way: $\mathbf{w} = \boldsymbol{A}\mathbf{v}$.

The matrix-style notation of the previous example is often used in practice for general linear operators. That is, if $\varphi : V \to W$ is a linear map, we often write simply $\mathbf{w} = \varphi\, \mathbf{v} := \varphi(\mathbf{v})$.

Example 2.3: (*pull-back of a function*)
Let S and $\tilde{S}$ be arbitrary sets and $\psi : S \to \tilde{S}$, an arbitrary function mapping one to the other. Then, we define a map $\psi_* : \mathscr{F}^0(\tilde{S}) \to \mathscr{F}^0(S)$ between the associated function spaces by

$$[\psi_* \tilde{f}](s) := \tilde{f}(\psi(s)). \tag{2.3a}$$ {vsFpback}

The right side here is a function on S defined by evaluating the function $\tilde{f}$ on $\tilde{S}$ at the point $\psi(s) \in \tilde{S}$. Thus, ψ_* maps functions on $\tilde{S}$ to functions on S, as advertised. Moreover, this mapping is linear. To see this, we observe

$$\begin{aligned} [\psi_*(\alpha \tilde{f} + \alpha' \tilde{f}')](s) &:= [\alpha \tilde{f} + \alpha' \tilde{f}'](\psi(s)) := \alpha \tilde{f}(\psi(s)) + \alpha' \tilde{f}'(\psi(s)) \\ &=: \alpha [\psi_* \tilde{f}](s) + \alpha' [\psi_* \tilde{f}'](s) =: [\alpha \psi_* \tilde{f} + \alpha' \psi_* \tilde{f}'](s). \end{aligned} \tag{2.3b}$$ {vsFpbLin}

Since this result holds for all points $s \in S$, we have shown that $\psi_*(\alpha \tilde{f} + \alpha' \tilde{f}') = \alpha \psi_* \tilde{f} + \alpha' \psi_* \tilde{f}'$ as functions on S for all functions $\tilde{f}, \tilde{f}'$ on $\tilde{S}$. In other words, ψ_* commutes with the formation of linear combinations; it is linear. We refer to the function $\psi_* \tilde{f}$ on S as the ***pull-back*** of the function $\tilde{f}$ on $\tilde{S}$ under the mapping $\psi : S \to \tilde{S}$. The direction of ψ_* is opposite to that of ψ itself, which is why this is called the pull-back.

Exercise 2.4: Show that a linear map $\varphi : V \to W$ must map the zero vector $\mathbf{0} \in V$ to the zero vector $\mathbf{0} \in W$.

Exercise 2.5: If $\varphi : U \to V$ and $\psi : V \to W$ are linear maps between vector spaces, show that the ***composition map*** $\psi \circ \varphi : U \to W$ defined by $\psi \circ \varphi(\mathbf{u}) := \psi(\varphi(\mathbf{u}))$ is also linear.

Exercise 2.6: (*shift operators*)
Let $V = \mathscr{S}^0$ be the vector space of countable sequences of scalars, and let $L : V \to V$ be the mapping

$$L : (x_1, x_2, \ldots) \mapsto (x_2, x_3, \ldots). \tag{2.6a}$$ {vsLShift}

Show that this mapping is linear. For obvious reasons, L is called the ***left-shift operator***. In addition, we can define the mappings

$$R_\alpha : (x_1, x_2, \ldots) \mapsto (\alpha, x_1, \ldots) \tag{2.6b}$$ {vsRShift}

on the same space of sequences. For which fixed value(s) of the scalar α is R_α linear? Again for obvious reasons, the linear R is called the ***right-shift operator***.

A general linear mapping $\varphi : V \to W$ is also called a ***homomorphism*** (of vector spaces). Alone, this is a rather pointless bit of mathematical nomenclature. However, linear maps with various special properties are given similar names which may be used to describe those properties succinctly. We collect these here in order to demonstrate the various possible features of linear maps.

- A ***monomorphism*** $\varphi : V \to W$ is linear and ***one-to-one***, meaning $\varphi(\mathbf{v}) = \varphi(\mathbf{v}')$ in W if and only if $\mathbf{v} = \mathbf{v}'$ in V.
- An ***epimorphism*** $\varphi : V \to W$ is linear and ***onto***, meaning that every $\mathbf{w} \in W$ satisfies $\mathbf{w} = \varphi(\mathbf{v})$ for at least one $\mathbf{v} \in V$.
- An ***isomorphism*** $\varphi : V \to W$ is linear and both one-to-one and onto.
- An ***endomorphism*** $\varphi : V \to V$ is an epimorhpism from a vector space to itself.
- An ***automorphism*** $\varphi : V \to V$ is an isomorphism from a vector space to itself.

It may seem at first that an endomorphism on a vector space V would have to be an automorphism. However, this is not actually the case.

Example 2.7: Let $\mathscr{S}^0$ denote the countable sequence space defined above, and let $L : (f_1, f_2, \ldots) \mapsto (f_2, f_3, \ldots)$ denote the left-shift operator thereon. This map has been shown to be linear in the previous exercise. It is also clearly an epimorphism since, for example, the vector $(0, f_1, f_2, \ldots)$ is mapped to $(f_1, f_2, \ldots)$ by L. However, other vectors are mapped to the same sequence by L. Rather than shifting the entries to the right and inserting zero, we could have inserted any other scalar α in the first argument and the subsequent left-shift would be the same. In short,

$$\forall \alpha \in \mathbb{F}, f \in \mathscr{S}^0 : \quad L \circ R_\alpha(f) = f, \tag{2.7a}$$ {vsLRSid}

where the notation has been defined in the exercise above. As a result, we see that L is not one-to-one. It is an epimorphism, but not an automorphism.

Exercise 2.8: Show that *every* epimorphism of a *finite-dimensional* vector space V *is* an automorphism.

Exercise 2.9: Why is there no word for a map from a vector space to itself which is one-to-one, but not onto?

Hint: I don't know either. Come up with an example of such a map, and show that such things exist only in infinite dimensions.

Exercise 2.10: Consider the ***derivative map*** $D : p(x) \mapsto p'(x) := \frac{\mathrm{d}}{\mathrm{d}x}\, p(x)$. Show that D is a linear map from the space $\mathscr{P}_3(x)$ of polynomials in a real variable x with order at most three, to the space $\mathscr{P}_2(x)$ of polynomials of order at most two. Is it one-to-one? Is it onto? If we instead view D as a linear *operator* on $\mathscr{P}_3(x)$, which we can, is it onto?

Exercise 2.11: Show that the ***multiplication map*** $X : p(x) \mapsto x\, p(x)$ is a linear map from $\mathscr{P}_2(x)$ to $\mathscr{P}_3(x)$. Is it one-to-one? Is it onto?

Exercise 2.12: Show explicitly that the composition map $X \circ D$ of the maps defined in the previous two exercises is a linear operator on $\mathscr{P}_3(x)$. Is it one-to-one? Is it onto? Which vectors does it preserve?

Exercise 2.13: An isomorphism, like any one-to-one and onto map, is invertible. Show that the inverse of an isomorphism is also linear, as well as one-to-one and onto. That is, any isomorphism is invertible, and the inverse is also an isomorphism. Since the identification of vectors in one space with those of another under an isomorphism is unique, two isomorphic vector spaces are essentially the same.

Exercise 2.14: Show that any two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension. Do this by actually constructing an isomorphism between the two spaces. Is your isomorphism natural?

2.1.1 Matrix Representations

Let $B = \{\mathbf{b}_i\}$ and $C = \{\mathbf{c}_j\}$ be given bases for V and W, respectively, and consider a linear map $\varphi : V \to W$. Let us assume, for simplicity, that both vector spaces involved here are finite-dimensional. Since arbitrary vectors can be expanded in these bases, we can write

$$\varphi(\mathbf{v}) = \varphi\left(\sum_i v^i\, \mathbf{b}_i\right) = \sum_i v^i\, \varphi(\mathbf{b}_i) =: \sum_i v^i \sum_j \varphi^j{}_i\, \mathbf{c}_j = \sum_{ij} \varphi^j{}_i\, v^i\, \mathbf{c}_j. \tag{2.15}$$ {lmapbas}

Here, we have defined the components $\varphi^j{}_i$ in the basis C on W of the vectors $\varphi(\mathbf{b}_i)$ gotten by acting φ on the given basis vectors in $B \subset V$. By (2.15), if we know all the coefficients $\varphi^j{}_i$, then we can easily calculate the action of φ on any vector $\mathbf{v}$. We can rewrite this result in a familiar, matrix form as

$$\varphi(\mathbf{v}) = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_m \end{pmatrix} \begin{pmatrix} \varphi^1{}_1 & \varphi^1{}_2 & & \varphi^1{}_n \\ \varphi^2{}_1 & \varphi^2{}_2 & & \varphi^2{}_n \\ & & \ddots & \\ \varphi^m{}_1 & \varphi^m{}_2 & & \varphi^m{}_n \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix} \tag{2.16}$$ {lmapmat}

where the component column vector on the right contains the components of $\mathbf{v}$ in the basis B on V. The matrix appearing here is called the ***matrix representation*** of φ with respect to the bases B and C. Note that these manipulations are well-defined even in infinite dimensions, where only finitely many entries in a given row or column of the the matrix will be non-zero. In an infinite-dimensional *topological* vector space, of course, one must worry (a lot!) about convergence.

Exercise 2.17: Work out the matrix representations of the derivative and multiplication maps $D : \mathscr{P}_3(x) \to \mathscr{P}_2(x)$ and $X : \mathscr{P}_2(x) \to \mathscr{P}_3(x)$ defined above. Also work out the matrix representation of the composition operator $X \circ D$, and show explicitly that this matrix is given by a product of the previous two. Use the ***monomial bases*** $\mathbf{b}_i := x^i$ and $\mathbf{c}_j := x^j$.

Exercise 2.18: How does the matrix appearing in (2.16) change if one changes one or both of the bases B and C according to (1.24)? Repeat the previous exercise in the ***Legendre bases*** $\mathbf{b}_i := L_i(x)$ and $\mathbf{c}_j := L_j(x)$, where the L_n denote the standard Legendre polynomials. Work out the matrix representations of D and X in this basis, and compare your results to what you would get simply using the change-of basis results derived here.

2.1.2 Kernel and Image

Any linear map $\varphi : V \to W$ naturally defines subspaces of the two vector spaces involved. There are the ***kernel*** and ***image*** spaces for φ, defined by

$$\ker\varphi := \{\mathbf{v} \in V \mid \varphi(\mathbf{v}) = \mathbf{0}\} \qquad \text{and} \qquad \operatorname{im}\varphi := \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V \ni \mathbf{w} = \varphi(\mathbf{v})\}, \quad (2.19)$$

respectively. In words, the kernel of φ is the set of vectors in V that get mapped to the zero vector in W by φ, while the image of φ is the set of vectors in W that actually result by applying φ to some vector in V. Both of these subsets are subspaces.

Exercise 2.20: Show that $\ker\varphi \subset V$ is a subspace for any linear map $\varphi : V \to W$.
Solution: Suppose that $\mathbf{v}, \mathbf{v}' \in \ker\varphi$, meaning $\varphi(\mathbf{v}) = \varphi(\mathbf{v}') = \mathbf{0}$. Forming any linear combination of these vectors, we have

$$\varphi(\alpha\mathbf{v} + \alpha'\mathbf{v}') = \alpha\varphi(\mathbf{v}) + \alpha'\varphi(\mathbf{v}') = \alpha\mathbf{0} + \alpha'\mathbf{0} = \mathbf{0}. \qquad (2.20a) \quad \{\texttt{vsKerSub}\}$$

Thus, $\ker\varphi$ must contain all linear combinations of its elements. It is a subspace.

Exercise 2.21: Show that $\operatorname{im}\varphi \subset W$ is a subspace for any linear map $\varphi : V \to W$.

Let $\pi_\varphi : V \to V/\ker\varphi$ denote the standard projection to the quotient space defined in the previous section. In addition, we can define the natural ***injection*** map $\iota_\varphi : \operatorname{im}\varphi \to W$ simply by $\iota_\varphi(\mathbf{w}) := \mathbf{w}$. To be clear here, we are viewing the subspace $\operatorname{im}\varphi \subset W$ as a vector space in its own right, independent of its identity with a particular collection of vectors in W. This identity is then captured in the natural map ι_φ, which arises precisely because vectors in $\operatorname{im}\varphi$ are already vectors in W. Now we define a new map $\tilde{\varphi} : V/\ker\varphi \to \operatorname{im}\varphi$ by

$$\tilde{\varphi}(\{\mathbf{v}\}) := \varphi(\mathbf{v}). \qquad (2.22) \quad \{\texttt{vsQuImIso}\}$$

To show that this definition makes sense entails proving that $\varphi(\mathbf{v})$ gives the same result in W for every vector $\mathbf{v}$ in a given equivalence class $\{\mathbf{v}\} \in V/\ker\varphi$. This is easy. For any pair of vectors $\mathbf{v}, \mathbf{v}' \in V$, we have

$$\varphi(\mathbf{v}') - \varphi(\mathbf{v}) = \varphi(\mathbf{v}' - \mathbf{v}), \qquad (2.23) \quad \{\texttt{vsQuImIsowd}\}$$

since φ is linear. As a result, we conclude that $\varphi(\mathbf{v}) = \varphi(\mathbf{v}')$ in W if and only if $\mathbf{v}' - \mathbf{v} \in \ker\varphi$ in V. This proves that the map $\tilde{\varphi}$ of (2.22) is well-defined since $\mathbf{v}' - \mathbf{v} \in \ker\varphi$ implies

$\varphi(\mathbf{v}') = \varphi(\mathbf{v})$, but actually allows us to go a bit further. If $\mathbf{v}$ and $\mathbf{v}'$ belong to *distinct* equivalence classes $\{\mathbf{v}\} \neq \{\mathbf{v}'\}$ in $V/\ker\varphi$, then we *must* have $\varphi(\mathbf{v}) \neq \varphi(\mathbf{v}')$ in W. Otherwise, by (2.23), we would have $\varphi(\mathbf{v}' - \mathbf{v}) = \mathbf{0}$, meaning that $\mathbf{v}' - \mathbf{v}$ belongs to the kernel of φ, and thus that $\{\mathbf{v}'\} = \{\mathbf{v}\}$ by definition of the quotient. This means that the map $\tilde{\varphi}$ of (2.22) is in fact *one-to-one*. Viewed as a map to $\mathrm{im}\,\varphi$, $\tilde{\varphi}$ is also *onto*, since certainly every vector $\mathbf{w} = \varphi(\mathbf{v})$ can also be written as $\mathbf{w} = \tilde{\varphi}(\{\mathbf{v}\})$. That is, the map $\tilde{\varphi}$ from the quotient space $V/\ker\varphi$ to the image space $\mathrm{im}\,\varphi$ *is an isomorphism*. Similarly, the map π_φ is an epimorphism by the definition above since it certainly maps onto $V/\ker\varphi$ but is not necessarily one-to-one, while ι_φ is a monomorphism since it is one-to-one by definition but not necessarily onto. Thus, we have proved that an arbitrary linear map is naturally decomposed into a product $\iota_\varphi \circ \tilde{\varphi} \circ \pi_\varphi$ of a monomorphism, an isomorphism and an epimorphism. This is a central result of linear algebra.

Exercise 2.24: Show that $\dim V = \dim(\ker\varphi) + \dim(\mathrm{im}\,\varphi)$ for any linear map $\varphi : V \to W$ defined on a finite-dimensional vector space.

Exercise 2.25: Let $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, and interpret the above result in terms of matrices. Show that an arbitrary $m \times n$ matrix M with rank p can be written as a product of the form $M = JAK$, where J is an $m \times p$ matrix with rank p, A is an *invertible* $p \times p$ matrix, and K is a $p \times n$ matrix with rank p. Argue that, by some judicious choice of bases, A can always be taken to be the *identity* matrix.

Exercise 2.26: Some pure math texts *define* a a subspace of a vector space V to be a vector space U together with a monomorphism $\iota : U \to V$. Show that this more abstract definition is completely equivalent to the "nuts-and-bolts" definition given above.

2.1.3 Direct Sums

Intuitively, the ***direct sum*** of two vector spaces, V and W, can be thought of as the space $V \oplus W$ of column vectors of the form

$$\begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix}, \tag{2.27}$$ {vsDScol}

where of course $\mathbf{v} \in V$ and $\mathbf{w} \in W$. This arrangement of vectors, in addition to being intuitively suggestive, allows the rules defining the direct sum to be captured using the ordinary laws of linear algebra:

$$\alpha \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} + \alpha' \begin{pmatrix} \mathbf{v}' \\ \mathbf{w}' \end{pmatrix} = \begin{pmatrix} \alpha\mathbf{v} + \alpha'\mathbf{v}' \\ \alpha\mathbf{w} + \alpha'\mathbf{w}' \end{pmatrix}. \tag{2.28}$$ {vsDScolLC}

Another way of defining the direct sum is as the vector space of elements of the Cartesian product $V \times W$, with addition and scalar multiplication defined such that

$$(\mathbf{v}, \mathbf{w}) + (\mathbf{v}', \mathbf{w}') := (\mathbf{v} + \mathbf{v}', \mathbf{w} + \mathbf{w}') \qquad \text{and} \qquad \alpha\,(\mathbf{v}, \mathbf{w}) := (\alpha\mathbf{v}, \alpha\mathbf{w}). \tag{2.29}$$ {vsDSasm}

In fact, this is generally the preferred notation for elements of a direct sum.

Exercise 2.30: Show that the Cartesian product $V \times W$ is indeed a vector space with the definitions (2.29) of vector addition and scalar multiplication.

Exercise 2.31: Show that the dimension of $V \oplus W$ is equal to the sum of those of V and W, if both of the latter are finite-dimensional.
Hint: Explicitly construct a basis for $V \oplus W$ from separate bases for V and W.

Exercise 2.32: Show that V and W are naturally identified with complementary subspaces of $V \oplus W$. Define explicitly monomorphisms $\iota_V : V \to V \oplus W$ and $\iota_W : W \to V \oplus W$, as well as epimorphisms $\pi_V : V \oplus W \to V$ and $\pi_W : V \oplus W \to W$.

The ***free vector space*** over an arbitrary set S is defined to be the set of functions $\mathscr{F}(S)$ on S that vanish at all but a *finite* number of its points. Contrast this definition with that of the function space $\mathscr{F}^0(S)$ above. The difference is that, while $\mathscr{F}^0(S)$ contains *all* functions on S, $\mathscr{F}(S)$ contains only those with finite ***support***. The support of a function f is the subset of its domain on which $f \neq 0$.

Exercise 2.33: Show that $\mathscr{F}(S)$ is a vector space, and a vector subspace of $\mathscr{F}^0(S)$. When are the two isomorphic?

Exercise 2.34: Define a natural basis on $\mathscr{F}(S)$, and show that its elements are in one-to-one correspondence with the points of S.
Hint: Consider the ***characteristic functions*** $\chi_s : S \to \mathbb{F}$, defined for each $s \in S$ such that $\chi_s(s) = 1$ and $\chi_s(s') = 0$ for all $s' \neq s \in S$.

Exercise 2.35: A common technique in the mathematics literature is to define objects by certain "universal properties." For example, we could define *a* free vector space over as set S to be a vector space F, together with a function $b : S \to F$ such that, for any *other* vector space V and function $c : S \to V$, there exists a *unique* linear map $\varphi : F \to V$ such that $c(s) = \varphi \circ b(s)$ for all $s \in S$. Mathematicians like to write this in the following diagrammatic form:

$$\begin{tikzcd} S \arrow[r, "b"] \arrow[dr, "c"'] & F \arrow[d, "\varphi"] \\ & V. \end{tikzcd} \tag{2.35a}$$ {vsFreeCD}

This is called a ***commutative diagram***, and it indicates that it doesn't matter whether one gets from the set S to the set V going directly via c, or the indirect map $\varphi \circ b$ that goes through F along the way. Show that $\mathscr{F}(S)$ is a free vector space over S under this definition, and that any *other* free vector space F over S must be isomorphic to $\mathscr{F}(S)$.

Exercise 2.36: Show that *every* vector space is isomorphic to the free vector space on some set B.

Exercise 2.37: Show that the direct sum of free vector spaces $\mathscr{F}(S) \oplus \mathscr{F}(S')$ is naturally isomorphic to the free vector space $\mathscr{F}(S \cup S')$, where the set-theoretic union here is understood to be *disjoint*.

Exercise 2.38: Show that the direct sum $V \oplus W$ of vector spaces could be defined "universally" by either of the commutative diagrams

$$\begin{tikzcd} V \arrow[dr, "\iota_V"] \arrow[ddr, "\varphi_V"'] & & W \arrow[dl, "\iota_W"'] \arrow[ddl, "\varphi_W"] \\ & V \oplus W \arrow[d, "\varphi"] & \\ & U & \end{tikzcd} \quad \text{or} \quad \begin{tikzcd} & U \arrow[d, "\psi"] \arrow[ddl, "\psi_V"'] \arrow[ddr, "\psi_W"] & \\ & V \oplus W \arrow[dl, "\pi_V"] \arrow[dr, "\pi_W"'] & \\ V & & W. \end{tikzcd} \tag{2.38a}$$

In the first case, the direct sum is defined as a vector space $V \oplus W$ together with linear maps $\iota_V : V \to V \oplus W$ and $\iota_W : W \to V \oplus W$ such that, for any other vector space U and pair of linear maps $\varphi_V : V \to U$ and $\varphi_W : W \to U$, there exists a unique linear map $\varphi : V \oplus W \to U$ that makes the diagram commute. In the later case, the direct sum is a vector space $V \oplus W$ together with linear maps $\pi_V : V \oplus W \to V$ and $\pi_W : V \oplus W \to W$ such that, for any other vector space U and pair of linear maps $\psi_V : U \to V$ and $\psi_W : U \to W$, there exists a unique linear map $\psi : U \to V \oplus W$ that makes the diagram commute. The latter diagram actually defines a ***direct product*** of vector spaces, but the two notions are the same in this case. Show that the direct sum is unique up to isomorphism in either case, and that the vector space defined by (2.29) is a direct sum.

Exercise 2.39: Give a precise mathematical meaning to the notion that the direct sum $V \oplus W$ is the "smallest" vector space containing V and W as *disjoint* subspaces.

2.2 THE DUAL SPACE

The set $\mathrm{Hom}(V, W)$ of *all* homomorphisms $\varphi : V \to W$ from one vector space to another is a vector space itself. Linear combinations in this space are defined in the obvious way:

$$[\alpha\varphi + \alpha'\varphi'](\mathbf{v}) := \alpha\,\varphi(\mathbf{v}) + \alpha'\,\varphi'(\mathbf{v}). \tag{2.40}$$ {vsHomVS}

This, of course, is similar to the usual rule (1.12a) which makes a function space a vector space. The difference arises because the values of the functions involved here are vectors in the space W, rather than merely scalars as in (1.12a). The key to showing that this definition induces a natural vector-space structure on $\mathrm{Hom}(V, W)$ is to show that the function $\alpha\varphi+\alpha'\varphi'$ from V to W defined in this way is, in fact, *linear*. It is straightforward to do so:

$$\begin{aligned}
[\alpha\varphi + \alpha'\varphi'](\beta\mathbf{v} + \beta'\mathbf{v}') &:= \alpha\,\varphi(\beta\mathbf{v} + \beta'\mathbf{v}') + \alpha'\,\varphi'(\beta\mathbf{v} + \beta'\mathbf{v}') \\
&= \alpha\,[\beta\,\varphi(\mathbf{v}) + \beta'\,\varphi(\mathbf{v}')] + \alpha'\,[\beta\,\varphi'(\mathbf{v}) + \beta'\,\varphi'(\mathbf{v}')] \\
&= \beta\,[\alpha\,\varphi(\mathbf{v}) + \alpha'\,\varphi'(\mathbf{v})] + \beta'\,[\alpha\,\varphi(\mathbf{v}') + \alpha'\,\varphi'(\mathbf{v}')] \\
&=: \beta\,[\alpha\varphi + \alpha'\varphi'](\mathbf{v}) + \beta'\,[\alpha\varphi + \alpha'\varphi'](\mathbf{v}').
\end{aligned} \tag{2.41}$$ {vsHomVSwd}

This calculation *proves* that a linear combination of linear maps, as defined above, from one vector space to another is another linear map between the same spaces. The set $\mathrm{Hom}(V, W)$ is therefore naturally a vector space for any pair of vector spaces V and W.

Exercise 2.42: When $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, give an interpretation of the above result in terms of matrices.

Exercise 2.43: Let $\varphi : V \to W$ be a linear map. Define an associated map $L_\varphi : \mathrm{Hom}(W, U) \to \mathrm{Hom}(V, U)$, where U denotes any other vector space. Similarly, define $R_\varphi : \mathrm{Hom}(U, V) \to \mathrm{Hom}(U, W)$. Prove that both associated maps are linear.

A crucial example of the vector space of homomorphisms discussed above arises when we simply take $W = \mathbb{F}$. This space consists of all linear scalar-valued functions on a vector space V, and is known as its ***dual space*** V^*. Elements of V^* are called ***dual vectors*** or ***co-vectors***, and satisfy

$$\boldsymbol{\omega} \in V^* \qquad \Leftrightarrow \qquad \boldsymbol{\omega} : V \to \mathbb{F} \quad \text{with} \quad \boldsymbol{\omega}(\alpha\mathbf{v} + \alpha'\mathbf{v}') = \alpha\,\boldsymbol{\omega}(\mathbf{v}) + \alpha'\,\boldsymbol{\omega}(\mathbf{v}'). \tag{2.44}$$ {vsDualDef}

Since dual vectors take values in the same field underlying V, V^* is always a vector space of the same type (real or complex) as V. Furthermore, when V is finite-dimensional, both vector spaces have the same dimension. If V is infinite-dimensional, then V^* is also infinite-dimensional, but whether they have the "same" dimension is uncertain at best. In any event, we pick a basis $B = \{\mathbf{b}_i\}$ on V and define the maps

$$\boldsymbol{\beta}^i(\mathbf{v}) := v^i \qquad \text{where} \qquad \mathbf{v} =: \sum_i v^i\,\mathbf{b}_i. \tag{2.45}$$ {vsDBasDef}

That is, the action of $\boldsymbol{\beta}^i$ on a given vector $\mathbf{v} \in V$ gives the component of that vector along $\mathbf{b}_i$ in the basis B. It is obvious that these functions on V are linear. However, when V is finite-dimensional, they also span V^*. We consider the infinite-dimensional case briefly below. In the finite-dimensional case, we consider the action of an arbitrary co-vector $\boldsymbol{\omega} \in V^*$ on an arbitrary vector $\mathbf{v} \in V$:

$$\boldsymbol{\omega}(\mathbf{v}) = \boldsymbol{\omega}\left(\sum_i v^i\,\mathbf{b}_i\right) = \sum_i v^i\,\boldsymbol{\omega}(\mathbf{b}_i) = \sum_i \boldsymbol{\beta}^i(\mathbf{v})\,\boldsymbol{\omega}(\mathbf{b}_i). \tag{2.46}$$ {vsDBasExp}

Since this result hold for *all* $\mathbf{v} \in V$, we can write

$$\boldsymbol{\omega} = \sum_i \omega_i\,\boldsymbol{\beta}^i \qquad \text{with} \qquad \omega_i := \boldsymbol{\omega}(\mathbf{b}_i), \tag{2.47}$$ {vsDBasCom}

where the linear combination in V^* is defined as in (2.40). Note the similarity between the definition of the components ω_i given here and the original definition of the $\boldsymbol{\beta}^i$ in (2.45). We have shown here that an arbitrary basis B on a finite-dimensional vector space V naturally gives rise to a basis $B^* := \{\boldsymbol{\beta}^i\}$ on V^*, known as the ***dual basis*** for B. Since the index i runs over the same set of values in each case, V and V^* must have the same dimension.

Exercise 2.48: Use the change-of-basis formulae (1.23) and (1.24) to work out a relationship between the bases B^* and $\tilde{B}^*$ on V^* dual to a given pair of bases B and $\tilde{B}$ on V. Give a matrix expression for one dual basis in terms of the other. Show that it has the same form as the transformation rule for the *components* of a vector $\mathbf{v}$ under a change of basis.

Exercise 2.49: Calculate and give a matrix expression for the relationship between the *components* (2.47) of a given dual vector $\boldsymbol{\omega}$ under a change of dual basis induced by a change of vector basis $B \to \tilde{B}$. Show that it has the same form as (1.24).

Exercise 2.50: When V is finite-dimensional, so V^* has the same finite dimension, they are isomorphic. However, is this isomorphism *natural*? Given a basis $B = \{\mathbf{b}_i\} \rightsquigarrow B^* = \{\boldsymbol{\beta}^i\}$, it makes sense to define the map

$$\mathbf{v} = \sum_i v^i\,\mathbf{b}_i \mapsto \mathbf{v}_B := \sum_i v^i\,\boldsymbol{\beta}^i \tag{2.50a}$$ {vsDBasIso}

from V to V^*. Show that $\mathbf{v} \mapsto \mathbf{v}_B$ is indeed an isomorphism for a given basis B, but that different bases lead to different isomorphisms.

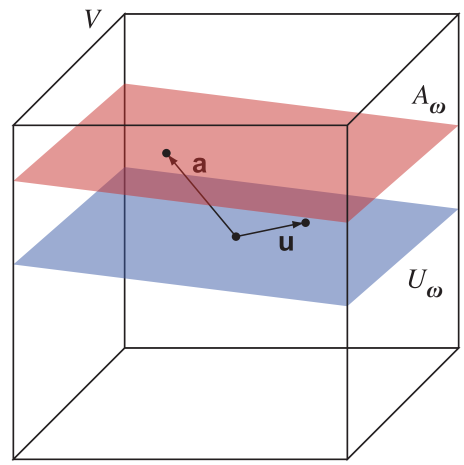


Figure 2.1: A dual vector $\boldsymbol{\omega} \in V^*$ defines a pair of parallel planes in a vector space V. One is the subspace U_ω of vectors $\mathbf{u} \in V$ such that $\boldsymbol{\omega}(\mathbf{u}) = 0$. The second is the affine space A_ω of vectors $\mathbf{a} \in V$ with $\boldsymbol{\omega}(\mathbf{a}) = 1$. Both have dimension one less than that of V. {vsdual}

Geometrically, one can identify a dual vector $\boldsymbol{\omega} \in V^*$ with the subset $A_{\boldsymbol{\omega}}$ of vectors $\mathbf{a} \in V$ for which $\boldsymbol{\omega}(\mathbf{a}) = 1$. This sub*set* of V is not a sub*space* since it cannot contain linear combinations of its elements. However, it is parallel to such a subspace, which might be denoted $U_{\boldsymbol{\omega}}$, containing all vectors $\mathbf{u} \in V$ for which $\boldsymbol{\omega}(\mathbf{u}) = 0$. Mathematically, $A_{\boldsymbol{\omega}}$ is said to have the structure of an ***affine space*** over the vector space $U_{\boldsymbol{\omega}}$. We will not elaborate this definition here, but its key properties are captured in the first exercise below. The geometry of a co-vector is depicted in Fig. 2.1.

Exercise 2.51: Show that A_ω is *not* a subspace of V, that U_ω *is*, and that the sum $\mathbf{a} + \mathbf{u}$ of any vector $\mathbf{a} \in A_\omega$ with any vector $\mathbf{u} \in U_\omega$ is again a vector in A_ω. Show that the "weighted average" $\alpha\,\mathbf{a} + (1-\alpha)\,\mathbf{a}'$ of any pair of vectors in A_ω once again gives a vector in A_ω. Extend this notion to arbitrary finite collections of vectors in A_ω.

Exercise 2.52: If V is finite-dimensional, show that U_ω has dimension one less than that of V for any $\boldsymbol{\omega} \in V^*$. If V is infinite-dimensional, show that U_ω must be so as well.

Exercise 2.53: Let $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ be dual vectors. Show that $A_\omega = A_{\omega'}$ if and only if $\boldsymbol{\omega} = \boldsymbol{\omega}'$. Can the same be said if $U_\omega = U_{\omega'}$?

Exercise 2.54: Working directly with the affine spaces A_ω and $A_{\omega'}$ associated with a pair of dual vectors, describe the laws of dual-vector addition and scalar multiplication geometrically.

2.2.1 Infinite-Dimensional Redux

When V is infinite-dimensional, one can still define the co-vectors $\boldsymbol{\beta}^i$ as in (2.45), but these fail to span V^* as a vector space. Recall that a basis for an infinite-dimensional vector space is a set whose *finite* linear combinations include every vector in the space. Accordingly, our $\boldsymbol{\beta}^i$ at least span a *subspace* $F \subset V^*$ consisting of finite linear combinations of the $\boldsymbol{\beta}^i$. Now consider a co-vector $\boldsymbol{\omega} \in F$ in this subsapce as a function on the basis *set* $B \subset V$. By (2.47), the values of this function are just the "components" ω_i of $\boldsymbol{\omega}$. Since $\boldsymbol{\omega} \in F$, only finitely many of these fail to vanish, each taking some finite value. Thus, $\boldsymbol{\omega}$ defines a *bounded* function on B. Since B is infinite, not all functions will have this property. Moreover, even if the set of components ω_i has no upper bound, the sums (2.46) will always converge because the vector $\mathbf{v}$ has a *finite* expansion in terms of the vectors in B; no infinite sums are involved. Thus, any *unbounded* function on B can be used to generate a co-vector $\boldsymbol{\eta} \in V^*$ lying outside the subspace F spanned by the $\boldsymbol{\beta}^i$. This means, of course, that B^* is *not* a basis. Note that this argument breaks down when V is finite-dimensional because *any* function on a finite set is bounded.

Example 2.55: The space of smooth (C^∞) functions of a real variable $x \in [0, 1]$ form a vector space in an obvious way. The key to showing this is to observe that the derivative operator $D := \frac{\mathrm{d}}{\mathrm{d}x}$ is linear, whence linear combinations of smooth functions must be smooth. We define a co-vector $\boldsymbol{\omega}$ on our space of smooth functions by $\boldsymbol{\omega}(f) := f'(1)$, the value of the derivative of f at $x = 1$. Again, this is obviously linear. The monomial functions x^n on the unit interval are smooth and linearly independent, but do not span the space of all such smooth functions. However, this partial basis can be completed to a full basis by Zorn's lemma. The key point is that $\boldsymbol{\omega}(x^n) = n$ and, since n can be arbitrarily large, the co-vector $\boldsymbol{\omega}$, viewed as a function on the completed basis, cannot be bounded. This is a concrete case of a co-vector on an infinite-dimensional vector space that cannot be written as a finite linear combination of dual-basis co-vectors.

Exercise 2.56: Although the $\boldsymbol{\beta}^i$ do not define a basis on V^* when V is infinite-dimensional, show that they "almost" do in the sense that the only $\boldsymbol{\omega} \in V^*$ with *all* $\omega_i = 0$ is the trivial, zero co-vector.

Exercise 2.57: When V is infinite-dimensional, it is of course naturally isomorphic to the *free* vector space over any one of its basis sets B. Show that V^* is naturally identified with the vector space of *all* functions over B. Explain why, although the free vector space over any set is naturally a subspace of the vector space of all functions over that set, this does not mean that V is naturally isomorphic to some subspace of V^* in the infinite-dimensional case.

The discussion above suggests that the dual-space V^* to an infinite-dimensional vector space V is *larger* than V in some intuitive sense. Since infinite is infinite, any literal interpretation of this idea is nonsense. However, we can make the notion somewhat more precise by considering the ***double dual*** space V^{**}. This is the dual space of the dual space to V. There is a *natural* map $\mu : V \to V^{**}$ defined by

$$[\mu(\mathbf{v})](\boldsymbol{\omega}) := \boldsymbol{\omega}(\mathbf{v}). \tag{2.58}$$ {vsDuDuMap}

It is straightforward to show that this map is linear, but that it maps onto *all* of V^{**} if and only if V is finite-dimensional. Unlike the maps from V to V^* defined above, which depend on one's choice of basis in V, this map is perfectly well-defined and requires no basis. This is quite evident in its definition (2.58). Thus, the failure of μ to give every vector in V^{**} when V is infinite-dimensional gives clear meaning to the intuitive notion that "taking duals enlarges an infinite-dimensional vector space." This intuition is perhaps the clearest expression of the distinction between finite- and infinite-dimensional spaces.

Exercise 2.59: Show that $\mu : V \to V^{**}$ is always a monomorphism, but is an isomorphism if and only if V is finite-dimensional.

Hint: When V is infinite-dimensional, pick an element $\boldsymbol{\omega} \in V^*$ corresponding to an unbounded function on the set $B \subset V$. Extend the subspace $F \subset V^*$ spanned by B^* to a subspace $U \subset V^*$ complementary to the one-dimensional subspace spanned by $\boldsymbol{\omega}$. Define a co-co-vector $\boldsymbol{\eta} \in V^{**}$ by $\boldsymbol{\eta}|_U = 0$ and $\boldsymbol{\eta}(\boldsymbol{\omega}) = 1$. Can $\boldsymbol{\eta} \in \operatorname{im}\mu$?

Comment 2.60: In infinite-dimensional *topological* vector spaces, one often imposes a condition of *continuity* on elements of the dual space in addition to linearity. The result is called the *topological* dual space, as opposed to the *algebraic* dual space constructed here. The difference can be profound, as the topological dual is often *iso*morphic to the space in question, though not naturally so for the reasons discussed in the exercises above. The topological double dual, accordingly, can often be identified with the vector space itself. A familiar example of this is the case of a Hilbert space, which actually is isomorphic to its dual, and therefore to its double dual, by the Riesz lemma. Recall that this lemma can be proved only using detailed properties of the Hilbert inner product, which induces the natural topology of Hilbert space. This more "finite-dimensional" behavior of the dual explains why topological vector spaces are so useful in functional analysis and, by extension, quantum theory. When vector spaces get bigger every time you take a dual, things can quickly get out of control.

This will be our last discussion of infinite-dimensional vector spaces. Throughout the remainder of this lecture, and the remainder of this course, we will assume all vector spaces to be finite-dimensional. We will attempt to note all results below that hold only under the assumption of finite dimensions, but do not guarantee that this has been done perfectly. Interested readers might try to figure out for themselves which results continue to hold in the infinite-dimensional case.

2.2.2 Dual Mappings

Let $\varphi : V \to W$ be a linear map. Given a co-vector $\boldsymbol{\omega} \in W^*$, we can define a co-vector $\varphi^*(\boldsymbol{\omega}) \in V^*$ by

$$[\varphi^*(\boldsymbol{\omega})](\mathbf{v}) := \boldsymbol{\omega}(\varphi(\mathbf{v})). \tag{2.61}$$

On the right side here, we map our given vector in V to a vector in W, and then evaluate the function $\boldsymbol{\omega}$ on that vector to give a number. Since any composition of linear maps is linear, the resulting scalar-valued function on V is linear, and therefore in V^*. Thus, associated to any $\varphi : V \to W$, we have a natural associated map $\varphi^* : W^* \to V^*$, called the ***dual map*** to φ or sometimes its ***transpose***. The latter name, of course, is borrowed from matrix theory. Note that the direction of the map is reversed in this association, which explains why φ^* is also sometimes called the ***pull-back map*** for φ and accordingly $\varphi^*(\boldsymbol{\omega})$ is called the ***pull-back*** of $\boldsymbol{\omega}$ under φ.

Exercise 2.62: Show that $\varphi^* : W^* \to V^*$ is linear.

Exercise 2.63: Let $\varphi : U \to V$ and $\psi : V \to W$ be linear. Show that $\psi^* \circ \varphi^* : W^* \to U^*$ is well-defined, and equal to $(\varphi \circ \psi)^*$.

Exercise 2.64: Show that φ^* is a monomorphism if and only if φ is an epimorphism. Conversely, show that φ^* is an epimorphism if and only if φ is a monomorphism. Use your results to show, in particular, that φ^* is an isomorphism if and only if φ is.

Exercise 2.65: When V and W are finite-dimensional and $\varphi : V \to W$ is linear, show that $\varphi^{**} = \varphi$.

The dual map φ^* defines its own kernel and image subspaces in W^* and V^*, respectively. Since φ^* is completely determined by φ, it is reasonable to associate these directly with the original map. When doing so, we refer to the ***co-kernel*** $\operatorname{coker}\varphi := \ker\varphi^* \subset W^*$ and ***co-image*** $\operatorname{coim}\varphi := \operatorname{im}\varphi^* \subset V^*$, respectively. Note that these are subspaces of the dual spaces to those related by the linear map $\varphi : V \to W$.

Exercise 2.66: In ordinary linear algebra, the co-kernel of a map $\varphi : V \to W$ between finite-dimensional vector spaces is often defined as the subspace of W *orthogonal* to the image subspace $\operatorname{im}\varphi \subset W$. The identification of the co-kernel subspace in W^* defined here with the conventional one in W can only be made with the help of the inner product usually assumed, which notably also defines the critical notion of orthogonality used in the conventional definition. Show nonetheless that a similar relation between the image and co-kernel obtains in our more general setting. Namely, show that $\mathbf{w} \in \operatorname{im}\varphi \subset W$ and $\boldsymbol{\omega} \in \operatorname{coker}\varphi \subset W^*$ imply $\boldsymbol{\omega}(\mathbf{w}) = 0$. Moreover, show that every $\boldsymbol{\omega} \in W^*$ such that $\boldsymbol{\omega}(\mathbf{w}) = 0$ for all $\mathbf{w} \in \operatorname{im}\varphi$ necessarily lies in $\operatorname{coker}\varphi$.

Exercise 2.67: Repeat the previous exercise, showing that $\boldsymbol{\eta} \in V^*$ satisfies $\boldsymbol{\eta}(\mathbf{v}) = 0$ for all $\mathbf{v} \in \ker\varphi \subset V$ if and only if $\boldsymbol{\eta} \in \operatorname{coim}\varphi$.

Exercise 2.68: The previous two exercises have introduced an important new concept, which is quite general. Given a subspace $U \subset V$, define the set $U^\perp$ of co-vectors $\boldsymbol{\eta} \in V^*$ such that $\boldsymbol{\eta}(\mathbf{u}) = 0$ for all $\mathbf{u} \in U$. Show that $U^\perp$ is a subspace of V^*. We refer to it as the ***normal subspace*** to U, and to the co-vectors it contains as ***normal*** to U. Thus, the previous exercises show that $\operatorname{coker}\varphi$ and $\operatorname{coim}\varphi$ are precisely the normal subspaces to $\operatorname{im}\varphi$ and $\ker\varphi$, respectively.

Exercise 2.69: If U is a subspace of a vector space V, show that $(V/U)^*$ is naturally isomorphic to $U^\perp$.

Exercise 2.70: When $U \subset V$ is a subspace of a finite-dimensional vector space, show that $U^{\perp\perp} = U$.

2.3 TENSOR ALGEBRA

There are a couple points in this lecture that may be new even to students with prior understanding of the mathematical theory of abstract vector spaces. Certainly, several perhaps unfamiliar applications to classical and quantum physics are presented. However, the most important sections for mathematically sophisticated students are likely those concerning Penrose's abstract index notation for tensors and, to a lesser extent, tensor densities. This notation serves as a convenient bridge between the traditional physicists' notation emphasizing components for calculational simplicity and the traditional mathematicians' notation emphasizing covariance. Also, the section on subspaces contains a slightly breathless attempt at motivating the introduction of all the abstract mathematical machinery described in this course into a physics curriculum.

2.4 TENSORS

Let $\varphi : V \to W$ be an arbitrary linear map, and pick a basis $\{\mathbf{b}_i\}$ on V. Then, the action of φ on an arbitrary vector $\mathbf{v} \in V$ can be written in the form

$$\varphi(\mathbf{v}) = \varphi\left(\sum_i v^i\, \mathbf{b}_i\right) = \sum_i v^i\, \varphi(\mathbf{b}_i) = \sum_i \boldsymbol{\beta}^i(\mathbf{v})\, \varphi(\mathbf{b}_i). \tag{2.71}$$ {vsTProdDec}

The action of φ on $\mathbf{v}$ is a linear combination of the vectors $\varphi(\mathbf{b}_i)$, in which the scalar coefficients are given by the action of the co-vectors $\boldsymbol{\beta}^i$ on $\mathbf{v}$. Another way of looking at this expression is to "factor out" the argument $\mathbf{v}$:

$$\varphi = \sum_i \mathbf{w}_i \otimes \boldsymbol{\beta}^i, \tag{2.72}$$ {vsTProd}

where we have defined $\mathbf{w}_i := \varphi(\mathbf{b}_i)$. The "$\otimes$" notation here is meant to emphasize the difference between the product of vectors and co-vectors contemplated here, which yields a linear operator, and more familiar products such as the natural action of a co-vector on a vector to yield a scalar. Indeed, the latter interpretation would make no sense here whatsoever since $\mathbf{w}_i$ is a vector in W, while $\boldsymbol{\beta}_i$ is a co-vector on V. A typical term in the sum (2.72) is called the ***tensor product*** of the vector $\mathbf{w}_i$ with the dual-basis element $\boldsymbol{\beta}_i$.

Exercise 2.73: Let $\varphi : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear map defined by the matrix

$$\varphi = \begin{pmatrix} 3 & -2 & 0 \\ 1 & 1 & -3 \end{pmatrix}. \tag{2.73a}$$

Using the obvious bases in $\mathbb{R}^3$ and $\mathbb{R}^2$, decompose this matrix into a sum of tensor products of the form (2.72). Give a matrix interpretation of the tensor product using column-vectors to denote elements of $\mathbb{R}^2$ and row-vectors to denote co-vectors on $\mathbb{R}^3$.

The tensor product easily generalizes to general vectors in W and co-vectors in V:

$$[\mathbf{w} \otimes \boldsymbol{\eta}](\mathbf{v}) := \boldsymbol{\eta}(\mathbf{v})\, \mathbf{w}. \tag{2.74}$$ {vsTPrSimp}

That is, the action of the tensor product on an arbitrary vector $\mathbf{v} \in V$ gives the vector $\mathbf{w}$, scaled by the value of $\boldsymbol{\eta}(\mathbf{v})$. Since $\boldsymbol{\eta}$ is linear, the tensor product must be so as well. As a linear map from V to W, the tensor product $\mathbf{w} \otimes \boldsymbol{\eta}$ has the unusual property that its image subspace in W is one-dimensional; every vector in it is proportional to $\mathbf{w}$. Certainly, not all linear maps $\varphi : V \to W$ have this property, and accordingly those which do are called ***simple***. The discussion above shows that *every* linear map $\varphi : V \to W$ can be written as a linear combination[1] of simple maps. This is a very important result, which highlights the central importance of the dual space in linear algebra. The point is that the complete

[1] Strictly speaking, this is true only in finite dimensions. Remember that "linear combination" really means "*finite* linear combination." While the expression (2.72) is formally linear in the case of an infinite-dimensional vector space, there is no guarantee that it is finite.

structure of an arbitrary space $\mathrm{Hom}(V,W)$ of homeomorphisms can be recovered from the vector space W and the dual space V^*, both of which are easy to analyze separately. The space of linear combinations of simple tensor products (2.74) is called the ***tensor product space*** $W \otimes V^*$. By (2.72), at least in finite dimensions, this space is naturally isomorphic to $\mathrm{Hom}(V,W)$.

Exercise 2.75: Show that the individual factors $\mathbf{w}$ and $\boldsymbol{\eta}$ in a given simple tensor product $\mathbf{w} \otimes \boldsymbol{\eta}$ are not unique.
Hint: What happens if you scale both arguments to the product in (2.74)?

Exercise 2.76: If $\boldsymbol{\eta} \in V^*$ is held fixed, show that the tensor product $\mathbf{w} \otimes \boldsymbol{\eta}$ defines a linear map from W to $\mathrm{Hom}(V,W)$. Show also that $\mathbf{w} \otimes \boldsymbol{\eta}$ defines a linear map from V^* to $\mathrm{Hom}(V,W)$ if $\mathbf{w} \in W$ is held fixed.

Exercise 2.77: The previous exercise has proved the identities

$$(\alpha\mathbf{w} + \alpha'\mathbf{w}') \otimes \boldsymbol{\eta} = \alpha(\mathbf{w} \otimes \boldsymbol{\eta}) + \alpha'(\mathbf{w}' \otimes \boldsymbol{\eta}) \qquad \text{and} \qquad \mathbf{w} \otimes (\beta\boldsymbol{\eta} + \beta'\boldsymbol{\eta}') = \beta(\mathbf{w} \otimes \boldsymbol{\eta}) + \beta'(\mathbf{w} \otimes \boldsymbol{\eta}') \tag{2.77a}$$ {vsTPrRel}

on the tensor-product space $W \otimes V^*$. These relations identify linear combinations of simple tensor products with other simple tensor products whenever one factor or the other is common to the two products being combined. Show that these are, in fact, the *only* linear relations on the tensor-product space. That is, show that

$$\mathbf{w} \otimes \boldsymbol{\eta} + \mathbf{w}' \otimes \boldsymbol{\eta}' = \mathbf{w}'' \otimes \boldsymbol{\eta}'' \tag{2.77b}$$

implies either that $\mathbf{w}' \propto \mathbf{w}$ or $\boldsymbol{\eta}' \propto \boldsymbol{\eta}$.

Exercise 2.78: If V and W have finite dimensions m and n, respectively, find the dimension of $W \otimes V^*$.
Hint: Construct a basis for the tensor product space from bases for W and V^*.

The discussion above gives one motivation for the definition of the tensor product. It is not the only one, however. There are many instances in physics where it is natural to consider a ***multi-linear map***. Such maps take collections of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ in (possibly distinct) vector spaces $V_1, \ldots, V_r$ as arguments and result in vectors $\mathbf{w}$ in another vector space W. We will see many examples of this sort of object below. Inserting a linear combination of vectors in any single argument of a mutli-linear map $\boldsymbol{T}$ gives the same linear combination of results:

$$\boldsymbol{T}(\mathbf{v}_1, \cdots, \alpha\mathbf{v}_i + \alpha'\mathbf{v}_i', \cdots, \mathbf{v}_r) = \alpha\,\boldsymbol{T}(\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_r) + \alpha'\,\boldsymbol{T}(\mathbf{v}_1, \cdots, \mathbf{v}_i', \cdots, \mathbf{v}_r), \tag{2.79}$$ {vsMLMap}

Note that all *other* arguments on the right side here are common to both terms. This is strongly reminiscent of the relations (2.77a) that define the tensor product: a linear combination of simple tensor products is itself simple only if they have factors in common. Therefore, we generalize the tensor product to arbitrary finite sets of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_r$ in arbitrary vector spaces $V_1, \cdots, V_r$, with the relations

$$\mathbf{v}_1 \otimes \cdots \otimes (\alpha\mathbf{v}_i + \alpha'\mathbf{v}_i') \otimes \cdots \otimes \mathbf{v}_r = \alpha(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{v}_r) + \alpha'(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_i' \otimes \cdots \otimes \mathbf{v}_r). \tag{2.80}$$ {vsTPrRel'}

With this definition, it is possible to view the *multi*-linear map $\boldsymbol{T}$ of (2.79) as a *linear* map from the tensor product space $V_1 \otimes \cdots \otimes V_r$ to W. The definition of the tensor product is chosen precisely so that multi-linear maps can be analyzed directly using previous results for linear maps. A multi-linear map is often called a ***tensor***.

Exercise 2.81: Show that any linear map $\boldsymbol{T} \in \mathrm{Hom}(U \otimes V, W)$ does indeed define a bi-linear map from the Cartesian product $U \times V$ to W.

Exercise 2.82: For any vector space V, show that there is a preferred linear map $\boldsymbol{\delta} : V \otimes V^* \to \mathbb{F}$. You may want to proceed as follows. First consider simple tensor products $\mathbf{v} \otimes \boldsymbol{\omega}$. There is only one natural action on such objects to produce a scalar. Show that it is bi-linear in $\mathbf{v}$ and $\boldsymbol{\omega}$, and thus that it extends to a linear map on $V \otimes V^*$. General elements of this tensor product are automorphisms of V. Viewing these a square matrices, what is the meaning of $\boldsymbol{\delta}$?

Exercise 2.83: Is there any natural monomorphism from V to $V \otimes W$?

2.5 ABSTRACT INDEX NOTATION

In our discussion of the dual space, we showed that any linear map $\varphi : V \to W$ naturally gives rise to a dual map $\varphi^* : W^* \to V^*$. In finite dimensions, there is a one-to-one correspondence $\varphi \leftrightarrow \varphi^*$ between maps and dual maps. In fact, we can go further and define yet another map $\bar{\varphi} : V \otimes W^* \to \mathbb{F}$. When acting on a simple tensor product, this map is defined by

$$\bar{\varphi}(\mathbf{v} \otimes \boldsymbol{\eta}) := \boldsymbol{\eta}(\varphi(\mathbf{v})). \tag{2.84}$$

The function $\bar{\varphi}$ is obviously bi-linear in $\mathbf{v}$ and $\boldsymbol{\eta}$, and therefore extends linearly to a map on the tensor-product space $V \otimes W^*$. Thus, the original linear map $\varphi : V \to W$ gives rise to two ancillary maps $\varphi^* : W^* \to V^*$ and $\bar{\varphi} : V \otimes W^* \to \mathbb{F}$.

Exercise 2.85: For finite-dimensional vector spaces V and W, show that any linear map $\bar{\varphi} : V \otimes W^* \to \mathbb{F}$ naturally gives rise to maps $\varphi : V \to W$ and $\varphi^* : W^* \to V^*$.
Hint: You will need to make use of the double-dual property $W^{**} = W$ of finite-dimensional vector spaces.

The discussion and exercise above have shown that there are many equivalent ways to define a linear map from one vector space to another, at least in finite dimensions. The situation is actually a bit worse with tensors. A tensor $\boldsymbol{T} : U \otimes V \to W$ can also be viewed fixed typos — CB as a linear map taking $U \otimes V \otimes W^* \to \mathbb{F}$, or $U \to V^* \otimes W$, or even $\mathbb{F} \to U^* \otimes V^* \otimes W$, among other possibilities. This proliferation of ancillary maps becomes more pronounced as the number of arguments to a tensor increases. However, the problem is entirely fictitious. Any one of the many linear maps associated with a given tensor suffices to define that tensor, and all the other maps associated to it. Penrose's ***abstract index notation*** uses a single, unified expression to denote all of these maps. The idea is to use indices to indicate how a given tensor acts on its arguments to produce a result in some vector space. Let's construct a short lexicon translating the notation used previously into abstract-index form

Vectors in the abstract index notation are denoted with a single superscript, while co-vectors get a single subscript:

$$\mathbf{v} \rightsquigarrow v^a \qquad \text{and} \qquad \boldsymbol{\omega} \rightsquigarrow \omega_a. \tag{2.86}$$ {ainVC}

Generally, we forego changing the font of the stem symbol in the abstract-index approach since the indices already indicate the geometric nature of a given tensor. The natural action of co-vectors on vectors is denoted with a repeated index:

$$\boldsymbol{\omega}(\mathbf{v}) \rightsquigarrow \omega_a \, v^a. \tag{2.87}$$ {ainVCcon}

The result here is a scalar. Since this scalar has no action on another vector $\tilde{\mathbf{v}}$ other than simple scaling, it would be confusing to use the same index to denote different vectors: should "$\omega_a \, v^a \, \tilde{v}^a$" mean "$\boldsymbol{\omega}(\mathbf{v}) \, \tilde{\mathbf{v}}$" or "$\boldsymbol{\omega}(\tilde{\mathbf{v}}) \, \mathbf{v}$"? Therefore, we attach *different* indices to denote different vectors in any calculation:

$$\boldsymbol{\omega}(\mathbf{v}) \, \tilde{\mathbf{v}} \rightsquigarrow \omega_a \, v^a \, \tilde{v}^b \qquad \text{and} \qquad \boldsymbol{\omega}(\tilde{\mathbf{v}}) \, \mathbf{v} \rightsquigarrow \omega_a \, \tilde{v}^a \, v^b. \tag{2.88}$$ {ainCVVdis}

One can think of these indices as labeling distinct copies $(V^a, \, V^b, \, \ldots)$ of the same vector space V. In any calculation, only one vector can ever be drawn from any one copy of a given space. Thus, for example, for $\mathbf{v}, \tilde{\mathbf{v}} \in V$ we would write

$$\mathbf{v} \otimes \tilde{\mathbf{v}} \rightsquigarrow v^a \, \tilde{v}^b. \tag{2.89}$$ {ainVVstp}

The abstract indices here cannot be the same since, for example, a co-vector $\boldsymbol{\omega} \in V^*$ has two natural actions on the tensor product $V \otimes V$: one where it acts on the first factor of a simple tensor product $\mathbf{v} \otimes \tilde{\mathbf{v}}$ to produce a scalar multiple of $\tilde{\mathbf{v}}$, and the other where it acts on the second factor to produce a scalar multiple of $\mathbf{v}$. These are just the two actions defined by (2.88). Both extend by linearity to actions on non-simple elements of the tensor product space, which also are denoted with two distinct indices:

$$\boldsymbol{B} \in V \otimes V \rightsquigarrow B^{ab}. \tag{2.90}$$ {ainVVgtp}

The abstract index notation for the two distinct natural actions of $\boldsymbol{\omega} \in V^*$ on elements of $V \otimes V$ is much simpler than one consistent with the conventions we have been using above:

$$\begin{aligned} \overset{1}{\boldsymbol{\omega}}(\mathbf{v} \otimes \tilde{\mathbf{v}}) &:= \boldsymbol{\omega}(\mathbf{v}) \, \tilde{\mathbf{v}} \quad \Rightarrow \quad \overset{1}{\boldsymbol{\omega}} : V \otimes V \to V \qquad \rightsquigarrow \qquad \omega_a \, B^{ab} \\ \overset{2}{\boldsymbol{\omega}}(\mathbf{v} \otimes \tilde{\mathbf{v}}) &:= \boldsymbol{\omega}(\tilde{\mathbf{v}}) \, \mathbf{v} \quad \Rightarrow \quad \overset{2}{\boldsymbol{\omega}} : V \otimes V \to V \qquad \rightsquigarrow \qquad \omega_b \, B^{ab}. \end{aligned} \tag{2.91}$$ {ainCVVrde}

The relative simplicity of the latter is the *raison d'être* of the abstract index notation.

Exercise 2.92: Write down the definition (2.74) of a simple tensor product in the abstract index notation. Check that the notation is internally consistent.

Exercise 2.93: A linear map $\varphi : V \to W$ is denoted $\varphi^a{}_b : V^b \to W^a$ in the abstract index notation. Write down abstract-index expressions for

a. the original action $\varphi(\mathbf{v})$ of φ on a vector $\mathbf{v} \in V$,

b. the dual action $\varphi^*(\boldsymbol{\eta})$ on a co-vector $\boldsymbol{\eta} \in W^*$,

c. the scalar action $\bar{\varphi}(\mathbf{v} \otimes \boldsymbol{\eta})$ on a simple tensor product in $V \otimes W^*$, and

d. the extension of that scalar action to an arbitrary element $T \in V \otimes W^*$.

Check that your notations are internally consistent.

Exercise 2.94: The abstract index notation draws various vectors in a given calculation from distinct copies (V^a, V^b, ...) of a given vector space V. Since these are all copies of the same space, however, they all must be canonically isomorphic to one another. Denote the canonical isomorphism by

$$\delta^a_b : V^b \to V^a \qquad \text{with} \qquad \delta^a_b\, v^b := v^a. \tag{2.94a}$$ {ainDelta}

Show that this map of vector spaces is naturally associated with the map $\boldsymbol{\delta} : V \otimes V^* \to \mathbb{F}$ discussed in an exercise in the previous section.

Exercise 2.95: Let $\varphi : V \to W$ be a linear mapping of vector spaces. Use the abstract index notation to define a natural associated mapping from $V \otimes V$ to $W \otimes W$. Describe its action on both simple and general tensor products.

2.5.1 Symmetry and Anti-Symmetry

If a tensor is defined in a product $V \otimes W$ of distinct vector spaces, then the abstract-index convention is that the order of indices does not matter. To emphasize this point, indices in different spaces are often drawn from different alphabets, $T^{aA} \in V^a \otimes W^A$, and their order is irrelevant, so $T^{aA} = T^{Aa}$. For tensors in spaces like $V \otimes V$, however, the order of indices generally *does* matter quite a bit. For example, the actions (2.91) of a covector $\boldsymbol{\omega}$ on an element $\boldsymbol{B} \in V \otimes V$ will not generally agree unless $\boldsymbol{B}$ is ***symmetric***. In the abstract-index notation, the condition of symmetry has the form

$$B^{ab} = B^{ba}. \tag{2.96}$$ {ainSymm}

Of course, not all tensors in $V \otimes V$ are symmetric. Some are ***anti-symmetric***,

$$B^{ab} = -B^{ba}, \tag{2.97}$$ {ainAnti}

but a general tensor in $V \otimes V$ has no particular symmetry.

The spaces of symmetric and anti-symmetric tensors are particularly useful in practice, and sometimes are given their own notations:

$$V \odot V := \left\{ \boldsymbol{S} \in V \otimes V \;\middle|\; S^{ba} = S^{ab} \right\} \quad \text{and} \quad V \wedge V := \left\{ \boldsymbol{A} \in V \otimes V \;\middle|\; A^{ba} = -A^{ab} \right\}. \tag{2.98}$$ {vsSAprod}

Each of these is a subspace of the full tensor product space $V \otimes V$. Given a pair of vectors $\mathbf{v}, \tilde{\mathbf{v}} \in V$, we can construct symmetric and anti-symmetric tensors in these two subspaces:

$$\mathbf{v} \odot \tilde{\mathbf{v}} \rightsquigarrow v^{(a}\, \tilde{v}^{b)} := \tfrac{1}{2}\left(v^a\, \tilde{v}^b + v^b\, \tilde{v}^a \right) \qquad \text{and} \qquad \mathbf{v} \wedge \tilde{\mathbf{v}} \rightsquigarrow v^{[a}\, \tilde{v}^{b]} := \tfrac{1}{2}\left(v^a\, \tilde{v}^b - v^b\, \tilde{v}^a \right). \tag{2.99}$$ {ainSAprod}

The two tensors constructed here are referred to as the ***symmetric*** and ***anti-symmetric parts*** of the simple tensor product $\mathbf{v} \otimes \tilde{\mathbf{v}}$. Both definitions extend to general tensors:

$$B^{(ab)} := \tfrac{1}{2}\left(B^{ab} + B^{ba} \right) \qquad \text{and} \qquad B^{[ab]} := \tfrac{1}{2}\left(B^{ab} - B^{ba} \right). \tag{2.100}$$ {ainSApart}

Further properties of these constructions are worked out in the exercises below.

Exercise 2.101: Show that the tensor product space $V \otimes V$ naturally decomposes as a direct sum of $V \odot V$ and $V \wedge V$. That is, show that the spaces of symmetric and anti-symmetric tensors are complementary subspaces of the full tensor-product space. Show that there are natural projection operators onto these two subspaces, which can be written in the form

$$\sigma^{ab}_{cd} := \delta^{(a}_c\,\delta^{b)}_d := \tfrac{1}{2}\left(\delta^a_c\,\delta^b_d + \delta^b_c\,\delta^a_d\right) \qquad \text{and} \qquad \alpha^{ab}_{cd} := \delta^{[a}_c\,\delta^{b]}_d := \tfrac{1}{2}\left(\delta^a_c\,\delta^b_d - \delta^b_c\,\delta^a_d\right), \tag{2.101a}$$ {ainSAproj}

respectively.
Hint: Show that a general $B^{ab} = B^{(ab)} + B^{[ab]}$, and that the two subspaces intersect only in the zero tensor.

Let $T^{a_1\cdots a_p}$ be a tensor in the p-fold tensor product $\otimes^p V$. The ***totally symmetric*** and ***totally anti-symmetric parts*** of this tensor are defined by

$$T^{(a_1\cdots a_p)} := \frac{1}{p!}\sum_{\sigma\in S_p} T^{a_{\sigma(1)}\cdots a_{\sigma(p)}} \qquad \text{and} \qquad T^{[a_1\cdots a_p]} := \frac{1}{p!}\sum_{\sigma\in S_p} (-1)^{|\sigma|}\, T^{a_{\sigma(1)}\cdots a_{\sigma(p)}}, \tag{2.102}$$ {ainTSApart}

respectively. Here, $\sigma \in S_p$ is an element of the permutation group on p symbols, and $|\sigma|$ denotes the number of elementary permutations needed to convert a given σ to the trivial permutation $\sigma(i) = i$. The role of $|\sigma|$ in (2.102) is merely to ensure odd permutations of indices are included in the sum for the totally anti-symmetric part with a minus sign. A given tensor $T^{a_1\cdots a_p}$ itself is called ***totally symmetric*** or ***totally anti-symmetric*** if it is equal to its own totally symmetric or totally anti-symmetric part. The spaces of such tensors are denoted $\odot^p V$ and $\wedge^p V$, respectively.

Exercise 2.103: Write down explicit expressions for $T^{(abc)}$ and $T^{[abc]}$ for a general three-index tensor T^{abc}. Suppose now that T^{abc} is already anti-symmetric in its first two indices, so $T^{abc} = T^{[ab]c}$. How do your explicit expressions simplify?

Exercise 2.104: The curvature tensor in Riemannian geometry has the symmetries

$$R_{abcd} = R_{[ab][cd]} \qquad \text{and} \qquad R_{[abc]d} = 0. \tag{2.104a}$$ {ainRieSymA}

Show that this is equivalent to asserting

$$R_{abcd} = R_{[ab][cd]} = R_{cdab} \qquad \text{and} \qquad R_{[abcd]} = 0. \tag{2.104b}$$ {ainRieSymB}

That is, show that the symmetries (2.104a) imply (2.104b) and *vice versa*.

Exercise 2.105: Show that the subspaces $\odot^p V$ and $\wedge^p V$ of $\otimes^p V$ are disjoint in the sense that they intersect only in the zero tensor. Are they complementary?

2.5.2 Concrete Indices

One reason that the abstract index notation is not favored by pure mathematicians, but *is* favored by applied mathematicians and physicists, is its similarity to an older notation once used by both. The previous notation represented vectors and tensors via their components relative to some basis, which could vary, and labeled those components with indices. Tensors were defined in terms of the laws relating their components in different bases. This is needlessly complicated, which is why the component notation has mostly fallen out of favor. Nonetheless, when one actually needs to calculate a number, such as the deflection of a light ray from a distant star by our sun, tensor components are one of the best tools available. The similarity of the abstract index notation to the component notation should therefore be construed as a feature, not a flaw.

A basis on an n-dimensional vector space V is denoted in the abstract index notation by $B = \{b^a_\alpha\}_{\alpha=1\cdots n}$. Here, a is an abstract index indicating that each b^a_α belongs to the same copy V^a of the vector space, while α is a ***concrete index*** taking n distinct integer values. The expansion (1.17a) of a vector in the abstract index notation takes the form

$$v^a = \sum_\alpha v^\alpha \, b^a_\alpha =: v^\alpha \, b^a_\alpha. \tag{2.106}$$ {ainVbas}

The second expression here suppresses the summation symbol for α. This is a convenient shorthand, often adopted when many such sums are involved, called the ***Einstein summation convention***. The similarity to an abstract-index expression is striking, of course. However, we will attempt to be consistent in drawing abstract and concrete indices from different alphabets, respectively lowercase latin and greek here. Thus, $v^a \in V^a$ represents an abstract vector in a given copy of the vector space V, while $v^\alpha \in \mathbb{F}^n$ represents its component scalars in a given basis.

The abstract-index notation for the dual basis differs slightly from that used previously, employing the same stem letter for both bases. A typical dual-basis element is b^α_a, rather than β^α_a as the previous notation would suggest, and (2.45) becomes

$$b^\alpha_a \, v^a := v^\alpha. \tag{2.107}$$ {ainDbas}

The possibility of distinguishing basis and dual-basis elements merely by the placement of indices allows us to conserve notation, which often becomes an issue in serious calculations.

Exercise 2.108: By the definition of a basis, show that we must have

$$b^\alpha_a \, b^a_\beta = \delta^\alpha_\beta := \begin{cases} 1 & \alpha = \beta \\ 0 & \alpha \neq \beta. \end{cases} \tag{2.108a}$$ {ainBasKr}

The $n \times n$ array of numbers δ^α_β appearing here is the familiar ***Kronecker symbol***.

Exercise 2.109: Since (2.106) and (2.107) must hold for *all* abstract vectors v^a, show that

$$b^a_\alpha \, b^\alpha_b = \delta^a_b, \tag{2.109a}$$ {ainBasId}

where the Einstein convention has been employed. Note that the right side here is *not* the Kronecker symbol, but the abstract identity transformation on V.

The ***components*** of a tensor $T^{a_1\cdots a_p}{}_{b_1\cdots b_q}$ are defined by its actions on the appropriate basis and dual-basis elements:

$$T^{\alpha_1\cdots\alpha_p}{}_{\beta_1\cdots\beta_q} := T^{a_1\cdots a_p}{}_{b_1\cdots b_q} \, b^{\alpha_1}_{a_1} \cdots b^{\alpha_p}_{a_p} \, b^{b_1}_{\beta_1} \cdots b^{b_q}_{\beta_q}. \tag{2.110}$$ {ainTBas}

For a tensor with no particular symmetries, each concrete index here takes n distinct values. Thus, a tensor in the space $(\otimes^p V) \otimes (\otimes^q V^*)$ considered here is defined by its n^{p+q} scalar components. Equivalently, we can view the various strings of basis elements contracted with T on the right side of (2.110) as a natural basis on the tensor space in question. Either way, we conclude that that tensor space has dimension n^{p+q}. If T does have symmetries, then the problem of counting its components can become very difficult indeed. Of course, the hardest math problems generally involve counting.

Exercise 2.111: Write down the formula for the abstract tensor T in (2.110) in terms of its components.

Exercise 2.112: Two bases $\{b^a_\alpha\}$ and $\{\tilde{b}^a_\alpha\}$ in the abstract index notation are joined by the entries in the change-of-basis matrix λ according to

$$\tilde{b}^a_\alpha = \lambda^\beta{}_\alpha\, b^a_\beta, \tag{2.112a}$$ {ainBasBas}

using the Einstein convention. Write down similar expressions for the associated changes in

a. the dual basis $\{b^\alpha_a\}$,

b. the components v^α of a vector v^a,

c. the components ω_α of a co-vector ω_a, and

d. the components $T^{\alpha_1\cdots\alpha_p}{}_{\beta_1\cdots\beta_q}$ of a tensor $T^{a_1\cdots a_p}{}_{b_1\cdots b_q}$.

Note the difference in the action of the array $\lambda^\beta{}_\alpha$ on superscript and subscript concrete indices. For this reason, the former are often called ***contravariant***, and the latter ***covariant***. This nomenclature has also been taken over to the abstract-index formalism.

Exercise 2.113: Let V be a vector space with finite dimension n, and consider a general $\boldsymbol{B} \in V \otimes V$. Show that

$$B^{ab} = B^{ba} \quad \Leftrightarrow \quad B^{ab} = B^{(ab)} \quad \Leftrightarrow \quad B^{[ab]} = 0 \tag{2.113a}$$ {ainSA0}

and

$$B^{ab} = -B^{ba} \quad \Leftrightarrow \quad B^{ab} = B^{[ab]} \quad \Leftrightarrow \quad B^{(ab)} = 0. \tag{2.113b}$$ {ainAS0}

Show explicitly that $V \otimes V$ has dimension n^2, while $V \odot V$ has dimension $n(n+1)/2$ and $V \wedge V$ has dimension $n(n-1)/2$.

Exercise 2.114: For an n-dimensional vector space V, show that the associated space $\odot^p V$ of totally symmetric tensors has dimension $(n+p-1)!/p!\,(n-1)!$.

Solution: If b^a_α is a basis for V, then a typical element of the associated basis on the space of totally symmetric tensors is $b^{(a_1}_{\alpha_1} \cdots b^{a_p)}_{\alpha_p}$. The symmetrization acts on all p abstract indices in the tensor product. Because of this symmetrization, many of the resulting basis tensors are the same. In fact, any permutation of a given set of values for concrete indices α_i will result in the same basis tensor; the expression is totally symmetric in the concrete indices as well as the abstract. Thus, all that matters is *how many times* a given concrete index value $k = 1 \cdots n$ arises in the list of concrete indices α_i for a given basis element. Denote this number c_k. Since a total of p indices are needed, our problem is equivalent to asking: How many sets of n non-negative integers c_k are there that add up to exactly p? Denote this number D^p_n. Separating out the first possible index value $k = 1$, for which c_1 can take any value from 0 to p, we find the recursion relation

$$D^p_n = \sum_{q=0}^{p} D^q_{n-1}. \tag{2.114a}$$ {vsDpnRecur}

That is, for each value of $c_1 = p - q$, we add the number of ways the $n-1$ remaining c_k can sum to q.

Playing the game fairly, we would have to start in low dimensions $n = 1, 2, 3, \ldots$ and build up enough intuition to guess at the correct formula. But here, the correct answer is given above, and we need only check it. To do so, we must show that

$$\sum_{q=0}^{p} \frac{(n+q-2)!}{q!\,(n-2)!} = \frac{(n+p-1)!}{p!\,(n-1)!}. \tag{2.114b}$$ {vsDnpRecur}

We do this by induction on p. The relation is obviously true for $p = 0$ since both sides are unity. Then, assuming the relation holds for $p-1$, we have

$$\frac{(n+p-2)!}{(p-1)!\,(n-1)!} + \frac{(n+p-2)!}{p!\,(n-2)!} = \frac{(n+p-2)!}{p!\,(n-1)!}\,(p+n-1) = \frac{(n+p-1)!}{p!\,(n-1)!}, \tag{2.114c}$$ {vsDnpConf}

as required. This proves the result.

Exercise 2.115: Show that the dimension of $\wedge^p V$ is $n!/p!\,(n-p)!$ if $0 \le p \le n$, and contains no non-trivial tensors for $p > n$. What is the dimension of $\wedge^n V$?

Exercise 2.116: Show that the Riemann curvature tensor R_{abcd} in n dimensions has $n^2\,(n^2-1)/12$ independent components. Do this twice, counting once based on the symmetries (2.104a) and again based on (2.104b).

2.6 DETERMINANTS AND DENSITIES

Let v_i^a denote a collection of p vectors in an n-dimensional vector space V, and consider their totally anti-symmetric tensor product $v_1^{[a_1} \cdots v_p^{a_p]}$. Suppose these vectors are linearly dependent, and substitute for one v_j^a in terms of the other vectors. Then, expanding the sum, every term will include an antisymmetric combination of the form $v_i^{[a_i}\, v_i^{a_j]}$. Since the same vector is repeated twice, the anti-symmetry implies that each term will vanish, and hence the sum. Conversely, if the v_i^a are linearly *in*dependent, then we can build a basis b_α^a whose first p elements are the v_i^a. Then, one can check that the contraction of the tensor product $b^1_{a_1} \cdots b^p_{a_p}$ with $v_1^{[a_1} \cdots v_p^{a_p]}$ gives exactly $1/p!$. Thus, when the v_i^a are linearly independent, their totally anti-symmetric product cannot vanish. This demonstrates that

$$v_1^{[a_1} \cdots v_p^{a_p]} = 0 \qquad \Leftrightarrow \qquad \{v_i^a\} \text{ are linearly dependent.} \tag{2.117}$$ {vsAntiDep}

This gives a nice tensor-algebraic criterion for linear independence, and also hints at the general importance of anti-symmetric tensors to be discussed below.

Exercise 2.118: Convince yourself of the argument of the previous paragraph in detail.

Given a basis $B = \{b_\alpha^a\}$ on V, we can define

$$\eta_{(B)}^{a_1 \cdots a_n} := n!\, b_1^{[a_1} \cdots b_n^{a_n]}. \tag{2.119}$$ {vsLCtens}

The result here is a totally anti-symmetric tensor with the maximum possible number of indices. We refer to such tensors as ***alternating***. Now, one of the exercises in the previous section has shown that $\wedge^n V$ is one-dimensional. Thus, every non-zero alternating tensor is some scalar multiple of every other one and, in particular, the tensors (2.119) in two different bases must be proportional. Using (2.112a), we can write

$$\begin{aligned} \eta_{(\tilde{B})}^{a_1 \cdots a_n} &= n!\, \tilde{b}_1^{[a_1} \cdots \tilde{b}_n^{a_n]} = n!\, (\lambda^{\alpha_1}{}_1\, b_{\alpha_1}^{[a_1}) \cdots (\lambda^{\alpha_n}{}_n\, b_{\alpha_n}^{a_n]}) \\ &= \sum_{\sigma \in S_n} (-1)^{|\sigma|}\, \lambda^{\sigma(1)}{}_1 \cdots \lambda^{\sigma(n)}{}_n\, \eta_{(B)}^{a_1 \cdots a_n}. \end{aligned} \tag{2.120}$$ {vsLCreln}

In going from the first line to the second here, we have used the definition (2.102) of anti-symmetrization and our freedom to rename any of the α_i indices. The coefficient of the tensor (2.119) is immediately recognizable as the determinant $\det\lambda$ of the square matrix appearing in (1.24). These tensors allow us to define the *non-tensorial* ***Levi-Civita symbol*** whose *components* in any given basis B are those of the associated alternating tensor (2.119):

$$\widetilde{\eta}^{a_1\cdots a_n}\, b^{\alpha_1}_{a_1}\cdots b^{\alpha_n}_{a_n} := \eta^{a_1\cdots a_n}_{(B)}\, b^{\alpha_1}_{a_1}\cdots b^{\alpha_n}_{a_n}. \tag{2.121}$$ {vsLCsymb}

This object is not a tensor because $\eta^{a_1\cdot a_n}_{(B)}$ varies with the basis B. Another way to see this non-tensorial character is to consider the relation between the components of the Levi-Civita symbol taken in two different bases:

$$\begin{aligned}\widetilde{\eta}^{a_1\cdots a_n}\, \tilde{b}^{\alpha_1}_{a_1}\cdots \tilde{b}^{\alpha_n}_{a_n} := \eta^{a_1\cdots a_n}_{(\tilde{B})}\, \tilde{b}^{\alpha_1}_{a_1}\cdots \tilde{b}^{\alpha_n}_{a_n} &= (\det\lambda)\, \eta^{a_1\cdots a_n}_{(B)}\, (\tilde{\lambda}^{\alpha_1}{}_{\beta_1} b^{\beta_1}_{a_1})\cdots(\tilde{\lambda}^{\alpha_n}{}_{\beta_n} b^{\beta_n}_{a_n}) \\ &= (\det\lambda)\, \tilde{\lambda}^{\alpha_1}{}_{\beta_1}\cdots\tilde{\lambda}^{\alpha_n}{}_{\beta_n}\, (\widetilde{\eta}^{a_1\cdots a_n}\, b^{\beta_1}_{a_1}\cdots b^{\beta_n}_{a_n}).\end{aligned} \tag{2.122}$$ {vsLCtran}

If the Levi-Civita symbol were actually a tensor, of course, the leading determinant factor here would be absent. This factor is the defining characteristic of a ***tensor density*** of ***weight*** one. In general, a tensor density of weight w can be defined as an object whose components transform with a factor of $(\det\lambda)^w$ in addition to the usual matrix actions of λ and $\tilde{\lambda}$ on its component indices. We denote tensor densities of integer weight w in the abstract-index notation with the corresponding number of tilde accents.

Exercise 2.123: Show that all numerical components of the Levi-Civita symbol $\widetilde{\eta}^{\alpha_1\cdots\alpha_n}$ in a given basis are either ±1 or zero. Thus, the Levi-Civita symbol is the object whose components in any given basis form the standard alternating array. Argue that this is consistent with (2.122).
Hint: Show that the effect of the string of $\tilde{\lambda}^{\alpha_i}{}_{\beta_i}$ matrices here is to produce a factor of $\det\tilde{\lambda}$ multiplying a given numerical component of $\widetilde{\eta}^{a_1\cdots a_n}$, which cancels the $\det\lambda$ factor in (2.122).

This classical definition of a density given above, based on the transformation of its components under a change of basis, is a bit of a mess. The modern version does not refer to components, but is less obviously connected to ordinary tensors. A tensor density of weight w is as a *homogeneous* map of degree w from the space of alternating tensors $\Xi^{a_1\cdots a_n}$ to the space of tensors of a given type:

$$[D(c\,\Xi)]^{a_1\cdots a_p}{}_{b_1\cdots b_q} = c^w\,[D(\Xi)]^{a_1\cdots a_p}{}_{b_1\cdots b_q}. \tag{2.124}$$ {vsDensHom}

In words, the action of D on an alternating tensor Ξ produces a tensor, and scaling Ξ by c produces the same tensor scaled by c^w. The definition of the Levi-Civita density in this scheme could not be simpler:

$$[\widetilde{\eta}(\Xi)]^{a_1\cdots a_n} := \Xi^{a_1\cdots a_n}. \tag{2.125}$$ {vsLChom}

This is obviously a homogeneous map of degree one giving results in the space of alternating tensors, as the index structure of the Levi-Civita symbol in (2.121) would suggest. Seeing the connection between the two definitions, however, will require a bit of work.

As mentioned above, the space of alternating tensors is one-dimensional, whence one can take ratios of them:

$$\Xi/\Upsilon := c \qquad \text{where} \qquad \Xi = c\,\Upsilon. \tag{2.126}$$ {vsATquot}

That is, the ratio Ξ/Υ is the scalar that must multiply Υ to yield Ξ. This is the key to the relation between the two definitions of densities put forward above. Starting with a homogeneous map D of weight w, and picking a basis B, we define the tensor

$$T_{(B)}(D) := D(\eta_{(B)}). \tag{2.127}$$ {vsDensHomTen

We will suppress indices in this calculation for simplicity; the index structure of (2.124) should be understood throughout. If we change the basis, then of course this tensor will change, and we have

$$T_{(\tilde{B})}(D) := D(\eta_{(\tilde{B})}) = D((\det\lambda)\,\eta_{(B)}) = (\det\lambda)^w\,T_{(B)}(D), \tag{2.128}$$ {vsDensTran}

where we have used (2.120). Thus, the homogeneous map D gives rise to a family of tensors $T_{(B)}(D)$, one for each basis B, which are related to one another by a factor of $(\det\lambda)^w$. The components of a given tensor $T_{(B)}(D)$ transform tensorially under a change of basis, of course, but the additional scaling of the tensor itself leads to an overall transformation like (2.122) for the components of the density defined by this family of tensors. Thus, the modern definition reproduces the classical one. Going the other way is even easier. Given a family of tensors $T_{(B)}$ with the scaling property (2.128), we set

$$D(\Xi) := (\Xi/\eta_{(B)})^w\,T_{(B)}. \tag{2.129}$$ {vsDensTenHom

If the tensors $T_{(B)}$ are indeed related as in (2.127), then it is easy to show that this is indeed a homogeneous map of degree w on the space of alternating tensors.

Exercise 2.130: Show explicitly that the definitions (2.121) and (2.125) agree.

Exercise 2.131: Given a non-zero alternating tensor $\Xi^{a_1\cdots a_n}$, define its ***inverse*** $\check{\Xi}_{a_1\cdots a_n}$ by

$$\Xi^{a_1\cdots a_n}\,\check{\Xi}_{a_1\cdots a_n} = n!. \tag{2.131a}$$ {vsATinv}

Show that the inverse always exists and is unique. Write an abstract-index expression for Υ/Ξ in terms of the inverse tensor $\check{\Xi}_{a_1\cdots a_n}$.

Exercise 2.132: Define the inverse $\underset{\sim}{\eta}_{a_1\cdots a_n}$ of the Levi-Civita symbol $\widetilde{\eta}^{a_1\cdots a_n}$. Show that it is a density of weight -1 in the classical sense. How should it be defined as a homogeneous map? This object, by an abuse of language, is also often called the Levi-Civita symbol.

Exercise 2.133: Show that

$$\Xi^{a_1\cdots a_n}\,\check{\Xi}_{b_1\cdots b_n} = n!\,\alpha^{a_1\cdots a_n}_{b_1\cdots b_n} := n!\,\delta^{[a_1}_{b_1}\cdots\delta^{a_n]}_{b_n}. \tag{2.133a}$$ {vsATepdel}

The tensor $\alpha^{a_1\cdots a_n}_{b_1\cdots b_n}$ appearing here is the natural projector from $\otimes^n V$ to its one-dimensional subspace $\wedge^n V$. More generally, we define

$$\alpha^{a_p\cdots a_n}_{b_p\cdots b_n} := \delta^{[a_p}_{b_p}\cdots\delta^{a_n]}_{b_n} \tag{2.133b}$$ {vsTASproj}

Show that

$$\delta^{b_p}_{a_p}\,\alpha^{a_p\cdots a_n}_{b_p\cdots b_n} = \frac{p}{n-p+1}\,\alpha^{a_{p+1}\cdots a_n}_{b_{p+1}\cdots b_n} \qquad \Rightarrow \qquad \Xi^{a_1\cdots a_n}\,\check{\Xi}_{a_1\cdots a_p b_{p+1}\cdots b_n} = p!\,(n-p)!\,\alpha^{a_{p+1}\cdots a_n}_{b_{p+1}\cdots b_n}. \tag{2.133c}$$ {vsATedcon}

These last results are known as the ***ϵ-δ identities***. This, of course, is because they are usually written using slightly different notation. See below.

2.6.1 Determinants of Tensors

Let $\varphi^a{}_b$ denote a linear operator on an n-dimensional vector space V, and define its ***determinant*** by

$$[\det \varphi](\Xi) := \frac{1}{n!}\, \varphi^{a_1}{}_{b_1} \cdots \varphi^{a_n}{}_{b_n}\, \check{\Xi}_{a_1 \cdots a_n}\, \Xi^{b_1 \cdots b_n}. \qquad (2.134)$$ {vsDetMixImp}

Because the inverse of the alternating tensor $\Xi^{a_1 \cdots a_n}$ makes this expression homogeneous of degree zero, this determinant is simply a scalar, and independent of $\Xi^{a_1 \cdots a_n}$.

Exercise 2.135: Show that the value of $\det \varphi$, as defined in (2.134), is equal to the determinant of the matrix (2.16) representing the linear map $\varphi^a{}_b$ in a given basis B. Show that this matrix determinant is basis-independent.
Hint: See the exercise above on the numerical values of the components $\widetilde{\eta}^{\alpha_1 \cdots \alpha_n}$ of the Levi-Civita density.

Exercise 2.136: Calculate the determinant of the linear operator $D \circ X : \mathscr{P}_n(x) \to \mathscr{P}_n(x)$, where $\mathscr{P}_n(x)$ is the space of polynomials of order at most n in a single real variable x and X and D denote the multiplication and derivative operators, respectively.

Now suppose we have a linear map A_{ab} mapping a vector space V to its dual V^*. We define the determinant of this operator by the homogeneous map

$$[\det \widetilde{\widetilde{A}}](\Xi) := \frac{1}{n!}\, A_{a_1 b_1} \cdots A_{a_n b_n}\, \Xi^{a_1 \cdots a_n}\, \Xi^{b_1 \cdots b_n} \qquad (2.137)$$ {vsDetDown}

of degree two. That is, the determinant of A_{ab} is naturally defined as a scalar density of weight two, as the notation suggests.

Exercise 2.138: Show that (2.137) is equivalent to the classical expression

$$\det \widetilde{\widetilde{A}} := \frac{1}{n!}\, A_{a_1 b_1} \cdots A_{a_n b_n}\, \widetilde{\eta}^{a_1 \cdots a_n}\, \widetilde{\eta}^{b_1 \cdots b_n} \qquad (2.138a)$$ {vsDetDownCl}

in terms of the Levi-Civita symbol.

Exercise 2.139: Show that the value of the scalar $(\det A)_{(B)}$ representing the density $\det \widetilde{\widetilde{A}}$ in some basis B is equal to the determinant of the square matrix of components $A_{\alpha\beta}$ representing the operator A_{ab} in the basis B. Show that this determinant changes with the basis B in just the right way to define a density of weight two in the old-fashioned sense.

Exercise 2.140: Show that $\det \widetilde{\widetilde{A}} \neq 0$ if and only if A_{ab} defines an isomorphism from V to V^*.

Exercise 2.141: Define the determinant of a tensor B^{ab} as a scalar density of weight -2. Exhibit your definition both as a homogeneous map and a a classical scalar density.

Exercise 2.142: Is it possible to define determinants of tensors with more than two indices?

2.6.2 Orientation

We have now seen a couple examples of densities, including the Levi-Civita symbols and various determinants. All have integer weight. However, the definition of a density naively seems to allow non-integer weights. It is indeed possible to consider such objects, and there are occasions in theoretical physics where it is natural to do so. However, an astute reader already may have noticed that a slight problem could arise in the definition (2.124) if w is not an integer: If a density is defined as a function on *all* alternating tensors, and $D(\Xi)$ is known, how should we define $D(-\Xi) = (-1)^w D(\Xi)$? If w is not an integer, it would seem that any answer to this question would have to embrace all the subtleties commonly associated with multi-valued functions in the complex plane. For this reason, densities of non-integer weight are typically defined only for vector spaces equipped with an additional structure, an orientation.

The ratio of two (non-zero) alternating tensors on a particular vector space can be either positive or negative. The set of all such tensors trivially breaks into two equivalence classes, each consisting of all alternating tensors whose ratios relative to one another are positive. However, this division into two sets is quite democratic: neither equivalence class is *intrinsically* "positive" or "negative." Each class is the "negative" of the other, but neither is preferred. An ***orientation*** is a specification of one of these two classes, which then is said to consist of ***positive*** or, more suggestively, ***right-handed*** alternating tensors. By extension, bases whose associated alternating tensors (2.119) belong to this class are themselves called ***right-handed***.

Exercise 2.143: A density D of arbitrary weight w on an oriented vector space V is defined by (2.124), with $\Xi^{a_1 \cdots a_n}$ restricted to be right-handed. Show that this definition is completely equivalent to the original one based on the set of *all* alternating tensors when w is an integer, but that it also extends to allow the consistent definition of densities with non-integer weight.

Exercise 2.144: Show that the equivalence relation

$$\Xi \sim \Xi' \qquad \Leftrightarrow \qquad \Xi/\Xi' > 0 \tag{2.144a}$$ {vsOrientER}

on the set of (non-zero) alternating tensors $\Xi = \Xi^{a_1 \cdots a_n}$ for a given vector space is ***reflexive*** ($\Xi \sim \Xi$), ***symmetric*** ($\Xi \sim \Xi'$ implies $\Xi' \sim \Xi$), and ***transitive*** ($\Xi \sim \Xi'$ and $\Xi' \sim \Xi''$ implies $\Xi \sim \Xi''$). These properties actually define an equivalence relation.

Lecture 3

Linear Geometries

The dual space V^* to any finite-dimensional vector space V has the same dimension as V itself. The two are therefore isomorphic, though not canonically so. A ***geometry***, in a general sense, is a *preferred* linear mapping from a V to its dual. Such a mapping allows vectors to act on other vectors to produce numbers, much as the dot product does in ordinary Newtonian mechanics. Indeed, the Newtonian dot product corresponds to the particular case of a ***Euclidean geometry*** on the vector space $\mathbb{R}^3$. However, the dot product has specific properties that are not common to all geometric structures of interest in physics.

In this lecture, we study the generalization of the dot product to inner products on real vector spaces. Like the dot product, general inner products are symmetric in their arguments, but they need not be positive definite. The most familiar example of an indefinite inner product in theoretical physics is that of Minkowski space, the arena of special relativity, and we begin with a motivation and discussion of this important example. As we will see, the key new element of our discussion is that indefinite inner product spaces admit subspaces whose intrinsic geometry is degenerate. This fact complicates their geometric analysis, and the purpose of this discussion is to sort out which aspects of the familiar geometry of Euclidean inner product spaces carry over to the general case, and which do not.

3.1 MINKOWSKI SPACE

The theory of special relativity was invented to fix an apparent paradox at the heart of classical physics. There is a fundamental tension between the ***principle of relativity*** on the one hand, which asserts that all *inertial* observers will agree as to the form of the equations of motion governing the dynamics of a given mechanical system, and the ***Maxwell equations*** on the other, which govern electromagnetic interactions of charged bodies. The former principle is familiar from elementary Newtonian mechanics, where observers in uniform rectilinear relative motion will agree that Newton's second law holds in the form $m\dot{\mathbf{v}} = \mathbf{F}$ for a given particle.

Principle of the Constancy of the Speed of Light: There exists an inertial frame of reference in which every light ray propagates in vacuum at the rate c.

Discussion

Principle of Relativity: Every frame of reference in uniform rectilinear motion relative to an inertial frame of reference is itself inertial.

Discussion

Principle of the Isotropy of Space Space in any inertial frame of reference is a three-dimensional Euclidean continuum.

Principle of the Homogeneity of Time Ideal clocks in different inertial frames of reference run at uniform rates relative to one another.

Discussion

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 4\pi\rho \qquad & \nabla \times \mathbf{H} - \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} &= \frac{4\pi}{c}\mathbf{j} \\ \nabla \cdot \mathbf{B} &= 0 \qquad & \nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} &= 0. \end{aligned} \tag{3.1}$$

3.2 METRICS ON REAL VECTOR SPACES

A ***metric*** on a real vector space V is a symmetric tensor of the type g_{ab}. It maps vectors to co-vectors according to

$$g_{ab} : v^a \mapsto v_b := v^a\, g_{ab}. \tag{3.2}$$ {metVtoCV}

Although V and V^* are isomorphic, the map $g_{ab} : V \to V^*$ may fail to define an isomorphism. In this case, g_{ab} is ***degenerate***. A non-degenerate metric g_{ab} is an isomorphism, and therefore admits a unique ***inverse metric*** g^{ab} defined by

$$g^{ab}\, g_{ac} = \delta^b_c. \tag{3.3}$$ {metInv}

It maps co-vectors to vectors according to

$$g^{ab} : \omega_b \mapsto \omega^a := g^{ab}\, \omega_b. \tag{3.4}$$ {metCVtoV}

Note that both here and in (3.2), we use the stem letter from the original vector or co-vector also to denote the dual object derived from it via the metric. Thus, because of its symmetry, a non-degenerate metric allows us to raise and lower contracted indices freely:

$$\omega_b\, v^b = \omega_b\, \delta^b_c\, v^c = \omega_b\, g^{ab}\, g_{ac}\, v^c = g^{ab}\, \omega_b\, v^c\, g_{ca} = \omega^a\, v_a. \tag{3.5}$$ {metIndRL}

This freedom to move indices around is very helpful in practice, but it is important to remember that it has its origin in the metric. If the metric is degenerate, or one considers other types of geometry, the rules can change.

Exercise 3.6: Show that the inverse metric is symmetric.

Even if a metric is non-degenerate, its may still differ the ordinary Euclidean dot product. To explore the various possibilities, it is convenient to introduce the following standard shorthands for the ***inner product*** and ***square norm*** associated to a metric g_{ab}:

$$v \cdot w := g_{ab}\, v^a\, w^b \qquad \text{and} \qquad \|v\|^2 := g_{ab}\, v^a\, v^b. \tag{3.7}$$ {vsMetIPN}

These definitions are obviously related. Mathematically, the inner product defines a ***symmetric bilinear form*** on V, which is just another way of saying it is a symmetric tensor with two covariant indices, while the square norm is its associated ***quadratic form***:

$$\|v\|^2 = v \cdot v \qquad \text{and} \qquad v \cdot w = \tfrac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right). \tag{3.8}$$ {metPolId}

The second expression here is known as the ***polarization identity***. The standard scheme for classifying metrics on a given vector space V studies these associated forms.

Exercise 3.9: Show explicitly that a given bilinear form $v \cdot w \mapsto \mathbb{F}$ gives rise to two maps from vectors to dual vectors, and that these maps agree if and only if the bilinear form is symmetric. Here, "symmetric" means $v \cdot w = w \cdot v$ for all $v^a, w^a \in V$, while "bilinear" means that the inner product is linear in each of its arguments separately.

Exercise 3.10: The polarization identity is often written in the more familiar form

$$\|v+w\|^2 = \|v\|^2 + 2v \cdot w + \|w\|^2\,. \tag{3.10a}$$ {metPolId'}

Show that this is equivalent to (3.8); each implies the other.

Exercise 3.11: A quadratic form $\|\cdot\|^2$ always defines a homogeneous function of degree two on a vector space V, but not every such function defines a quadratic form. Show that a quadratic homogeneous function, also denoted $\|\cdot\|^2$ for convenience, gives rise to a symmetric bilinear form if and only if

$$\|v+w\|^2 + \|v-w\|^2 = 2\left(\|v\|^2 + \|w\|^2\right) \qquad \forall v^a, w^a \in V. \tag{3.11a}$$ {metParLaw}

This condition is known as the ***parallelogram law***. You may find it convenient to proceed as follows:

1. Use the polarization identity to *define* the inner product, and then use parallelogram law to prove additivity: $(u+v)\cdot w = u \cdot w + v \cdot w$.
2. Use the previous result to prove linearity under scaling by rational factors: $(m/n\, v)\cdot w = m/n\,(v \cdot w)$.
3. Argue that any homogeneous function of degree 2 on a *finite-dimensional* vector space must be continuous, and thereby extend linearity to general scalars.

A quadratic form $\|\cdot\|^2$ is called ***positive*** if $\|v\|^2 \geq 0$ for all $v^a \in V$, and similarly ***negative*** if $\|v\|^2 \leq 0$. A positive or negative form is called ***definite*** if $\|v\|^2$ vanishes only for $v^a = 0$, and ***semi-definite*** otherwise. A quadratic form taking both positive and negative values on V is called ***indefinite***. In this last case, especially, it is useful to classify vectors using nomenclature borrowed from relativity theory. A given vector $v^a \in V$ is called ***time-like*** if $\|v\|^2 < 0$, ***space-like*** if $\|v\|^2 > 0$, and ***light-like*** or ***null*** if $\|v\|^2 = 0$. A positive-definite metric, meaning a metric that gives rise to a positive-definite quadratic form on a vector space V, is called ***Euclidean***, and has all the familiar properties of the dot product in Newtonian physics. However, there are many other metrics of interest in theoretical physics. The Lorentzian metric on four-dimensional Minkowski space-time, which we will examine in detail below, is just one of these.

Exercise 3.12: A pair of vectors $v^a, w^a \in V$ are said to be ***orthogonal*** in the metric g_{ab} if their inner product vanishes: $v \cdot w = 0$. Show that g_{ab} is degenerate if and only if V contains *null* vectors v^a that are orthogonal to *every* other vector w^a in V. Find a simple way to express degeneracy of g_{ab} directly in terms of the quadratic form $\|\cdot\|^2$, with no direct reference to the metric or inner product. Such quadratic forms are of course called ***degenerate*** themselves.

Exercise 3.13: Given any pair of vectors v^a, w^a in a vector space V carrying a metric g_{ab}, with w^a non-null, prove the ***Pythagorean theorem***

$$\|v\|^2 = \frac{(v \cdot w)^2}{\|w\|^2} + \left\| v - \frac{v \cdot w}{\|w\|^2}\, w \right\|^2 . \tag{3.13a}$$ {metPyth}

Specialize this general formula to the cases where w^a is either a space-like or a time-like unit vector, $\|w\|^2 = \pm 1$. Discuss the similarities and differences between the signs of the terms in the present result and those in the familiar expression for Euclidean space. Does the result extend to the case where w^a is null?

Exercise 3.14: Show that the metric g_{ab} associated to a semi-definite quadratic form $\|\cdot\|^2$ on a vector space V is non-degenerate if and only if the form is actually definite.

Hint: First, use the Pyhtagorean theorem to show that any non-zero null vector $v^a \in V$ must be orthogonal to any non-null vector $w^a \in V$. Then, use the polarization identity to show that v^a must also be orthogonal to other null vectors w^a.

Exercise 3.15: Show that any indefinite quadratic form $\|\cdot\|^2$ must admit a non-trivial null vector $v^a \neq 0$.

Hint: Such a norm must admit at least one time-like vector v^a and one space-like vector w^a. Consider $\|(1-\alpha)v + \alpha w\|^2$ as a function of $0 \leq \alpha \leq 1$. Is this function continuous?

Exercise 3.16: In an arbitrary vector space V equipped with a quadratic form $\|\cdot\|^2$, show that any pair vectors of different types (space-like, time-like, or null) must be linearly independent. Further, show that any two orthogonal space-like vectors are linearly independent, and that any two orthogonal time-like vectors are linearly independent. Does a similar result apply to a pair of orthogonal null vectors?

Exercise 3.17: Give specific examples to show that the sum of two linearly independent time-like vectors can be time-like, space-like or null. Do likewise for sums of two space-like or two null vectors. Show, however, that a general linear combination of *mutually orthogonal* time-like vectors must be time-like, and similarly for such combinations of space-like or null vectors.

3.2.1 Orthonormal Bases

A basis $\{e^a_\alpha\}$ on a d-dimensional vector space V carrying a metric g_{ab} is called ***orthonormal*** if every basis vector is either ***unit*** or null, and orthogonal to every other basis vector:

$$g_{\alpha\beta} := g_{ab}\, e^a_\alpha\, e^b_\beta = \begin{cases} 0 & \text{if } \alpha \neq \beta \\ \pm 1 \text{ or } 0 & \text{if } \alpha = \beta. \end{cases} \tag{3.18}$$ {metONbas}

The three possibilities when $\alpha = \beta$ allow for indefinite and degenerate metrics. Any orthonormal basis will contain a number k of space-like unit vectors, a further number ℓ of time-like unit vectors, and finally a number n of null vectors. In fact, these three numbers are the same for every orthonormal basis on a given space with a given metric and together define the ***signature***, written $k + \ell + n$, of that metric. Clearly, the sum of these numbers must give the total dimension d of the vector space in question. For this reason, n is usually omitted, and the signature is often expressed simply as $k + \ell$, even when the sum is not d. In a sense to be made precise below, the signature provides the only invariant data characterizing a metric.

Exercise 3.19: Show that an orthonormal basis contains null vectors if and only if the metric g_{ab} is degenerate. Further, show that the null vectors in an orthonormal basis can be taken to be *any* basis for the ***null space*** $N := \ker(g_{ab})$, where the metric is viewed as a linear map from V to its dual. Conclude that the ***nullity*** n in the signature of a metric g_{ab} is simply the dimension of its kernel.

Exercise 3.20: Let g_{ab} be a degenerate metric on a vector space V, and let $N \subset V$ be its null space. Show that g_{ab} projects to define a *non-degenerate* metric $\check{g}_{ab}$ on the quotient space V/N. Conversely, given a vector space V, a subspace $N \subset V$, and a non-degenerate metric q_{ab} on the quotient V/N, define a unique metric $\hat{q}_{ab}$ on V with null space N. Thus, conclude that the ***quotient metric*** $\check{g}_{ab}$ and the null space $N \subset V$ completely characterize a degenerate metric g_{ab} on V.

A metric g_{ab} on a vector space V is said to ***induce*** a metric h_{ab} on any subspace $U \subset V$. The definition is the obvious one: since vectors in U are of course vectors in V, we set

$$h_{ab}\, u^a\, \tilde{u}^b := g_{ab}\, u^a\, \tilde{u}^b. \tag{3.21}$$ {metInduc}

The induced metric on U may have rather different characteristics than the original metric on V. In particular, every space V with indefinite metric g_{ab} admits subspaces U on which the induced metric h_{ab} is definite. Obviously, the one-dimensional subspaces of V spanned by a single space-like or time-like vector have this property, but there may be higher-dimensional examples as well. A ***maximal positive-*** or ***negative-definite subspace*** is a subspace $U \subset V$ on which the induced metric h_{ab} is either positive- or negative-definite, and which is not contained within any larger subspace $U' \subset V$ with this property.

Exercise 3.22: Use Zorn's lemma to show that any vector space V with a metric g_{ab} admits both maximal positive- and negative-definite subspaces, although unless g_{ab} is indefinite one or both of these will be the trivial subspace consisting only of the zero vector.

Hint: Order the positive-definite subspaces of V, for instance, by inclusion. Bound any nested sequence of such spaces by their union, and show that this union is a positive-definite subspace of V.

Comment: Since V is finite-dimensional by assumption, this line of argument is a bit like using a sledgehammer to open a walnut. However, we have here a slick approach to a problem which otherwise would require laying a great deal of groundwork to resolve. Read the rest of this subsection, and then return to approach this problem in a more pedestrian way.

Exercise 3.23: In contrast to the null space N, which is always unique, show that the maximal positive- and negative-definite subspaces of V are unique only in the trivial case that g_{ab} is positive- or negative-definite. Identify both maximal subspaces in either case.

In ordinary Euclidean geometry, we are quite accustomed to thinking of the ***orthogonal complement*** $U^\perp$ of any subspace $U \subset V$. By definition, $U^\perp$ consists of all vectors in V orthogonal to *every* vector $u^a \in U$ in the metric g_{ab}. Note the similarity in notation to the *normal* subspace $U_\perp \subset V^*$ associated to any subspace $U \subset V$; we shall return to this shortly. When g_{ab} is positive-definite, $U^\perp \subset V$ is always a subspace, is always disjoint from U, and together with U always spans V. That is, in Euclidean geometry, $U^\perp$ intersects U

only in the zero vector, and every vector $v^a \in V$ can be written as a linear combination of components in U and $U^\perp$. Together, these properties guarantee that $U^\perp$ complements U, whence the term "orthogonal complement." When g_{ab} is indefinite, however, orthogonal complements do not always exist. In fact, this peculiar feature of indefinite metrics is arguably their only new characteristic, apart from the possibility of degeneracy discussed above. Indefinite metrics need not be degenerate, no more than degenerate metrics need be indefinite. Null spaces are the hallmark of degenerate metrics, while the subtlety in defining orthogonal complements is the hallmark of indefinite metrics. Things get amusingly, though not hopelessly, complicated when a metric is both indefinite and degenerate.

Exercise 3.24: If $U \subset V$ is a subspace of a vector space V with metric g_{ab}, show that the set $U^\perp$ of vectors orthogonal to *every* $u^a \in U$ is also a subspace of U.

Exercise 3.25: Show that $U^\perp$ intersects U in the previous exercise precisely in the null space $N^U \subset U \subset V$ of the metric h_{ab} induced on U by g_{ab}. Thus, conclude that U and $U^\perp$ are disjoint if and only if the induced metric on U is non-degenerate.

Exercise 3.26: Continuing, show that U and $U^\perp$ together span U if and only if the null space N^U of the induced metric h_{ab} on U is a subspace of the null space N of the original metric g_{ab} on V.
Hint: If N^U is not a subspace of N, then there must exist a null vector $n^a \in N^U$ and a vector $v^a \in V$ such that $n \cdot v \neq 0$. Show that the latter *cannot* be written in the form $v^a = u^a + r^a$, with $u^a \in U$ and $r^a \in U^\perp$.

Exercise 3.27: Following the previous exercises, show that the subspace $U^\perp \subset V$ orthogonal to a given subspace $U \subset V$ always exists, but is an orthogonal *complement* if and only if the induced metric h_{ab} on U is non-degenerate. In particular, show that any positive- or negative-definite subspace of V admits an orthogonal complement, but that a genuinely *semi*-definite subspace does not. If the metric h_{ab} induced on U is *indefinite*, is $U^\perp$ an orthogonal complement to U? Give examples to support your answer.

Exercise 3.28: When g_{ab} is non-degenerate, show that $U^\perp \subset V$ is the image of the normal subspace $U_\perp \subset V^*$ under the inverse metric g^{ab} for any given subspace $U \subset V$.

Now we are in a position to prove the key result needed to establish both existence of orthonormal bases and invariance of the signature. Let $S \subset V$ be a maximal positive-definite subspace, and consider any vector $r^a \in S^\perp$ in its orthogonal complement. This r^a cannot be space-like since otherwise every vector u^a in the subspace $U \subset V$ spanned by S and r^a could be written as a linear combination of *orthogonal* space-like vectors $s^a \in S$ and r^a. By an exercise above, every such u^a would itself be space-like, and U would therefore be positive-definite. This would contradict our assumption that S is a *maximal* positive-definite subspace, whence every $r^a \in S^\perp$ must be either time-like or null. Equivalently, $S^\perp$ must be negative semi-definite. Continuing, this subspace $S^\perp$ must admit a maximal negative-definite subspace $T \subset S^\perp \subset V$. The orthogonal complement N to T *within the subspace* $S^\perp$ can contain only null vectors since $S^\perp$ is negative semi-definite. Moreover, each $n^a \in N$ must be orthogonal not only to vectors $t^a \in T$, but also to vectors $s^a \in S$ since $N \subset S^\perp$. Thus, choosing any maximal positive-definite subspace $S \subset V$ and any maximal negative-definite subspace $T \subset S^\perp \subset V$, we get a unique totally null space $N \subset V$ orthogonal to both, and unique decomposition of any $v^a \in V$ into components

$$v^a = s^a + t^a + n^a \qquad \text{with} \qquad s^a \in S, \quad t^a \in T \quad \text{and} \quad n^a \in N. \tag{3.29}$$ {metSTdec}

We now need to mop up a bit, showing that this sort of decomposition $V = S \oplus T \oplus U$ is universal in the sense that it is always possible to select *mutually orthogonal* maximal positive- and negative-definite subspaces S and T without loss of generality and that N, as the notation suggests, is indeed the null space of the metric g_{ab} on V. We do this in the following series of exercises.

Exercise 3.30: Show that a maximal negative-definite subspace $T \subset S^\perp$, with $S \subset V$ a maximal positive-definite subspace, as defined above, is also a maximal negative-definite subspace of the total vector space V.

Hint: Consider T as a subspace of V, and use the decomposition (3.38) to show that $T^\perp \subset V$ consists precisely of vectors $v^a \in V$ lying in the subspace $S \oplus N \subset V$ spanned by vectors in S and N. Reverse the argument above to show that, since $S \oplus N$ is positive semi-definite, $T \subset V$ must be maximal.

Exercise 3.31: Show that *every* maximal negative-definite subspace $T \subset V$ arises as above, as a maximal negative-definite subspace of the orthogonal complement $S^\perp$ to some maximal positive-definite subspace $S \subset V$.

Hint: Given T, modify the arguments above slightly to pick a maximal positive-definite subspace $S \subset T^\perp \subset V$, and show that S is a maximal definite subspace of V. Then, show that $T \subset S^\perp$ is a maximal negative-definite subspace of the orthogonal complement of S in V.

Exercise 3.32: Show that the totally null space N orthogonal to the subspace $S \oplus T$ spanned by any pair of mutually orthogonal maximal positive- and negative-definite subspaces S and T must be precisely the null space N of the metric g_{ab} defined above.

Hint: Denote the null space of g_{ab} temporarily by $\tilde{N}$. Since $n^a \in N$ as defined above is orthogonal to *every* $v^a \in V$ by (3.38), we know immediately that $N \subset \tilde{N}$. To show that $\tilde{N} \subset N$, and establish equality, show that any null vector $n^a \in V$ with non-zero components along S and/or T cannot lie in the kernel of g_{ab}.

The arguments above have shown that any vector space V can be decomposed into a trio of mutually orthogonal subspaces—S, T and N—with S maximal and positive-definite, T maximal and negative-definite, and N the null space of g_{ab}. Apart from N, this decomposition is not unique unless g_{ab} is strictly positive- or negative-definite, in which case two of these three subspaces are trivial. However, such decompositions always exist. We now establish our key results using techniques from ordinary linear algebra.

Exercise 3.33: The

3.2.2 Isometry

Exercise 3.34: When is $U^{\perp\perp} \subset V$ equal to U? When is $U^\perp \subset V$ equal to U?

Exercise 3.35: Suppose that $U \subset V$ has a non-degenerate induced metric $\overset{U}{g}_{ab}$, so that we can write any $v^a \in V$ uniquely in the form

$$v^a = u^a + r^a \qquad \text{with} \qquad u^a \in U \quad \text{and} \quad r^a \in U^\perp. \tag{3.35a}$$ {metOrthDec}

Show that

$$\|v\|^2 = \|u\|^2 + \|r\|^2 = \|u\|^2 + \|v-u\|^2 \,. \tag{3.35b}$$ {metPythGen}

This result generalizes the Pythagorean theorem to orthogonal complements.

Exercise 3.36: Show that every vector space V with indefinite metric g_{ab} admits both a maximal positive-definite subspace S and a maximal negative-definite subspace T. Show, however, that these are not unique. *Hint*: To construct a maximal positive-definite subspace of a d-dimensional vector space V, start by picking any unit space-like vector $s_1^a \in V$. Then, consider the metric g_{ab}^1 induced on the subspace U^1 of vectors $v^a \in V$ orthogonal to s_1^a. Show that this metric is either indefinite or negative semi-definite. In the former case, pick a unit space-like vector $s_2^a \in U^1$, and induce a metric g_{ab}^2 on the subspace $U^2 \subset U^1$ of vectors in U^1 orthogonal to s_2^a in the metric g_{ab}^1. Continue in this way to produce a series of vectors $s_1^a, \cdots, s_k^a$ and nested subspaces $U^k \subset \cdots \subset U^1 \subset V$, with U^k carrying a negative semi-definite metric g_{ab}^k. Show that the vectors $s_1^a, \cdots s_k^a \in V$ are orthonormal in the original metric g_{ab}, and therefore must be linearly independent by an exercise above, spanning a k-dimensional subspace $S \subset V$. Show that this subspace is both positive-definite and, using (3.35a), maximal.

Exercise 3.37: Show that every maximal positive-definite subspace $S \subset V$ admits an orthogonal complement $S^\perp$, and that $S^\perp$ contains a maximal negative-definite subspace $T \subset V$. Show that T is unique if and only if the nullity n of the original metric g_{ab} on V vanishes. What is T in this case?

Following the previous exercises, we see that it is always possible to express a vector space V with indefinite metric g_{ab} as a direct sum $V = S \oplus T \oplus N$, where S is a maximal positive-definite subspace, T is a maximal negative-definite subspace, and N is the null space of g_{ab}. That is, given choices of S and T, and recalling that $N = \ker(g_{ab})$ is specified once and for all by the metric, every vector $v^a \in V$ can be expressed uniquely in the form

$$v^a = s^a + t^a + n^a \qquad \text{with} \qquad s^a \in S, \quad t^a \in T \quad \text{and} \quad n^a \in N. \tag{3.38}$$ {metSTdec}

Moreover, because we can always choose $T \subset S^\perp$, we will have

$$\|v\|^2 = \|s\|^2 + \|t\|^2 . \tag{3.39}$$ {metSTnorm}

Every n-dimensional vector space V with a metric admits orthonormal bases of the form $\{e_1^a, \cdots, e_k^a, e_{k+1}^a, \cdots, e_\ell^a, e_{\ell+1}^a, \cdots, e_n^a\}$
. We prove this by induction, assuming a non-degenerate metric for simplicity. In general, one of three situations will hold:

1. All vectors $v^a \in V$ are null. This case is impossible for a non-degenerate metric because g_{ab} must be invertible. Indeed, if *all* vectors are null, then

$$4v \cdot \tilde{v} = \|v + \tilde{v}\|^2 - \|v - \tilde{v}\|^2 = 0 \quad \forall \tilde{v}^b \in V \qquad \Rightarrow \qquad g_{ab}\, v^a = 0 \in V^*. \tag{3.40}$$ {metNulIdf}

   That is, if every vector is null, then every $v^a \in V$ belongs to the kernel of $g_{ab} : V \to V^*$. However, since a non-degenerate g_{ab} is an isomorphism, no such kernel can exist.

2. The metric g_{ab} on V is positive semi-definite, so every vector in $v^a \in V$ is either null or space-like.

3. There exists at least one time-like vector $v^a \in V$ with $\|v\|^2 < 0$.

We assume the third case for generality. Then, choose any time-like vector $e_1^a \in V$ normalized so that $g_{ab}\, e_1^a\, e_1^b = -1$. This will be our first basis vector. Next, we consider the subspace $U_1 \subset V$ orthogonal to e_1^a. This is the subspace of vectors $u^b \in V$ normal to the non-zero co-vector $g_{ab}\, e_1^a$, and therefore has dimension $n-1$. The metric g_{ab} on V induces a symmetric tensor g^1_{ab} on U_1 because any two vectors in U_1 are in particular vectors in V:

$$g^1_{ab}\, u^a\, \tilde{u}^b := g_{ab}\, u^a\, \tilde{u}^b. \tag{3.41}$$ {metInduc}

This tensor defines a non-degenerate map from U_1 to U_1^* if and only if, for every $u^a \in U^1$, there exists at least one $\tilde{u}^b \in U^1$ such that $g^1_{ab}\, u^a\, \tilde{u}^b \neq 0$. We know that a $\tilde{v}^b \in V$ must exist such that $g_{ab}\, u^a\, \tilde{v}^b \neq 0$ because g_{ab} is a metric, and we set

$$\tilde{u}^b := \tilde{v}^b + (e_1 \cdot \tilde{v})\, e_1^b. \tag{3.42}$$

This vector has the same non-zero inner product with u^a as $\tilde{v}^b$ since u^a is orthogonal to e_1^b, and is itself orthogonal to e_1^a by construction. Thus, g^1_{ab} is non-degenerate on U_1.

The proof proceeds by substituting the non-degenerate metric g^1_{ab} on the vector space U_1 for the non-degenerate metric g_{ab} on the vector space V. Once again, either the second or third case listed above will apply. Suppose it is the third again, and pick a vector $e_2^a \in U_1$ normalized so that $g^1_{ab}\, e_2^a\, e_2^b = -1$. This e_2^a is also a unit time-like vector in the original vector space by the definition (3.41) of the metric g^1_{ab}, and moreover is orthogonal to the first basis vector e_1^a by the definition of U_1. Thus, we now have two orthonormal basis vectors, e_1^a and e_2^a, on V and a corresponding nested sequence of subspaces $U_2 \subset U_1 \subset V$. The smallest of these can be regarded either as the subspace of U_1 orthogonal to e_2^a, or as the subspace of V orthogonal to both e_1^a and e_2^a. It has dimension $n-2$ and carries a non-degenerate metric g^2_{ab} defined as in (3.41). We carry on in this way until we have a set of orthonormal time-like basis vectors $\{e_1^a, \cdots, e_k^a\}$ such that the subspace U_k orthogonal to all of them carries a positive semi-definite metric g^k_{ab}. Then, we pick a space-like basis vector e_{k+1}^a and normalize it so that $g^k_{ab}\, e_{k+1}^a\, e_{k+1}^b = +1$. Apart from the change of sign in the normalization convention, everything proceeds as before. Eventually, we will have an orthonormal set $\{e_1^a, \cdots, e_k^a, e_{k+1}^a, \cdots, e_\ell^a\}$ of time-like and space-like unit vectors, and the subspace U_ℓ orthogonal to all of them will contain no non-null vectors. However, this must be the trivial subspace of V since otherwise any $u^a \in U_\ell$ is orthogonal to any linear combination of the basis elements e_α^a by definition of U_ℓ, and to every vector in U_ℓ by the argument in the first case above. Thus, U_ℓ is the kernel of g_{ab} on V, and must be trivial since the metric is an isomorphism. We therefore conclude that $\ell = n$ and $\{e_\alpha^a\}$ is an orthonormal basis with k time-like and $n-k$ space-like unit vectors.

Exercise 3.43: Generalize the discussion above to allow for degenerate metrics g_{ab}. Show that one can construct orthonormal bases of the form $\{e_1^a, \cdots, e_k^a, e_{k+1}^a, \cdots, e_\ell^a, e_{\ell+1}^a, \cdots, e_n^a\}$ with the first k basis vectors time-like, the next $\ell - k$ space-like, and the remainder null.

Any vector space with a metric will admit multiple orthonormal bases, and it is useful to understand how they are related. The most important feature of this problem is that, for a given metric g_{ab} of V, the numbers k and ℓ of time-like and non-null basis vectors are the same for every orthonormal basis $\{e_1^a, \cdots, e_k^a, e_{k+1}^a, \cdots, e_\ell^a, e_{\ell+1}^a, \cdots e_n^a\}$. The second of these results is immediate. One of the exercises above has shown that the null basis vectors $e_{\ell+1}^a, \cdots, e_n^a$ can be taken to be *any* basis on the subspace $N := \ker(g_{ab}) \subset V$. Thus, $n - \ell = \dim(\ker(g_{ab}))$, and therefore ℓ, must be independent of the orthonormal basis. To show that k is also invariant, we can focus on the special case of non-degenerate metrics. Then, the time-like and space-like basis vectors in a given basis separately span complementary subspaces of V. Call these T and S, respectively, because every vector in T is time-like, while every vector in S is space-like. Any subspace $U \subset V$ containing T as a *proper* subspace must contain at least one vector in S, and therefore at least one space-like and one time-like vector. The metric induced on U by g_{ab} as in (3.41) is therefore indefinite. In general, we call any $T \subset V$ of this type a ***maximal negative-definite subspace*** of V. Since every orthonormal basis on V defines such a subspace, we need to show that any two have the same dimension. We do this in the following series of exercises.

Exercise 3.44: Use the Pythagorean theorem to show that any positive semi-definite metric is either positive definite or degenerate.

Solution: Let $v^a \in V$ be a non-zero null vector. Then, the Pythagorean theorem gives

$$0 = \|v\|^2 = \frac{(v \cdot w)^2}{\|w\|^2} + \left\| v - \frac{v \cdot w}{\|w\|^2}\, w \right\|^2 \tag{3.44a}$$

for any space-like vector $w \in V$. But both terms here are non-negative because w^a is space-like in the first place and the norm of *any* vector in V is non-negative in the second. Thus, both terms must vanish, so $v \cdot w = 0$ for every space-like $w^a \in V$. If $w^a \in V$ is also null, then we have

$$v \cdot w = \tfrac{1}{2}\left(\|v + w\|^2 - \|v\|^2 - \|w\|^2\right) = \tfrac{1}{2}\,\|v + w\|^2 \geq 0, \tag{3.44b}$$

again using the assumed positive semi-definite character of the metric. But this must hold also under the substitution $w^a \mapsto -w^a$. The only possibility again that $v \cdot w = 0$. Thus, v^a is orthogonal to all vectors $w^a \in V$, which means that $g_{ab} v^a = 0 \in V^*$ and the metric is degenerate. The only alternative to degeneracy is that there is no non-zero null vector $v^a \in V$, which of course means that the metric is strictly positive definite.

Exercise 3.45: Show that any subspace $U \subset V$ defines a *unique* ***orthogonal subspace*** $U^\perp \subset V$ comprising all vectors in V orthogonal to *every* $u^a \in U$. When g_{ab} is non-degenerate, show that the dimension of $U^\perp$ plus that of U is always equal to that of V. However, show that U and $U^\perp$ are not always complementary to one another, generically intersecting in the kernel subspace $\ker(g_{ab}^U) \subset U \subset V$ for the metric g_{ab}^U induced on U by g_{ab} as in (3.41). Conclude that $U^\perp$ complements U if and only if the intrinsic metric g_{ab}^U on U is non-degenerate. In this case, we call $U^\perp$ the ***orthogonal complement*** to U.

Exercise 3.46: Show that the intrinsic metric g_{ab}^T on any maximal negative-definite subspace is non-degenerate, and that the orthogonal complement $S := T^\perp$ is a ***maximal positive-definite subspace*** as long as the full metric g_{ab} is non-degnerate. (The meaning here should be transparent.)

Hint: The orthogonal complement S cannot contain any time-like vectors since then T would not be maximal. Thus, the intrinsic metric g_{ab}^S on it must be positive semi-definite.

.

Exercise 3.47: Now consider general decompositions of V into ***orthogonal complement*** subspaces U and W. That is, U and W are complementary subspaces of V such that every vector in U is orthogonal to every vector in W. Argue that every vector $v^a \in V$ admits a decomposition $v^a = u^a + w^a$ with $u^a \in U$ and $w^a \in W$, and show that

$$\|v\|^2 = \|u\|^2 + \|w\|^2 \,. \tag{3.47a}$$

This result generalizes the Pythagorean theorem to higher-dimensional decompositions of the vector space V. Check that you reproduce (3.13a) in the cases where W is a one-dimensional sub-space of W consisting either of time-like or space-like vectors.

Exercise 3.48: Let $T \subset V$ be any maximal negative-definite subspace of a vector space V with non-degenerate metric g_{ab}. Define the ***orthogonal complement*** subspace $T^\perp \subset V$ of all vectors in V orthogonal to every vector in T. Show that $T^\perp$ must be a ***maximal positive-definite subspace*** of V. That is, show that every non-zero vector $s^a \in T^\perp$ must have strictly positive norm squared $\|s\|^2 > 0$, and that no other subspace $U \subset V$ containing $T^\perp$

Time for class. More to come.

Exercise 3.49: An ***orthonormal set*** in a Euclidean vector space is a collection of vectors $\{e^a_\alpha\}$ such that

$$g_{ab}\, e^a_\alpha\, e^b_\beta = \delta_{\alpha\beta} := \begin{cases} 1 & \alpha = \beta \\ 0 & \alpha \neq \beta. \end{cases} \tag{3.49a}$$ {vsMetOrthSet}

Generalize the Pythagorean theorem to prove ***Bessel's inequality***

$$\|v\|^2 \geq \sum_\alpha (e_\alpha \cdot v)^2 \tag{3.49b}$$ {vsMetBess}

for any vector v^a and orthonormal set $\{e^a_\alpha\}$.

Exercise 3.50: Show that a basis $\{b^a_\alpha\}$ on a Euclidean vector space V can always be taken to be orthonormal without loss of generality.

Hint: Start with an arbitrary basis $\{b^a_\alpha\}$ and normalize its first element to define $e^a_1 := b^a_1 / \|b_1\|$. Then, use its second element to define $e^a_2 := [b^a_2 - (e_1 \cdot b_2)\, e^a_1] / \|b_2 - (e_1 \cdot b_2)\, e_1\|$. Show that this vector is non-zero, orthogonal to e^a_1, and normalized so that $\|e_2\| = 1$. Then, iterate through the rest of the vectors in $\{b^a_\alpha\}$, always subtracting off components along all the e^a_β with $\beta < \alpha$ and then normalizing to find an orthonormal basis $\{e^a_\alpha\}$. This is known as the ***Gram–Schmidt orthogonalization process***. It also works in infinite-dimensional Hilbert spaces, but the proof is more difficult.

Exercise 3.51: Show that the inverse metric must be symmetric, $g^{ab} = g^{ba}$.

Exercise 3.52: Show that any non-trivial linear combination of linearly-independent null vectors v^a and w^a in an n-dimensional vector space V with indefinite metric g_{ab} must be non-null.

Solution: Define the one-dimensional subspaces $\{v\}$ and $\{w\}$ of V consisting of vectors proportional to the two given null vectors, and then their orthogonal complements $\{v\}^\perp$ and $\{w\}^\perp$. The latter consist, respectively, of all vectors orthogonal to v^a and w^a in the metric g_{ab}. Both of these orthogonal complements must have dimension $n-1$ since g_{ab} is invertible; the co-vectors $g_{ab}\, v^b$ and $g_{ab}\, w^b$ cannot vanish. Moreover, since v^a and w^a are linearly independent, these co-vectors cannot be proportional.

3.2.3 Symplectic Geometry

3.3 COMPLEX AND KÄHLER SPACES

Lecture 4

Manifolds

The previous lectures have looked at the *algebraic* generalization from $\mathbb{R}^d$ to abstract vector spaces. In this lecture, we begin exploring the natural *analytic* generalization of $\mathbb{R}^d$, the differentiable manifold. The vector-space structure is fairly rigid because of the homogeneity imposed by the algebraic laws of vector addition and scalar multiplication. Indeed, every finite-dimensional vector space is isomorphic to $\mathbb{R}^d$ (or $\mathbb{C}^d$) for some d. Differentiable manifolds, in contrast, offer a much larger generalization. The price of this generalization is a loss of the sort of homogeneity that makes vector spaces so easy to deal with. However, from the physicist's point of view, this loss is arguably a good thing; experimental evidence suggests that many spaces appropriate to model real-world phenomena ought to be fundamentally inhomogeneous. Nonetheless, one must still be able to differentiate and integrate, and this leads naturally to the idea of a manifold as a sort of natural arena for differential and integral calculus in multiple dimensions.

4.1 COORDINATE CHARTS

Every-day experience tells us that events in the real world form a four-dimensional continuum, which traditionally we imagine as three dimensions of space and one of time. Relativity theory tells us that these are fused together in a single entity, space-time, whose geometry is dictated by the gravitational field. From this point of view, it is quite clear that space-time cannot be a vector space in the presence of gravitating bodies. Even if it supports the algebraic operations of vector addition and scalar multiplication, its geometry cannot be adapted to them in any sense since then it would be homogeneous, and the gravitational field trivial. Thus, even neglecting the issue of the global structure of our universe, the framework for the calculus underlying classical theoretical physics must be generalized to allow for non-trivial gravitational fields.

It is a tremendous idealization to assert that our actual universe forms a continuum, four-dimensional or otherwise. This may be one principle on which practically all researchers in the perpetually nascent field of "quantum gravity" could agree. However, it is not our purpose to explore this issue here. What is clear is that, in some limit or approximation, whatever "fundamental" theory describes the world gives way to a description in terms of functions varying with some degree of smoothness over space-time. We aim only to describe phenomena in this regime.

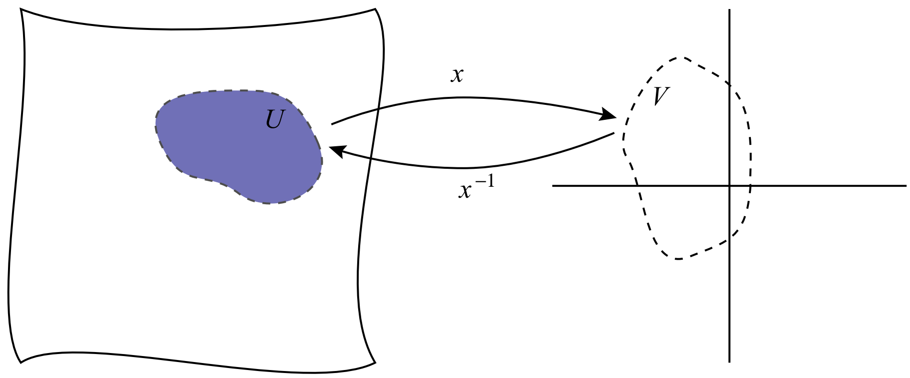


Figure 4.1: A coordinate chart (x, U) on a manifold M. The coordinates x define a one-to-one map from a given subset $U \subset M$ onto an open subset $V \subset \mathbb{R}^d$. Restricted to the image V, this map is invertible. {mfdChart}

A ***smooth manifold*** of ***dimension*** d is a set M, together with a family $\mathscr{A}$ of pairs (x, U), where U is a subset of M and $x : U \to \mathbb{R}^d$ assigns a definite set of values $x^\alpha(m)$ to each point of $m \in U$. Here, α is a concrete index taking values from 1 to d. The pair (x, U) is called a ***coordinate chart*** on M, while the collection $\mathscr{A}$ of all such charts is known as an ***atlas***. These must satisfy the following conditions:

1. Any coordinate map $x : U \to \mathbb{R}^d$ is *one-to-one*, and its image $V := x(U)$ in in the coordinate space is *open*. The first condition here means that the coordinate map x has an inverse $x^{-1} : V \to U$, so the sets $U \subset M$ and $V \subset \mathbb{R}^d$ are effectively identified. A subset $V \subset \mathbb{R}^d$ is called ***open*** if, for every $p \in V$, there exists a radius $\epsilon > 0$ small enough that every point $p' \in \mathbb{R}^d$ within the open ball $\|p' - p\| < \epsilon$ is included in V. This condition means that V, and thus U, cannot contain isolated or exceptional points, and gives the precise sense in which a manifold is "locally" a regular d-dimensional continuum. It is illustrated in Fig. 4.1.

2. Whenever two coordinate charts (x_1, U_1) and (x_2, U_2) in the atlas $\mathscr{A}$ overlap in M, the ***transition function*** $\phi_{12} := x_1 \circ (x_2)^{-1}$, giving the x_1-coordinates of a given $m \in M$ in the overlap region as functions of the x_2-coordinates of that same point, is *smooth* as a function from one subset of $\mathbb{R}^d$ to another. These subsets of $\mathbb{R}^d$ are the images $V_{12} := x_2(U_1 \cap U_2)$ and $V_{21} := x_1(U_1 \cap U_2)$ of the overlap region in M in either coordinate space. The function ϕ_{12} is ***smooth*** if it has well-defined partial derivatives to *all* orders with respect to all combinations of the variables $x_2 \in \mathbb{R}^d$. That is, the correspondence $x_2^\alpha \mapsto x_1^\beta$ of sets of coordinate values established by mapping through the manifold defines a collection of d scalar functions $x_1^\beta(x_2^1, \cdots, x_2^d)$, each of which is continuous, differentiable with respect to each x_2^α, second-differentiable with respect to every pair of x_2^α, and so on *ad infinitum*. In a sense to be made precise below, a differential calculus is naturally induced on the domain U of any coordinate chart

(x, U) on M by the ordinary differential calculus on $\mathbb{R}^d$. This condition, illustrated in Fig. 4.2, implies that the various calculi on M associated with different charts are consistent with one another.

3. The atlas $\mathscr{A}$ ***covers*** M in the sense that every point $m \in M$ lies within at least one coordinate chart. It is also ***maximal*** in the sense that every pair (x, U) of the type described above, with x a one-to-one mapping from $U \subset M$ to an open subset of $\mathbb{R}^d$ having smooth transition functions with every coordinate chart in $\mathscr{A}$, is itself in $\mathscr{A}$. That is, every pair (x, U) satisfying the first two conditions above is admissible as a coordinate chart M, and we insist that all such admissible charts are actually contained in the atlas. This avoids a potential issue where two intuitively identical manifolds would differ mathematically only because the particular collections of charts in their atlases differ. The covering condition, meanwhile, asserts that every point of M is covered by *some* coordinate chart. In conjunction with the first condition described above, this implies that every point of M has a local neighborhood that forms a d-dimensional continuum.

The definition above is heuristic. It captures the key ideas that every point of a smooth manifold is part of a local continuum, and that the relations allowing one to pass from one local continuum to another are smooth. Thus, one can take derivatives of suitable functions on a manifold M in a consistent way, even though M is only *locally* like $\mathbb{R}^d$ and may be quite different globally. In this sense, a smooth manifold is merely a natural arena for differential calculus. There are some additional subtleties in passing from this intuitive picture to a mathematically rigorous definition. In particular, M must carry a separate and more fundamental structure, a topology, before it can be rigorously defined as a manifold, and moreover that topology must be "second countable" and "Hausdorff." These ideas, and their implications for manifolds, are the subject of a good mathematics course on topology, but are not really needed here. The intuitive content of the topological structure underlying the manifold structure of M is suggested by the following exercise.

Exercise 4.1: The Euclidean distance $\|\cdot\|$ between points in $\mathbb{R}^d$ was used in the first condition above to define the notion of an open set. There is therefore a natural way to measure the (coordinate) distance between any two points $m, m' \in U \subset M$ both lying in a single coordinate chart (x, U) of M. Does it follow that there is a natural measure of distance between points of M? What natural structure on M is associated with the Euclidean norms in its various coordinate systems?

Hint: What happens under a change of coordinates x on a fixed domain $U \subset M$? Show that two different coordinate systems have the property that an open ball in one always contains, and is contained in, an open ball in the other.

The definition of a manifold given here may look a bit daunting. However, from a practical point of view, showing that a particular set M is a manifold is actually quite straightforward. One need only find a small set, almost always finite, of coordinate charts covering M, and then show that the transition functions between each pair are smooth. There is a certain art to the first step in this process: there are many small sets of charts that cover M, and some may be easier to work with than others. But the second step is

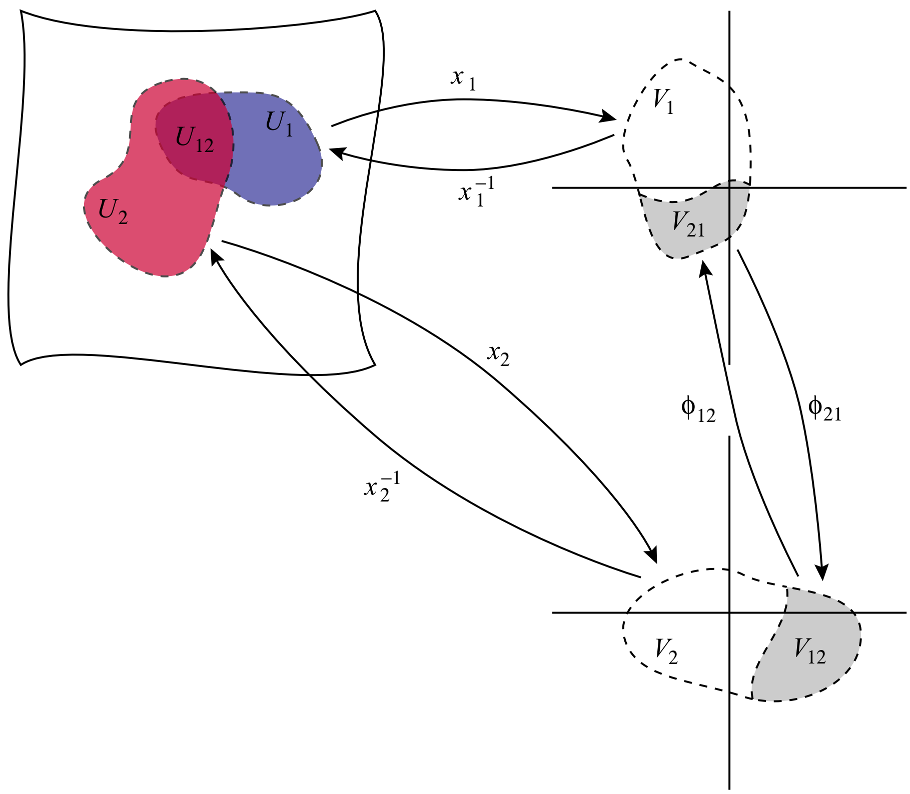


Figure 4.2: Conditions on the overlap of two coordinate charts (x_1, U_1) and (x_2, U_2) on a manifold M. As in Fig. 4.1, the coordinate system x_1 defines an invertible map from a subset $U_1 \subset M$ to an open subset $V_1 \subset \mathbb{R}^d$. Likewise, the coordinates x_2 define a map from another subset $U_2 \subset M$ to an open subset V_2 in a *distinct* copy of $\mathbb{R}^d$. If U_1 and U_2 intersect in M, then we define the images $V_{21} \subset V_1 \subset \mathbb{R}^d$ and $V_{12} \subset V_2 \subset \mathbb{R}^d$ of that intersection in either coordinate space. Composing coordinate maps and inverses on these subsets where both are defined, we get the transition functions $\phi_{12} := x_1 \circ x_2^{-1}$ and $\phi_{21} := x_2 \circ x_1^{-1}$. These functions are inverse to one another, and together define a one-to-one correspondence between points in V_{12} and V_{21}. The manifold M is smooth if all such transition functions are smooth in the conventional sense for maps between open subsets of $\mathbb{R}^d$. {mfdSmooth}

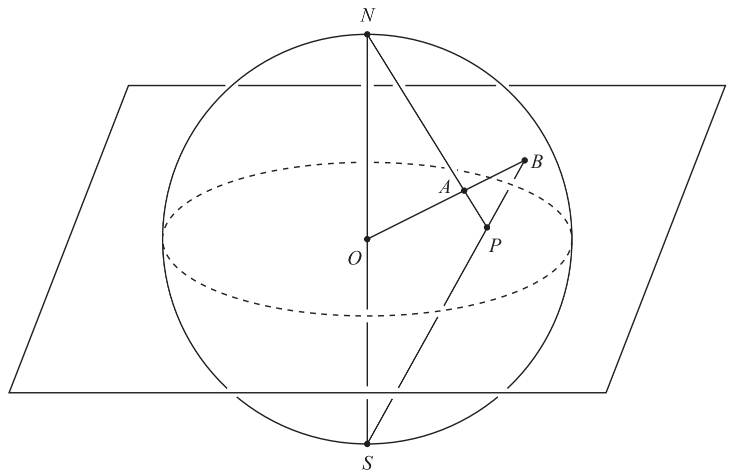


Figure 4.3: Geometry of the stereographic projection. Under the projection that maps the south pole S to the origin O, the point P on the sphere gets mapped to the point A defined by the intersection of the line joining the north pole N to P. The north pole itself is "mapped to infinity" under this projection. Under the other stereographic projection, mapping the north pole N to the origin O, the same point P is mapped to the point B where the line joining the south pole S to P intersects the equatorial plane. The vectors OA and OB in the equatorial plane are parallel, and the product of their lengths is unity. {mfdStereo}

mechanical. The transition functions from $\mathbb{R}^d$ to itself are determined by the first step, and one must merely check that they are smooth everywhere they need to be.

Exercise 4.2: Show that the ***n*-sphere**, defined as the subset

$$S^n := \left\{(x^0, \cdots, x^n) \mid (x^0)^2 + \cdots + (x^n)^2 = 1\right\} \tag{4.2a}$$ {mfdSphere}

of $\mathbb{R}^{n+1}$, is a smooth manifold.

Solution: Describe a general vector in $\mathbb{R}^{n+1}$ by a pair $(x, \mathbf{x})$, where $\mathbf{x}$ is a vector in $\mathbb{R}^n$. The n-sphere is then the set where $x^2 + \|\mathbf{x}\|^2 = 1$. Define the ***north*** and ***south poles*** of S^n to be the points $p_\pm := (\pm 1, \mathbf{0})$, which obviously lie in the sphere, and the ***stereographic projections***

$$\psi_\pm(x, \mathbf{x}) := \frac{\mathbf{x}}{1 \pm x} \tag{4.2b}$$ {mfdSterDef}

to $\mathbb{R}^n$. The function ψ_+ is well-defined everywhere on S^n except at the south pole, which is the only point where $x = -1$, while ψ_- is well-defined everywhere except the north pole. Thus, the coordinate patches of the stereographic projections are the sets $U_\pm := S^n - \{p_\mp\}$ missing only one pole. Moreover, each chart $\psi_\pm$ maps the pole $p_\pm$ it *does* contain to the zero vector $\mathbf{0} \in \mathbb{R}^d$. The image $V_{\pm\mp}$ of the overlap region $U_+ \cap U_-$ in the chart $(\psi_\pm, U_\pm)$ is therefore the annulus $R^n - \{\mathbf{0}\}$. The stereographic projections for S^2 are illustrated in Fig. 4.3.

The two stereographic charts $(\psi_\pm, U_\pm)$ obviously cover the entire sphere. To find the inverses of these coordinate charts, we calculate

$$\boldsymbol{\xi}_\pm := \frac{\mathbf{x}}{1 \pm x} \qquad \Rightarrow \qquad \|\boldsymbol{\xi}_\pm\|^2 = \frac{\|\mathbf{x}\|^2}{(1 \pm x)^2} = \frac{1 - x^2}{(1 \pm x)^2} = \frac{1 \mp x}{1 \pm x} \qquad \Rightarrow \qquad x = \pm\left(\frac{1 - \|\boldsymbol{\xi}_\pm\|^2}{1 + \|\boldsymbol{\xi}_\pm\|^2}\right). \tag{4.2c}$$ {mfdSterPre}

Here, we have defined the n-dimensional vector $\boldsymbol{\xi}_\pm$ to be the value of the coordinate chart $\psi_\pm$ at a particular point of S^n, and solved algebraically for the exceptional variable x describing that point of S^n in terms of this vector. With this value in hand, we can easily solve for the vector part of the original point in S^n:

$$\mathbf{x} = (1 \pm x)\,\boldsymbol{\xi}_\pm = \left(1 + \frac{1 - \|\boldsymbol{\xi}_\pm\|^2}{1 + \|\boldsymbol{\xi}_\pm\|^2}\right)\boldsymbol{\xi}_\pm \qquad \rightsquigarrow \qquad \psi_\pm^{-1}(\boldsymbol{\xi}_\pm) = \left(\pm\frac{1 - \|\boldsymbol{\xi}_\pm\|^2}{1 + \|\boldsymbol{\xi}_\pm\|^2}, \frac{2\,\boldsymbol{\xi}_\pm}{1 + \|\boldsymbol{\xi}_\pm\|^2}\right). \tag{4.2d}$$ {mfdSterInv}

Inserting this result into the *other* coordinate chart, we find the transition functions

$$\boldsymbol{\xi}_\mp = \phi_{\mp\pm}(\boldsymbol{\xi}_\pm) := \psi_\mp \circ \psi_\pm^{-1}(\boldsymbol{\xi}_\pm) = \psi_\mp\left(\pm\frac{1 - \|\boldsymbol{\xi}_\pm\|^2}{1 + \|\boldsymbol{\xi}_\pm\|^2}, \frac{2\,\boldsymbol{\xi}_\pm}{1 + \|\boldsymbol{\xi}_\pm\|^2}\right) = \frac{\dfrac{2\boldsymbol{\xi}_\pm}{1 + \|\boldsymbol{\xi}_\pm\|^2}}{1 - \dfrac{1 - \|\boldsymbol{\xi}_\pm\|^2}{1 + \|\boldsymbol{\xi}_\pm\|^2}} = \frac{\boldsymbol{\xi}_\pm}{\|\boldsymbol{\xi}_\pm\|^2}. \tag{4.2e}$$ {mfdSterTrn}

That is, the transition function $\phi_{\mp\pm}$ maps a coordinate vector $\boldsymbol{\xi}_\pm$ with norm r to a coordinate vector $\boldsymbol{\xi}_\mp$ in the other coordinate system lying in the same direction, but with norm $1/r$. This map is not smooth on all of $\mathbb{R}^d$, but its only singularity is at the origin $\boldsymbol{\xi}_\pm = \mathbf{0}$. This, of course, is precisely the point omitted from the overlap region $V_{\mp\pm}$ within either coordinate system. Thus, the transition functions are smooth everywhere they need to be, and S^n is a smooth manifold. This construction is sketched in Fig. 4.3.

Exercise 4.3: Define the ***orthogonal projection*** coordinate charts

$$\psi_{\pm i}(x^0, \cdots, x^n) := \underbrace{(x^0, \cdots, x^{i-1}, x^{i+1}, \cdots, x^n)}_{\text{missing the } i^{\text{th}} \text{ coordinate}} \tag{4.3a}$$ {mfdOrthProj}

from the hemisphere $U_{\pm i} := \left\{(x^0, \cdots, x^n) \;\middle|\; \pm x^i > 0\right\}$ in S^n to the open unit disk in $\mathbb{R}^n$. Show that this collection of coordinate charts covers S^n and that all transition functions are smooth on the appropriate overlap regions in the coordinate spaces.

Hint: Argue that you need only consider explicitly the overlap of one pair of coordinate charts, say (ψ_{+0}, U_{+0}) and (ψ_{+1}, U_{+1}).

Exercise 4.4: Show that any vector space V of dimension d is trivially a manifold of dimension d.

Hint: Pick any basis to define a *global* coordinate chart, and argue that the resulting manifold structure is independent of the basis chosen.

Exercise 4.5: Show that the ***cylinder***

$$C^2 := \left\{(x, y, z) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 = 1\right\} \tag{4.5a}$$ {mfdCyl2}

is a manifold.

The previous exercise can be done in a variety of ways. However, the easiest is to couple the stereographic projections (4.2b) on S^1, otherwise known as the circle, with the natural coordinate z along the axis of the cylinder. The proof that C^2 is a manifold then follows immediately from that for S^1. This idea extends rather nicely.

Exercise 4.6: Argue that the Cartesian product $M_1 \times M_2$ of any pair of manifolds with dimensions d_1 and d_2, respectively, naturally has the structure of a manifold with dimension $d_1 + d_2$.

Hint: Take charts (x_1, U_1) and (x_2, U_2) on M_1 and M_2 separately, and form the product charts $(x_1 \times x_2, U_1 \times U_2)$ mapping $U_1 \times U_2$ to $\mathbb{R}^{d_1+d_2}$. Show that these cover the product manifold, and argue that the overlap functions must be smooth.

Exercise 4.7: Generalize your proof that C^2 is a manifold to general spaces of the form $\mathbb{R}^m \times S^n$.

Exercise 4.8: Show that the ***n-torus***

$$T^n := \underbrace{S^1 \times \cdots \times S^1}_{n \text{ copies}} \tag{4.8a}$$ {mfdTnDef}

is a manifold. Sketch T^2 as a surface in $\mathbb{R}^3$.
Hint: Use induction on n.

4.2 SMOOTH MAPS

We have made the point several times above that the definition of a manifold M has been chosen deliberately to allow a natural extension of differential calculus to spaces more general than $\mathbb{R}^d$. This generalization uses coordinate charts on M do define the operation of taking derivatives of functions on M, and the smoothness of the transition functions from one chart to another guarantees that the rules for doing calculus in distinct, but overlapping charts, are compatible with one another. This section makes these ideas more precise.

The most elementary analytical question one can ask on $\mathbb{R}^d$ is whether a given function $f(x^1, \cdots, x^d)$ is differentiable. How should we address this issue when $f : M \to \mathbb{R}$ is a function on a manifold? The obvious answer is to turn f into a function on $\mathbb{R}^d$ using some coordinate chart (x, U_x) on M:

$$f : M \to \mathbb{R} \qquad \rightsquigarrow \qquad f_x := f \circ x^{-1} : V_x \subset \mathbb{R}^d \to \mathbb{R}, \tag{4.9}$$ {mfdSFCoord}

where $V_x := x(U_x)$, the image of the coordinate chart in $\mathbb{R}^d$, is open by definition. We say that $f : M \to \mathbb{R}$ is ***differentiable*** at a point $m \in M$ if there exists a coordinate chart (x, U_x) covering m in which the partial derivative

$$\partial_\alpha f := \left.\frac{\partial f_x}{\partial x^\alpha}\right|_{x(m)} \tag{4.10}$$ {mfdSFDiff}

exists for each $\alpha = 1, \cdots, d$. This partial derivative can be calculated straightforwardly since f_x is just a function on $\mathbb{R}^d$. Moreover, although this definition of differentiability seems to prefer one particular coordinate system on M, this is not actually the case. The coordinate expressions for a particular f in various charts on M are related by the transition functions:

$$f_{\tilde{x}} := f \circ \tilde{x}^{-1} = f \circ x^{-1} \circ x \circ \tilde{x}^{-1} = f_x \circ \phi_{x\tilde{x}}. \tag{4.11}$$ {mfdSFTrans}

Thus, using the standard chain rule for partial derivatives on $\mathbb{R}^d$, we immediately find

$$\tilde{\partial}_\alpha f := \left.\frac{\partial f_{\tilde{x}}}{\partial \tilde{x}^\alpha}\right|_{\tilde{x}(m)} = \left.\frac{\partial (\phi_{x\tilde{x}})^\beta}{\partial \tilde{x}^\alpha}\right|_{\tilde{x}(m)} \left.\frac{\partial f_x}{\partial x^\beta}\right|_{x(m)} =: \frac{\partial x^\beta}{\partial \tilde{x}^\alpha}\, \partial_\beta f, \tag{4.12}$$ {mfdSFComp'}

where the Einstein convention has been assumed. In the intermediate step here, we have taken the partial derivative of the scalar function $(\phi_{x\tilde{x}})^\beta$ defined by taking the appropriate component in the x-coordinate space of the vector $\phi_{x\tilde{x}}(\tilde{x})$ resulting from the action of the transition function on a variable point in the $\tilde{x}$-coordinate space. The final result has recast this using a more compact and intuitive notation, which however suppresses the origin of the functional dependence of x^β on $\tilde{x}^\alpha$ via the abstract manifold M. The compact notation is indeed very convenient to use in practice, but it is worthwhile to recognize that its real meaning inheres in the definition of the abstract manifold M support the two coordinate charts in question.

Exercise 4.13: Pick any coordinate chart (x, U) on M, and select any single coordinate x^β therein. Show that x^β defines a smooth function on M at every point of U.
Hint: What is $\partial_\alpha x^\beta$?

Exercise 4.14: Show that (4.12) can be rewritten in the form

$$\tilde{\partial}_\alpha f = \tilde{\partial}_\alpha x^\beta \, \partial_\beta f. \tag{4.14a}$$ {mfdSFComp}

Define the quantities appearing on the right side here carefully on the abstract manifold M before stating and proving the result.

The discussion of differentiability above helps elucidate the definition of a manifold M. As we have discussed above, requiring coordinate charts to extend over open sets effectively eliminates the possibility of exotic points in a chart where the notion of differentiability induced on the manifold by the coordinates would not make proper sense. Meanwhile, the covering condition implies that there is always a coordinate chart available at a given $m \in M$ to discriminate between differentiable functions and non- at that point. Finally, we can now understand the requirement that the transition functions be smooth to imply, via the chain rule, that whenever overlapping coordinate charts offer two possible definitions of differentiability at a point $m \in M$, those definitions *must* agree. Thus, a manifold structure on a given set M is effectively equivalent to a consistent, global specification of a preferred notion of differentiability. In fact, we can go a little further. The transition functions are not merely *differentiable*, but *smooth* on $\mathbb{R}^d$; there is no reason to stop at first derivatives in our discussion. We can argue that if every mixed partial derivative of one coordinate expression $f_x : V_x \to \mathbb{R}$ for a given function $f : M \to \mathbb{R}$ exists at some point $m \in U_x \subset M$, then every mixed partial derivative of the alternate coordinate expression $f_{\tilde{x}} : V_{\tilde{x}} \to \mathbb{R}$ will exist in an overlapping chart $m \in U_{\tilde{x}} \subset M$. The smoothness criterion for overlapping patches can therefore be understood to assert that all coordinate charts on M will agree on the class $C^\infty(M)$ of ***smooth functions*** $f : M \to \mathbb{R}$ whose coordinate expressions are infinitely continuously differentiable. In a loose sense, if one knows the exact set of functions $C^\infty(M)$ that are smooth throughout M, then one knows the manifold structure of M; a manifold *is* a set of smooth functions. This reductionist point of view can be useful in practice since it allows to define other smooth objects on a manifold by the way they interact with smooth functions. However, the reader is cautioned that it may not be literally true in a rigorous mathematical sense that every manifold M determines and is determined by its associated set $C^\infty(M)$ of smooth functions. This, in fact, is a rather deep question related to current research on "non-commutative geometry."

Exercise 4.15: Give a condition for second-differentiablity of $f : M \to \mathbb{R}$ by defining $\partial_\alpha \partial_\beta f$ *á la* (4.10) in a particular coordinate chart (x, U_x). Find the analogue of (4.14a) relating these second derivatives in overlapping coordinate charts.

The discussion above has defined smooth functions $f : M \to \mathbb{R}$ by their interaction with the smooth coordinate charts on the manifold M. A sort of conjugate idea is to look at maps $\sigma : \mathbb{R} \to M$. It is natural to think of such a map as defining a ***parameterized curve*** $\sigma(t) \in M$, and to ask whether or not such a curve is smooth. As in the case of smooth functions, the answer is almost immediate: a curve is smooth when it *looks* smooth in coordinates. That is, given $\sigma : \mathbb{R} \to M$, we define

$$\sigma^x := x \circ \sigma \qquad \text{or} \qquad \sigma^x(t) := x(\sigma(t)) \in \mathbb{R}^d. \tag{4.16}$$ {mfdSCCoord}

By definition, a map $\sigma^x : \mathbb{R} \to \mathbb{R}^d$ is smooth at "time" t_0 if and only the derivatives with respect to t of each component $[\sigma^x(t)]^\alpha$ of the resulting vector in $\mathbb{R}^d$ exist when $t = t_0$. Thus, we *define* $\sigma : \mathbb{R} \to M$ to be a ***smooth curve*** at $t_0 \in \mathbb{R}$ if there exists a coordinate chart (x, U) on M covering $m_0 := \sigma(t_0)$ such that all of the derivatives

$$\overset{n}{\sigma}{}^\alpha := \left. \frac{\mathrm{d}^n [\sigma^x]^\alpha}{\mathrm{d}t^n} \right|_{t_0} \tag{4.17}$$ {mfdSCDiffN}

exist. When only a few derivatives with respect to "time" are taken, we will denote this with the traditional physicists' notation using dot accents: $\dot{\sigma}^\alpha$, $\ddot{\sigma}^\alpha$, etc. As before, it is straightforward to prove that a curve is smooth with respect to one chart covering $m_0 \in M$ if and only if it is smooth with respect to every such chart. Once again, the definition of a smooth manifold is just such that the notions of smooth curves associated with various coordinate charts will agree.

Exercise 4.18: Show that the first derivatives of the coordinate representations of a smooth curve in two charts (x, U) and $(\tilde{x}, \tilde{U})$ that overlap at $\sigma(t_0) \in U \cap \tilde{U} \subset M$ are related by

$$\dot{\tilde{\sigma}}^\beta = \dot{\sigma}^\alpha \, \partial_\alpha \tilde{x}^\beta \qquad \text{where} \qquad \dot{\sigma}^\alpha := \left. \frac{\mathrm{d}[\sigma^x]^\alpha}{\mathrm{d}t} \right|_{t=t_0} \quad \text{and} \quad \dot{\tilde{\sigma}}^\beta := \left. \frac{\mathrm{d}[\sigma^{\tilde{x}}]^\beta}{\mathrm{d}t} \right|_{t=t_0}. \tag{4.18a}$$ {mfdSCComp}

Note the difference between the relationships (4.14a) and (4.18a) among derivatives of coordinate expressions of smooth functions and smooth curves, respectively, in different coordinate charts on M.

Exercise 4.19: Find a relation of the form (4.18a) between the second derivatives $\ddot{\tilde{\sigma}}^\beta$ and $\ddot{\sigma}^\alpha$ of the coordinate representations for a smooth curve $\sigma : \mathbb{R} \to M$ at a point $\sigma(t_0) \in U \cap \tilde{U} \subset M$ of the curve covered by two coordinate charts (x, U) and $(\tilde{x}, \tilde{U})$.

Exercise 4.20: Show that a curve $\sigma : \mathbb{R} \to M$ is smooth if and only if, for every smooth function $f : M \to \mathbb{R}$, the function $f_\sigma := f \circ \sigma$ on $\mathbb{R}$ is smooth.
Hint: Any coordinate chart (x, U) defines a specific collection of smooth functions on $U \subset M$, the coordinate functions themselves.

Exercise 4.21: Show that the quantity

$$\dot{f}_\sigma := \dot{\sigma}^\alpha \, \partial_\alpha f, \tag{4.21a}$$ {mfdSCFdot}

although defined in terms of coordinate derivatives $\dot{\sigma}^\alpha$ and $\partial_\alpha f$ at $m_0 := \sigma(t_0) \in M$, is independent of the coordinates used on M to define it. How would you design a similar coordinate-invariant quantity $\ddot{f}_\sigma$ involving second derivatives of $\sigma(t)$ and $f(m)$?

Hint: Consider the previous exercise.

Suppose we ***reparameterize*** the curve in the above discussion. That is, suppose we change the "time" parameter by setting $t = \rho_{ts}(s)$, where $\rho_{ts} : \mathbb{R} \to \mathbb{R}$ is an invertible function on the real line expressing the original time parameter t in terms of a new one, s. We then can define a new curve $\sigma_s := \sigma_t \circ \rho_{ts}$, or $\sigma_s(s) = \sigma_t(\rho_{ts}(s))$, where we have used the notation σ_t to denote the original parameterization of the curve discussed previously. The image of σ_s in M comprises the same points as that of σ_t, so in a sense they are the same curve. However, the reparamterization will affect the derivatives of σ.

Exercise 4.22: Let $\dot{\sigma}_s^\alpha$ and $\ddot{\sigma}_s^\alpha$ denote the first and second derivatives, respectively, at a particular point $\sigma_s(s)$ of the curve $\sigma_s : \mathbb{R} \to M$. Similarly, let $\dot{\sigma}_t^\alpha$ and $\ddot{\sigma}_t^\alpha$ denote the first and second derivatives at the same point $\sigma_t(\rho_{ts}(s))$ of that same curve in its original parameterization $\sigma_t : \mathbb{R} \to M$. Both are defined with respect to a single coordinate chart (x, U) on M. Use the chain rule to show that

$$\dot{\sigma}_s^\alpha = \dot{\sigma}_t^\alpha \, \dot{\rho}_{ts}(s) := \dot{\sigma}_t^\alpha \, \dot{t}_s. \tag{4.22a}$$ {mfdSCRepar}

Find a similar expression relating $\ddot{\sigma}_s^\alpha$ to $\ddot{\sigma}_t^\alpha$.

Exercise 4.23: Argue that a curve is smooth in one parameterization $\sigma_t : \mathbb{R} \to M$ if and only if it is smooth in all such smooth parameterizations. Thus, a smooth curve σ in M could properly be viewed as a smooth parameterized curve up to reparameterization.

Exercise 4.24: Argue that there is no combination of derivatives of a parameterized curve $\sigma_t : \mathbb{R} \to M$ at one of its points that gives the same value in all possible parameterizations of that curve. That is, there are no reparameterization invariants for a particular smooth curve in a general manifold.

We have now seen two examples of smooth maps, functions $f : M \to \mathbb{R}$ and curves $\sigma : \mathbb{R} \to M$. Both of these are special cases of the more general notion of smooth maps $\psi : M \to N$ between manifolds M and N, where one or the other manifold is taken to be $\mathbb{R}$. However, the results already found indicate the general patterns. We define a ***smooth map*** $\psi : M \to N$ between manifolds as a function between the sets M and N such that there exist coordinate systems (x, U) on M and (y, V) on N with $n := \psi(m) \in N$, in which the mapping $y \circ \psi \circ x^{-1} : \mathbb{R}^{\dim M} \to \mathbb{R}^{\dim N}$ is smooth in the ordinary sense. The existence of one pair of charts covering m and n such that this coordinate function is smooth implies that similar coordinate representations of ψ in any other pair of charts will be smooth as well. The definition of a smooth map is illustrated in Fig. 4.4.

Exercise 4.25: Define the first derivative of a map $\psi : M \to N$ between manifolds by

$$\partial_\alpha^\mu \psi_x^y := \left. \frac{\partial(y^\mu \circ \psi \circ x^{-1})}{\partial x^\alpha} \right|_{x=x(m)}, \tag{4.25a}$$ {mfdSMDiff}

where (x, U) and (y, V) are coordinate charts covering $m \in M$ and $n := \psi(m) \in N$, respectively. Express the analogous derivative $\tilde{\partial}_\gamma^\nu \psi_{\tilde{x}}^{\tilde{y}}$ in alternate charts $(\tilde{x}, \tilde{U})$ and $(\tilde{y}, \tilde{V})$ in terms of the derivatives (4.25a). Show that in the special cases of a smooth function or paramaterized curve, your result reduces to either (4.14a) or (4.18a), respectively. Also show that your results reproduce the reparameterization formulae (4.22a).

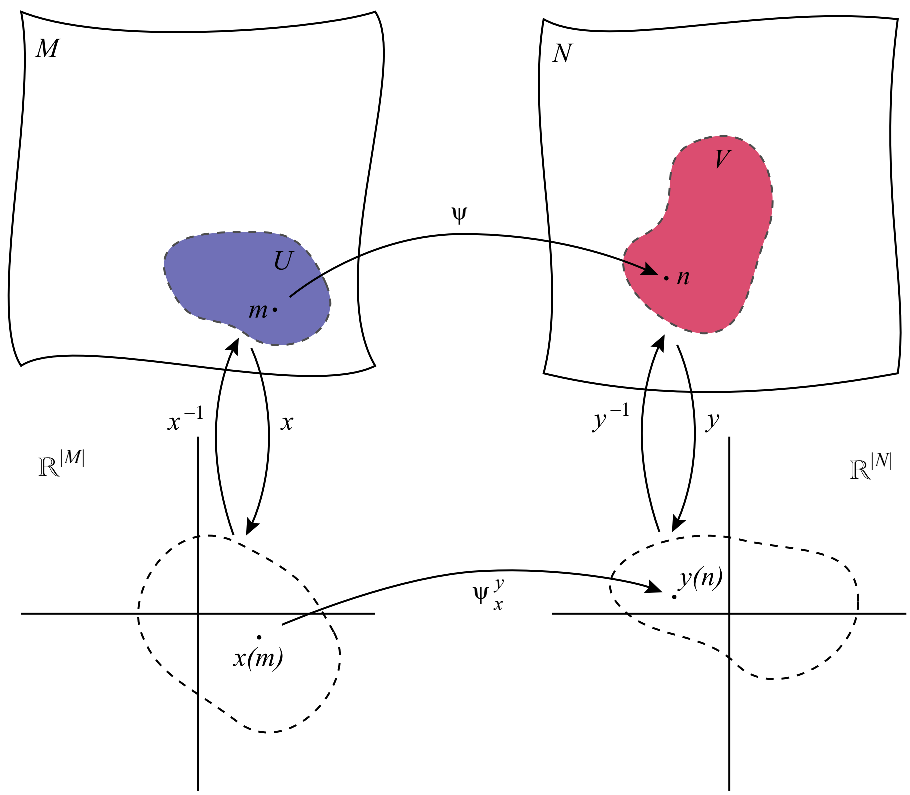


Figure 4.4: A smooth mapping from a manifold M to a manifold N has a smooth coordinate representation $\psi_x^y := y \circ \psi \circ x^{-1}$ from $\mathbb{R}^{|M|}$ to $\mathbb{R}^{|N|}$. Here, (x, U) and (y, V) are coordinate charts covering $m \in M$ and $n := \psi(m) \in N$, respectively. If the coordinate representation is smooth in one such pair of charts, it will be smooth in all others covering those points by the chain rule. {mfdMapping}

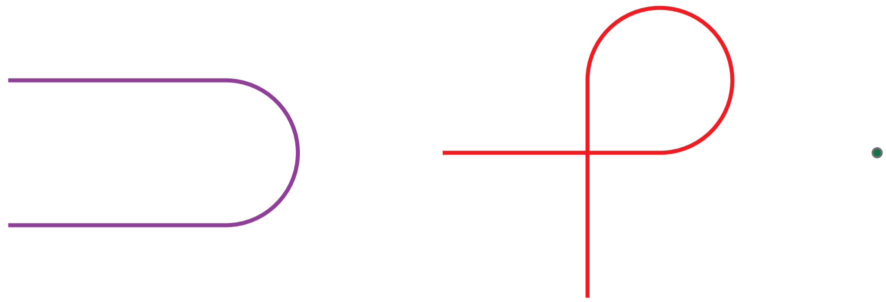


Figure 4.5: The maps $\psi : \mathbb{R} \to \mathbb{R}^2$. (a) The first is an embedding because it is both locally and globally regular. The image in $\mathbb{R}^2$ is a faithful representation of the line. (b) The second is an immersion. It is locally, but not globally, regular since it intersects itself non-trivially and does not faithfully represent the line. (c) The third is not an immersion. It does not represent the line faithfully, even locally. {mfdEmbedR}

Exercise 4.26: Repeat the previous exercise for the second derivative of $\psi : M \to N$, defined by

$$\partial^\mu_{\alpha\beta}\psi^y_x := \frac{\partial^2(y^\mu \circ \psi \circ x^{-1})}{\partial x^\alpha\, \partial x^\beta}\bigg|_{x=a} . \tag{4.26a}$$ {mfdSMDiff2}

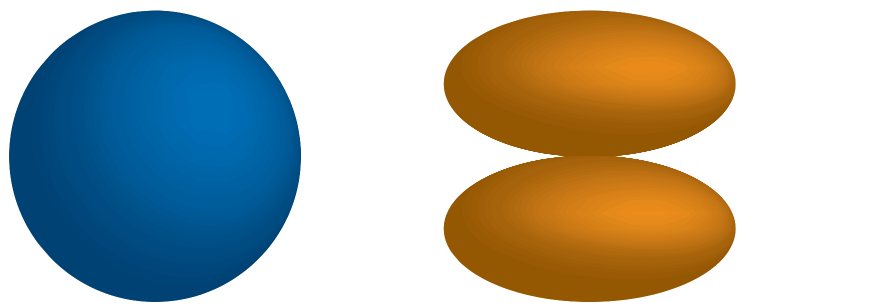
Once again, show that your results reduce to the previous expressions when either M or N is taken to be the real line $\mathbb{R}$.

As with vector spaces, there are several important examples of smooth maps between manifolds. However, there is one important new feature of the discussion, which distinguishes local from global properties of maps. The possibilities are indicated by the examples of Fig. 4.5, showing three possible maps from the line $\mathbb{R}$ to the plane $\mathbb{R}^2$. The first is an ***embedding*** since it is locally regular and globally non-intersecting. The second is an ***immersion*** since it is locally regular as a map from $\mathbb{R}$ to $\mathbb{R}^2$, but not so globally. This immersion does not respect the intrinsic manifold structure of the line since it identifies two *distinct* points. This identification is not detectable locally, for example, to an observer trapped on the line, but only globally to an observer in $\mathbb{R}^2$ or to a collection of observers on the line who can recognize that every smooth function on their line happens to take the same value at what appear *a priori* to be two distinct points. Finally, Fig. 4.5 shows an example of a map taking every point of the line to a single point of $\mathbb{R}^2$. Its is neither an embedding nor an immersion since it does not represent the proper manifold structure of $\mathbb{R}$ within $\mathbb{R}^2$ even locally. Note that this is a maximally degenerate case. A more general example of a non-immersive smooth curve in $\mathbb{R}^2$ could be described, for example, by a particle that moves around for a while before coming to rest for some finite period of time and then resuming motion. The rest period violates the local immersion property of the map $\mathbb{R} \to \mathbb{R}^2$. Finally, and perhaps most importantly, there is the notion of a ***diffeomorphism***. Two manifolds M and N are diffeomorphic if there is a smooth mapping $\psi : M \to N$ that

(a) is one-to-one and onto, and (b) has a smooth inverse ψ^{-1}. Typically, one proves that two manifolds are diffeomorphic by first finding a natural invertible mapping ψ between the them, and then introducing enough coordinate systems to cover both manifolds before checking that both ψ and ψ^{-1} are smooth relative to every pair of such coordinate charts. Like isomorphic vector spaces, diffeomorphic manifolds are essentially the same. Generically, however, diffeomorphic manifolds may not be naturally so.

Exercise 4.27: Show that a curve $\sigma : \mathbb{R} \to M$ is an immersion if and only if its velocity $\dot{\sigma}^\alpha$ vanishes only at isolated times. Does satisfying this condition imply that σ is an embedding?

Exercise 4.28: Describe the following maps from S^2 to $\mathbb{R}^3$:



(a) (b) (c)

Say whether each is an immersion, and embedding, or neither. Is it possible to draw an immersion of S^2 into $\mathbb{R}^3$ that is not an embedding?

Exercise 4.29: Consider the set T^2 of equivalence classes $[x, y]$ of vectors in $\mathbb{R}^2$ generated by the relation

$$(x', y') \sim (x, y) \qquad \Leftrightarrow \qquad x' - x = a \quad \text{and} \quad y' - y = b \quad \text{with} \quad a,\, b \text{ integers.} \tag{4.29a}$$

Show that this space is a manifold, and that it is diffeomorphic to $S^1 \times S^1$.

4.3 TANGENT VECTORS

The Newtonian dynamics of a particle with fixed mass moving around in space relate its acceleration to a force dictated by its position and velocity at a particular time. All of these quantities—position, velocity, acceleration and force—are vectors in the standard Newtonian theory. However, when we generalize to the situation where space is a smooth manifold rather than 3-dimensional Euclidean space, it is not immediately clear in what sense any of these remain vectors. Indeed, the very reason for introducing the notion of a manifold is to allow for situations where the various possible positions of a particle do not form a vector space in any natural sense. Thus, position is certainly not a vector in a general manifold and, if any of the remaining quantities are to remain vectors, we must clearly indicate in what vector space they lie. That space cannot be the manifold itself since, as we have indicated, it is not a vector space. In this section, we will see that the velocity of a particle moving along a parameterized curve through a particular point in a manifold is indeed a vector in a particular vector space naturally associated with that point. However, the other quantities above, most notably the acceleration, are not vectors in that space in any natural sense.

Consider a particle moving along a parameterized path $\sigma(t)$ through a d-dimensional manifold M. If we introduce a chart (x, U) on M, then the particle's velocity at time t relative to those coordinates is given by the derivatives $\dot{\sigma}^\alpha(t)$ of the components of its position vector in the natural basis on $\mathbb{R}^d$. This offers a reasonable way to define the velocity of a particle moving in a general manifold in terms of components relative to coordinate charts. However, we will not follow this line of argument just yet, preferring instead to take a more direct and geometric route to the same destination. Suppose there is a smooth scalar potential $f(m)$ defined throughout M. Then, an observer moving with the particle could measure that potential at the location of the particle at each moment t of time, thereby building the function $f_\sigma(t)$ whose time-derivative $\dot{f}_\sigma(t)$ is given in terms of the components $\dot{\sigma}^\alpha$ of the coordinate velocity by (4.21a). Now, if our observer could do this simultaneously for *every* such scalar potential f, she would have enough information to reconstruct her velocity in the coordinate-component sense from her measurements of various potentials along her trajectory. In particular, applying this procedure to the potential $f(m) := x^\alpha(m)$, which is guaranteed to be a smooth function on M at least in a neighborhood of any point m covered by the chart (x, U), will naturally reproduce the coordinate-velocity component $\dot{\sigma}^\alpha$ in that chart. However, the rule for producing the derivative $\dot{f}_\sigma$ from a given smooth function f on M is rather more general and, most importantly, coordinate-independent. This idea, which was prefigured in the result (4.21a) of an exercise above, generalizes most easily to generic manifolds.

A ***tangent vector*** v on a manifold M is defined to be a mapping from smooth functions f on M to scalars $v(f)$ that

1. is linear in its argument f, so

$$v(\alpha f + \beta g) = \alpha\, v(f) + \beta\, v(g) \tag{4.30}$$ {mfdVFlin}

   for arbitrary pairs of smooth functions f and g and (constant) scalars α and β, and

2. has the ***Leibniz property***

$$v(fg) = f(m)\, v(g) + g(m)\, v(f), \tag{4.31}$$ {mfdVFLieb}

   where $m \in M$ is called the ***base point*** for the tangent vector v on M. This rule relates the value of v on the product function fg on M, which of course must be smooth, to its values on f and g separately and the scalars $f(m)$ and $g(m)$ gotten by evaluating the f and g at the base point m for v.

Together, these conditions on tangent vectors capture the intuitive features of a *first-order directional derivative* on M. Clearly, *any* derivative operator on the space of smooth functions will be linear since every derivative of a constant vanishes. The Leibniz property is what singles out the first-order behavior of the derivative operators we call tangent vectors. It is obviously related to the product rule for derivatives in ordinary differential calculus.

Exercise 4.32: Show that the linear and Leibniz properties of a tangent vector v based at $m \in M$ imply that $v(f)$ vanishes for any smooth function f on M that is constant throughout some open neighborhood of m. We describe this property of v by saying it ***annihilates constants***.

Hint: Suppose that $f(m') = 1$ for every $m' \in M$ in an open neighborhood of m, and show that $f^2(m') = f(m')$ within that neighborhood. Then, calculate the Leibniz rule and use linearity to establish the result.

Exercise 4.33: Show that any parameterized curve $\sigma : \mathbb{R} \to M$ defines a family of tangent vectors $\dot{\sigma}(t)$, one based at each point $\sigma(t) \in M$ along the curve, given by

$$[\dot{\sigma}(t)](f) := \left. \frac{\mathrm{d} f(\sigma(t'))}{\mathrm{d}t'} \right|_{t'=t}. \qquad (4.33\text{a}) \quad \{\texttt{mfdSCvel}\}$$

We call this tangent vector the ***velocity*** of the curve at time t.

Exercise 4.34: Let $\sigma : \mathbb{R} \to M$ again be a parameterized curve in a manifold M, and define its ***acceleration*** at time t by

$$[\ddot{\sigma}(t)](f) := \left. \frac{\mathrm{d}^2 f(\sigma(t'))}{\mathrm{d}t'^2} \right|_{t'=t}. \qquad (4.34\text{a}) \quad \{\texttt{mfdSCacc}\}$$

Show that the acceleration is *not* a tangent vector at $\sigma(t) \in M$.

Hint: Is this mapping of smooth functions to scalars linear? Is it Leibniz?

Exercise 4.35: Show that the ***position*** $\sigma(t)$ of a parameterized curve at time t naturally defines a linear map from smooth functions on M to scalars, but that this mapping is not Leibniz and therefore not a tangent vector on M.

We have seen in the first exercise above that the velocity of a curve passing through a given point of a manifold is naturally a tangent vector at that point. We could have *defined* tangent vectors at a given $m \in M$ in terms of such velocities, rather than as first-order directional derivatives. However, the correspondence between parameterized curves and tangent vectors obviously should not be one-to-one; many parameterized curves passing through m will have the same velocity at that point.

Exercise 4.36: Formulate an equivalence relation among parameterized curves passing through a common point$m \in M$ that equates curves with the same velocity at that point. Show that the corresponding equivalence classes of parameterized curves are naturally in one-to-one correspondence with tangent vectors in the directional-derivative sense given above.

Hint: This should not be difficult. The equivalence relation in question *is*, more or less, that the intrinsic derivatives along two equivalent parameterized curves take the same value for *every* smooth function on M.

The set $T_m M$ of tangent vectors based at a given point $m \in M$ naturally form a vector space, called the ***tangent space*** to M at m. Linear combinations in this space are defined in the obvious way:

$$[\alpha v + \beta w](f) := \alpha\, v(f) + \beta\, w(f). \qquad (4.37) \quad \{\texttt{mfdTVlcDef}\}$$

We have seen many times that such a linear combination of linear functionals on a vector space, which the set $C^\infty(M)$ of smooth functions on M certainly is, is again a linear functional on that vector space. However, to be a tangent vector, we must show that this particular linear combination is also Leibniz:

$$\begin{aligned}[\alpha v+\beta w](fg) &= \alpha\, v(fg)+\beta\, w(fg)\\ &= \alpha[f(m)\, v(g)+g(m)\, v(g)]+\beta\,[f(m)\, w(g)+g(m)\, w(f)]\\ &= f(m)\,[\alpha v+\beta w](g)+g(m)\,[\alpha v+\beta w](f).\end{aligned} \tag{4.38}$$ {mfdTVlcLieb}

We have used the Leibniz properties of v and w separately here to expand the action on the product function in the intermediate step here before rearranging and collecting terms to prove the Leibniz property of the linear combination. Thus, as advertised, the set of tangent vectors based at a common of M is naturally a vector space. Note, however, that if we consider tangent vectors v and w based at *different* points of M, then their linear combination is certainly *not* a tangent vector on M; the critical rearrangement of terms used to prove the Leibniz property of the linear combination in (4.38) will be impossible.

The above result, that T_mM is a vector space, is interesting in its own right, but becomes much more so when we observe that it is *finite-dimensional*. This may seem rather surprising at first since the original vector space $C^\infty(M)$ is clearly infinite-dimensional, and thus the set of *all* linear functions on it will be so as well. It is the Leibniz rule that whittles the enormous dual space $[C^\infty(M)]^*$ down to its much better-behaved, finite-dimensional subspace T_mM. We see this tremendous simplification by actually constructing a basis on T_mM. Let (x, U) be any coordinate chart covering $m\in M$, and define the ***coordinate basis*** vectors ∂_α by

$$\partial_\alpha(f) := \left.\frac{\partial(f\circ x^{-1})}{\partial x^\alpha}\right|_{x(m)}. \tag{4.39}$$ {mfdTScbas}

That is, we turn the function $f: M\to\mathbb{R}$ into a function of coordinates using the given chart, and then take the partial derivative of the resulting function on $\mathbb{R}^d$ along one of the coordinate axes. Note the similarity to (4.10). It is easy to see that this procedure defines a linear, Leibniz operator, a tangent vector, at m. To show that these vectors actually form a basis, we must show that they are both linearly independent and span T_mM. To show linear independence, we suppose a non-trivial family of scalars v^α exists such that

$$[v^\alpha\,\partial_\alpha](f) := v^\alpha\,\partial_\alpha(f) = 0 \qquad \forall f\in C^\infty(M). \tag{4.40}$$ {mfdTScbasLI}

However, we can choose f to be one of the coordinates here and easily calculate $\partial_\alpha(x^\beta)=\delta^\beta_\alpha$ to show that (4.40) implies $v^\beta=0$. Thus, the ∂_α are linearly independent. To show that they also span T_mM, we take an arbitrary smooth function $f: M\to\mathbb{R}$ and map it over using the chart (x, U) to give the coordinate representation function $F := f_x := f\circ x^{-1}$ of (4.9). Let $a := x(m)$ be the point of $\mathbb{R}^d$ corresponding to $m\in M$ in these coordinates, choose an open ball of points surrounding a that lies entirely within the image $x(U)$ of the chart in $\mathbb{R}^d$, and let b be any point of that ball. The definition of a coordinate chart guarantees that such a ball exists, and the ordinary properties of open balls in $\mathbb{R}^d$ imply that it contains every point on the straight line in $\mathbb{R}^d$ connecting a to b. Parameterize this line to have constant velocity in the coordinate space and use the fundamental theorem of calculus to find

$$F(b) - F(a) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\big[F((1-t)\,a + t\,b)\big]\,\mathrm{d}t = (b^\alpha - a^\alpha)\underbrace{\int_0^1 \frac{\partial F}{\partial x^\alpha}((1-t)\,a + t\,b)\,\mathrm{d}t}_{:=G_\alpha(b)}. \tag{4.41}$$ {mfdTScbasSp'

The integrand in the final result denotes the algebraic expression for the partial derivative of $F : \mathbb{R}^d \to \mathbb{R}$ with respect to its α^{th} argument, evaluated at the indicated position along the line. The key point is that, mapping everything back over to the manifold, we have proved the existence of smooth functions $g_\alpha := G_\alpha \circ x$ on M, defined at all points $p := x^{-1}(b)$ in an open neighborhood of $m \in M$, such that

$$f(p) = f(m) + (x^\alpha(p) - x^\alpha(m))\,g_\alpha(p). \tag{4.42}$$ {mfdTScbasSp}

From this, the following exercise allows us to conclude that every tangent vector v at m can be written as a linear combination of the coordinate basis vectors ∂_α dictated by the coordinates (x, U). That is, the ∂_α form a basis for T_mM, which therefore must have dimension equal to that of the manifold M itself.

Exercise 4.43: Let $v \in T_mM$ be an arbitrary tangent vector at $m \in M$, and let f be an arbitrary smooth function on M. Show that

$$v(f) = v(x^\alpha)\,g_\alpha(m) = v(x^\alpha)\,\partial_\alpha(f), \tag{4.43a}$$ {mfdTScbasExp

where the functions g_α are defined on M as in (4.42). Conclude that we can express v as the linear combination $v = v^\alpha\,\partial_\alpha$, where $v^\alpha := v(x^\alpha)$ are the ***coordinate components*** of v in the chart (x, U).

Hint: Establish the first equality using the Leibniz property of v. The second follows from the Leibniz property of ∂_β.

Exercise 4.44: Find the change-of-basis matrix relating the coordinate bases ∂_α and $\tilde{\partial}_\alpha$ at each point of a pair of overlapping coordinate charts (x, U) and $(\tilde{x}, \tilde{U})$. Calculate the relation between the components v^α and $\tilde{v}^\alpha$ of a given $v \in T_mM$ in these bases.

Hint: See (4.14a).

The above discussion has shown that the tangent space T_mM at a given point of a d-dimensional manifold M is naturally a vector space of dimension d. It is perhaps worthwhile to digress for a moment here and consider why the acceleration (4.34a) of a given parameterized curve in M should not be a tangent vector. That it *is* not is perfectly clear from the exercise above, but why it *must* not be is perhaps not so clear. In ordinary Newtonian mechanics, the acceleration is calculated by taking an additional derivative of the velocity, considered as a vector-valued function of time. However, while the velocity of a parameterized curve $\sigma(t)$ in a manifold M is certainly defined as a function of time, its vector values at various moments belong to *different* vector spaces $T_{\sigma(t)}M$. These vector spaces all have the same finite dimension d, and therefore are isomorphic, *but not naturally so*. Indeed, as we have argued after (4.38) above, a linear combination of tangent vectors based at *different* points of M *cannot* be identified with a tangent vector at *any* point of

M. As we will see in the next lecture, the task of picking a preferred isomorphism between the various vector spaces tangent to a given manifold M, and therefore of equating the acceleration of a curve with a tangent vector on M, is performed by a *physical* entity: the gravitational field.

4.4 TENSOR FIELDS

A ***vector field*** on a manifold M is a function assigning a particular vector $v(m) \in T_m M$ to each point $m \in M$. Note that the vector $v(m)$ assigned to a given point is required to be based at that point. This is consistent with our ordinary notion of a vector field. The electric field, for example, is a vector field over space that describes the force on a momentarily stationary charged particle at a particular position. When space is a Euclidean vector space, it is still natural to think of this force vector as though it were based at the position of the charge being acted upon, even though it is not strictly necessary to do so. In the context of manifolds, there is no option. We *define* a vector field v such that its vector value $v(m)$ at a given point belongs to the tangent space at that point.

The laws of electromagnetism, for example, require that we be able to differentiate vector fields. Thus, to do physics on manifolds, we must be able first of all to say which vector fields are differentiable, or smooth. We have classified the smooth functions f on a manifold above using coordinate charts, and could do the same here for vector fields. However, a simpler option is available. For each point $m \in M$ of the manifold, a vector field gives a vector $v(m)$ based at that point, which then may be applied to a general smooth function f. The result is a number. As we vary the point m, and therefore the vector $v(m)$, this number will vary, and we will find a function

$$v(f) : m \mapsto [v(m)](f). \tag{4.45}$$ {mfdVFact}

A ***smooth vector field*** v on M has the additional property that the function $v(f)$ defined on M in this way is *smooth* for every smooth function $f \in C^\infty(M)$. Thus, the set of smooth vector fields on a manifold is dictated by its set of smooth functions.

Exercise 4.46: Define the coordinate basis vector ∂_α at each point $m \in U$ in the domain of a coordinate chart (x, U) on M, as in (4.39). Show that, if one varies m, these vectors form a smooth vector field over U.

Exercise 4.47: Let v be a smooth vector field, and let (x, U) be a coordinate chart on a manifold M. Use (4.43a) to expand v into components relative to the coordinate basis at each point of U. Show that the component functions $v^\alpha(m)$ vary smoothly with $m \in U$.
Hint: The coordinates x^α are themselves smooth functions on U.

Exercise 4.48: Every smooth vector field v on a manifold M defines a map from the space of smooth functions $C^\infty(M)$ to itself according to (4.45). Show that this map is necessarily both linear and Leibniz:

$$v(\alpha f + \beta g) = \alpha\, v(f) + \beta\, v(g) \qquad \text{and} \qquad v(fg) = f\, v(g) + g\, v(f), \tag{4.48a}$$ {mfdVFderiv}

where f and g are arbitrary smooth functions on M, while α and β are arbitrary (constant) scalars. Mathematicians refer to a map $v : C^\infty(M) \to C^\infty(M)$ with these two properties as a ***derivation*** on the ring of smooth functions. The terminology is directly related to the intuitive picture of a tangent vector as a first-order directional derivative operator. Show that every such derivation on $C^\infty(M)$ defines a smooth vector field on M.

Hint: How do you show that a derivation defines a tangent vector $v(m)$ is each tangent space $T_m M$?

In our discussions of vector spaces V above, we found it very useful to define and explore the dual vector space V^*. In the theory of manifolds developed above, we have found a natural vector space $T_m M$ of tangent vectors at each point $m \in M$. The dual spaces to these are naturally called the ***co-tangent spaces*** on M, and are denoted $T^*_m M$. That is, ***co-tangent vector*** $\omega \in T^*_m M$ is a linear map from vectors $v \in T_m M$ to scalars $\omega(v)$. We have reviewed before just why such objects are generally interesting, and will not recapitulate those arguments here. Rather, we move straight on to the issue of defining co-vector *fields* on a manifold M. Obviously, a co-tangent vector field will assign a co-tangent vector $\omega(m) \in T^*_m M$ to each point $m \in M$ of our manifold, and this co-vector knows how to act on vectors $v \in T_m M$ to produce a number. A ***smooth co-vector field*** on a manifold M has the property that, for every smooth vector *field* v on M, the *function*

$$\omega(v) : m \mapsto [\omega(m)](v(m)) \tag{4.49}$$ {mfdCFact}

is smooth. That is, the value of the function $\omega(v)$ at a given point $m \in M$ is defined to be the contraction of the co-vector $\omega(m) \in T^*_m M$ with the vector $v(m) \in T_m M$. Thus, the set of smooth co-vector fields on M is fixed by the sets of smooth vector fields and smooth functions.

Exercise 4.50: Let f be a smooth function on M, and define a co-vector field $\mathrm{d}f$ by

$$[\mathrm{d}f](v) := v(f). \tag{4.50a}$$ {mfdCFgrad}

That is, the contraction of the co-vector $\mathrm{d}f$ with a vector v at any point of M is simply the value of the function $v(f)$ at that point. Show that $\mathrm{d}f$, which we call the ***gradient*** of f, is a smooth co-vector field.

Hint: To show that $\mathrm{d}f$ is a co-vector at each point $m \in M$, that is that it is linear on $T_m M$, recall the definition (4.37) of the vector structure of the tangent space. To show that $\mathrm{d}f$ is a *smooth* co-vector field is then almost immediate.

Exercise 4.51: Show that the basis on $T^*_m M$ dual to the coordinate basis ∂_α induced on $T_m M$ by a given coordinate chart (x, U) on M consists precisely of the gradients $\mathrm{d}x^\alpha$ of the coordinate functions at m.

Hint: Recall the definitions (2.45) of the dual basis for a general vector space and (4.43a) of the coordinate components of a vector $v \in T_m M$.

Exercise 4.52: Show that any smooth co-vector field ω on a manifold M can be expanded in the form

$$\omega = \omega_\alpha \,\mathrm{d}x^\alpha \qquad \text{with} \qquad \omega_\alpha := \omega(\partial_\alpha) \tag{4.52a}$$ {mfdCFbasExp}

throughout the domain U of a given coordinate chart (x, U) on M. We refer to the values $\omega_\alpha(m)$ at each point $m \in U$ as the ***coordinate components*** of the co-vector $\omega(m)$ at that point. Show that, as we vary $m \in U$, each coordinate component ω_α of a smooth co-vector field ω describes a smooth function.

Exercise 4.53: Show that the change-of-basis formula relating the coordinate components of a given co-vector $\omega \in T^*_n M$ in two overlapping coordinate systems (x, U) and $(\tilde{x}, \tilde{U})$ at $m \in M$ can be written in the form

$$\tilde{\omega}_\alpha = \tilde{\partial}_\alpha(x^\beta)\,\omega_\beta. \tag{4.53a}$$

Compare this result with your formula in the case of a vector field.

A smooth co-vector field gives a map from the space of smooth vector fields on a manifold M to its space of smooth functions. But is every such map associated with a smooth co-vector field? That is, suppose we have a mapping $F : v \mapsto F(v)$ taking smooth vector fields to smooth functions on M. Is there necessarily a smooth co-vector field ω such that $\omega(v) = F(v)$ as functions on M? The answer is certainly no. A co-vector at a point depends linearly on its vector argument, and the function F need not.

4.5 LIE DERIVATIVES AND THE DIFFEOMORPHISM GROUP

push-forward

Lecture 5

Covariant Derivatives

5.1 PARALLEL TRANSPORT