

Linear Algebra Prerequisites for Studying Brownian Motion of Manifolds

Jean-Jacques St Leroux

May 2026

1 Introduction

Linear algebra is a crucial component for studying differential geometry, hence we shall cover its most relevant components in this document. This document assumes a knowledge of linear algebra topics that one who has taken introductory linear algebra would have. The topics covered will be as follows: In Section 2, we recap vector spaces and subspaces. In Section 3, we shall cover inner products and dot products, norms, orthogonality, projection, and orthogonal complements, and conclude with projection onto a subspace with an orthonormal basis. In Section 4, we discuss matrices and shall cover determinants, eigenvalues and eigenvectors, matrix representation of linear maps, symmetric matrices, positive definite matrices, and changes of basis. In Section 5, we tie things together with quadratic forms and conclude with the metric tensor (the first fundamental form).

Throughout, we will work over the field of real numbers $\mathbb{R}$, as that is the setting relevant to the differential geometry of real manifolds. Where a definition or result extends naturally to the complex case, we shall note it briefly, but we will not dwell on it.

2 Vector Spaces and Subspaces

I will assume that you have some linear algebra, so we won't build up our intuition from the ground-up on vector spaces and subspaces. Instead we'll use it to lay the foundation of the later sections.

2.1 Vector Spaces

A vector v is simply an element of a vector space, V . What is a vector space, then? Well, a vector space is a nonempty set of objects (vectors) equipped with two operations – addition and *scalar* multiplication – that satisfy the axioms below. Note that a vector space is defined on a field K , where a field is a set equipped with two binary operations – addition and multiplication – such

that both operations are associative and commutative, both have identity elements (typically denoted 0 and 1, with $0 \neq 1$), every element has an additive inverse, every nonzero element has a multiplicative inverse, and multiplication distributes over addition. In short, a field is a set on which one can add, subtract, multiply, and divide (by nonzero elements) the way one is used to. The most straightforward example of a field is the real numbers, $\mathbb{R}$, so you can read the below axioms with that in mind. Just know you can also define vector spaces on other fields, too, like the complex numbers $\mathbb{C}$ and the rational numbers $\mathbb{Q}$. We call the vector space over a field K a K -vector space.

For addition, the axioms are:

1. **Closure Under Addition.** If $v, w \in V$, then $v + w \in V$.
2. **Commutative Property of Addition.** For $v, w \in V$, it follows $v + w = w + v$.
3. **Associative Property of Addition.** For $v, w, x \in V$, it follows $(v + w) + x = v + (w + x)$.
4. **The Additive Identity.** There exists an element $0 \in V$ such that for any vector $v \in V$, $v + 0 = v$.
5. **The Additive Inverse.** For each vector $v \in V$ there exists a vector $v' \in V$ such that $v + v' = 0$. It follows $v' = (-v)$.

And for scalar multiplication, the axioms are:

1. **Closure Under Scalar Multiplication.** For an element $a \in K$ and a vector $v \in V$, $av \in V$.
2. **Distributivity of Scalar Multiplication on Vector Addition.** For an element $a \in K$ and vectors $v, w \in V$, it follows $a(v + w) = av + aw$.
3. **Distributivity of Scalar Multiplication on Field Addition.** For elements $a, b \in K$ and a vector $v \in V$, it follows $(a + b)v = av + bv$.
4. **Field Multiplication Compatibility.** For $a, b \in K$ and $v \in V$, it follows $a(bv) = (ab)v$.
5. **The Multiplicative Identity.** For the multiplicative identity $1 \in K$ and every vector $v \in V$, $1v = v$.

A few useful consequences follow at once from these axioms. The additive identity $0 \in V$ is unique (if 0 and $0'$ both worked, then $0 = 0 + 0' = 0'$), and so is the additive inverse of each vector. Moreover, $0_K v = 0_V$ for every $v \in V$ (since $0_K v = (0_K + 0_K)v = 0_K v + 0_K v$, and cancelling gives the result), and $(-1)v = -v$. We will use these facts without further comment.

As an exercise, you may wish to prove that $\mathbb{R}^n$, or the space of n -dimensional real-valued vectors, is itself a vector space (spoiler alert, it is!). $\mathbb{R}^n$, of course, is the most traditional vector space that one encounters, and is the vector space we work with for the sake of our project.

2.2 Subspaces and Bases

With a vector space, we can define a few different properties. All of the definitions below build off each other, so we introduce them in order, and do so briefly and mostly informally.

- **Linear combinations.** Given a finite set $G = \{g_1, g_2, \dots, g_m\}$ of vectors in a K -vector space V , a linear combination of these vectors in G is an element of V of the form $a_1g_1 + a_2g_2 + \dots + a_mg_m$ where $a_1, a_2, \dots, a_m \in K$ are scalars called the coefficients of the linear combination. You may be familiar with the fact that we can construct a system of linear equations, where we form a nonempty set of linear combinations on the same set of vectors in G , such that each linear combination is set equal to an element of the field K . We can then construct the *coefficient matrix* of this system, organizing the coefficients into rows and columns.
- **Linear independence.** The vectors of a finite subset $G = \{g_1, \dots, g_m\}$ of a K -vector space V are said to be linearly independent if no element of G can be written as a linear combination of the other elements of G . Equivalently, the vectors are linearly independent if a linear combination $a_1g_1 + a_2g_2 + \dots + a_mg_m$ results in the zero vector if and only if all its coefficients are zero. A set that is not linearly independent is called linearly dependent.
- **Linear subspace.** A linear subspace W of a vector space V is a nonempty subset of V that is closed under the addition and scalar multiplication of V . Equivalently, W is a subspace if and only if W is nonempty and $av + bw \in W$ for every $v, w \in W$ and $a, b \in K$. Therefore, every linear combination of (finitely many) vectors in W is also an element of W . Note that every subspace contains the zero vector, since taking $a = b = 0$ with any $v, w \in W$ gives $0 \in W$.
- **Linear span.** Given a subset G of a vector space V , the linear span of G , denoted $\text{span}(G)$, is the set of all (finite) linear combinations of vectors in G . Equivalently, and perhaps more elegantly, it is the smallest linear subspace of V containing G , that is, the intersection of all linear subspaces of V that contain G . The two descriptions describe the same set.
- **Basis and Dimension.** A subset B of a vector space V is a basis of V if its elements are linearly independent and span V . Equivalently, B is a basis if every $v \in V$ can be written uniquely as a finite linear combination of elements of B . Any two bases of a vector space have the same cardinality; this common cardinality is the dimension of the vector space, written $\dim V$. We will be concerned with finite-dimensional vector spaces, and most often with $\mathbb{R}^n$, whose dimension is n as witnessed by the standard basis $\{e_1, \dots, e_n\}$.

3 Inner Products and Projections

Surely, if you have taken a linear algebra or multivariable calculus you have encountered the dot product. However, the dot product is a specialized version of an inner product, which you may too already know. We will recap it in this section, and also discuss orthogonality and projections of vectors with the notions of inner products.

3.1 What is an Inner Product?

An inner product is a function that takes in a pair of vectors v, w in the K -vector space V and outputs a scalar. Here, the field K is restricted to being the real or complex numbers; we will focus on the case $K = \mathbb{R}$, so the output is real. The inner product is denoted here as $\langle v, w \rangle$. Like vector spaces, the inner product must satisfy a set of axioms. For $K = \mathbb{R}$, these axioms are as follows:

1. $\langle v, w \rangle \in \mathbb{R}$ for all $v, w \in V$.
2. $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$.
3. $\langle v + u, w \rangle = \langle v, w \rangle + \langle u, w \rangle$ for all $u, v, w \in V$.
4. $\langle v, w + u \rangle = \langle v, w \rangle + \langle v, u \rangle$ for all $u, v, w \in V$.
5. $\langle rv, w \rangle = r\langle v, w \rangle$ for all $r \in \mathbb{R}$ and $v, w \in V$. By symmetry, $\langle v, rw \rangle = r\langle v, w \rangle$ as well.
6. $\langle v, v \rangle \geq 0$, with equality if and only if $v = 0$.

Three important properties follow from these axioms. Axioms 3, 4, and 5 constitute the property of the inner product being *bilinear* (linear in its first and second argument by both addition and scaling). Axiom 2 constitutes the property of the inner product being *symmetric* (regardless of which vector comes first in the argument, the inner product is the same). Axiom 6 constitutes the property of the inner product being *positive definite* (the inner product can only be zero if its arguments are zero – it can't be zero with nonzero arguments). One should observe that, on a real vector space, symmetry (Axiom 2) together with linearity in the first argument (Axioms 3 and 5) already forces linearity in the second argument (Axiom 4); we list Axiom 4 separately for emphasis, and so that the list extends without surprise to the complex case, where Axiom 2 is replaced by conjugate symmetry $\langle v, w \rangle = \overline{\langle w, v \rangle}$ and the scaling axiom becomes $\langle rv, w \rangle = r\langle v, w \rangle$ together with $\langle v, rw \rangle = \bar{r}\langle v, w \rangle$.

Another interesting notion arising from inner products is an **inner product space**. A vector space is a set with two operations (vector addition and scalar multiplication) satisfying certain axioms. It has no notion of length, angle, or distance built in; it only knows linear combinations. When you say a vector space is equipped with an inner product, you are adding a third operation — a map $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$ (where F is $\mathbb{R}$ or $\mathbb{C}$) — that satisfies the inner product

axioms. You now have a set with three operations that all coexist and interact. Intuitively, we can see an inner product space is just a vector space equipped with an inner product.

3.2 The Dot Product (algebraically)

As mentioned above, the dot product is a type of inner product. Algebraically, the dot product is the sum of the products of the corresponding entries of the two sequences of numbers. By the coordinate definition, the dot product of a vector pair $a, b \in V$, specified on some basis, is defined algebraically as

$$a \cdot b = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n,$$

where n is the dimension of the vector space and a_i, b_i are the components of a, b in the chosen basis. If vectors are specified as column vectors, then one may take the product of the transpose of one vector with the other to obtain a 1×1 matrix, which corresponds to a scalar value on K . This matrix product is notated by $a^\top b$. Of course, the properties of inner products apply to the dot product: it is bilinear, symmetric, and positive definite. We will verify all of these properties here.

The dot product is bilinear. The dot product is linear in its first argument, as

$$\begin{aligned} (a + ru) \cdot b &= \sum_{i=1}^n (a_i + ru_i) b_i \\ &= \sum_{i=1}^n a_i b_i + r \sum_{i=1}^n u_i b_i \\ &= a \cdot b + r(u \cdot b). \end{aligned}$$

The argument in the second argument is identical (or, alternatively, follows from symmetry, established next).

The dot product is symmetric. We work through the case on real vector spaces and do not address conjugate symmetry. It follows

$$a \cdot b = \sum_{i=1}^n a_i b_i = \sum_{i=1}^n b_i a_i = b \cdot a$$

by commutativity of multiplication in the field.

The dot product is positive definite. Consider the dot product $a \cdot a$. Then,

$$a \cdot a = a_1 a_1 + a_2 a_2 + \cdots + a_n a_n = a_1^2 + a_2^2 + \cdots + a_n^2,$$

which is a sum of squares of real numbers and is therefore non-negative. If $a = 0$, the sum of zeros is zero. Conversely, if $a \cdot a = 0$, then each summand a_i^2 must individually be zero (since a sum of non-negative reals is zero only if every term is zero), so $a_i = 0$ for every i , which means $a = 0$.

3.3 The Dot Product (geometrically)

In Euclidean geometry, a vector is a geometric object in Euclidean space that possesses both a magnitude and a direction. A vector can be pictured as an arrow. Its magnitude is its length, and its direction is the direction to which the arrow points. We consider the absolute value of the magnitude, denoting the magnitude of a vector v as $\|v\|$.

Consider two vectors a, b which start at the same point. The line connecting the tips of the two vectors is given by $a - b$, which forms a triangle. The angle θ between a and b can be solved for using the law of cosines, namely, $\|a - b\|^2 = \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos(\theta)$. Expanding the left-hand side algebraically (using the bilinearity and symmetry we just established), we have:

$$\begin{aligned}(a - b) \cdot (a - b) &= \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos(\theta) \\ a \cdot a - 2(a \cdot b) + b \cdot b &= \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos(\theta) \\ \|a\|^2 - 2(a \cdot b) + \|b\|^2 &= \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos(\theta) \\ a \cdot b &= \|a\|\|b\|\cos(\theta).\end{aligned}$$

The dot product induces what is called the **norm**. Consider the case where both arguments are the vector a . The angle between these vectors is zero, and $\cos(0) = 1$. Thus we have $a \cdot a = \|a\|^2$, or $\sqrt{a \cdot a} = \|a\|$. $\|a\|$ is called the norm of the vector a .

Precisely, a norm on a real or complex vector space V is a function $\|\cdot\| : V \rightarrow \mathbb{R}_{\geq 0}$ that behaves in certain ways like the distance from the origin: it is absolutely homogeneous ($\|rv\| = |r|\|v\|$ for any scalar r), it obeys the triangle inequality ($\|v + w\| \leq \|v\| + \|w\|$), and it vanishes only at the origin ($\|v\| = 0$ if and only if $v = 0$). Every inner product induces a norm via $\|v\| := \sqrt{\langle v, v \rangle}$, and the dot product induces the familiar Euclidean norm $\|v\| = \sqrt{v_1^2 + \cdots + v_n^2}$ on $\mathbb{R}^n$. The triangle inequality for this induced norm is a consequence of the Cauchy-Schwarz inequality $|\langle v, w \rangle| \leq \|v\|\|w\|$, which one may verify by expanding $\|v - tw\|^2 \geq 0$ in t and choosing t to minimize the resulting quadratic.

3.4 Orthogonality

By the previous result, we may conclude $\cos(\theta) = \frac{a \cdot b}{\|a\|\|b\|}$ whenever a and b are nonzero. If the angle θ between these vectors is $\frac{\pi}{2}$ radians, they are perpendicular, or more generally they are **orthogonal** on the bilinear form given by the dot product. An interesting property rises from these statements: two nonzero vectors are orthogonal if and only if their dot product is zero.

Proof: If nonzero vectors a, b are orthogonal, then $\theta = \pi/2$, so $\cos(\theta) = 0$, and it is clear from $a \cdot b = \|a\|\|b\|\cos(\theta)$ that $a \cdot b = 0$. Conversely, suppose $a \cdot b = 0$ with both a, b nonzero. Then $\|a\|\|b\|\cos(\theta) = 0$, and since $\|a\|, \|b\| > 0$, we have $\cos(\theta) = 0$. Since the range of θ is $[0, \pi]$, we have $\theta = \frac{\pi}{2}$, and the vectors are thus orthogonal. $\square$

Note this proof does not account for the zero vector, but doing so is considered unnecessary as the zero vector is orthogonal to every vector ($0 \cdot v = 0$ for

every v). More generally, given any inner product $\langle \cdot, \cdot \rangle$ on a vector space V , we say v and w are orthogonal (with respect to that inner product) if $\langle v, w \rangle = 0$. The dot product case is then the special case where $V = \mathbb{R}^n$ and the inner product is the standard dot product. This more general viewpoint will be important when we get to the metric tensor.

3.5 Orthogonal Projection

Given a vector x and a subspace W , what is the "closest" point in W to x ? This is an interesting question, and we can use the familiar definitions of projection to answer it.

We have discussed that for a dot product, $a \cdot b = 0$ means the two vectors are perpendicular — they share no component in common. One can turn this into a definition of "how much of a component is there in common" by using scalar projection, which as you may know is given by $\text{comp}_b a = \frac{a \cdot b}{\|b\|}$, read as "the scalar component of the projection of a onto b ". If one wanted to know the vector representation of this value, they would simply multiply the value by the unit vector in the direction of b (i.e. $b/\|b\|$), which yields the presumably familiar expression $\text{proj}_b a = \frac{a \cdot b}{\|b\|^2} b$, read as "the vector projection of a onto b ". Scalar/vector projection is typically introduced geometrically as "the shadow of a onto b ." Orthogonal projection is the same operation, but now understood as the solution to a minimization problem, and generalized from a single vector to an arbitrary subspace.

To answer our question above, let W be a subspace of $\mathbb{R}^n$ and let x be a vector in $\mathbb{R}^n$. Denote the closest vector to x in W by x_W . By "closest", we mean the unique vector $x_W \in W$ minimizing $\|x - x_W\|$; equivalently, the unique vector $x_W \in W$ such that the residual $x_{W^\perp} := x - x_W$ is orthogonal to every vector in W (see below). Here $W^\perp = \{v \in \mathbb{R}^n : v \cdot w = 0 \text{ for all } w \in W\}$, called the orthogonal complement, is itself a subspace of $\mathbb{R}^n$ — the orthogonal subspace with respect to W .

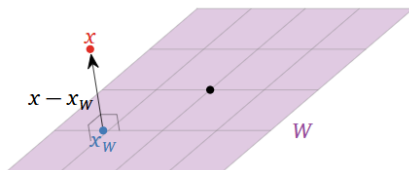


Figure 1: Margalit, Rabinoff, "Interactive Linear Algebra"

The above vector equation is equivalent to $x = x_W + x_{W^\perp}$, where $x_W \in W$ and $x_{W^\perp} \in W^\perp$. Such an expression is called the orthogonal decomposition of x , and the "closest" vector x_W is called the **orthogonal projection** of x onto W .

We previously defined projection between two vectors. Such is the one-dimensional case of orthogonal projection (where $W = \text{span}(u)$), and it can be re-derived here. Suppose we want to find some scalar value c that makes the vector $v - cu$ orthogonal to u , such that $(v - cu) \cdot u = 0$. It follows

$$\begin{aligned}(v - cu) \cdot u &= 0 \\ v \cdot u - c(u \cdot u) &= 0 \\ \implies c &= \frac{v \cdot u}{u \cdot u}, \\ \text{proj}_u v = cu &= \frac{v \cdot u}{u \cdot u} u.\end{aligned}$$

An intuitive result which follows is that the component of one vector that is orthogonal to another is the difference between the initial vector and its projection onto the other. Precisely, $x_\perp = x - \text{proj}_u x$.

One interesting fact arises from this notion, which is that $\mathbb{R}^n = W \oplus W^\perp$. (Recall: given two vector spaces V and W over the same field F , their external direct sum $V \oplus W$ is defined as the set of ordered pairs (v, w) with $v \in V$ and $w \in W$, with componentwise addition and scalar multiplication. When V and W are subspaces of a common vector space with $V \cap W = \{0\}$, we identify $V \oplus W$ with the internal sum $V + W = \{v + w : v \in V, w \in W\}$ inside the ambient space.) The proof of this statement is as follows.

Suppose W is a subspace of $\mathbb{R}^n$ of dimension j with an orthonormal basis $\{e_1, e_2, \dots, e_j\}$. We can extend this orthonormal basis to an orthonormal basis of $\mathbb{R}^n$ by adding additional basis vectors $\{e_{j+1}, e_{j+2}, \dots, e_n\}$, all of which are orthonormal to the preceding vectors (this is just the Gram-Schmidt process applied to any extension to a basis of $\mathbb{R}^n$). Then, for any vector $v \in \mathbb{R}^n$, we may write

$$v = \sum_{i=1}^n (v \cdot e_i) e_i = \underbrace{\sum_{i=1}^j (v \cdot e_i) e_i}_{=: w \in W} + \underbrace{\sum_{i=j+1}^n (v \cdot e_i) e_i}_{=: w^\perp \in W^\perp}.$$

The first sum lies in W by construction; the second lies in $W^\perp$ because each e_i with $i > j$ is orthogonal to every e_k with $k \leq j$, and hence orthogonal to every vector in $W = \text{span}(e_1, \dots, e_j)$. To see uniqueness, suppose $v = w + w^\perp = w' + (w')^\perp$ with $w, w' \in W$ and $w^\perp, (w')^\perp \in W^\perp$. Then $w - w' = (w')^\perp - w^\perp \in W \cap W^\perp$, and the only vector orthogonal to itself is the zero vector (by positive definiteness of the dot product), so $w = w'$ and the decomposition is unique. Thus every $v \in \mathbb{R}^n$ can be uniquely decomposed as $v = w + w^\perp$ with $w \in W$ and $w^\perp \in W^\perp$, so $\mathbb{R}^n = W \oplus W^\perp$. $\square$

3.6 Projection Onto a Subspace with an Orthonormal Basis

The only requirements for a set of vectors to form a basis for a vector space is that its elements are linearly independent and span the vector space. There are no requirements on mutual orthogonality or a standardized length of unit 1, but if we impose these conditions onto the basis vectors then it is said to be an **orthonormal basis**. Explicitly, a basis $\{q_1, q_2, \dots, q_k\}$ of a subspace $W \subseteq \mathbb{R}^n$ is orthonormal if $q_i \cdot q_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta (equal to 1 when $i = j$ and 0 otherwise).

Suppose that W is the span of two orthonormal basis vectors q_1, q_2 . We aim to find the orthogonal decomposition of x on W . In this instance, we aim to find the sum of the projection on each vector, which we can do because the basis vectors are orthonormal and do not interfere. The norm of each q_i is 1, so the denominator $q_i \cdot q_i$ in the one-dimensional projection formula simplifies to 1, and we have

$$\text{proj}_W x = (x \cdot q_1) q_1 + (x \cdot q_2) q_2.$$

And by the same logic as the 1D case, the residual $x - \text{proj}_W x$ is orthogonal to the subspace, which you can verify by taking its dot product with q_1 and q_2 separately and confirming you get zero both times. This generalizes to an n -dimensional subspace as we have seen above: if W has orthonormal basis $\{q_1, \dots, q_k\}$, then

$$\text{proj}_W x = \sum_{i=1}^k (x \cdot q_i) q_i.$$

4 Matrices as Linear Maps

There's no doubt you have seen matrices and know how to add, subtract, and multiply them, as well as compute inverses and find eigenvalues alongside other delightful properties. But what exactly is going on under the hood? And what precisely does a matrix represent?

First, it is important to understand the notion of linear maps. A linear map is a function $T : V \rightarrow W$ between vector spaces V and W , which respects the basic operations of vector addition and scalar multiplication. In other words, the function preserves linear combinations. If you fix a basis in V and in W , then T has a matrix representation. For example, if A is an $m \times n$ matrix of real numbers, then $f(\mathbf{x}) = A\mathbf{x}$ describes a linear map $\mathbb{R}^n \rightarrow \mathbb{R}^m$.

Let's work through an example of how exactly a matrix is a representation of a linear map. Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for V . Then every vector $\mathbf{v} \in V$ is uniquely determined by the coefficients $c_1, \dots, c_n$ in the field $\mathbb{R}$, so we may represent these vectors in the form $\mathbf{v} = c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n$.

If $f : V \rightarrow W$ is a linear map, it follows that

$$f(\mathbf{v}) = f(c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n) = c_1 f(\mathbf{v}_1) + \dots + c_n f(\mathbf{v}_n).$$

We see here the function f is entirely determined by the vectors $f(\mathbf{v}_1), \dots, f(\mathbf{v}_n)$. Now, let $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ be a basis for W . Then we can represent each vector $f(\mathbf{v}_j)$ in terms of this basis with its own respective coefficients, because each output corresponds to a vector in W . Namely:

$$f(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + \dots + a_{mj}\mathbf{w}_m.$$

Thus, since we see each vector in W is represented by its own coordinates with respect to its basis vectors, the function f is entirely determined by the values of a_{ij} . If we put these values into an $m \times n$ matrix M , then we can conveniently use it to compute the vector output of f for any vector in V . To get M , every column j of M is a vector

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

corresponding to $f(\mathbf{v}_j)$ as defined above. To define it more clearly, for some column j that corresponds to the mapping $f(\mathbf{v}_j)$,

$$\mathbf{M} = \begin{pmatrix} \cdots & a_{1j} & \cdots \\ & \vdots & \\ & a_{mj} & \end{pmatrix}$$

where M is the matrix of f . In other words, every column $j = 1, \dots, n$ has a corresponding vector $f(\mathbf{v}_j)$ whose coordinates $a_{1j}, \dots, a_{mj}$ are the elements of column j . A single linear map may be represented by many matrices. This is because the values of the elements of a matrix depend on the bases chosen. Now, we'll go through the sections below with this element of rigor in mind.

4.1 Determinants, Eigenvalues, Eigenvectors, and the Trace

Determinants. You likely know determinant topics such as cofactor expansion, the Leibniz formula, 2×2 expansion and other tactics. However, one of the most important takeaways from determinants is that they are a signed volume distortion factor of a linear map. That is, for a square matrix A , the unit cube maps to a parallelepiped whose volume is oriented by $\det(A)$. This makes understanding key determinant properties much more intuitive.

- If a matrix isn't invertible, it represents collapsing a vector space onto a subspace of strictly smaller dimension, which means the determinant is zero (akin to taking the unit cube and flattening it onto a line).
- Matrix multiplication corresponds to the composition of linear maps. Hence, for two matrices A, B , $\det(AB) = \det(A) \cdot \det(B)$. Geometrically, you can think of this as such: If matrix A scales volumes by $\det(A)$ and matrix B scales volumes by $\det(B)$, then applying B followed by A scales volumes by $\det(A) \cdot \det(B)$.

- The determinant of a matrix's transpose is equal to the determinant of the initial matrix. You may think of this as the parallelepiped spanned by the columns of A having the same volume as the parallelepiped spanned by its rows.

Eigenvalues and Eigenvectors. When you think of eigenvalues and eigenvectors, your mind may jump to the fact that, for a square matrix A , $Av = \lambda v$ and there exists a corresponding characteristic polynomial. With our understanding of linear maps, however, we can think of this fact more geometrically. An eigenvector is a direction of a linear map that doesn't rotate; it only scales (or reverses) by a factor λ , the eigenvalue. This is what is meant by the expression $Av = \lambda v$. To find these eigenvectors, we solve $(A - \lambda I)v = 0$ (where I is the identity matrix of equal size as A) for a nontrivial v which forces $\det(A - \lambda I) = 0$ and gives us the characteristic polynomial. In case this isn't clear, we can provide a sketch of why: it follows $Av = \lambda v \implies Av - \lambda v = 0 \implies (A - \lambda I)v = 0$, and we may denote $A - \lambda I$ as M . For M to have a nontrivial kernel, there must be linear dependence between its columns, and its determinant is zero.

- **Example with a two-dimensional shear matrix.** A shear matrix is a type of linear transformation that distorts shapes by shifting one axis relative to the other while keeping the origin fixed. A common 2×2 shear matrix is of the form

$$S = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

for some $a \in \mathbb{R}$ (in this case, with $a \neq 0$ for the shear to be nontrivial). By substituting S into $(A - \lambda I)v = 0$, we obtain $\lambda = 1$ with algebraic multiplicity 2, and the corresponding eigenvector $\mathbf{v} = (x \ 0)^T$, $x \neq 0$. This vector corresponds to the x -axis, which tells us it remains unchanged beyond scaling in the linear map. Note the geometric multiplicity here is 1, which is strictly less than the algebraic multiplicity of 2, so the shear matrix is not diagonalizable.

- **Example with a two-dimensional rotation matrix.** The most general rotation matrix in two-dimensions is of the form

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

where $\theta \in [0, 2\pi)$. This represents a counterclockwise rotation in the x - y plane. By substituting R into $(A - \lambda I)v = 0$, we obtain $\lambda = \cos(\theta) \pm i \sin(\theta) = e^{\pm i\theta}$ and the corresponding eigenvectors $\mathbf{v}_1 = (1 \ -i)^T$ (associated with $\lambda = e^{i\theta}$), $\mathbf{v}_2 = (1 \ i)^T$ (associated with $\lambda = e^{-i\theta}$) under the assumption $\theta \notin \{0, \pi\}$. This is an interesting example, as we see the eigenvalues are complex numbers.

- **Algebraic and Geometric Multiplicity.** Two quick notes to mention: The algebraic multiplicity of an eigenvalue is the number of times it appears as a root of the matrix's characteristic polynomial. The geometric

multiplicity of an eigenvalue is the dimension of its eigenspace, i.e., the number of linearly independent eigenvectors corresponding to that eigenvalue. Geometric multiplicity is always less than or equal to algebraic multiplicity, but they do not differ for symmetric matrices.

The trace. For a square matrix A , its trace is the sum of its diagonal entries: $\text{tr}(A) = \sum_{i=1}^n a_{ii}$. The trace satisfies linearity, is invariant under transposition, and is cyclic ($\text{tr}(AB) = \text{tr}(BA)$). There are many interesting facts about the trace. One is that if a matrix is diagonalizable, its trace is equal to the sum of its eigenvalues. Another fact is that the trace has a connection to the determinant, which itself is the product of a matrix's eigenvalues. The determinant tells you the *total* volume scaling factor. The trace tells you the *infinitesimal rate of volume change* for maps near the identity. Like the determinant, the trace is independent of basis, and itself is equal to the sum of all eigenvalues.

4.2 Symmetric Matrices

A matrix A is symmetric if and only if it is equal to its transpose. This is a well-known fact with an interesting property. If $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{R}^n$, we also say that it is a *symmetric bilinear form*, which is simply terminology that follows from the inner product being symmetric and bilinear (precisely, it is a positive-definite symmetric bilinear form). If we fix a basis $\{v_1, v_2, \dots, v_n\}$, we find that the entries $A_{ij} := \langle v_i, v_j \rangle$ form a symmetric matrix, since $A_{ij} = \langle v_i, v_j \rangle = \langle v_j, v_i \rangle = A_{ji}$. So, symmetric matrices are matrix representations of symmetric bilinear forms – inner products in particular.

Another noteworthy property of symmetric matrices is that they satisfy the spectral theorem, which we shall prove next. Perhaps the most important take-away from this information is that every symmetric matrix is, after an orthogonal change of basis, just a diagonal matrix. The spectral theorem states the following.

The Spectral Theorem (real case). Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then:

1. All eigenvalues of A are real.
2. Eigenvectors of A corresponding to distinct eigenvalues are orthogonal.
3. There exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A .

Equivalently, there exists an orthogonal matrix Q (for which $Q^\top Q = I$) and a diagonal matrix D such that $A = QDQ^\top$, where the columns of Q are the orthonormal eigenvectors and the diagonal entries of D are the corresponding eigenvalues.

Claim (1) may be non-obvious: the characteristic polynomial of a general $n \times n$ real matrix can certainly have complex roots (we saw this with the rotation matrix in Section 4.1). The spectral theorem says that for symmetric matrices, this never happens. Claim (2) carries the notion of orthogonality we built up in Section 3 into the eigenvector setting: as long as two eigenvalues differ, their

corresponding eigenvectors are automatically perpendicular, with no additional work required to verify. Claim (3) is the strongest of the three, and is essentially what allows the diagonalization to happen. It says that the whole space $\mathbb{R}^n$ has a basis made entirely of eigenvectors of A , each one orthogonal to the others and of unit length. The matrix form $A = QDQ^\top$ is then a direct consequence of claim (3), since the columns of Q can be taken to be these orthonormal eigenvectors and the diagonal entries of D to be the corresponding eigenvalues.

Before we prove this, it is worth seeing an example. Consider the matrix

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

Its characteristic polynomial is $\det(A - \lambda I) = (3 - \lambda)^2 - 1$, whose roots are $\lambda_1 = 2$ and $\lambda_2 = 4$. Both are indeed real. Solving $(A - \lambda_i I)v = 0$ for each i gives the eigenvectors $\mathbf{v}_1 = (1 \ -1)^\top$ and $\mathbf{v}_2 = (1 \ 1)^\top$, which are orthogonal, as well. Normalizing them to unit length and arranging them as columns yields the orthogonal matrix

$$Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

and one may verify directly that $Q^\top A Q = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} = D$, so $A = QDQ^\top$.

This example works, but we shall now prove that this isn't a coincidence.

Proof of (1). The plan to prove (1) is to find a single scalar quantity that we can compute in two different ways, with one expression involving λ and the other involving its complex conjugate $\bar{\lambda}$. If both expressions describe the same scalar, then λ must equal $\bar{\lambda}$. Only real numbers satisfy that, so λ would have to be real.

Let λ be an eigenvalue of A , which could be complex, with corresponding eigenvector $v \in \mathbb{C}^n$ (also possibly complex). By $\bar{v}$, we mean the vector whose entries are the complex conjugates of the entries of v . The quantity $\bar{v}^\top v$ is then

$$\bar{v}^\top v = \sum_i \bar{v}_i v_i = \sum_i |v_i|^2 = \|v\|^2,$$

which is always a real, non-negative number, and equals zero only when $v = 0$.

We compute the quantity $\bar{v}^\top A v$ in two ways. On one hand, using $A v = \lambda v$,

$$\bar{v}^\top A v = \bar{v}^\top (\lambda v) = \lambda \bar{v}^\top v = \lambda \|v\|^2.$$

On the other hand, using $A^\top = A$ and the fact that A has real entries (so $A\bar{v} = \overline{A v}$),

$$\bar{v}^\top A v = (A^\top \bar{v})^\top v = (A \bar{v})^\top v = (\overline{A v})^\top v = (\bar{\lambda} v)^\top v = \bar{\lambda} \bar{v}^\top v = \bar{\lambda} \|v\|^2.$$

Let's quickly explain the logic used above. The first equality uses the general matrix identity $u^\top A v = (A^\top u)^\top v$, which is a consequence of how transposition interacts with multiplication. The second equality uses $A^\top = A$, which is the

symmetry hypothesis. The third equality uses the fact that A has real entries: when A has nothing complex to conjugate, applying A to the conjugated vector $\bar{v}$ gives the same result as conjugating after applying A , namely $A\bar{v} = \overline{Av}$. The fourth equality substitutes the eigenvalue equation $Av = \lambda v$. The remaining steps just pull the scalar $\bar{\lambda}$ out front and reuse $\bar{v}^\top v = \|v\|^2$.

Equating both expressions and noting $\|v\|^2 \neq 0$ (since v is an eigenvector, hence nonzero), we obtain $\lambda = \bar{\lambda}$, which forces λ to be real. The trick here, which is worth emphasizing, is moving A from one side of the inner product to the other – a maneuver permitted exactly because A is symmetric.

Proof of (2). Suppose $Av = \lambda v$ and $Aw = \mu w$ with $\lambda \neq \mu$. Using $A^\top = A$,

$$\lambda(v \cdot w) = (Av) \cdot w = v^\top A^\top w = v^\top Aw = v \cdot (Aw) = \mu(v \cdot w).$$

Hence $(\lambda - \mu)(v \cdot w) = 0$. Since $\lambda \neq \mu$, we conclude $v \cdot w = 0$, so v and w are orthogonal. As in (1), this proof works because we can shift A across the inner product.

Sketch of (3). The proof is by induction on the dimension n . What we shall prove explicitly is that if v is an eigenvector of a symmetric matrix A , then A maps the subspace $v^\perp$ to itself. To see this, suppose $w \in v^\perp$, so $w \cdot v = 0$. Then

$$(Aw) \cdot v = w \cdot (A^\top v) = w \cdot (Av) = \lambda(w \cdot v) = 0,$$

so $Aw \in v^\perp$.

This lemma enables the induction. By isolating one eigenvector v_1 and looking at its orthogonal complement $v_1^\perp$, we reduce an n -dimensional problem to an $(n - 1)$ -dimensional problem on which the inductive hypothesis applies directly.

The induction thus proceeds as follows. The characteristic polynomial of A has at least one root, by the fundamental theorem of algebra, and that root is real by part 1, so A has at least one real eigenvector. Pick one and normalize it; call it v_1 . The lemma tells us A restricts to a linear map on the $(n - 1)$ -dimensional subspace $v_1^\perp$, and on this subspace it is again symmetric. By inductive hypothesis, A restricted to $v_1^\perp$ admits an orthonormal basis of eigenvectors of its own, say $\{v_2, \dots, v_n\}$. Combining v_1 with this basis yields an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A . $\square$

Geometrically, the decomposition $A = QDQ^\top$ describes the action of any symmetric matrix as a three-step process: rotate by $Q^\top$ so that the eigenvectors of A become the standard basis, stretch each axis by the corresponding eigenvalue λ_i (with reflection if λ_i is negative), then rotate back by Q . In other words, a symmetric matrix is just a diagonal stretch performed in a rotated coordinate system. The unit sphere is sent to an ellipsoid whose axes are the eigenvectors of A and whose semi-axis lengths are $|\lambda_i|$. This picture will become important in the next section, when we ask which symmetric matrices properly correspond to inner products.

4.3 Positive Definite Matrices

From Section 4.2, we know inner products on $\mathbb{R}^n$ correspond to symmetric matrices. However, exactly which symmetric matrices correspond to inner products? An inner product is a positive-definite symmetric bilinear form, so the answer depends on the requirement of positive definiteness.

To get a feel for what positive definiteness asks of a matrix, consider the expression $x^\top Ax$. This takes a vector x and produces a single scalar. When $A = I$ (the identity matrix), the result is $x^\top Ix = x^\top x = \|x\|^2$, which is always strictly positive for any nonzero x . The condition of positive definiteness asks whether a more general symmetric matrix A shares this property of the identity. Does $x^\top Ax$ stay strictly positive whenever the input x is nonzero? A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if

$$x^\top Ax > 0 \quad \text{for every nonzero } x \in \mathbb{R}^n.$$

If the strict inequality is replaced by ≥ 0 , allowing equality for some nonzero x , we call A positive semi-definite instead.

With this definition, the answer to our motivating question is clear: a symmetric matrix A corresponds to an inner product on $\mathbb{R}^n$ if and only if A is positive definite. The standard dot product corresponds to the case $A = I$, and every inner product on $\mathbb{R}^n$ arises in this way from a unique positive definite symmetric matrix. So positive definite symmetric matrices are exactly the inner-product matrices of $\mathbb{R}^n$.

There are several equivalent ways to recognize a positive definite matrix, and each is useful in different situations. A few are:

1. $x^\top Ax > 0$ for every nonzero $x \in \mathbb{R}^n$ (the definition).
2. All eigenvalues of A are positive.
3. Every leading principal minor of A is positive (this is *Sylvester's criterion*).
The k th leading principal minor is the determinant of the top-left $k \times k$ submatrix.
4. There exists an invertible matrix B such that $A = B^\top B$.
5. There exists a unique lower-triangular matrix L with positive diagonal entries such that $A = LL^\top$ (the **Cholesky decomposition**).

Each of these five conditions is a different way of expressing the same property, and which one is easiest to check depends on what we already know about A . Condition (1) is the definition itself. Condition (2) ties positive definiteness to the spectrum, which the spectral theorem has already made accessible. Condition (3) is computational: only basic determinants are needed, with no eigenvalues. Conditions (4) and (5) give two factorizations of A , both of which simplify many later computations once the factorization is in hand.

The equivalence between (1) and (2) is from the spectral theorem. By that theorem, we may write $A = QDQ^\top$ with Q orthogonal and D diagonal, where

the diagonal entries of D are the eigenvalues $\lambda_1, \dots, \lambda_n$ of A . Substituting $y := Q^\top x$ yields

$$x^\top A x = x^\top Q D Q^\top x = (Q^\top x)^\top D (Q^\top x) = y^\top D y = \sum_{i=1}^n \lambda_i y_i^2.$$

Since Q is invertible, x ranges over all nonzero vectors of $\mathbb{R}^n$ exactly when y does. The right-hand side is a weighted sum of squares with the eigenvalues as weights, so it is positive for every nonzero y if and only if every λ_i is strictly positive. Hence (1) $\iff$ (2).

The equivalence with (4) is similar. If $A = Q D Q^\top$ with all eigenvalues positive, define $B := D^{1/2} Q^\top$, where $D^{1/2}$ is the diagonal matrix whose entries are the positive square roots of the entries of D . Then $B^\top B = Q D^{1/2} D^{1/2} Q^\top = Q D Q^\top = A$, and B is invertible because $D^{1/2}$ and Q both are. Conversely, if $A = B^\top B$ with B invertible, then $x^\top A x = (Bx)^\top (Bx) = \|Bx\|^2$, which is positive whenever x is nonzero, since B is invertible. The remaining equivalences (3) and (5) are results not proved here.

Positive definite matrices appear all over mathematics, but the one that matters for our purposes is the metric tensor of a manifold, which at each point is encoded by a positive definite symmetric matrix. We'll revisit this in 5.2.

4.4 Change-of-Basis Matrices

Recall from the introduction of this section that the matrix representation of a linear map depends on the choice of basis. The question we now address is the following: given a vector, a linear map, or a bilinear form, and a change to a new basis, how does the corresponding coordinate vector or matrix transform? This question matters because in differential geometry one constantly changes coordinate charts and we must know how our objects behave under that change.

Let V be a finite-dimensional vector space (think $\mathbb{R}^n$ for concreteness), and let $\mathcal{B} = \{v_1, \dots, v_n\}$ and $\mathcal{B}' = \{v'_1, \dots, v'_n\}$ be two bases of V . Since $\mathcal{B}$ is a basis, each new basis vector v'_j can be written uniquely as

$$v'_j = \sum_{i=1}^n P_{ij} v_i,$$

for some scalars $P_{ij} \in \mathbb{R}$. The matrix $P = (P_{ij}) \in \mathbb{R}^{n \times n}$ is the **change-of-basis matrix from $\mathcal{B}'$ to $\mathcal{B}$** , and its j th column consists of the coordinates of v'_j in the basis $\mathcal{B}$. Because $\mathcal{B}'$ is itself a basis, P is invertible.

Transformation of coordinate vectors. Let $v \in V$ be a vector with coordinate column x in $\mathcal{B}$ and coordinate column x' in $\mathcal{B}'$, so $v = \sum_i x_i v_i = \sum_j x'_j v'_j$. Substituting the relation $v'_j = \sum_i P_{ij} v_i$ into the second expression yields

$$v = \sum_j x'_j \sum_i P_{ij} v_i = \sum_i \left(\sum_j P_{ij} x'_j \right) v_i.$$

Comparing with $v = \sum_i x_i v_i$ and using uniqueness of the representation in a basis, we obtain

$$x_i = \sum_j P_{ij} x'_j, \quad \text{or in matrix form} \quad x = Px', \quad x' = P^{-1}x.$$

Note that the basis vectors transform by P , while the coordinates transform by P^{-1} . This may at first feel backward, but it is precisely the right behavior: a vector is an intrinsic object whose identity does not depend on the basis, so when the basis vectors grow, the coordinates must shrink to keep the represented vector unchanged.

Transformation of linear maps. Let $T : V \rightarrow V$ be a linear map, with matrix A in $\mathcal{B}$ and matrix A' in $\mathcal{B}'$. For any $v \in V$ with coordinates x in $\mathcal{B}$ and x' in $\mathcal{B}'$, the image $T(v)$ has coordinates Ax in $\mathcal{B}$ and $A'x'$ in $\mathcal{B}'$. Since coordinates transform by P^{-1} , we have

$$A'x' = P^{-1}(Ax) = P^{-1}A(Px') = (P^{-1}AP)x'.$$

As this holds for every x' , we conclude $A' = P^{-1}AP$. Matrices related in this way are called similar, and they share basis-independent invariants such as the determinant, the trace, the characteristic polynomial, and the eigenvalues.

Transformation of symmetric bilinear forms. Let $B : V \times V \rightarrow \mathbb{R}$ be a symmetric bilinear form, with matrix A in $\mathcal{B}$ (so $B(v, w) = x^\top Ay$, where x, y are the $\mathcal{B}$ -coordinates of v, w) and matrix A' in $\mathcal{B}'$. Then

$$B(v, w) = x^\top Ay = (Px')^\top A(Py') = (x')^\top (P^\top AP)y'.$$

Comparing with $B(v, w) = (x')^\top A'y'$, we conclude

$$A' = P^\top AP.$$

Matrices related in this way are called **congruent**. Congruence preserves symmetry and positive definiteness, but it doesn't necessarily preserve eigenvalues. The basis-independent invariants of a symmetric bilinear form under congruence are its rank and the number of positive, negative, and zero eigenvalues (called its signature). This statement is known as Sylvester's law of inertia.

The two transformation rules $A' = P^{-1}AP$ (for linear maps) and $A' = P^\top AP$ (for bilinear forms) look similar but are different objects; it is worth mentioning why. A linear map takes in a vector and outputs a vector, so P must act once on the input side and once on the output side, and the two actions go in opposite directions (which is why P and P^{-1} both appear in the similarity transformation). A bilinear form, on the other hand, eats two vectors and produces a scalar, so P must act on both input sides in the same direction, hence the two P 's on the same side of A in the congruence transformation. Two inputs, no output, two P 's; one input, one output, one P and one P^{-1} . Once the distinction is internalized this way, the two rules stop being confusable.

A special case worth highlighting is when P is orthogonal, that is, $P^\top P = I$, so that $P^{-1} = P^\top$. In this case the similarity transformation $P^{-1}AP$ and the

congruence transformation $P^\top AP$ coincide. This is exactly the situation in the spectral theorem: $A = QDQ^\top$ can be read both as a similarity (preserving eigenvalues of A) and as a congruence (preserving the bilinear-form structure of A). This is why the spectral theorem produces such a clean simultaneous diagonalization: the orthogonal change of basis respects both perspectives at once.

5 Quadratic Forms

We now will comprehensively bring the pieces of Section 4 together. Quadratic forms in a sense "package the information" of a symmetric bilinear form into a single scalar-valued function of one vector argument, and they will provide the most straightforward language for talking about the metric tensor of a manifold as we will see later.

5.1 Quadratic Form of a Symmetric Matrix

At several points already, in Section 4 and in the inner product discussion, we have encountered expressions of the form $x^\top Ax$. These expressions are quadratic, in the sense that scaling the input vector x by some scalar r scales the output by r^2 , since $(rx)^\top A(rx) = r^2 x^\top Ax$. Because such expressions show up so often, it is worth giving them their own name and a short study.

A **quadratic form** on $\mathbb{R}^n$ is a function $q : \mathbb{R}^n \rightarrow \mathbb{R}$ of the form

$$q(x) = x^\top Ax = \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j,$$

where $A \in \mathbb{R}^{n \times n}$ is a square matrix.

To make this concrete in two dimensions, take $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

Computing $x^\top Ax$ directly gives

$$q(x) = ax_1^2 + (b+c)x_1x_2 + dx_2^2.$$

Every term has total degree two in the components of x , which is what we mean by "quadratic" here. So a quadratic form is just a polynomial in the entries of x whose every term is of degree exactly two.

At first glance there is no requirement that A be symmetric. However, since $q(x)$ is a scalar, it equals its own transpose: $q(x) = q(x)^\top = (x^\top Ax)^\top = x^\top A^\top x$. Adding both expressions and dividing by two yields

$$q(x) = x^\top \left(\frac{A + A^\top}{2} \right) x.$$

The matrix $(A + A^\top)/2$ is symmetric and produces the exact same quadratic form q . So without loss of generality, every quadratic form can be associated

with a unique symmetric matrix, and we shall take the matrix of a quadratic form to be symmetric by convention. This gives a one-to-one correspondence between quadratic forms on $\mathbb{R}^n$ and symmetric matrices in $\mathbb{R}^{n \times n}$.

The symmetric bilinear form $B(x, y) := x^\top Ay$ associated with q can be recovered from the quadratic form alone via the polarization identity:

$$B(x, y) = \frac{1}{2}(q(x + y) - q(x) - q(y)),$$

which one can verify by direct expansion. So quadratic forms and symmetric bilinear forms encode exactly the same information, and we may pass freely between them.

What the polarization identity is saying, in words, is that we can recover the bilinear form $B(x, y)$, which is a function of two vectors, from the quadratic form $q(x)$, which is a function of just one vector. So no information is actually lost when we express a bilinear form as a quadratic form, even though one of the two arguments has seemingly vanished.

We classify a quadratic form $q(x) = x^\top Ax$ by the behavior of its values:

- **Positive definite:** $q(x) > 0$ for every nonzero x . Equivalently, A is positive definite.
- **Positive semi-definite:** $q(x) \geq 0$ for every x , but $q(x) = 0$ may occur for some nonzero x .
- **Negative definite / semi-definite:** defined analogously with the inequalities reversed.
- **Indefinite:** q takes both positive and negative values.

The spectral theorem gives us a particularly clean way to understand the geometry of a quadratic form. Since A is symmetric, we can write $A = QDQ^\top$ with Q orthogonal and $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. Setting $y := Q^\top x$, the quadratic form becomes

$$q(x) = y^\top Dy = \sum_{i=1}^n \lambda_i y_i^2.$$

In the eigenbasis then, a quadratic form is just a weighted sum of squares with the eigenvalues as the weights. The classification above is now immediate from the signs of the λ_i : all positive means positive definite, all negative means negative definite, mixed signs mean indefinite, and so on. The level sets $\{x : q(x) = c\}$ are ellipsoids, hyperboloids, paraboloids, or degenerate quadrics, depending on the signature. The case of greatest interest to us is the positive definite one, in which every level set $\{q = c\}$ with $c > 0$ is a (possibly skewed) ellipsoid centered at the origin.

5.2 Wrapping Up: The First Fundamental Form

We are at last in a position to assemble the geometric object that motivated this whole document.

Let $S \subset \mathbb{R}^3$ be a smooth surface, parametrized locally by a smooth map $\mathbf{r} : U \rightarrow \mathbb{R}^3$ on an open set $U \subseteq \mathbb{R}^2$ with coordinates (u^1, u^2) .

To clarify what this setup is saying, a surface in $\mathbb{R}^3$ is a two-dimensional curved object sitting in three-dimensional space. To work with it analytically, we describe each point on the surface using two parameters, u^1 and u^2 , which vary over some flat region $U \subseteq \mathbb{R}^2$. The map $\mathbf{r}$ takes each parameter pair $(u^1, u^2) \in U$ and returns the corresponding point $\mathbf{r}(u^1, u^2)$ in $\mathbb{R}^3$, which sits on the surface. A standard example is the unit sphere, where we might choose u^1 and u^2 to be the latitude and longitude, and $\mathbf{r}(u^1, u^2)$ would then give the point on the sphere at those coordinates.

At a point $p = \mathbf{r}(u^1, u^2) \in S$, the partial derivatives

$$\mathbf{r}_1 := \frac{\partial \mathbf{r}}{\partial u^1}, \quad \mathbf{r}_2 := \frac{\partial \mathbf{r}}{\partial u^2}$$

are vectors in $\mathbb{R}^3$, and they span the tangent plane $T_p S$ to the surface at p (provided $\mathbf{r}_1$ and $\mathbf{r}_2$ are linearly independent, which is the regularity condition we always assume of a parametrization).

The tangent plane at p is the flat plane that best approximates the surface near p . The vector $\mathbf{r}_1$ gives the direction the surface moves when only u^1 is varied (with u^2 held fixed), and $\mathbf{r}_2$ gives the direction when only u^2 is varied. Together, these two vectors span the tangent plane, and any tangent vector at p can be written as a linear combination of them.

A tangent vector $\xi \in T_p S$ may then be written as

$$\xi = \xi^1 \mathbf{r}_1 + \xi^2 \mathbf{r}_2.$$

Because S sits inside $\mathbb{R}^3$, it inherits a notion of length and angle from the ambient Euclidean structure simply by restricting the dot product on $\mathbb{R}^3$ to tangent vectors. This restriction is called the **first fundamental form** of S . Explicitly, for tangent vectors $\xi = \xi^i \mathbf{r}_i$ and $\eta = \eta^j \mathbf{r}_j$ at the same point p (we now use the Einstein summation convention, where any repeated index, one upper and one lower, is summed over), we have

$$\langle \xi, \eta \rangle = (\xi^i \mathbf{r}_i) \cdot (\eta^j \mathbf{r}_j) = \xi^i \eta^j (\mathbf{r}_i \cdot \mathbf{r}_j) = g_{ij} \xi^i \eta^j,$$

where the components

$$g_{ij} := \mathbf{r}_i \cdot \mathbf{r}_j$$

are the entries of the **metric tensor** (or first fundamental form) of S in the chosen parametrization.

The Einstein summation convention is just shorthand. Any time you see an index appearing once as a superscript and once as a subscript in the same expression, sum over that index from 1 to the appropriate upper bound. So

$g_{ij} \xi^i \eta^j$ stands for $\sum_{i=1}^2 \sum_{j=1}^2 g_{ij} \xi^i \eta^j$ in our two-parameter surface setting. The benefit of the convention is that in tensor calculations summation symbols pile up quickly, so dropping them keeps the notation easy to read.

The entries of the metric tensor are simply the dot products of the tangent-plane basis vectors $\mathbf{r}_i$ with each other, one such dot product for each pair of indices. From these entries alone, every notion of length, angle, and area on the surface can be computed using only the parameters (u^1, u^2) and the values g_{ij} , with no further reference to the ambient $\mathbb{R}^3$. This is what makes the metric tensor the right object for studying intrinsic geometry, because once we have it the surface can be studied as an object in its own right.

The associated quadratic form gives the squared length of a tangent vector,

$$\|\xi\|^2 = \langle \xi, \xi \rangle = g_{ij} \xi^i \xi^j.$$

In classical notation, with $E := g_{11}$, $F := g_{12} = g_{21}$, and $G := g_{22}$, the first fundamental form is written

$$I(\xi, \xi) = E(\xi^1)^2 + 2F \xi^1 \xi^2 + G(\xi^2)^2,$$

which is the form one usually encounters first in a differential geometry course.

The matrix

$$g = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}$$

is symmetric, because the dot product is symmetric, and positive definite, because the dot product is positive definite and the linear independence of $\mathbf{r}_1, \mathbf{r}_2$ ensures that any nonzero tangent vector ξ is a nonzero vector in $\mathbb{R}^3$, with $\|\xi\|^2 > 0$. So at every point of the surface, the metric tensor is a positive definite symmetric matrix, which means the entire apparatus we developed in Section 4 – symmetric matrices, positive definite matrices, the spectral theorem, congruence transformation – applies directly.

A few immediate consequences of the first fundamental form deserve mention:

- **Arc length** of a curve $\gamma(t) = \mathbf{r}(u^1(t), u^2(t))$ on S between $t = a$ and $t = b$ is

$$L(\gamma) = \int_a^b \sqrt{g_{ij}(u(t)) \dot{u}^i(t) \dot{u}^j(t)} dt.$$

- **Angle** between tangent vectors ξ and η at a point is given by

$$\cos \theta = \frac{g_{ij} \xi^i \eta^j}{\sqrt{g_{ij} \xi^i \xi^j} \sqrt{g_{ij} \eta^i \eta^j}}.$$

- **Surface area element** is

$$dA = \sqrt{\det g} du^1 du^2 = \sqrt{EG - F^2} du^1 du^2.$$

If we reparametrize S from (u^1, u^2) to $(\tilde{u}^1, \tilde{u}^2)$, the Jacobian matrix $P^i_j := \partial u^i / \partial \tilde{u}^j$ encodes the change of basis on each tangent plane, and the metric tensor transforms as

$$\tilde{g}_{ij} = P^k_i P^\ell_j g_{k\ell}, \quad \text{or in matrix form} \quad \tilde{g} = P^\top g P.$$

This is exactly the congruence transformation we identified in Section 4.4 as the correct transformation law for a symmetric bilinear form. The fact that the metric tensor transforms by congruence (rather than by similarity) is what makes it a well-defined geometric object on the surface, independent of any particular choice of coordinates.

This brings us to the threshold of the differential geometry we shall need for studying Brownian motion on manifolds. The first fundamental form encodes everything required to do intrinsic geometry on a surface: lengths, angles, areas, and distance. From here we can build the Laplace–Beltrami operator, the geodesic equation, and ultimately the diffusion processes that define Brownian motion in this setting. But that is a story for another document.