Math 510 **Fall 2022**
Notes for Homework 4 **Drew Armstrong**

Contents

1 Linear and Bilinear Forms

1.1 Linear Forms

Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). A linear function

$$\varphi : V \to \mathbb{R}$$

is called a *linear form*. If V is an infinite dimensional space of functions such as L^2 then a linear form is usually called a *linear functional*.

Linear forms on $\mathbb{R}^n$ are particularly simple. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a linear form and for each basis vector $\mathbf{e}_i$ define the scalar

$$b_i := \varphi(\mathbf{e}_i).$$

Then for any vector $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n \in \mathbb{R}^n$ we have

$$\begin{aligned}\varphi(\mathbf{x}) &= \varphi(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n)\\ &= x_1\varphi(\mathbf{e}_1) + \cdots + x_n\varphi(\mathbf{e}_n)\\ &= x_1b_1 + \cdots + x_nb_n.\end{aligned}$$

If we write $\mathbf{b} = (b_1, \ldots, b_n)$ then this becomes

$$\varphi(\mathbf{x}) = \mathbf{b}^T\mathbf{x}.$$

We will denote the function $\mathbf{x} \mapsto \mathbf{b}^T\mathbf{x}$ by $\varphi_{\mathbf{b}} : \mathbb{R}^n \to \mathbb{R}$. Thus we obtain a bijection between vectors and linear forms:

$$\begin{array}{rcl} \mathbb{R}^n & \to & \text{linear forms on } \mathbb{R}^n \\ \mathbf{b} & \mapsto & \varphi_{\mathbf{b}}. \end{array}$$

Indeed, the function $\varphi_{\mathbf{b}}(\mathbf{x}) = \mathbf{b}^T\mathbf{x}$ is linear, and we have just seen that every linear function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is equal to $\varphi_{\mathbf{b}}$ for some $\mathbf{b} \in \mathbb{R}^n$.

More abstractly, let V be an inner product space over $\mathbb{R}$ (or a Hermitian space over $\mathbb{C}$). Then for any vector $\mathbf{u} \in V$ we can define a linear form

$$\varphi_{\mathbf{u}}(\mathbf{v}) := \langle \mathbf{u}, \mathbf{v} \rangle.$$

Again, this gives a map[1] from V to the set of linear forms on V:

$$\begin{array}{rcl} V & \to & \text{linear forms on } V \\ \mathbf{u} & \mapsto & \varphi_{\mathbf{u}}. \end{array}$$

But this need not be a bijection in general. To investigate this, suppose that vectors $\mathbf{u}_1, \mathbf{u}_2 \in V$ correspond to the same functional, so that for all $\mathbf{v} \in V$ we have

$$\begin{aligned} \varphi_{\mathbf{u}_1}(\mathbf{v}) &= \varphi_{\mathbf{u}_2}(\mathbf{v}) \\ \langle \mathbf{u}_1, \mathbf{v} \rangle &= \langle \mathbf{u}_2, \mathbf{v} \rangle \\ \langle \mathbf{u}_1, \mathbf{v} \rangle - \langle \mathbf{u}_2, \mathbf{v} \rangle &= 0 \\ \langle \mathbf{u}_1 - \mathbf{u}_2, \mathbf{v} \rangle &= 0. \end{aligned}$$

Since this applies to any $\mathbf{v}$ we can take $\mathbf{v} = \mathbf{u}_1 - \mathbf{u}_2$ to obtain

$$\langle \mathbf{u}_1 - \mathbf{u}_2, \mathbf{u}_1 - \mathbf{u}_2 \rangle = 0.$$

But it is an axiom of (Hermitian) inner products that $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ implies $\mathbf{x} = \mathbf{0}$, hence we must have

$$\begin{aligned} \mathbf{u}_1 - \mathbf{u}_2 &= 0 \\ \mathbf{u}_1 &= \mathbf{u}_2. \end{aligned}$$

This shows that the map from V to the set of linear forms on V is always injective. However, it is not necessarily surjective. This is the subject of the Riesz Representation Theorem.

Theorem (Riesz Representation). Let V be a *Hilbert space*. This means that V is an inner product space over $\mathbb{R}$ (or a Hermitian space over $\mathbb{C}$), and that Cauchy sequences with respect to the norm $\| - \| = \sqrt{\langle -, - \rangle}$ converge.[2] Let $\varphi : V \to \mathbb{R}$ be a linear functional. Then

$$\varphi = \varphi_{\mathbf{u}} \text{ for some } \mathbf{u} \in V \quad \Longleftrightarrow \quad \varphi \text{ is continuous with respect to } \| - \|.$$

[1] So many different words for "function". The purpose is to avoid confusion when discussing many different kinds of functions at the same time.

[2] Recall: We say that $\mathbf{v}_1, \mathbf{v}_2, \ldots$ is a Cauchy sequence if for all $k \geq \ell \geq N$ we have $\|\mathbf{v}_k - \mathbf{v}_\ell\| \to 0$ as $N \to \infty$.

If V is finite dimensional then every linear functional is continuous. If V is infinite dimensional then there exist **discontinuous** functionals, but they are often ignored.

Let me introduce a some jargon. Given a vector space V over $\mathbb{R}$ (or $\mathbb{C}$) we define its *dual space* as the set of linear forms:[3]

$$\begin{aligned} V^\vee &= \text{the dual space} \\ &= \{\text{all linear forms } V \to \mathbb{R}\}. \end{aligned}$$

As the name suggests, the set $V^\vee$ is also a vector space over $\mathbb{R}$. For a given list of forms $\varphi_i : V \to \mathbb{R}$ and scalars $a_i \in \mathbb{F}$ we define the form $\sum a_i\varphi_i : V \to \mathbb{R}$ "pointwise":

$$\left(\sum a_i\varphi_i\right)(\mathbf{v}) := \sum a_i\varphi_i(\mathbf{v}) \text{ for all } \mathbf{v} \in V.$$

I claim that this definition makes the map $V \to V^\vee$ into a linear map. To see this, let's give the map a name. Let Φ denote the map that sends the vector $\mathbf{u} \in V$ to the form $\varphi_{\mathbf{u}} \in V^\vee$:

$$\begin{array}{rcl} \Phi : V & \to & V^\vee \\ \mathbf{u} & \mapsto & \varphi_{\mathbf{u}} \end{array}$$

Then for any linear combination of vectors $\sum a_i\mathbf{u}_i \in V$, I claim that

$$\Phi\left(\sum a_i\mathbf{u}_i\right) = \sum a_i\Phi(\mathbf{u}_i),$$

where each side of the equation is a linear form. To show that two forms are equal we must show that they define the same function $V \to \mathbb{R}$. So consider any vector $\mathbf{v} \in V$. Then since $\Phi(\mathbf{u})$ is just another name for $\varphi_{\mathbf{u}}$, we have

$$\begin{aligned} \left[\Phi\left(\sum a_i\mathbf{u}_i\right)\right](\mathbf{v}) &= \varphi_{\sum a_i\mathbf{u}_i}(\mathbf{v}) \\ &= \left\langle \sum a_i\mathbf{u}_i, \mathbf{v}\right\rangle \\ &= \sum a_i\langle \mathbf{u}_i, \mathbf{v}\rangle \\ &= \sum a_i\varphi_{\mathbf{u}_i}(\mathbf{v}) \\ &= \left[\sum a_i\Phi(\mathbf{u}_i)\right](\mathbf{v}). \end{aligned}$$

Thus $\Phi : V \to V^\vee$ is an injective linear map, and if V is finite dimensional then it is also surjective, hence it is an isomorphism $V \cong V^\vee$. When V is infinite dimensional then Φ is **not surjective**, however it is common to restrict the definition of $V^\vee$ as follows:

$$V^\vee = \{\text{the set of } \textbf{continuous} \text{ linear functionals } V \to \mathbb{R}\}.$$

Then from the Riesz Reprentation Theorem we will still have $V \cong V^\vee$.

[3]It is more common to write V^* for the dual space, but I am already using that notation for the conjugate transpose.

Another piece of jargon is the Dirac *bra-ket notation* from quantum physics. To motivate this, consider the isomorphism between $\mathbb{R}^n$ and its dual:

$$\begin{array}{ccc} \mathbb{R}^n & \cong & (\mathbb{R}^n)^\vee \\ \mathbf{b} & \leftrightarrow & \varphi_{\mathbf{b}}, \end{array}$$

where the form $\varphi_{\mathbf{b}} : \mathbb{R}^n \to \mathbb{R}$ corresponding to the column vector $\mathbf{b}$ is defined by $\varphi_{\mathbf{b}}(\mathbf{x}) = \mathbf{b}^T\mathbf{x}$. But every linear function corresponds to a matrix, and the linear function $\varphi_{\mathbf{b}} : \mathbb{R}^n \to \mathbb{R}$ corresponds to the $1 \times n$ row vector $\mathbf{b}^T$. In the language of Chapter 2, we have

$$[\varphi_{\mathbf{b}}] = \mathbf{b}^T.$$

Thus it makes sense to identify the dual space $(\mathbb{R}^n)^\vee$ with the space of row vectors, and the isomorphism $\mathbb{R}^n \cong (\mathbb{R}^n)^\vee$ with transposition:[4]

$$\begin{array}{ccc} \mathbb{R}^n & \cong & (\mathbb{R}^n)^\vee \\ \mathbf{b} & \leftrightarrow & \mathbf{b}^T. \end{array}$$

For infinite dimensional spaces we can no longer use matrices. However, if V is an infinite dimensional Hilbert space of functions, such as $L^2(\mathbb{C})$, and $V^\vee$ is its dual space of **continuous** functionals, Dirac introduced the following notation:

$$\begin{array}{ccc} V & \cong & V^\vee \\ |f\rangle & \leftrightarrow & \langle f|. \end{array}$$

This notation is compatible with the inner product notation $\langle -, - \rangle$ since, by definition, the functional $\langle f| \in V^\vee$ acts on the vector $|g\rangle \in V$ by

$$\langle f| \text{ acting on } |g\rangle = \langle f, g\rangle.$$

Hence in the physics notation the inner product is written as $\langle f|g\rangle$.

1.2 Bilinear Forms

Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). A *bilinear form* is a function

$$\varphi : V \times V \to \mathbb{R}$$

that is linear in each coordinate:

- $\varphi(\mathbf{u}, \sum a_i \mathbf{v}_i) = \sum a_i \varphi(\mathbf{u}, \mathbf{v}_i)$,
- $\varphi(\sum a_i \mathbf{u}_i, \mathbf{v}) = \sum a_i \varphi(\mathbf{u}_i, \mathbf{v})$.

[4] Another piece of jargon: Sometimes the elements of $(\mathbb{R}^n)^\vee$ are called *co-vectors*.

Remark: Over $\mathbb{C}$ we want one of the coordinates to be conjugate linear. In this course I have picked the first coordinate:

$$\varphi(\sum a_i \mathbf{u}_i, \mathbf{v}) = \sum a_i^* \varphi(\mathbf{u}_i, \mathbf{v}).$$

In this case we say that φ is *sesquilinear* (one-and-a-half times linear) instead of bilinear. For example, an inner product is a bilinear function and a Hermitian inner product is a sesquilinear function.

As with linear forms, we begin with the case of Euclidean space. Let $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear form, and for any two basis vectors $\mathbf{e}_i, \mathbf{e}_j \in \mathbb{R}^n$ define the scalar

$$b_{ij} := \varphi(\mathbf{e}_i, \mathbf{e}_j).$$

Then for any vectors $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$ and $\mathbf{y} = y_1\mathbf{e}_1 + \cdots + y_n\mathbf{e}_n$ we have

$$\begin{aligned}\varphi(\mathbf{x}, \mathbf{y}) &= \varphi(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n, y_1\mathbf{e}_1 + \cdots + y_n\mathbf{e}_n) \\ &= \sum x_i y_j \varphi(\mathbf{e}_i, \mathbf{e}_j) \\ &= \sum x_i y_i b_{ij}.\end{aligned}$$

If we let B be the $n \times n$ matrix with ij entry b_{ij} then this becomes

$$\varphi(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T B \mathbf{y} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Exercise: Verify this. Conversely, for any $n \times n$ matrix B we can define a bilinear form φ_B by

$$\varphi_B(\mathbf{x}, \mathbf{y}) := \mathbf{x}^T B \mathbf{y}.$$

If B has ij entry b_{ij} then it follows that

$$\varphi_B(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{e}_i^T B \mathbf{e}_j = (i\text{th row of } B)\mathbf{e}_j = b_{ij}.$$

Hence for any $n \times n$ matrices B and C we have

$$\begin{aligned}\varphi_B = \varphi_C \quad &\Longrightarrow \quad \varphi_B(\mathbf{x}, \mathbf{y}) = \varphi_C(\mathbf{x}, \mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \\ &\Longrightarrow \quad \varphi_B(\mathbf{e}_i, \mathbf{e}_j) = \varphi_C(\mathbf{e}_i, \mathbf{e}_j) \text{ for all } i, j \\ &\Longrightarrow \quad b_{ij} = c_{ij} \text{ for all } i, j \\ &\Longrightarrow \quad B = C.\end{aligned}$$

In summary, we obtain a bijection between $n \times n$ matrices and bilinear forms:

$$\begin{aligned}\text{square matrices } \mathbb{R}^{n\times n} \quad &\leftrightarrow \quad \text{bilinear forms } \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \\ B \quad &\leftrightarrow \quad \varphi_B.\end{aligned}$$

We can also view this as an isomorphism of vector spaces, since bilinear forms can be added and multiplied by scalars, as can any kind of functions with values in $\mathbb{R}$. The following result compares properties of the form φ_B to properties of the matrix B.

Theorem (Properties of Bilinear Forms). Let B be an $n \times n$ matrix over $\mathbb{R}$ (or $\mathbb{C}$) and consider the bilinear (or sesquilinear) form φ_B defined by[5]

$$\varphi_B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T B \mathbf{y} \text{ over } \mathbb{R} \qquad \text{or} \qquad \varphi_B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^* B \mathbf{y} \text{ over } \mathbb{C}.$$

(a) **Symmetric.** We have

$$\begin{aligned} \varphi_B(\mathbf{x}, \mathbf{y}) = \varphi_B(\mathbf{y}, \mathbf{x}) \text{ for all } \mathbf{x}, \mathbf{y} \quad &\Longleftrightarrow \quad B^T = B, \\ \varphi_B(\mathbf{x}, \mathbf{y})^* = \varphi_B(\mathbf{y}, \mathbf{x}) \text{ for all } \mathbf{x}, \mathbf{y} \quad &\Longleftrightarrow \quad B^* = B. \end{aligned}$$

In the first case we say that the form φ_B and the matrix B are *symmetric*. In second case we say they are *Hermitian*.

(b) **Positive Semi-Definite.** We have

$$\varphi_B(\mathbf{x}, \mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \quad \Longleftrightarrow \quad B = A^T A \text{ (or } B = A^* A) \text{ for some matrix } A.$$

In this case the form φ_B and the matrix B are called *positive semi-definite*.[6]

(c) **Positive Definite.** Let φ_B be positive semi-definite, so that $B = A^T A$ (or $B = A^* A$) as in part (b). Then we have

$$\varphi_B(\mathbf{x}, \mathbf{x}) = 0 \text{ implies } x = \mathbf{0} \quad \Longleftrightarrow \quad \text{the matrix } A \text{ has independent columns.}$$

In this case the form φ_B and the matrix B are called *positive definite*.

(d) **Negative.** If $B = -A^T A$ (or $B = -A^* A$) for some matrix A then we have

$$\varphi_B(\mathbf{x}, \mathbf{x}) \leq 0 \text{ for all } x,$$

in which case we say that φ_B and B are *negative semi-definite*. If, in addition, the matrix A has independent columns then

$$\varphi_B(\mathbf{x}, \mathbf{x}) = 0 \text{ implies } \mathbf{x} = \mathbf{0},$$

in which case we say that φ_B and B are *negative definite*.

(e) **Indefinite.** If B is not of the form $\pm A^T A$ (or $\pm A^* A$) for some matrix A, then there exist points $\mathbf{x}$ and $\mathbf{y}$ such that

$$\varphi_B(\mathbf{x}, \mathbf{x}) > 0 \quad \text{and} \quad \varphi_B(\mathbf{y}, \mathbf{y}) < 0.$$

In this case we say that φ_B and B are *indefinite*.

[5]Recall: For any matrix A with complex entries, A^* denotes the conjugate transpose matrix. If $\mathbf{x}$ is a column vector then $\mathbf{x}^*$ is a row vector.

[6]Some books use the alternate term *non-positive definite*.

Example: The identity matrix I corresponds to the standard dot product $\varphi_I(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{y}$ on $\mathbb{R}^n$ and the standard Hermitian product $\varphi_I(\mathbf{x},\mathbf{y}) = \mathbf{x}^*\mathbf{y}$ on $\mathbb{C}^n$, both of which are positive definite. Indeed, we can write $I = I^T I$, where I has independent columns.

Remark: Many problems in applied mathematics seek to minimize an expression of the form $\mathbf{x}^T B\mathbf{x}$ (or $\mathbf{x}^* B\mathbf{x}$). If we know that $B = A^T A$ (or $B = A^* A$) for some matrix A with independent columns then we are guaranteed that a unique minimum exists. Indeed, from part (b) we know that $\mathbf{x}^T B\mathbf{x} \geq 0$ for all $\mathbf{x}$ and from part (c) we know that $\mathbf{x}^T B\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.

Proof. We only prove the complex versions, since the real versions are just a special case. Furthermore, we will only prove one direction of (b) and (c). The other directions are harder and we will prove them after discussing the Spectral Theorem.

(a): If b_{ij} is the ij entry of the matrix B then we have seen that $\varphi_B(\mathbf{e}_i,\mathbf{e}_j) = b_{ij}$ where $\mathbf{e}_i$ and $\mathbf{e}_j$ are standard basis vectors. Suppose that $\varphi_B(\mathbf{x},\mathbf{y})^* = \varphi_B(\mathbf{y},\mathbf{x})$ for all $\mathbf{x},\mathbf{y} \in \mathbb{C}^n$, then in particular we must have

$$b_{ij}^* = \varphi_B(\mathbf{e}_i,\mathbf{e}_j)^* = \varphi_B(\mathbf{e}_j,\mathbf{e}_j) = b_{ij},$$

and hence $B^* = B$. Conversely, suppose that $B^* = B$. Then for all $\mathbf{x},\mathbf{y} \in \mathbb{C}^n$ we have

$$\begin{aligned}\varphi_B(\mathbf{x},\mathbf{y})^* &= (\mathbf{x}^* B\mathbf{y})^* \\ &= \mathbf{y}^* B^*(\mathbf{x}^*)^* \\ &= \mathbf{y}^* B\mathbf{x} \\ &= \varphi_B(\mathbf{y},\mathbf{x}).\end{aligned}$$

(b): Suppose that $B = A^* A$ for some matrix A, and let $\|\mathbf{v}\| = \sqrt{\mathbf{v}^*\mathbf{v}}$ be the standard Hermitian norm on $\mathbb{C}^n$. Then for all $\mathbf{x} \in \mathbb{C}^n$ we have

$$\begin{aligned}\varphi_B(\mathbf{x},\mathbf{x}) &= \mathbf{x}^* B\mathbf{x} \\ &= \mathbf{x}^* A^* A\mathbf{x} \\ &= (A\mathbf{x})^*(A\mathbf{x}) \\ &= \|A\mathbf{x}\|^2 \geq 0.\end{aligned}$$

(c): Continuing from (b), suppose that $\varphi_B(\mathbf{x},\mathbf{x}) = 0$, so that $\|A\mathbf{x}\|^2 = 0$. This implies that $A\mathbf{x} = \mathbf{0}$ because of properties of the standard Hermitian norm.[7] But if A has independent columns then this implies that $\mathbf{x} = \mathbf{0}$. There are many ways to see this. One method uses the fact that $(A^T A)^{-1}$ exists to get

$$\begin{aligned}A\mathbf{x} &= \mathbf{0} \\ A^T A\mathbf{x} &= A^T\mathbf{0}\end{aligned}$$

[7]Recall that $\|\mathbf{v}\|^2 = |v_1|^2 + \cdots + |v_n|^2$, so that $\|\mathbf{v}\| = 0$ if and only if $|v_i| = 0$ (and hence $v_i = 0$) for all i.

$$\begin{aligned} A^T A\mathbf{x} &= \mathbf{0} \\ \mathbf{x} &= (A^T A)^{-1}\mathbf{0} \\ \mathbf{x} &= \mathbf{0}. \end{aligned}$$

(d): This follows from (b) and (c), and the fact that

$$\varphi_{-B}(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T(-B)\mathbf{x} = -\mathbf{x}^T B\mathbf{x} = -\varphi_B(\mathbf{x}, \mathbf{x}).$$

(e): This follows from (b), (c) and (d). □

As with linear forms, it is also possible to define bilinear (sesquilinear) forms on infinite dimensional vector spaces. Let V be any Hermitian inner product space over $\mathbb{C}$ and let $B : V \to V$ be any linear operator.[8] Then we can define a function $\varphi_B : V \times V \to \mathbb{C}$ by

$$\varphi_B(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, B\mathbf{y}\rangle.$$

In the finite dimensional case this corresponds to $\langle \mathbf{x}, B\mathbf{y}\rangle = \mathbf{x}^* B\mathbf{y}$, where B is a matrix. If B^* is the conjugate transpose matrix, then we observe that

$$\langle B^*\mathbf{x}, \mathbf{y}\rangle = (B^*\mathbf{x})^*\mathbf{y} = \mathbf{x}^*(B^*)^*\mathbf{y} = \mathbf{x}^* B\mathbf{y} = \langle \mathbf{x}, B\mathbf{y}\rangle.$$

This computation suggests a way to define a "conjugate transpose operator" $B^* : V \to V$, even when V is infinite dimensional. The definition is really a theorem.

Theorem (Adjoint Operators). Let V be a complex Hilbert space and consider a linear operator $B : V \to V$. If B is **continuous** with respect to the standard norm $\| - \| = \sqrt{\langle -, - \rangle}$ then there exists a unique linear operator $B^* : V \to V$, which is also continuous, satisfying

$$\langle B^*\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, B\mathbf{v}\rangle \text{ for all } \mathbf{u}, \mathbf{v} \in V.$$

The operator B^* is called the *adjoint* of B.[9]

These ideas are particularly important in quantum mechanics. In the standard statistical interpretation, a nonzero vector in Hilbert space $\psi \in V$ corresponds to the *state* of a quantum system. An operator $Q : V \to V$ satisfyiing $Q^* = Q$ corresponds to an *observable quantity.* The outcome of a measurement is random but the *expected value* of quantity Q on state ψ is

$$\langle \psi, Q\psi\rangle \quad \text{ or } \langle \psi|Q|\psi\rangle \text{ in Dirac notation.}$$

Those who study quantum mechanics will notice that it is mostly linear algebra, but the notation is different and the vectors and operators are sometimes just pretend.[10]

[8] Yet another fancy word that just means "function".

[9] An operator is continuous if and only if it is bounded

[10] Indeed, we have seen that the "functions" $\delta(x)$ and $e^{2\pi i x}$ are treated as elements of $L^2(\mathbb{C})$, even though $e^{2\pi i x}$ is not square integrable and $\delta(x)$ doesn't really exist. Furthermore, the theorem on adjoints applies to **continuous** operators, but many operators of interest in quantum mechanics, such as position and momentum, are not continuous.

1.3 Quadratic Forms

Let V be a vector space over $\mathbb{R}$. Given a bilinear form $\varphi : V \times V \to \mathbb{R}$ we define the corresponding *quadratic form* $Q : V \to \mathbb{R}$ by

$$Q(\mathbf{x}) := \varphi(\mathbf{x}, \mathbf{x}).$$

In the case of Euclidean space $V = \mathbb{R}^n$ suppose that $\varphi(\mathbf{x}, \mathbf{y}) = \varphi_B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T B \mathbf{y}$ for a square matrix B. Then the corresponding quadratic form is

$$Q_B(\mathbf{x}) = \mathbf{x}^T B \mathbf{x}.$$

Quadratic forms give a relationship between polynomials of degree 2 and linear algebra. For example, consider a polynomial in two variables:

$$f(x, y) = 2 + x - y + 3x^2 + 2xy + 4y^2.$$

We can express this in terms of linear algebra as follows:

$$f(x,y) = 2 + \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Indeed, for any 2×2 matrix B we observe that

$$\begin{aligned}
\mathbf{x}^T B \mathbf{x} &= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} \\
&= x(ax + by) + y(cx + dy) \\
&= ax^2 + bxy + cyx + dy^2 \\
&= ax^2 + (b + c)xy + dy^2.
\end{aligned}$$

This formula shows that the choice of b and c is not unique. It is common to choose $b = c$ so that the corresponding matrix B is symmetric. Thus we can express any polynomial $\alpha x^2 + \beta xy + \gamma y^2$ in terms of a symmetric matrix:

$$\alpha x^2 + \beta xy + \gamma y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \alpha & \beta/2 \\ \beta/2 & \gamma \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

And we can rewrite the polynomial $f(x, y)$ above using a symmetric matrix:

$$f(x,y) = 2 + \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

More generally, let $\mathbf{x} = (x_1, \ldots, x_n)$ be a vector of n unknowns. Then any polynomial $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ of degree 2 has a unique expression of the form

$$f(\mathbf{x}) = b + \mathbf{b}^T \mathbf{x} + \mathbf{x}^T B \mathbf{x},$$

where b is a scalar, $\mathbf{b}^T$ is a row vector and B is a symmetric matrix. For example, in the case $n = 3$ it is common to write $\mathbf{x} = (x, y, z)$ instead of $\mathbf{x} = (x_1, x_2, x_3)$. Then we have

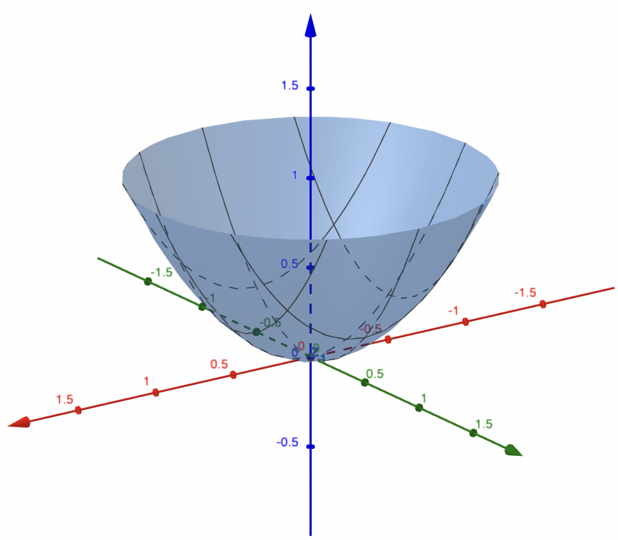
$$\begin{aligned}
f(x,y,z) &= b + b_1 x + b_2 y + b_3 z + b_{11}x^2 + b_{22}y^2 + b_{33}z^2 + b_{12}xy + b_{13}xz + b_{23}yz \\
&= b + \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} b_{11} & b_{12}/2 & b_{13}/2 \\ b_{12}/2 & b_{22} & b_{23}/2 \\ b_{13}/2 & b_{23}/2 & b_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\
&= b + \mathbf{b}^T\mathbf{x} + \mathbf{x}^T B \mathbf{x}.
\end{aligned}$$

Thus the degree zero terms correspond to a scalar b, the degree 1 terms correspond to a vector $\mathbf{b}^T$,[11] and the degree 2 terms correspond to a matrix B. To describe higher degree polynomials we would need cubes of numbers, hypercubes of numbers, etc. Such objects are called “tensors” and they are more difficult to work with. Luckily, degree 2 polynomials are sufficient for most applications.[12]

Here are three simplest examples of quadratic forms. Let

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{so that } Q_B(\mathbf{x}) = \mathbf{x}^T B \mathbf{x} = x^2 + y^2.$$

The graph of $Q_B(x, y)$ in $\mathbb{R}^3$ looks like a paraboloid with a unique minimum at $(0, 0)$:



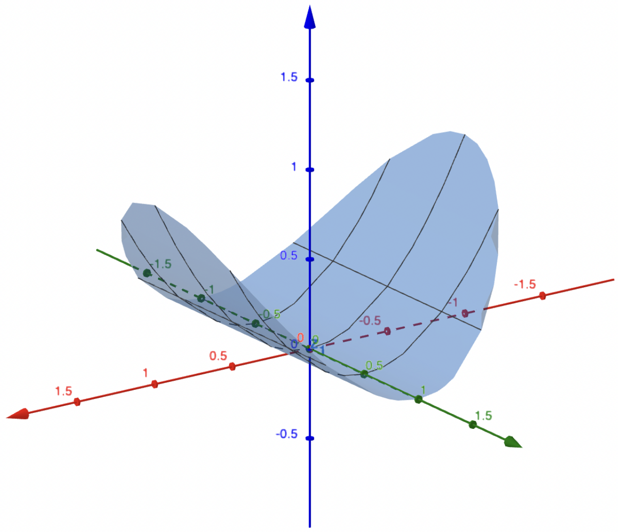
Indeed, this matrix is positive definite because it can be factored as $B = I^T I$, where I is the identity matrix, which has independent columns. Next, let

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{so that } Q_B(\mathbf{x}) = \mathbf{x}^T B \mathbf{x} = x^2.$$

[11]It doesn’t matter whether we write the degree 1 terms as $\mathbf{b}^T\mathbf{x}$ or $\mathbf{x}^T\mathbf{b}$. I am simply following the convention from Section 1.1, where linear forms correspond to row vectors.

[12]It is a curious fact that most physical laws can be expressed in terms of first and second derivatives. Higher derivatives are almost never useful.

The graph of $Q_B(x, y)$ in $\mathbb{R}^3$ is a parabolic cylinder:



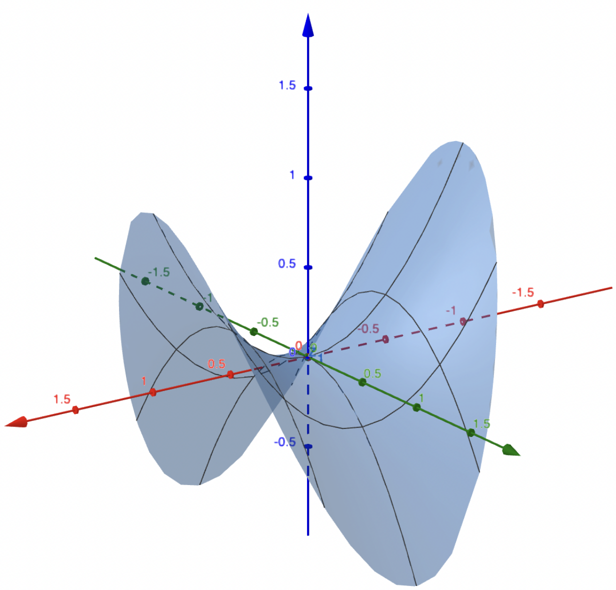
This time the minimum is not unique, since $Q_B(0, y) = 0$ for any value of y. Indeed, this matrix can be factored as

$$B = A^T A = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix},$$

where the matrix A does **not** have independent columns. Finally, let

$$B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ so that } Q_B(\mathbf{x}) = \mathbf{x}^T B \mathbf{x} = x^2 - y^2.$$

This time the graph of $Q_B(x, y)$ in $\mathbb{R}^3$ is a saddle:



Since Q_B takes both positive and negative values, it follows from the previous section that B cannot be factored as $B = A^T A$ for any matrix A, although this is a bit hard to see directly.

In the next chapter we will prove the Spectral Theorem, which makes the analysis of quadratic forms much easier. As a preview, we will prove the following results. Let B be a square matrix satisfying $B^T = B$. Then:

- The eigenvalues of B are real.
- B is positive semi-definite if and only if all eigenvalues are ≥ 0.
- B is positive definite if and only if all eigenvalues are > 0.
- B if indefinite if and only if there exist both positive and negative eigenvalues.

1.4 Multivariable Taylor Expansion

From calculus we are familiar with the idea of a Taylor series. Suppose that a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable k times at the point $p \in \mathbb{R}$. Then for small values of x we have

$$f(p + x) = f(p) + f'(p)x + \frac{1}{2}f''(p)x^2 + \cdots + \frac{1}{k!}f^{(k)}(p)x^k + \text{higher terms},$$

where the higher terms are vanishingly small.[13]

The concept of Taylor series can be generalized to higher dimensions using a little bit of linear algebra. Consider a real valued function $f : \mathbb{R}^n \to \mathbb{R}$ written as

$$f(\mathbf{x}) = f(x_1, \ldots, x_n),$$

where $\mathbf{x} \in \mathbb{R}^n$ is the input vector. We will denote first partial derivatives by

$$f_i = \frac{\partial}{\partial x_i} f,$$

and second partial derivatives by

$$f_{ij} = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f.$$

Note that f_i and f_{ij} are themselves functions from $\mathbb{R}^n$ to $\mathbb{R}$. Suppose that the first and second partials exist and are continuous at some point $\mathbf{p} \in \mathbb{R}^n$. Then Clairaut's theorem tells us that

$$f_{ij}(\mathbf{p}) = f_{ji}(\mathbf{p}) \text{ for all } i, j.$$

Furthermore, we define the gradient vector at $\mathbf{p}$:

$$(\nabla f)_{\mathbf{p}} = \begin{pmatrix} f_1(\mathbf{p}) \\ \vdots \\ f_n(\mathbf{p}) \end{pmatrix}$$

[13] The exact nature of the higher terms will not concern us; we don't do analysis in this course.

and the Hessian matrix at $\mathbf{p}$:

$$(Hf)_{\mathbf{p}} = \begin{pmatrix} f_{11}(\mathbf{p}) & \cdots & f_{1n}(\mathbf{p}) \\ \vdots & & \vdots \\ f_{n1}(\mathbf{p}) & \cdots & f_{nn}(\mathbf{p}) \end{pmatrix}.$$

Note that the Hessian matrix is symmetric. Then for small vectors $\mathbf{x} \in \mathbb{R}^n$, the multivariable Taylor series tells us that

$$f(\mathbf{p} + \mathbf{x}) = f(\mathbf{p}) + (\nabla f)_p^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T (Hf)_{\mathbf{p}} \mathbf{x} + \text{higher terms},$$

where the higher terms are vanishingly small. Note the relationship to linear and bilinear forms. The linear part of the Taylor series is a linear form

$$\mathbf{x} \mapsto (\nabla f)_{\mathbf{p}}^T \mathbf{x} = f_1(\mathbf{p})x_1 + f_2(\mathbf{p})x_2 + \cdots + f_n(\mathbf{p})x_n,$$

and the quadratic part of the Taylor series is a quadratic form

$$\mathbf{x} \mapsto \frac{1}{2}\mathbf{x}^T (Hf)_{\mathbf{p}} \mathbf{x} = \frac{1}{2} \sum f_{ij}(\mathbf{p}) x_i x_j.$$

Higher terms of the Taylor series can be described by multilinear forms, but, as I said, these don't come up much in applications.

For example, consider again the polynomial function, with $(x_1, x_2) = (x, y)$:

$$f(x, y) = 2 + x - y + 3x^2 + 2xy + 4y^2.$$

We compute the first and second partial derivatives:

$$\begin{aligned} f_1 &= 1 + 6x + 2y, \\ f_2 &= -1 + 2x + 8y, \\ f_{11} &= 6, \\ f_{12} &= 2, \\ f_{21} &= 2, \\ f_{22} &= 8. \end{aligned}$$

This gives the following gradient vector and Hessian matrix:

$$\nabla f = \begin{pmatrix} 1 + 6x + 2y \\ -1 + 2x + 8y \end{pmatrix} \quad \text{and} \quad Hf = \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix}.$$

The Taylor expansion at $\mathbf{p} = (0, 0)$ is

$$f(0 + x, 0 + y) = f(0, 0) + (\nabla f)_{(0,0)}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T (Hf)_{(0,0)} \mathbf{x}$$

$$= 2 + \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 6 & 2 \\ 2 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

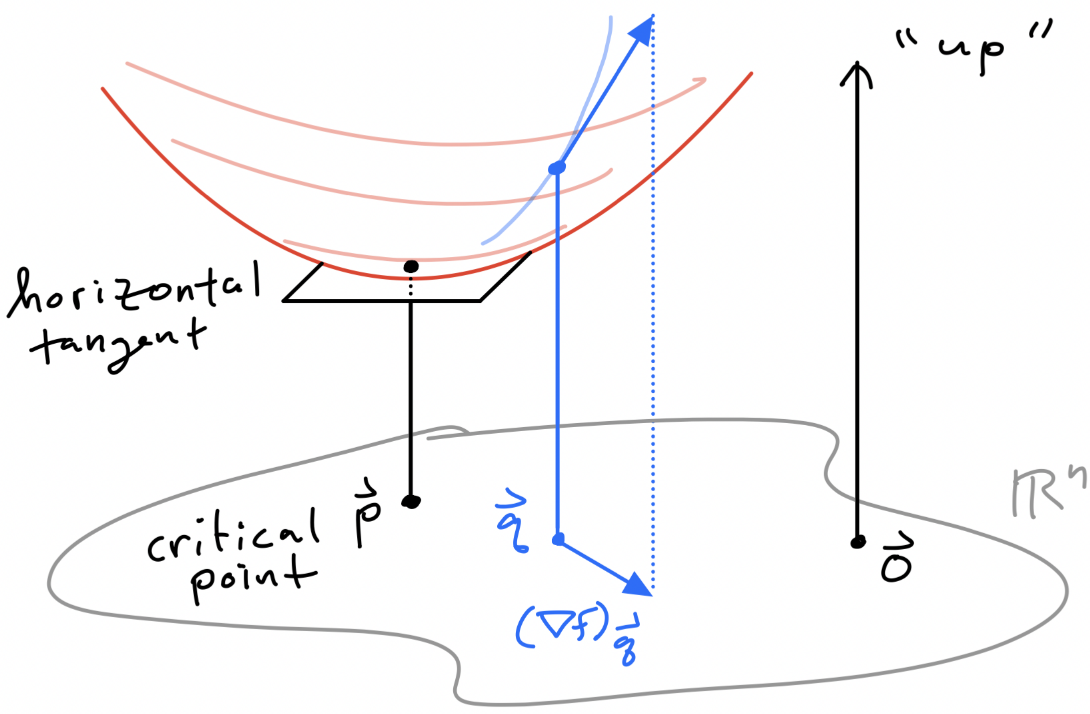
which we already computed in the previous section. The Taylor expansion at $\mathbf{p} = (1, 1)$ is

$$\begin{aligned} f(1+x, 1+y) &= f(1,1) + (\nabla f)^T_{(1,1)} \mathbf{x} + \frac{1}{2} \mathbf{x}^T (Hf)_{(1,1)} \mathbf{x} \\ &= 11 + \begin{pmatrix} 9 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 6 & 2 \\ 2 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

And the Taylor expansion at $\mathbf{p} = (\frac{-10}{44}, \frac{8}{44})$ is

$$\begin{aligned} f\left(\frac{-10}{44} + x, \frac{8}{44} + y\right) &= f\left(\frac{-10}{44}, \frac{8}{44}\right) + (\nabla f)^T_{(\frac{-10}{44}, \frac{8}{44})} \mathbf{x} + \frac{1}{2} \mathbf{x}^T (Hf)_{(\frac{-10}{44}, \frac{8}{44})} \mathbf{x} \\ &= \frac{79}{44} + \begin{pmatrix} 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 6 & 2 \\ 2 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \frac{79}{44} + \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

Note that $\mathbf{p} = (\frac{-10}{44}, \frac{8}{44})$ is a *critical point* of f, since the gradient vector vanishes: $(\nabla f)_\mathbf{p} = \mathbf{0}$. Recall that $(\nabla f)_\mathbf{p}$ is the direction of greatest increase of f near the point $\mathbf{p}$. If $(\nabla f)_\mathbf{p} = \mathbf{0}$ then the function is in equilibrium because it can't decide which way is "up". Here is a picture:



A multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ near a critical point $\mathbf{p}$ is approximately a quadratic form:

$$f(\mathbf{p}+\mathbf{x}) = f(\mathbf{p}) + \frac{1}{2}\mathbf{x}^T(Hf)_{\mathbf{p}}\mathbf{x} + \text{higher terms}.$$

Thus we have the following facts, which are sometimes called the *multivariable second derivative test*. Assume that $(\nabla f)_{\mathbf{p}} = \mathbf{0}$. Then:

- f has a local minimum at $\mathbf{p}$ if and only if $(Hf)_{\mathbf{p}}$ is positive definite.
- f has a local maximum at $\mathbf{p}$ if and only if $(Hf)_{\mathbf{p}}$ is negative definite.

Indeed, if $(Hf)_{\mathbf{p}}$ is positive definite then we have

$$\mathbf{x}^T(Hf)_{\mathbf{p}}\mathbf{x} \geq 0 \text{ for all } \mathbf{x}, \text{ and } \mathbf{x}^T(Hf)_{\mathbf{p}}\mathbf{x} = 0 \text{ if and only if } \mathbf{x} = \mathbf{0}.$$

so that[14]

$$f(\mathbf{p}+\mathbf{x}) \geq f(\mathbf{p}) \text{ for all } \mathbf{x}, \text{ and } f(\mathbf{p}+\mathbf{x}) = f(\mathbf{p}) \text{ if and only if } \mathbf{x} = \mathbf{0},$$

If $(Hf)_{\mathbf{p}}$ is positive (or negative) semi-definite then there is a local minimum (or maximum) in some directions, but in some directions the function is constant. Otherwise, if $(Hf)_{\mathbf{p}}$ is indefinite then there exist small $\mathbf{x}$ and $\mathbf{y}$ such that $f(\mathbf{p}+\mathbf{x}) > f(\mathbf{p})$ and $f(\mathbf{p}+\mathbf{y}) < f(\mathbf{p})$. Geometrically, this is a higher dimensional saddle point.

In the previous example, it happens that

$$\text{the matrix } B = \begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix} \text{ is positive definite,}$$

so the function $f(x,y) = f(x,y) = 2 + x - y + 3x^2 + 2xy + 4y^2$ has a local minimum at $\mathbf{p} = \left(\frac{-10}{44}, \frac{8}{44}\right)$. To verify that B is positive definite, I computed the eigenvalues $7+\sqrt{5}$ and $7-\sqrt{5}$, which are both positive. Later I will show you how to find a matrix A with independent columns such that $B = A^T A$. Such a matrix is not unique; here is one example, called the *Cholesky decomposition*:

$$A = \begin{pmatrix} \sqrt{3} & \sqrt{3}/3 \\ 0 & \sqrt{33}/3 \end{pmatrix}.$$

Check:

$$A^T A = \begin{pmatrix} \sqrt{3} & 0 \\ \sqrt{3}/3 & \sqrt{33}/3 \end{pmatrix} \begin{pmatrix} \sqrt{3} & \sqrt{3}/3 \\ 0 & \sqrt{33}/3 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix}.$$

[14] Remember, the higher order terms are vanishingly small, so they don't affect the inequality.

2 Determinants

2.1 Multilinear Forms

We have studied linear and bilinear forms. Now we discuss the general situation. Let V be a vector space over $\mathbb{R}$, and recall the notation for Cartesian product:

$$V^k := V \times V \times \cdots \times V = \{(\mathbf{x}_1, \ldots, \mathbf{x}_k) : \mathbf{x}_i \in \mathbb{R}^n \text{ for all } i\}.$$

A *multilinear k-form* is a function

$$\varphi : V^k \to \mathbb{R}$$

that is linear in each input. In other words, for any index i we have

$$\varphi\left(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \sum a_j \mathbf{u}_j, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k\right) = \sum a_i \varphi(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{u}_j, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k).$$

(This time we don't bother with Hermitian forms, since it's not clear where to put the complex conjugates.) Just as with linear and bilinear forms, k-forms can be added and multiplied by scalars. That is, given k-forms φ, ψ and scalar a, we define the k-form $\varphi + a\psi$ by

$$(\varphi + a\psi)(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \varphi(\mathbf{v}_1, \ldots, \mathbf{v}_k) + a\psi(\mathbf{v}_1, \ldots, \mathbf{v}_k).$$

Thus we obtain a vector space of multilinear k-forms:[15]

$$\mathcal{T}^k(V) = \{\text{multilinear } k\text{-forms } \varphi : V^k \to \mathbb{R}\}.$$

In the case $k = 1$ we also use the notation of the dual space

$$V^\vee = \mathcal{T}^1(V) = \{\text{linear forms } V \to \mathbb{R}\}.$$

For example, consider Euclidean space $V = \mathbb{R}^n$. In the previous section we proved that $\mathcal{T}^1(\mathbb{R}^n)$ is isomorphic to the vector space of row vectors:

$$\mathcal{T}^1(\mathbb{R}^n) \cong \{1 \times n \text{ row vectors}\} = \mathbb{R}^{1\times n},$$

and hence

$$\dim \mathcal{T}^1(\mathbb{R}^n) = n.$$

We also proved that $\mathcal{T}^2(\mathbb{R}^n)$ is isomorphic to the vector space of $n \times n$ matrices:

$$\mathcal{T}^2(\mathbb{R}^n) \cong \{n \times n \text{ matrices}\} = \mathbb{R}^{n\times n},$$

and hence[16]

$$\dim \mathcal{T}^1(\mathbb{R}^n) = n^2.$$

[15] The letter $\mathcal{T}$ is for "tensor".

[16] An $n \times n$ matrix is uniquely determined by its n^2 entries. More formally, let E_{ij} the the $n \times n$ matrix with 1 in the ij position and zeros elsewhere. Then the set of matrices E_{ij} with $1 \leq i, j \leq n$ is a basis for $\mathbb{R}^{n\times n}$. More generally, one can show that $\mathbb{R}^{m\times n}$ has dimension mn.

More generally, I claim that

$$\dim \mathcal{T}^k(\mathbb{R}^n) = n^k.$$

In order to prove this we will construct a "standard basis" for $\mathcal{T}^k(\mathbb{R}^n)$.

Theorem (The Dual Standard Basis). Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis for $\mathbb{R}^n$. Now we will construct a corresponding "standard basis" for the dual space $(\mathbb{R}^n)^\vee = \mathcal{T}^1(\mathbb{R}^n)$. For all $1 \leq i \leq n$, let $\varepsilon_i : \mathbb{R}^n \to \mathbb{R}$ be the linear form defined by picking out the ith coordinate:

$$\varepsilon_i(\mathbf{x}) = \varepsilon_i \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_i.$$

To see that this ε_i is linear, consider any linear combination $\sum a_j \mathbf{x}_j \in \mathbb{R}^n$, where x_{ij} is the ith entry of the vector $\mathbf{x}_j \in \mathbb{R}^n$. Then we have

$$\varepsilon_i \left(\sum a_j \mathbf{x}_j \right) = \varepsilon_i \begin{pmatrix} \sum a_j x_{1j} \\ \vdots \\ \sum a_j x_{nj} \end{pmatrix} = \sum a_j x_{ij} = \sum a_j \varepsilon_i(\mathbf{x}_j).$$

In the previous section we showed that every linear form $\varphi : \mathbb{R}^n \to \mathbb{R}$ can be expressed as $\varphi(\mathbf{x}) = \mathbf{b}^T \mathbf{x}$ for some unique vector $\mathbf{b} = (b_1, \ldots, b_n)$. Equivalently, each linear form φ can be expressed as

$$\varphi = b_1 \varepsilon_1 + \cdots + b_n \varepsilon_n,$$

for some unique scalars $b_1, \ldots, b_n$. This shows that $\varepsilon_1, \ldots, \varepsilon_2$ is indeed a basis for $(\mathbb{R}^n)^\vee$. In terms of matrices, note that

$$\varepsilon_i(\mathbf{x}) = \varepsilon_i \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_i = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

which shows that the linear function ε_i corresponds to a standard row vector:

$$[\varepsilon_i] = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}.$$

Finally, we say that the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ and $\varepsilon_1, \ldots, \varepsilon_n \in (\mathbb{R}^n)^\vee$ are "dual" because

$$\varepsilon_i(\mathbf{e}_j) = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

If we were only going to talk about row vectors and column vectors then this level of abstraction is completely unnecessary. However, it becomes necessary when we talk about k-forms.

Tensor Product of Forms. Let V be a vector space over $\mathbb{R}$. Consider a k-form $\varphi : V^k \to \mathbb{R}$ and an ℓ-form $\psi : V^\ell \to \mathbb{R}$. Then the *tensor product* $\varphi \otimes \psi$ is a $(k+\ell)$-form defined as follows:

$$(\varphi \otimes \psi)(\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_{k+\ell}) := \varphi(\mathbf{v}_1, \ldots, \mathbf{v}_k) \cdot \psi(\mathbf{v}_{k+1}, \ldots, \mathbf{v}_{k+\ell}).$$

It is straightforward to check that this function is linear, and hence $\varphi \otimes \psi \in \mathcal{T}^{k+\ell}(V)$. One can also check that the tensor product is associative, hence if φ, ψ, ω are k, ℓ, m-forms, respectively, then we obtain a $(k+\ell+m)$-form:

$$\varphi \otimes \psi \otimes \omega = (\varphi \otimes \psi) \otimes \omega = \varphi \otimes (\psi \otimes \omega).$$

For example, for any standard 1-forms ε_i and ε_j we obtain a 2 form $\varepsilon_i \otimes \varepsilon_j$ defined as follows:

$$(\varepsilon_i \otimes \varepsilon_j)(\mathbf{v}_1, \mathbf{v}_2) = \varepsilon_i(\mathbf{v}_1) \cdot \varepsilon_j(\mathbf{v}_2).$$

And for any standard 1-forms $\varepsilon_i, \varepsilon_j, \varepsilon_k$ we obtain a 3-form $\varepsilon_i \otimes \varepsilon_j \otimes \varepsilon_k$ by

$$(\varepsilon_i \otimes \varepsilon_j \otimes \varepsilon_k)(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \varepsilon_i(\mathbf{v}_1) \cdot \varepsilon_j(\mathbf{v}_2) \cdot \varepsilon_k(\mathbf{v}_3).$$

To be more explicit let's consider an example with $V = \mathbb{R}^3$. Then we have

$$(\varepsilon_1 \otimes \varepsilon_2)\left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}\right) = \varepsilon_1 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \cdot \varepsilon_2 \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = (1)(3) = 3$$

and

$$(\varepsilon_2 \otimes \varepsilon_1)\left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}\right) = \varepsilon_2 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \cdot \varepsilon_1 \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = (-1)(2) = -2,$$

which shows that $\varepsilon_1 \otimes \varepsilon_2$ and $\varepsilon_2 \otimes \varepsilon_1$ define different bilinear functions. In other words, we see that the **tensor product is not commutative**.

Theorem (The Standard Basis of k-Forms). Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the standard basis of $\mathbb{R}^n$ and let $\varepsilon_1, \ldots, \varepsilon_n$ be the dual standard basis of $\mathcal{T}^1(\mathbb{R}^n)$. Then I claim that the following set is a basis for the vector space of k-forms:

$$\{\varepsilon_{i_1} \otimes \varepsilon_{i_2} \otimes \cdots \otimes \varepsilon_{i_k} : i_1, i_2, \ldots, i_k \in \{1, 2, \ldots, n\}\}.$$

Note that this basis contains n^k elements, and hence

$$\dim \mathcal{T}^k(\mathbb{R}^n) = n^k.$$

We won't bother to prove this since we have already proved the cases $k = 1$ and $k = 2$ in the previous section. The general proof is similar, but with more horrible notation. To see how this works, we will repeat our proof for $k = 2$ in the new language. Let B be an $n \times n$ matrix with ij entry b_{ij} and consider the 2-form

$$\varphi_B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T B \mathbf{y}.$$

Note that for any basis vectors $\mathbf{e}_i, \mathbf{e}_j$ we have

$$\varphi_B(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{e}_i^T B \mathbf{e}_j = b_{ij}.$$

Furthermore, for any vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ we have

$$\varphi_B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T B \mathbf{y} = \sum b_{ij} x_i y_j.$$

On the other hand, since $(\varepsilon_i \otimes \varepsilon_j)(\mathbf{x}, \mathbf{y}) = x_i y_j$, we can express this as

$$\varphi_B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T B y = \sum b_{ij} x_i y_j = \sum b_{ij} (\varepsilon_i \otimes \varepsilon_j)(\mathbf{x}, \mathbf{y}) = \left(\sum b_{ij} (\varepsilon_i \otimes \varepsilon_j)\right)(\mathbf{x}, \mathbf{y}),$$

and hence

$$\varphi_B = \sum b_{ij} (\varepsilon_i \otimes \varepsilon_j).$$

More generally, any 3-form $\varphi \in \mathcal{T}^3(\mathbb{R}^n)$ corresponds to an $n \times n \times n$ cube of numbers b_{ijk}:

$$\varphi = \sum b_{ijk} (\varepsilon_i \otimes \varepsilon_j \otimes \varepsilon_k).$$

These are some kind of "higher dimensional matrices", but they are much harder to work with. In this course we will focus only on very special kinds of k-forms.

Symmetric and Alternating k-Forms. We say that a k-form $\varphi \in \mathcal{T}^k(V)$ is *symmetric* if switching any two inputs leaves the output unchanged. For example, if φ is symmetric then

$$\varphi(\mathbf{v}_2, \mathbf{v}_1, \mathbf{v}_3, \ldots, \mathbf{v}_k) = \varphi(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_k).$$

We say that a k-form φ is *alternating* if switching any two inputs multiplies the output by -1. For example, if φ is alternating then

$$\varphi(\mathbf{v}_2, \mathbf{v}_1, \mathbf{v}_3, \ldots, \mathbf{v}_k) = -\varphi(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_k).$$

To be more explicit, let's consider $V = \mathbb{R}^3$. I claim that the 2-form $\varphi = \varepsilon_1 \otimes \varepsilon_2 + \varepsilon_2 \otimes \varepsilon_1$ is symmetric. Indeed, for any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ we observe that

$$\begin{aligned}
\varphi(\mathbf{x}, \mathbf{y}) &= (\varepsilon_1 \otimes \varepsilon_2 + \varepsilon_2 \otimes \varepsilon_1)\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}\right) \\
&= (\varepsilon_1 \otimes \varepsilon_2)\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}\right) + (\varepsilon_2 \otimes \varepsilon_1)\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}\right) \\
&= \varepsilon_1 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \varepsilon_2 \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} + \varepsilon_2 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \varepsilon_1 \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\
&= x_1 y_2 + x_2 y_1
\end{aligned}$$

is equal to

$$\varphi(\mathbf{y}, \mathbf{x}) = (\varepsilon_1 \otimes \varepsilon_2 + \varepsilon_2 \otimes \varepsilon_1)\left(\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right)$$

$$
\begin{aligned}
&= (\varepsilon_1 \otimes \varepsilon_2)\left(\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) + (\varepsilon_2 \otimes \varepsilon_1)\left(\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) \\
&= \varepsilon_1 \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \cdot \varepsilon_2 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \varepsilon_2 \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \cdot \varepsilon_1 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
&= y_1 x_2 + y_2 x_1 \\
&= x_1 y_2 + x_2 y_1.
\end{aligned}
$$

On the other hand, the 2-form $\psi = \varepsilon_1 \otimes \varepsilon_2 - \varepsilon_2 \otimes \varepsilon_1$ is alternating since

$$\psi(\mathbf{x}, \mathbf{y}) = (\varepsilon_1 \otimes \varepsilon_2 - \varepsilon_2 \otimes \varepsilon_1)(\mathbf{x}, \mathbf{y}) = x_1 y_2 - x_2 y_1$$

and

$$\psi(\mathbf{y}, \mathbf{x}) = (\varepsilon_1 \otimes \varepsilon_2 - \varepsilon_2 \otimes \varepsilon_1)(\mathbf{y}, \mathbf{x}) = y_1 x_2 - y_2 x_1 = -(x_1 y_2 - x_2 y_1) = -\psi(\mathbf{x}, \mathbf{y}).$$

Since the sum of symmetric forms is symmetric, and the sum of alternating forms is alternating, we can define the following vector spaces[17]

$$
\begin{aligned}
\mathcal{S}^k(V) &= \text{the space of symmetric } k\text{-forms } V^k \to \mathbb{R}, \\
\mathcal{A}^k(V) &= \text{the space of alternating } k\text{-forms } V^k \to \mathbb{R}.
\end{aligned}
$$

For small k and n, it is not too hard to write down a basis for $\mathcal{S}^k(\mathbb{R}^n)$ in terms of the standard basis of $\mathcal{T}^k(\mathbb{R}^n)$. To save space, let's write

$$\varepsilon_{ij} = \varepsilon_i \otimes \varepsilon_j, \quad \varepsilon_{ijk} = \varepsilon_i \otimes \varepsilon_j \otimes \varepsilon_k, \quad \text{etc.}$$

Then, for example, we have

$$
\begin{aligned}
\mathcal{S}^1(\mathbb{R}^2) &= \operatorname{span}\{\varepsilon_1, \varepsilon_2\}, \\
\mathcal{S}^2(\mathbb{R}^2) &= \operatorname{span}\{\varepsilon_{11}, \varepsilon_{12} + \varepsilon_{21}, \varepsilon_{22}\}, \\
\mathcal{S}^3(\mathbb{R}^2) &= \operatorname{span}\{\varepsilon_{111}, \varepsilon_{112} + \varepsilon_{121} + \varepsilon_{211}, \varepsilon_{122} + \varepsilon_{212} + \varepsilon_{211}, \varepsilon_{222}\},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathcal{S}^1(\mathbb{R}^3) &= \operatorname{span}\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}, \\
\mathcal{S}^2(\mathbb{R}^3) &= \operatorname{span}\{\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, \varepsilon_{12} + \varepsilon_{21}, \varepsilon_{13} + \varepsilon_{31}, \varepsilon_{23} + \varepsilon_{32}\}, \\
\mathcal{S}^3(\mathbb{R}^3) &= \operatorname{span}\{\varepsilon_{111}, \varepsilon_{222}, \varepsilon_{333}, \\
&\qquad\qquad \varepsilon_{112} + \varepsilon_{121} + \varepsilon_{211}, \varepsilon_{113} + \varepsilon_{131} + \varepsilon_{311}, \varepsilon_{223} + \varepsilon_{232} + \varepsilon_{322}, \\
&\qquad\qquad \varepsilon_{221} + \varepsilon_{212} + \varepsilon_{122}, \varepsilon_{331} + \varepsilon_{313} + \varepsilon_{133}, \varepsilon_{332} + \varepsilon_{323} + \varepsilon_{233},
\end{aligned}
$$

[17] Alternating forms are also called *anti-symmetric.* In advanced calculus, a *differential form* is an alternating k-form whose coefficients can change from point to point. More precisely, a differential form on a k-dimensional manifold assigns an alternating k-form to the tangent space at each point.

$$\varepsilon_{123}+\varepsilon_{132}+\varepsilon_{213}+\varepsilon_{231}+\varepsilon_{312}+\varepsilon_{321}\}.$$

In particular, we have

$$\dim\mathcal{S}^1(\mathbb{R}^3)=2,\quad \dim\mathcal{S}^2(\mathbb{R}^3)=6,\quad \dim\mathcal{S}^3(\mathbb{R}^3)=10.$$

Maybe you can see a pattern here. In general, one can use a combinatorial argument[18] to show that

$$\dim\mathcal{S}^k(\mathbb{R}^n)=\binom{n+k-1}{k}.$$

Let's test this on the special case $k=2$. Recall from the previous section that a symmetric bilinear form is the same thing as a symmetric $n\times n$ matrix, hence $\mathcal{S}^2(\mathbb{R}^n)$ can be identified with the space of symmetric $n\times n$ matrices. A symmetric matrix is uniquely determined by the n diagonal elements and the $n(n-1)/2$ elements above the diagonal. (We don't need to specify the entries below the diagonal because they are equal to the above-diagonal elements.) Hence we must have

$$\dim\mathcal{S}^2(\mathbb{R}^n)=n+\frac{n(n-1)}{2}=\frac{2n+n(n-1)}{2}=\frac{n^2+n}{2}=\frac{(n+1)n}{2},$$

which agrees with the formula

$$\binom{n+2-1}{2}=\binom{n+1}{2}=\frac{(n+1)n}{2}.$$

It is trickier to find a basis for the space of alternating k-forms. Here are some small examples:

$$\begin{aligned}
&\mathcal{A}^1(\mathbb{R}^2)=\operatorname{span}\{\varepsilon_1,\varepsilon_2\},\\
&\mathcal{A}^2(\mathbb{R}^2)=\operatorname{span}\{\varepsilon_{12}-\varepsilon_{21}\},\\
&\mathcal{A}^k(\mathbb{R}^2)=\{0\}\text{ for }k>2,\\
&\mathcal{A}^1(\mathbb{R}^3)=\operatorname{span}\{\varepsilon_1,\varepsilon_2,\varepsilon_3\},\\
&\mathcal{A}^2(\mathbb{R}^3)=\operatorname{span}\{\varepsilon_{12}-\varepsilon_{21},\varepsilon_{13}-\varepsilon_{31},\varepsilon_{23}-\varepsilon_{32}\},\\
&\mathcal{A}^3(\mathbb{R}^3)=\operatorname{span}\{\varepsilon_{123}+\varepsilon_{231}+\varepsilon_{312}-\varepsilon_{132}-\varepsilon_{213}-\varepsilon_{321}\},\\
&\mathcal{A}^k(\mathbb{R}^3)=\{0\}\text{ for }k>3.
\end{aligned}$$

You will prove on the homework that $\dim\mathcal{A}^k(\mathbb{R}^n)=0$ for all $k>n$. That is, if $k>n$ then any alternating k-form on $\mathbb{R}^n$ must be the zero function that sends any k-tuple of vectors in $\mathbb{R}^n$ to zero. For $0\le k\le n$ I claim that[19]

$$\dim\mathcal{A}^k(\mathbb{R}^n)=\binom{n}{k}.$$

[18]There is one basis element of $\mathcal{S}^k(\mathbb{R}^n)$ for each weakly increasing sequence $1\le i_1\le i_2\le\cdots\le i_k\le n$ of k numbers between 1 and n. Such a weakly increasing sequence can be encoded as a word of length $n+k-1$ containing k "stars" and $n-1$ "bars". For example, the word $**|*||***$ corresponds to $1\le 1\le 2\le 4\le 4\le 4$. Such a word has length $k+(n-1)=n+k-1$. The number of such words is $\binom{n+k-1}{k}$ since from $n+k-1$ possible positions, we must choose k positions to place the stars.

[19]The definition of "alternating" doesn't really apply to 0-forms and 1-forms. However, it is convenient to define $\mathcal{A}^0:=\mathcal{T}^0:=\{0\}$ and $\mathcal{A}^1:=\mathcal{T}^1$, so the dimension formula is still correct when $k=0$ and $k=1$.

We won't prove this theorem in general, but we will prove the special case when $k = n$:

$$\dim \mathcal{A}^n(\mathbb{R}^n) = \binom{n}{n} = 1.$$

In other words, there exists a unique (up to scalar multiplication) alternating n-form on $\mathbb{R}^n$. At the risk of spoiling the surprise, I will tell you right now that this unique form is called the *determinant*.

According to the examples listed above, we have

$$\begin{aligned}
\mathcal{A}^2(\mathbb{R}^2) &= \operatorname{span}\{\varepsilon_{12} - \varepsilon_{21}\}, \\
\mathcal{A}^3(\mathbb{R}^3) &= \operatorname{span}\{\varepsilon_{123} + \varepsilon_{231} + \varepsilon_{312} - \varepsilon_{132} - \varepsilon_{213} - \varepsilon_{321}\}.
\end{aligned}$$

Recall that $\varepsilon_{12} - \varepsilon_{21}$ represents the 2-form $\varepsilon_1 \otimes \varepsilon_2 - \varepsilon_2 \otimes \varepsilon_1$, which we have already discussed. When applied to two vectors $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ in $\mathbb{R}^2$ it gives

$$(\varepsilon_{12} - \varepsilon_{21})(\mathbf{x}, \mathbf{y}) = \varepsilon_{12}(\mathbf{x}, \mathbf{y}) - \varepsilon_{21}(\mathbf{x}, \mathbf{y}) = x_1 y_2 - x_2 y_1.$$

In general, if $\varphi \in \mathcal{T}^k(\mathbb{R}^n)$ is a k-form on $\mathbb{R}^n$ and if A is a $n \times k$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$, it is convenient to define

$$\varphi(A) := \varphi(\mathbf{a}_1, \ldots, \mathbf{a}_k).$$

Thus for any 2×2 matrix we have

$$(\varepsilon_{12} - \varepsilon_{21})(A) = (\varepsilon_{12} - \varepsilon_{21}) \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = \varepsilon_{12} \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} - \varepsilon_{21} \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = a_1 b_2 - a_2 b_1,$$

and for any 3×3 matrix

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

we have

$$\begin{aligned}
&(\varepsilon_{123} + \varepsilon_{231} + \varepsilon_{312} - \varepsilon_{132} - \varepsilon_{213} - \varepsilon_{321})(A) \\
&\qquad = \varepsilon_{123}(A) + \varepsilon_{231}(A) + \varepsilon_{312}(A) - \varepsilon_{132}(A) - \varepsilon_{231}(A) - \varepsilon_{321}(A), \\
&\qquad = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_1 b_3 c_2 - a_2 b_3 c_1 - a_3 b_2 c_1.
\end{aligned}$$

You may recognize these formulas from your previous linear algebra course. But where do they come from? And how do we know that there are no other alternating 2-forms on $\mathbb{R}^2$ and no other alternating 3-forms on $\mathbb{R}^3$?

2.2 Uniqueness of the Determinant

As we have seen, the formula for the determinant of a 3×3 matrix is rather complicated. I could give a general formula right now, but it is actually more useful to work with the **properties** of the determinant. Explicit formulas for the determinant are messy, but the properties of the determinant are easy to describe.

As before, we will think of a k-form $\varphi \in \mathcal{T}^k(\mathbb{R}^n)$ as a function sending $n \times k$ matrices to scalars. That is, for any matrix A with columns $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$ we will write

$$\varphi(A) := \varphi(\mathbf{a}_1, \ldots, \mathbf{a}_k).$$

This function is "multilinear in the columns of A". For example, consider some $n \times 3$ matrices

$$A = \left(\mathbf{u} \mid \mathbf{v} \mid \mathbf{a} \right), \quad B = \left(\mathbf{u} \mid \mathbf{v} \mid \mathbf{b} \right), \quad C = \left(\mathbf{u} \mid \mathbf{v} \mid \mathbf{a} + \lambda \mathbf{b} \right),$$

with $\mathbf{u}, \mathbf{v}, \mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then for any 3-form $\varphi \in \mathcal{T}^3(\mathbb{R}^3)$ we have

$$\varphi(C) = \varphi(A) + \lambda \cdot \varphi(B).$$

Warning: Multilinear functions are not linear. For example, consider any bilinear function $\varphi \in \mathcal{T}^2(\mathbb{R}^n)$, and consider any two $n \times 2$ matrices

$$A = \left(\mathbf{a}_1 \mid \mathbf{a}_2 \right) \text{ and } B = \left(\mathbf{b}_1 \mid \mathbf{b}_2 \right), \text{ hence } A + B = \left(\mathbf{a}_1 + \mathbf{b}_1 \mid \mathbf{a}_2 + \mathbf{b}_2 \right).$$

Then we have

$$\begin{aligned} \varphi(A + B) &= \varphi(\mathbf{a}_1 + \mathbf{b}_1, \mathbf{a}_2 + \mathbf{b}_2) \\ &= \varphi(\mathbf{a}_1, \mathbf{a}_2) + \varphi(\mathbf{b}_1, \mathbf{b}_2) + \varphi(\mathbf{a}_1, \mathbf{b}_2) + \varphi(\mathbf{b}_1, \mathbf{a}_2) \\ &= \varphi(A) + \varphi(B) + \varphi(\mathbf{a}_1, \mathbf{b}_2) + \varphi(\mathbf{b}_1, \mathbf{a}_2), \end{aligned}$$

which is **not** equal to $\varphi(A) + \varphi(B)$.[20]

As mentioned in the previous section, there exists a unique (up to scalar multiplication) alternating n-form on $\mathbb{R}^n$, which can be interpreted as the determinant of an $n \times n$ matrix. In this section we will prove that there is no more than one such function, so that

$$\dim \mathcal{A}^n(\mathbb{R}^n) \leq 1,$$

and in the next section we will show that there is at least one such function, so that

$$\dim \mathcal{A}^n(\mathbb{R}^n) \geq 1.$$

Theorem (Uniqueness of the Determinant). Let φ be a function sending $n \times n$ matrices to scalars. We say that φ is a *determinant function* if it satisfies the following three properties:

[20] For the same reason, we will have $\det(A + B) \neq \det(A) + \det(B)$.

(1) *Multilinear.* The function φ is linear in each individual column.

(2) *Alternating.* If A' is obtained from A by swapping two columns, then $\varphi(A') = -\varphi(A)$.

(3) *Normalized.* The function φ sends the identity matrix I_n to 1.

In other words, a determinant function is an alternating n-form $\varphi \in \mathcal{A}^n(\mathbb{R}^n)$ that is appropriately normalized so that

$$\varphi(I_n) = \varphi(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1.$$

I claim that

there is at most one determinant function.

In order to streamline the proof I will isolate several lemmas, which have independent interest.

Lemma A. Let φ be a determinant function. If A has a repeated column then

$$\varphi(A) = 0.$$

Proof. Suppose that the ith and jth columns are equal and let A' be the matrix obtained from A by switching the ith and jth columns. On the one hand we have $A' = A$. On the other hand, property (2) tells us that

$$\begin{aligned} \varphi(A') &= -\varphi(A) \\ \varphi(A) &= -\varphi(A) \\ 2\varphi(A) &= 0 \\ \varphi(A) &= 0. \end{aligned}$$

□

Lemma B. Let φ be a determinant function. If A has dependent columns then

$$\varphi(A) = 0.$$

Proof. Let A have columns $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$. If these columns are dependent then there exists some i such that $\mathbf{a}_i$ can be expressed as a linear combination of the other columns. Without loss of generality,[21] suppose that $i = 1$, so we can write

$$\mathbf{a}_1 = b_1\mathbf{a}_2 + \cdots + b_n\mathbf{a}_n,$$

for some scalars $b_2, \ldots, b_n$. Now let $\hat{A}_1(\mathbf{a}_j)$ denote the matrix A with the first column replaced by $\mathbf{a}_j$. From property (1) we have

$$\varphi(A) = b_1 \cdot \varphi(\hat{A}_1(\mathbf{a}_2)) + \cdots + b_n \cdot \varphi(\hat{A}_1(\mathbf{a}_n)).$$

[21] By applying property (2) we can swap the 1st and ith columns, which does not affect whether the determinant is zero or nonzero.

But each matrix $\hat{A}_1(\mathbf{a}_j)$ with $j \neq 1$ has a repeated column, so from Lemma A we must have

$$\begin{aligned}\varphi(A) &= b_1 \cdot \varphi(\hat{A}_1(\mathbf{a}_2)) + \cdots + b_n \cdot \varphi(\hat{A}_1(\mathbf{a}_n)) \\ &= b_1 \cdot 0 + b_2 \cdot 0 + \cdots + b_n \cdot 0 \\ &= 0.\end{aligned}$$

□

The next lemma refers to the elementary matrices, which we discussed in the previous chapter:

$$D_i(\lambda) = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \lambda & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix},$$

$$L_{ij}(\lambda) = \begin{pmatrix} 1 & & & & \\ & 1 & \cdots & \lambda & \\ & & 1 & \vdots & \\ & & & 1 & \\ & & & & 1 \end{pmatrix},$$

$$T_{ij} = \begin{pmatrix} 1 & & & & \\ & 0 & \cdots & 1 & \\ & \vdots & 1 & \vdots & \\ & 1 & \cdots & 0 & \\ & & & & 1 \end{pmatrix}.$$

Lemma C. Let φ be a determinant function. Then for any square matrix A we have

$$\begin{aligned}\varphi(AD_i(\lambda)) &= \lambda \cdot \varphi(A), \\ \varphi(AL_{ij}(\lambda)) &= \varphi(A), \\ \varphi(AT_{ij}) &= -\varphi(A).\end{aligned}$$

Proof. First, note that $AD_i(\lambda)$ has the same columns as A except that the ith column has been scaled by λ, hence $\varphi(AD_i(\lambda)) = \lambda \cdot \varphi(A)$ follows from property (1). Next, note that AT_{ij} is obtained from A by switching columns i and j, hence the identity $\varphi(AT_{ij}) = -\varphi(A)$ is just a restatement of (2). Finally, note that kth column of $AL_{ij}(\lambda)$ is equal to the kth column of A, except in the case $k = j$, in which case

$$(j\text{th column of } AL_{ij}(\lambda)) = (j\text{th column of } A) + \lambda \cdot (i\text{th column of } A).$$

To simplify notation, let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be the columns of A and let $\hat{A}_j(\mathbf{v})$ denote the matrix A with the jth column replaced by vector $\mathbf{v}$. Then from property (1) we have

$$\varphi(AL_{ij}(\lambda)) = \varphi(A) + \lambda \cdot \varphi(\hat{A}_j(\mathbf{a}_i)).$$

But the matrix $\hat{A}_j(\mathbf{a}_i)$ has a repeated column, so it follows from Lemma A that

$$\varphi(AL_{ij}(\lambda)) = \varphi(A) + \lambda \cdot 0 = \varphi(A).$$

$\square$

Lemma D. Let φ be a determinant function. Then we have

$$\varphi(D_i(\lambda)) = \lambda, \quad \varphi(L_{ij}(\lambda)) = 1, \quad \varphi(T_{ij}) = -1.$$

Proof. Taking $A = I$ in Lemma C and using property (3) gives[22]

$$\begin{aligned} \varphi(D_i(\lambda)) &= \varphi(ID_i(\lambda)) = \lambda \cdot \varphi(I) = \lambda, \\ \varphi(L_{ij}(\lambda)) &= \varphi(IL_{ij}(\lambda)) = \varphi(I) = 1, \\ \varphi(T_{ij}) &= \varphi(IT_{ij}) = -\varphi(I) = -1. \end{aligned}$$

$\square$

Lemma E. Let φ be a determinant function. For elementary matrices $E_1, \ldots, E_k$ we have

$$\varphi(E_1 E_2 \cdots E_k) = \varphi(E_1)\varphi(E_2) \cdots \varphi(E_k).$$

Proof. By applying Lemma D, we can rephrase Lemma C as saying that

$$\varphi(AE) = \varphi(A)\varphi(E) \text{ for any elementary matrix } E.$$

If $E_1, \ldots, E_k$ are elementary matrices, then it follows by induction that

$$\begin{aligned} \varphi(E_1 \cdots E_k) &= \varphi(E_1 \cdots E_{k-1})\varphi(E_k) \\ &= \varphi(E_1) \cdots \varphi(E_{k-1})\varphi(E_k). \end{aligned}$$

$\square$

Proof of the Theorem. Let δ_1 and δ_2 be any two determinant functions. Our goal is to show that $\delta_1 = \delta_2$. If A is not invertible then the columns of A are dependent and it follows from Lemma B that $\delta_1(A) = 0 = \delta_2(A)$. So let us suppose that A is invertible. In this case we can apply column operations to reduce A to the identity matrix:

$$AE_1 E_2 \cdots E_k = I.$$

Since elementary matrices are invertible, this becomes

$$A = E_k^{-1} \cdots E_1^{-1}.$$

[22] This is our first and only use of property (3).

If E is elementary then E^{-1} is also elementary, so Lemma D implies that $\delta_1(E^{-1}) = \delta_2(E^{-1})$. Finally, by Lemma E we have

$$\begin{aligned}
\delta_1(A) &= \delta_1(E_k^{-1} \cdots E_1^{-1}) \\
&= \delta_1(E_k^{-1}) \cdots \delta_1(E_1^{-1}) \\
&= \delta_2(E_k^{-1}) \cdots \delta_2(E_1^{-1}) \\
&= \delta_2(E_k^{-1} \cdots E_1^{-1}) \\
&= \delta_2(A).
\end{aligned}$$

□

Thus we have proved that there exists at most one determinant function. From this point on, we will use the notation $\det(A)$ to refer to this function.

We end this section by giving a new criterion for invertibility of square matrices.

Theorem. For any square matrix A we have

$$A \text{ is invertible} \quad \Longleftrightarrow \quad \det(A) \neq 0.$$

Proof. If A is not invertible then A has dependent columns and it follows from Lemma B that $\det(A) = 0$. Conversely, suppose that A is invertible. In the previous chapter we showed that a square matrix is invertible if and only if its Reduced Row Echelon Form is an identity matrix, so that

$$E_k \cdots E_2 E_1 A = I$$

for some elementary matrices $E_1, \ldots, E_k$. From Lemma E it follows that

$$\begin{aligned}
A &= E_1^{-1} E_2^{-1} \cdots E_k^{-1} \\
\det(A) &= \det(E_1^{-1}) \det(E_2^{-1}) \cdots \det(E_k^{-1}) \neq 0.
\end{aligned}$$

□

Note that we only use elementary matrices $D_i(\lambda)$ with $\lambda \neq 0$ so that $\det(E) \neq 0$ for every elementary matrix E.

2.3 Algebraic Properties of the Determinant

In the previous section we studied the application of determinant functions to elementary matrices, and we used this to prove that there exists at most one determinant function. In this section we will apply the same lemmas to prove some interesting algebraic properties of determinants. Only in the next section will we finally prove that determinants exist!

Theorem. For any square matrices A and B we have

(a) $\det(A^T) = \det(A)$,

(b) $\det(AB) = \det(A)\det(B)$,

(c) $\det(A^{-1}) = 1/\det(A)$.

Proof. (a): Note that A^T is invertible if and only if A is invertible, hence $\det(A^T) = 0$ if and only if $\det(A) = 0$. If $\det(A) \neq 0$ then A is invertible and we can write

$$A = E_1 \cdots E_k$$

for some elementary matrices $E_1, \ldots, E_k$. Note that the transpose E^T of an elementary matrix E is also elementary, and from Lemma C we have $\det(E^T) = \det(E)$. It follows that

$$\begin{aligned}
A^T &= E_k^T \cdots E_1^T \\
\det(A^T) &= \det(E_k^T \cdots E_1^T) \\
&= \det(E_k^T) \cdots \det(E_1^T) \\
&= \det(E_k) \cdots \det(E_1) \\
&= \det(E_1) \cdots \det(E_k) \\
&= \det(E_1 \cdots E_k) \\
&= \det(A).
\end{aligned}$$

(b): Note that AB is invertible if and only if both of A and B are invertible, so that $\det(AB) = 0$ if and only if $\det(A)\det(B) = 0$. If $\det(A) \neq 0$ and $\det(B) \neq 0$ then A and B are both invertible, hence we can write

$$\begin{aligned}
A &= E_1 \cdots E_k, \\
B &= F_1 \cdots F_\ell,
\end{aligned}$$

for some elementary matrices $E_1, \ldots, E_k$ and $F_1, \ldots, F_\ell$. It follows that

$$\begin{aligned}
\det(AB) &= \det(E_1 \cdots E_k F_1 \cdots F_\ell) \\
&= \det(E_1) \cdots \det(E_k) \det(F_1) \cdots \det(F_\ell) \\
&= [\det(E_1) \cdots \det(E_k)][\det(F_1) \cdots \det(F_\ell)] \\
&= \det(E_1 \cdots E_k) \det(F_1 \cdots F_\ell) \\
&= \det(A)\det(B).
\end{aligned}$$

(c): If A is invertible then $\det(A) \neq 0$ and from (b) we obtain

$$\begin{aligned}
A^{-1}A &= I \\
\det(A^{-1}A) &= \det(I) \\
\det(A^{-1})\det(A) &= 1 \\
\det(A^{-1}) &= 1/\det(A).
\end{aligned}$$

□

As you see, the elementary matrices are quite useful.

2.4 Formulas for the Determinant

I hope you have developed an appreciation for the remarkable properties of determinants. In this section I will prove that determinants actually exist, and in the next section I will finally tell you what determinants "really are". I guess I could have told you that first, but it didn't fit the narrative.

There are several equivalent ways to define the determinant of an $n \times n$ matrix. If A is not invertible then we must have $\det(A) = 0$, so let us suppose that A is invertible. In this case we can perform row (or column) operations to transform A into the identity matrix, which allows us to write A as a product of elementary matrices:

$$A = E_1 \cdots E_k.$$

Then from Lemma E in Section 2.2 we must have

$$\det(A) = \det(E_1) \cdots \det(E_k),$$

where the determinants of elementary matrices are trivial to compute. You might think we could use this formula to **define** the determinant, but the factorization of A into elementary matrices is not unique, and it's not clear that we wouldn't get different values of $\det(A)$ from different factorizations of A. Essentially this has to do with the uniqueness of the RREF, but I don't want to prove this. Instead I'll just give an example computation.

Computing the Determinant by Elimination. Consider again the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix}.$$

First we perform down-elimination steps to put A in upper triangular form:

$$L_{31}(-2)L_{21}(-1)A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -5 \end{pmatrix}. \tag{$*$}$$

Then we scale the rows to turn the pivots into ones:

$$D_3(-1/5)D_2(-1)L_{31}(-2)L_{21}(-1)A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then we perform up-elimination to obtain an identity matrix:

$$L_{12}(-2)L_{13}(-3)L_{23}(-1)D_3(-1/5)D_2(-1)L_{31}(-2)L_{21}(-1)A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Taking the elementary matrices to the other side gives

$$\begin{aligned} A &= L_{21}(-1)^{-1}L_{31}(-2)^{-1}D_2(-1)^{-1}D_3(-1/5)^{-1}L_{23}(-1)^{-1}L_{13}(-3)^{-1}L_{12}(-2)^{-1} \\ &= L_{21}(1)L_{31}(2)D_2(-1)D_3(-5)L_{23}(1)L_{13}(3)L_{12}(2), \end{aligned}$$

and taking the determinant of each side gives

$$\begin{aligned} \det(A) &= 1 \cdot 1 \cdot (-1) \cdot (-5) \cdot 1 \cdot 1 \cdot 1 \\ &= 5. \end{aligned}$$

Note that this is the product of the pivot entries in step $(*)$. Hence we could have stopped there. In general, if no row transpositions are required, then the determinant is just the product of the diagonal entries after down-elimination.

Next I will give the traditional definition of the determinant, which expresses it as an "alternating sum" over permutations. After that I will give a recursive formula, which is more useful.

Permutation Definition of the Determinant. Let S_n denote the set of *permutations*, i.e., the set of bijective functions $\{1, \ldots, n\} \to \{1, \ldots, n\}$. It is convenient to express a permutation by listing the sequence of values:

$$\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n)).$$

Each permutation $\sigma \in S_n$ has a well-defined *sign*, or *parity*:

$$\operatorname{sgn}(\sigma) \in \{1, -1\}.$$

Essentially this tells us the number of swaps necessary to obtain the list $(\sigma(1), \ldots, \sigma(n))$ from the list $(1, \ldots, n)$, or vice versa. The number of swaps is not unique, but it turns out that it is always even, or always odd. For example, we can get from $(1,2,3)$ to $(3,2,1)$ using 3 swaps:

$$(1,2,3) \to (2,1,3) \to (2,3,1) \to (3,2,1),$$

Or we can get there using 5 swaps:

$$(1,2,3) \to (1,3,2) \to (3,1,2) \to (2,1,3) \to (2,3,1) \to (3,2,1).$$

But we could never get there using an even number of swaps.[23] Since we can get from $(1,2,3)$ to $(3,2,1)$ using only odd numbers of swaps, we define

$$\operatorname{sgn}(3,2,1) = -1.$$

[23]It is a bit tricky to prove this so we won't bother. It fits better in a course on "group theory".

Of the $n!$ permutations in S_n, it turns out that exactly half are "even" and half are "odd". For example, here is the sign table for S_3:

σ	$\mathrm{sgn}(\sigma)$
$(1,2,3)$	$+1$
$(2,3,1)$	$+1$
$(3,1,2)$	$+1$
$(1,3,2)$	-1
$(2,1,3)$	-1
$(3,2,1)$	-1

Finally, recall the standard basis of k-forms:

$$\varepsilon_{i_1} \otimes \varepsilon_{i_2} \otimes \cdots \otimes \varepsilon_{i_k} \text{ for all } i_1, \ldots, i_k \in \{1, \ldots, n\}.$$

Then we define the determinant function $\det \in \mathcal{A}^n(\mathbb{R}^n)$ as follows:

$$\boxed{\det = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \cdot \varepsilon_{\sigma(1)} \otimes \varepsilon_{\sigma(2)} \otimes \cdots \otimes \varepsilon_{\sigma(n)}.}$$

For example, when $n = 3$, the above table of signs gives

$$\begin{aligned} \det = {} & \varepsilon_1 \otimes \varepsilon_2 \otimes \varepsilon_3 + \varepsilon_2 \otimes \varepsilon_3 \otimes \varepsilon_1 + \varepsilon_3 \otimes \varepsilon_1 \otimes \varepsilon_2 \\ & - \varepsilon_1 \otimes \varepsilon_3 \otimes \varepsilon_2 - \varepsilon_2 \otimes \varepsilon_1 \otimes \varepsilon_3 - \varepsilon_3 \otimes \varepsilon_2 \otimes \varepsilon_1. \end{aligned}$$

Equivalently, if A is an $n \times n$ matrix with ij entry a_{ij} then we define

$$\boxed{\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \cdot a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}.}$$

One can check that this function satisfies the three properties of a determinant function, but to do so requires a more thorough study of permutations than we have time for.

Laplace Expansion. The permutation definition of the determinant is explicit but it's mostly useless. Another, recursive, definition called *Laplace expansion* or *expansion by cofactors* has many applications.

For any $n \times n$ matrix A we let $\hat{A}_{ij}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and the jth column. To **expand along the ith row**, we fix some i and then compute

$$\boxed{\det(A) = \sum_j (-1)^{i+j} a_{ij} \det(\hat{A}_{ij}).}$$

To **expand along the jth column** we fix some j and compute

$$\boxed{\det(A) = \sum_i (-1)^{i+j} a_{ij} \det(\hat{A}_{ij}).}$$

One must check that these formulas agree with the permutation definition of the determinant. Alternatively, one could prove that these formulas obey the three rules for determinant functions. But I'm not going to do this. Instead I will just give some examples.

First we compute a general 3×3 determinant by expanding along the second row:

$$\begin{aligned}\det\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} &= -a_2 \cdot \det\begin{pmatrix} b_1 & c_1 \\ b_3 & c_3 \end{pmatrix} + b_2 \cdot \det\begin{pmatrix} a_1 & c_1 \\ a_3 & c_3 \end{pmatrix} - c_2 \cdot \det\begin{pmatrix} a_1 & b_1 \\ a_3 & b_3 \end{pmatrix} \\ &= -a_2(b_1c_3 - b_3c_1) + b_2(a_1c_3 - a_3c_1) - c_2(a_1b_3 - a_3b_1) \\ &= a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1.\end{aligned}$$

Next we expand a specific our favorite matrix along the second column:

$$\begin{aligned}\det\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix} &= -2 \cdot \det\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} + 1 \cdot \det\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} - 4 \cdot \det\begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix} \\ &= -2(1-2) + 1(1-6) - 4(1-3) \\ &= -2(-1) + 1(-5) - 4(-2) \\ &= 2 - 5 + 8 \\ &= 5.\end{aligned}$$

And also along the first row:

$$\begin{aligned}\det\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix} &= 1 \cdot \det\begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} - 2 \cdot \det\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} + 3 \cdot \det\begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix} \\ &= 1(1-4) - 2(1-2) + 3(4-2) \\ &= 1(-3) - 2(-1) + 3(2) \\ &= -3 + 2 + 6 \\ &= 5.\end{aligned}$$

2.5 Cramer's Rule (Optional)

While we're on the subject, there is a famous trick relating determinants to solutions of linear systems. Let A be a square $n \times n$ matrix and consider the linear system

$$A\mathbf{x} = \mathbf{b}$$
$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Assume that A is invertible, so the system has a unique solution $\mathbf{x} = (x_1, \ldots, x_n)$. Then the ith coordinate of the solution is given by

$$\boxed{x_i = \frac{\det(\hat{A}_i(\mathbf{b}))}{\det(A)},}$$

where $\hat{A}_i(\mathbf{b})$ is the matrix obtained from A by replacing its ith column with $\mathbf{b}$:

$$\hat{A}_i(\mathbf{b}) = \left(\begin{array}{c|c|c|c|c|c|c} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{array} \right).$$

Proof. Consider the matrix

$$X_i := \hat{I}_i(\mathbf{x}) = \left(\begin{array}{c|c|c|c|c|c|c} \mathbf{e}_1 & \cdots & \mathbf{e}_{i-1} & \mathbf{x} & \mathbf{e}_{i+1} & \cdots & \mathbf{e}_n \end{array} \right) = \begin{pmatrix} 1 & & x_1 & & \\ & 1 & \vdots & & \\ & & x_i & & \\ & & \vdots & 1 & \\ & & x_n & & 1 \end{pmatrix}.$$

By Laplace expansion along the ith column, we observe that[24]

$$\det(X_i) = (-1)^{i+i} x_i \det(I_{n-1}) = x_i.$$

Next we observe that

$$\begin{aligned} AX_i &= A \left(\begin{array}{c|c|c|c|c|c|c} \mathbf{e}_1 & \cdots & \mathbf{e}_{i-1} & \mathbf{x} & \mathbf{e}_{i+1} & \cdots & \mathbf{e}_n \end{array} \right) \\ &= \left(\begin{array}{c|c|c|c|c|c|c} A\mathbf{e}_1 & \cdots & A\mathbf{e}_{i-1} & A\mathbf{x} & A\mathbf{e}_{i+1} & \cdots & A\mathbf{e}_n \end{array} \right) \\ &= \left(\begin{array}{c|c|c|c|c|c|c} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{array} \right) \\ &= \hat{A}_i(\mathbf{b}), \end{aligned}$$

and hence

$$\begin{aligned} AX_i &= \hat{A}_i(\mathbf{b}) \\ \det(A)\det(X_i) &= \det(\hat{A}_i(\mathbf{b})) \\ \det(X_i) &= \det(\hat{A}_i(\mathbf{b}))/\det(A) \\ x_i &= \det(\hat{A}_i(\mathbf{b}))/\det(A). \end{aligned}$$

□

For example, let A be the 3×3 matrix from the previous section and consider the linear system

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

[24] The matrix obtained by deleting the ith row and column of X_i is the $(n-1) \times (n-1)$ identity matrix I_{n-1}. Every other $(n-1) \times (n-1)$ matrix in the expansion has a row (also a column) of zeros, hence its determinant is zero.

Then we have

$$x_1 = \det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 4 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix} = \frac{-3}{5},$$

$$x_2 = \det\begin{pmatrix} 1 & 1 & 3 \\ 1 & 0 & 1 \\ 2 & 0 & 1 \end{pmatrix} \Big/ \det\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix} = \frac{1}{5},$$

$$x_3 = \det\begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 0 \\ 2 & 4 & 0 \end{pmatrix} \Big/ \det\begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix} = \frac{2}{5}.$$

Cramer's Rule is useful when we want to pick out a specific coordinate of the solution. We can use this idea to give an explicit formula for the entries of an inverse matrix. Let A be an invertible $n \times n$ square matrix and let $X = \left(\mathbf{x}_1 \mid \cdots \mid \mathbf{x}_n \right)$ be a matrix whose columns $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^n$ are unknown vectors. If X is the inverse of A then we must have

$$\begin{aligned} AX &= I \\ A\left(\mathbf{x}_1 \mid \cdots \mid \mathbf{x}_n \right) &= \left(\mathbf{e}_1 \mid \cdots \mid \mathbf{e}_n \right) \\ \left(A\mathbf{x}_1 \mid \cdots \mid A\mathbf{x}_n \right) &= \left(\mathbf{e}_1 \mid \cdots \mid \mathbf{e}_n \right), \end{aligned}$$

which is equivalent to n matrix equations: $A\mathbf{x}_i = \mathbf{e}_i$ for each i. Let x_{ij} be the ij entry of the unknown matrix X, which is the ith entry of the jth column vector $\mathbf{x}_j$. Then Cramer's Rule says that

$$\begin{aligned} x_{ij} &= i\text{th coordinate of } \mathbf{x}_j \\ &= i\text{th coordinate of the solution to } A\mathbf{x}_j = \mathbf{e}_j \\ &= \det(\hat{A}_i(\mathbf{e}_j))/\det(A), \end{aligned}$$

where $\hat{A}_i(\mathbf{e}_j)$ is the matrix obtained from A by replacing its ith column with $\mathbf{e}_j$. By Laplace expansion along the ith column we have

$$\det(\hat{A}_i(\mathbf{e}_j)) = (-1)^{i+j}\det(\hat{A}_{ji}),$$

where $\hat{A}_{ji}$ is the $(n-1) \times (n-1)$ matrix obtained from A by deleting the jth row and ith column. If $\det(A) \neq 0$ then we conclude that

$$\boxed{(ij \text{ entry of } A^{-1}) = \frac{1}{\det(A)}(-1)^{i+j}\det(\hat{A}_{ji}).}$$

Warning: Note that the positions of i and j are switched in this formula![25]

[25] I have forgotten this many times.

For example, suppose that

$$AX = I,$$
$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then we have

$$\begin{aligned} x_{11} &= (-1)^{1+1} \det(\hat{A}_{11})/\det(A) = a_{22}/\det(A), \\ x_{12} &= (-1)^{1+2} \det(\hat{A}_{21})/\det(A) = -a_{12}/\det(A), \\ x_{21} &= (-1)^{2+1} \det(\hat{A}_{12})/\det(A) = -a_{21}/\det(A), \\ x_{22} &= (-1)^{2+2} \det(\hat{A}_{22})/\det(A) = a_{11}/\det(A), \end{aligned}$$

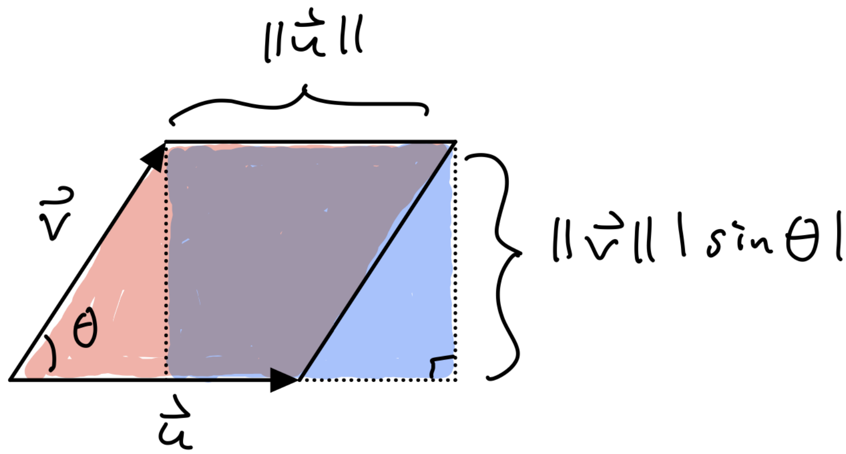
which is just the usual formula for the inverse of a 2×2 matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

2.6 Geometric Interpretation

My bias is that algebra is based on geometry,[26] hence for me the "true meaning" of the determinant is its geometric interpretation.

Consider two vectors in the plane, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ with angle θ between them. The area of the parallelogram they generate is $\|\mathbf{u}\|\|\mathbf{v}\||\sin\theta|$.[27] Indeed, in the following picture the red parallelogram and the blue rectangle have the same area:



[26] And geometry is based on physics. I believe that physics is the true foundation of mathematics, not axiomatic set theory.

[27] The absolute value accounts for negative angles.

On the other hand, we can interpret this area as a determinant. Let A be the 2×2 matrix with columns $\mathbf{u}$ and $\mathbf{v}$:

$$A = \left(\begin{array}{c|c} \mathbf{u} & \mathbf{v} \end{array} \right).$$

I claim that the area of the parallelogram equals (the absolute value of) the determinant of A. To prove this we use a clever trick. First we observe that

$$\sqrt{\det(A^T A)} = \sqrt{\det(A^T)\det(A)} = \sqrt{\det(A)\det(A)} = \sqrt{\det(A)^2} = |\det(A)|.$$

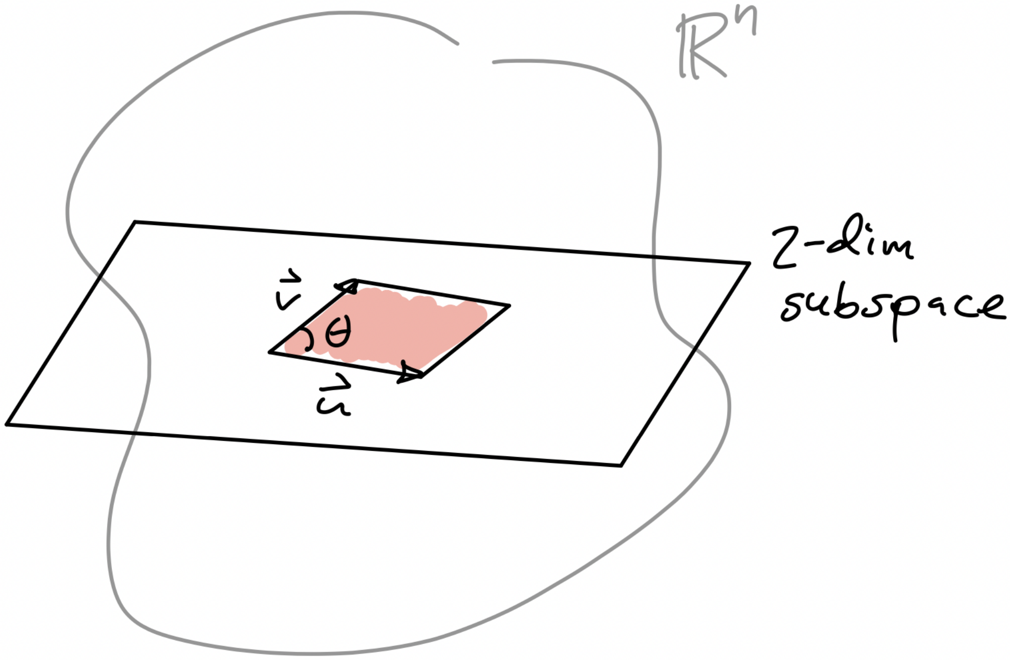
But the determinant of $A^T A$ can also be computed as follows:

$$\begin{aligned} A^T A &= \begin{pmatrix} \mathbf{u}^T \\ \hline \mathbf{v}^T \end{pmatrix} \left(\begin{array}{c|c} \mathbf{u} & \mathbf{v} \end{array} \right) \\ A^T A &= \begin{pmatrix} \|\mathbf{u}\|^2 & \mathbf{u} \bullet \mathbf{v} \\ \mathbf{u} \bullet \mathbf{v} & \|\mathbf{v}\|^2 \end{pmatrix} \\ \det(A^T A) &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \bullet \mathbf{v})^2 \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\|\mathbf{u}\| \|\mathbf{v}\| \cos\theta)^2 \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 (1 - \cos^2\theta) \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2\theta. \end{aligned}$$

So we conclude that

$$|\det(A)| = \sqrt{\det(A^T A)} = \sqrt{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2\theta} = \|\mathbf{u}\| \|\mathbf{v}\| |\sin\theta|.$$

This trick is much more important than it looks. Suppose now that our parallelogram lives in n-dimensional space, generated by vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ with angle θ:

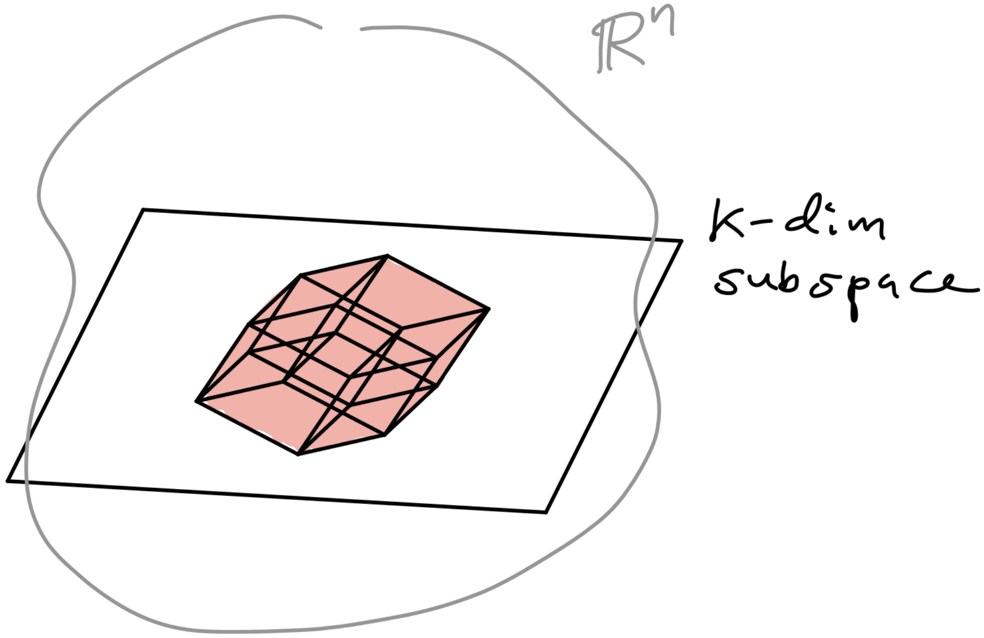


For geometric reasons, the area of the parallelogram is still $\|\mathbf{u}\|\|\mathbf{v}\||\sin\theta|$, but now the $n \times 2$ matrix $A = \left(\mathbf{u} \mid \mathbf{v} \right)$ is **not square**, so $\det(A)$ is not defined. However, the matrix $A^T A$ is still square, so we may still consider $\det(A^T A)$, and the same calculation as above shows that

$$\sqrt{\det(A^T A)} = \|\mathbf{u}\|\|\mathbf{v}\||\sin\theta|.$$

In general, we have the following theorem.

Theorem (Geometric Interpretation of the Determinant). Let A be an $n \times k$ matrix with columns $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$, which generate a k-*parallelotope* living in n-dimensional space:



Let $\mathrm{Vol}_k(A)$ denote the volume of this k-parallelotope, measured within the k-dimensional subspace that it spans. We call this the *k-volume* of the k-parallelotope. Then we have[28]

$$\boxed{\mathrm{Vol}_k(A) = \sqrt{\det(A^T A)}.}$$

If $k = n$, then we are measuring the full n-dimensional volume of an n-parallelotope in $\mathbb{R}^n$. In this case the matrix A is square, and we obtain

$$\boxed{\mathrm{Vol}_n(A) = |\det(A)|.}$$

Note that we already proved this theorem in the case $k = 2$. The proof of the general case proceeds in four steps:

(1) For $n \times n$ matrices A we have $\mathrm{Vol}_n(A) = |\det(A)|$.

(2) For $n \times n$ matrices A we have $|\det(A)| = \sqrt{\det(A^T A)}$.

(3) It follows from (1) and (2) that the n-volume of an n-parallelotope in $\mathbb{R}^n$ depends only on the lengths and angles between its generating vectors.

(4) Hence we also have $\mathrm{Vol}_k(A) = \sqrt{\det(A^T A)}$, even when $k \neq n$.

[28]This volume can very well be zero, which happens when the columns of A are not independent. In this case, the k-parallelotope generated by $\mathbf{a}_1, \ldots, \mathbf{a}_k$ is "flat", i.e., it lives in a smaller-dimensional subspace of $\mathbb{R}^n$. For example, a 3-parallelogram generated by dependent vectors is actually some kind of 2-dimensional hexagon. I guess there is a recursive formula for the lower-dimensional volume but I don't want to work it out.

The hardest part is (1), which we will prove below. The proof of (2) is a simple calculation, which was given above. For the proof of (3) let A be $n \times n$. We observe that the ij entry of the $n \times n$ matrix $A^T A$ is

$$\mathbf{a}_i^T \mathbf{a}_j = \mathbf{a}_i \bullet \mathbf{a}_j = \|\mathbf{a}_i\| \|\mathbf{a}_j\| \cos\theta_{ij},$$

where θ_{ij} is the angle between $\mathbf{a}_i$ and $\mathbf{a}_j$. Since from (1) we have

$$\mathrm{Vol}_n(A) = |\det(A)| = \sqrt{\det(A^T A)},$$

and since the entries of $A^T A$ only depend on the lengths $\|\mathbf{a}_i\|$ and angles θ_{ij}, it follows that the volume $\mathrm{Vol}_n(A)$ only depends on the lengths and angles. But now suppose that A is $k \times n$ with columns $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$. In this case the ij entry of $A^T A$ is still given by

$$\mathbf{a}_i^T \mathbf{a}_j = \mathbf{a}_i \bullet \mathbf{a}_j = \|\mathbf{a}_i\| \|\mathbf{a}_j\| \cos\theta_{ij},$$

hence $\det(A^T A)$ has exactly the same formula in terms of $\|\mathbf{a}_i\|$ and θ_{ij} as it does when A is a $k \times k$ square matrix. Then from the square case we conclude that

$$\mathrm{Vol}_k(A) = \text{some formula involving the lengths } \|\mathbf{a}_i\| \text{ and angles } \theta_{ij} = \sqrt{\det(A^T A)}.$$

This completes the proof, except for part (1). $\square$

Before diving into the proof of (1), we consider the case $k = 3$. The technical name for a 3-parallelogram is a *parallelepiped*.

Volume of a Parallelepiped. Let A be an $n \times 3$ matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \in \mathbb{R}^n$, and let θ_{ij} be the angle between vectors $\mathbf{a}_i$ and $\mathbf{a}_j$, which can be computed via the dot product:

$$\theta_{ij} = \arccos\left(\frac{\mathbf{a}_i \bullet \mathbf{a}_j}{\|\mathbf{a}_i\| \|\mathbf{a}_j\|}\right).$$

Then the volume (i.e., the 3-volume) of the parallelepiped generated by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ is given by

$$\begin{aligned}
\mathrm{Vol}_3(A)^2 &= \det(A^T A) \\
&= \det\begin{pmatrix} \|\mathbf{a}_1\|^2 & \mathbf{a}_1 \bullet \mathbf{a}_2 & \mathbf{a}_1 \bullet \mathbf{a}_3 \\ \mathbf{a}_1 \bullet \mathbf{a}_2 & \|\mathbf{a}_2\|^2 & \mathbf{a}_2 \bullet \mathbf{a}_3 \\ \mathbf{a}_1 \bullet \mathbf{a}_3 & \mathbf{a}_2 \bullet \mathbf{a}_3 & \|\mathbf{a}_3\|^2 \end{pmatrix} \\
&= \det\begin{pmatrix} \|\mathbf{a}_1\|^2 & \|\mathbf{a}_1\|\|\mathbf{a}_2\|\cos\theta_{12} & \|\mathbf{a}_1\|\|\mathbf{a}_3\|\cos\theta_{13} \\ \|\mathbf{a}_1\|\|\mathbf{a}_2\|\cos\theta_{12} & \|\mathbf{a}_2\|^2 & \|\mathbf{a}_2\|\|\mathbf{a}_3\|\cos\theta_{23} \\ \|\mathbf{a}_1\|\|\mathbf{a}_3\|\cos\theta_{13} & \|\mathbf{a}_2\|\|\mathbf{a}_3\|\cos\theta_{23} & \|\mathbf{a}_3\|^2 \end{pmatrix},
\end{aligned}$$

which after some simplification becomes

$$\mathrm{Vol}_3(A) = \|\mathbf{a}_1\|\|\mathbf{a}_2\|\|\mathbf{a}_3\|\sqrt{(1 + 2\cos\theta_{12}\cos\theta_{13}\cos\theta_{23} - (\cos^2\theta_{12} + \cos^2\theta_{13} + \cos^2\theta_{23}))}.$$

This formula is much more difficult to derive without determinants.[29]

Proof of (1). For an $n \times n$ matrix A we need to prove that

$$|\det(A)| = \mathrm{Vol}_n(A).$$

Actually, we will prove that

$$\det(A) = \pm\mathrm{Vol}_n(A),$$

where the sign depends on the ordering of the columns, and is not relevant to the geometry. Thus we will show that the determinant can be interpreted as a "signed volume".[30] According to Section 2.2, we only need to show that the function Vol_n from $(\mathbb{R}^n)^n$ to $\mathbb{R}$ satisfies the three rules of a determinant function:

- *Multilinear.* The function $\mathrm{Vol}_n(A)$ is linear in each individual column of A.
- *Alternating.* If A' is obtained from A by switching two columns, then

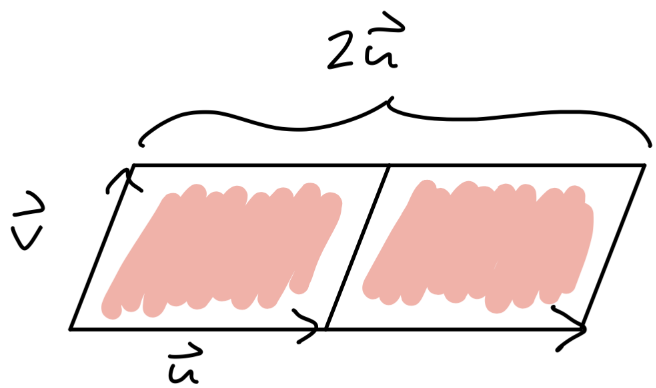
$$\mathrm{Vol}_n(A') = -\mathrm{Vol}_n(A).$$

- *Normalized.* We have $\mathrm{Vol}_n(I_n) = 1$.

The third property is part of the **definition** of volume. It just says that the unit n-cube has n-volume 1. And we can just assume that the second property is true, since we don't care about the sign of the volume. Thus we only need to show that Vol_n is multilinear.

There is a subtle difficulty here, since to prove a theorem about volume, one must have a formal definition of volume, which we don't. In fact, the most common formal definition of volume is based the determinant! But any proof using this formalization would be circular.

Instead of developing a rigorous "measure theory",[31] we will proceed intuitively. It is intuitively obvious that scaling one of the columns scales the volume by the same amount. For example, doubling one side of a parallelogram doubles the area:

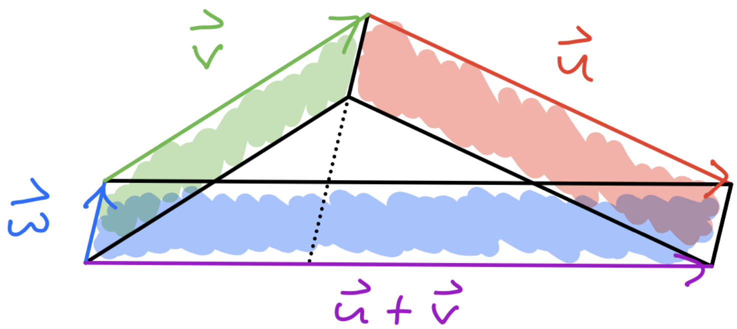


[29] If $n = 3$ then we can also express the volume in terms of the cross product, but doing so breaks the symmetry, and the cross product doesn't generalize to higher dimensions.

[30] This should be familiar from Calculus, since the area under a curve is actually a "signed area", with regions below the x-axis having "negative area". See the next section.

[31] Measure theory is the term for the modern, rigorous, theory of integration.

Thus we only need to show that Vol_n preserves addition in each column. In the case of parallelograms, we need to show that the areas of the parallelograms generated by $\mathbf{u}, \mathbf{w}$ and $\mathbf{v}, \mathbf{w}$ add to the area of the parallelogram generated by $\mathbf{u} + \mathbf{v}$ and $\mathbf{w}$. For example, in the following picture we need to show that the areas of the red and green parallelograms add to the area of the blue parallelogram:



The proof uses the dotted line, which is parallel to $\mathbf{w}$. This line divides the blue parallelogram into two pieces, which have the same areas as the red and green parallelograms. This follows because parallelograms with the same base and height have the same area.

In higher dimensions the scaling argument is still plausible but the addition argument is harder to visualize. Instead of trying to generalize the above picture, we will base our argument on a general geometric principle called Cavalieri's Principle, which we take as an axiom.[32]

Cavalieri's Principle. An *n-prism* in $\mathbb{R}^n$ has the following form. Let $V \subseteq \mathbb{R}^n$ be an $(n-1)$-dimensional subspace. For any subset $S \subseteq V$ and for any vector $\mathbf{a} \in \mathbb{R}^n$ that is **not** in V, we define the "prism over S generated by $\mathbf{a}$":

$$\mathrm{Prism}_S(\mathbf{a}) = \{\mathbf{p} + t\mathbf{a} : \mathbf{p} \in S \text{ and } 0 \leq t \leq 1\}.$$

Then Cavalieri's principle says that

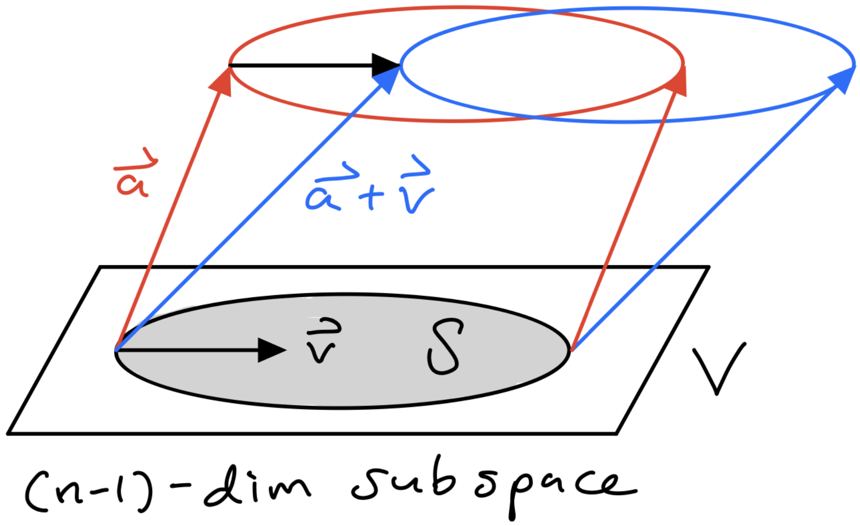
$$\mathrm{Vol}_n(\mathrm{Prism}_S(\mathbf{a})) = \mathrm{Vol}_n(\mathrm{Prism}_S(\mathbf{a} + \mathbf{v})) \quad \text{for any vector } \mathbf{v} \in V.$$

More colloquially:

two prisms with the same base and the same height have the same volume.

Here is a picture:

[32] This principle is often taken as an axiom, for example when deriving the volume of a sphere in $\mathbb{R}^3$ without calculus.



For any $n \times n$ matrix A we will show that applying an elementary matrix of the form $L_{ij}(\lambda)$ to A does not change the volume of the n-parallelotope:

$$\mathrm{Vol}_n(AL_{ij}(\lambda)) = \mathrm{Vol}_n(A).$$

To be precise, let A have columns $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ and let S_j be the $(n-1)$-parallelogram living in the $(n-1)$-dimensional subspace $V \subseteq \mathbb{R}^n$ generated by the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, except for $\mathbf{a}_j$. We can view the n-parallelotope generated by A as $\mathrm{Prism}_{S_j}(\mathbf{a}_j)$. If A' is obtained from A by replacing column $\mathbf{a}_j$ by itself plus any vector $\mathbf{v} \in V$, then Cavalieri's principle says

$$\mathrm{Vol}_n(A') = \mathrm{Vol}_n(\mathrm{Prism}_{S_j}(\mathbf{a}_j + \mathbf{v})) = \mathrm{Vol}_n(\mathrm{Prism}_{S_j}(\mathbf{a}_j)) = \mathrm{Vol}_n(A).$$

We are interested in the special case when $\mathbf{v} = \lambda \mathbf{a}_i$ for some $i \neq j$, in which case $A' = AL_{ij}(\lambda)$.

And that's enough about that.

2.7 Application to Calculus

In the previous section we showed that the a determinant can be viewed as the n-volume of an n-parallelotope living in $\mathbb{R}^n$. Now we apply this idea to volumes of arbitrary shapes.

Scaling Factor. Consider an $n \times n$ matrix A with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$. We can think of A as the linear function $\mathbb{R}^n \to \mathbb{R}^n$ that sends $\mathbf{x} \mapsto A\mathbf{x}$. Hence A sends the unit n-cube generated by the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ to the n-parallelotope generated by the vectors $A\mathbf{e}_i = \mathbf{a}_i$. Since the unit n-cube has volume 1 (by definition), we see that

$$\begin{aligned}
\mathrm{Vol}_n(A) &= |\det(A)| \\
\mathrm{Vol}_n(A) &= |\det(A)| \cdot 1 \\
\mathrm{Vol}_n(n\text{-parallelotope generated by } \mathbf{a}_1, \ldots, \mathbf{a}_n) &= |\det(A)| \cdot \mathrm{Vol}_n(\text{unit } n\text{-cube}).
\end{aligned}$$

More generally, consider an $n \times n$ matrix B with columns $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbb{R}^n$. Then A sends the n-parallelotope generated by $\mathbf{b}_1, \ldots, \mathbf{b}_n$ to the n-parallelotope generated by $A\mathbf{b}_1, \ldots, A\mathbf{b}_n$, which are the columns of AB. Hence we have

$$\begin{aligned}
&\mathrm{Vol}_n(\text{image of the parallelotope } \mathbf{b}_1, \ldots, \mathbf{b}_n \text{ under the function } A)\\
&\qquad = \mathrm{Vol}_n(\text{parallelotope generated by } A\mathbf{b}_1, \ldots, A\mathbf{b}_n)\\
&\qquad = \mathrm{Vol}_n(A\mathbf{b}_1, \ldots, A\mathbf{b}_n)\\
&\qquad = \mathrm{Vol}_n(AB)\\
&\qquad = |\det(AB)|\\
&\qquad = |\det(A)\det(B)|\\
&\qquad = |\det(A)| \cdot |\det(B)|\\
&\qquad = |\det(A)| \cdot \mathrm{Vol}_n(B)\\
&\qquad = |\det(A)| \cdot \mathrm{Vol}_n(\text{parallelotope } \mathbf{b}_1, \ldots, \mathbf{b}_n).
\end{aligned}$$

For example, the unit n-cube corresponds to the identity matrix $B = I_n$. It is also worth mentioning the case when $A = \lambda I_n$ for some scalar λ, so that A is the function that dilates $\mathbb{R}^n$ by a factor of λ. In this case we have[33]

$$\det(\lambda I_n) = \lambda^n,$$

so the function A scales volumes in $\mathbb{R}^n$ by a factor of λ^n. Indeed, if you double the side length of a cube in $\mathbb{R}^3$ then its volume gets multiplied by $8 = 2^3$.

We can think of a square matrix A in two ways. If we think of it as a collection of numbers then the determinant is the (signed) volume of the parallelotope generated by the columns. On the other hand, if we think of A as a linear function $\mathbb{R}^n \to \mathbb{R}^n$ then we should think of $|\det(A)|$ as a "volume scaling factor". Indeed, we have shown that applying A to any parallelotope in $\mathbb{R}^n$ scales its volume by $|\det(A)|$. I claim that the same idea holds for arbitrary[34] subsets of $\mathbb{R}^n$. To be precise, for any subset $S \subseteq \mathbb{R}^n$ we define the image set

$$A(S) := \{\text{the set of points } A\mathbf{p} \text{ for all } \mathbf{p} \in S\}.$$

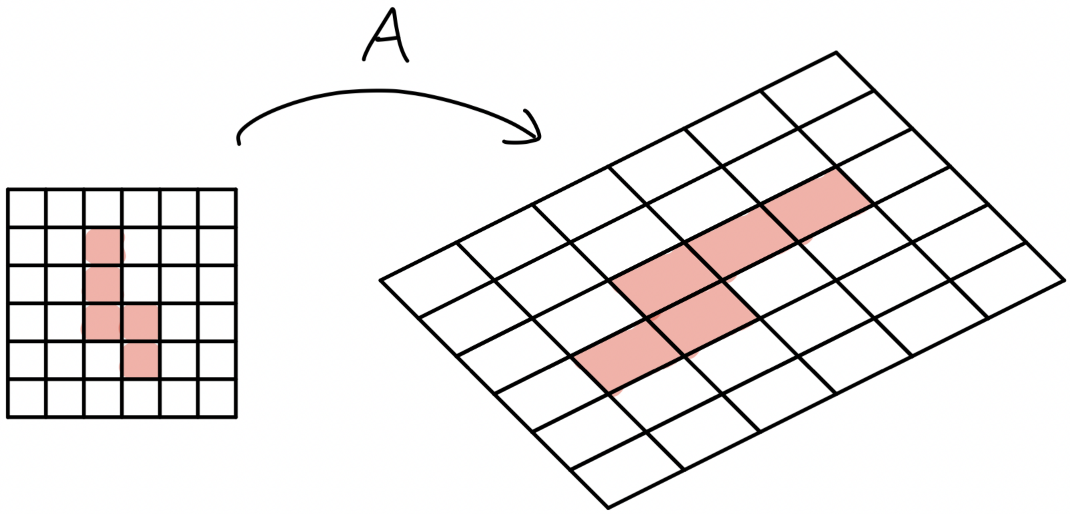
In this case I claim that

$$\boxed{\mathrm{Vol}_n(A(S)) = |\det(A)| \cdot \mathrm{Vol}_n(S).}$$

The idea of the proof is that any reasonable subset of $\mathbb{R}^n$ can be approximated as a union of tiny parallelotopes. To simplify the discussion we will use tiny cubes. Suppose that the set $S \subseteq \mathbb{R}^n$ is a union of tiny cubes. Then the image $A(S) \subseteq \mathbb{R}^n$ is a union of tiny parallelotopes, each of whose volume has been scaled by $|\det(A)|$. But the total volume is just the sum of the volumes of the tiny pieces. Hence the total volume is also scaled by $|\det(A)|$. Here is a picture:

[33] This follows from multilinearity. Multiplying one column by λ multiplies the determinant λ. Multiplying each of the n columns by λ multiplies the determinant by λ^n.

[34] Arbitrary "measurable" subsets. The real numbers are wild enough that they admit pathological examples such as "sets whose volume cannot be defined". I am happy to ignore such things.



Thinking of determinants as volume scaling factors of linear functions gives an intuitive explanation for the identity $\det(AB) = \det(A)\det(B)$. Indeed, for any subset $S \subseteq \mathbb{R}^n$ and for any $n \times n$ matrices A, B we have

$$\mathrm{Vol}_n((AB)(S)) = |\det(AB)| \cdot \mathrm{Vol}_n(S),$$

but also

$$\begin{aligned}\mathrm{Vol}_n((AB)(S)) &= \mathrm{Vol}_n(A(B(S)) \\ &= |\det(A)| \cdot \mathrm{Vol}_n(B(S)) \\ &= |\det(A)| \cdot |\det(B)| \cdot \mathrm{Vol}_n(S),\end{aligned}$$

which implies that $|\det(AB)| = |\det(A)| \cdot |\det(B)|$. (The sign is a bit trickier to handle.) This idea also gives meaning to the determinant of an abstract linear function $f : V \to V$, independent of choosing a basis for V.

Linear Approximation. We have seen that a linear function $A : \mathbb{R}^n \to \mathbb{R}^n$ scales the n-volume of an arbitrary shape $S \subseteq \mathbb{R}^n$ by a factor of $|\det(A)|$. In this section we will generalize from linear to **non-linear functions**.

A general function $\mathbf{r} : \mathbb{R}^m \to \mathbb{R}^n$ has the form

$$\mathbf{r}(x_1, \ldots, x_n) = \mathbf{r}(\mathbf{x}) = (\mathbf{r}_1(\mathbf{x}), \ldots, \mathbf{r}_m(\mathbf{x})),$$

where each component function $\mathbf{r}_i(x_1, \ldots, x_n)$ sends $\mathbb{R}^n \to \mathbb{R}$. Suppose that each $\mathbf{r}_i$ has continuous partial derivatives near some point $\mathbf{p} \in \mathbb{R}^n$, and consider the Taylor expansion:

$$\mathbf{r}_i(\mathbf{p} + \mathbf{x}) = \mathbf{r}_i(\mathbf{p}) + (\nabla \mathbf{r}_i)_{\mathbf{p}}^T \mathbf{x} + \text{higher terms},$$

where the higher terms are small when $\mathbf{x}$ is close to $\mathbf{0}$. Then we collect the components into a column vector:

$$\begin{aligned}
\mathbf{r}(\mathbf{p}+\mathbf{x}) &= \begin{pmatrix} \mathbf{r}_1(\mathbf{p}+\mathbf{x}) \\ \vdots \\ \mathbf{r}_m(\mathbf{p}+\mathbf{x}) \end{pmatrix} \\
&\approx \begin{pmatrix} \mathbf{r}_1(\mathbf{p}) + (\nabla \mathbf{r}_1)_{\mathbf{p}}^T \mathbf{x} \\ \vdots \\ \mathbf{r}_m(\mathbf{p}) + (\nabla \mathbf{r}_m)_{\mathbf{p}}^T \mathbf{x} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{r}_1(\mathbf{p}) \\ \vdots \\ \mathbf{r}_m(\mathbf{p}) \end{pmatrix} + \begin{pmatrix} (\nabla \mathbf{r}_1)_{\mathbf{p}}^T \mathbf{x} \\ \vdots \\ (\nabla \mathbf{r}_m)_{\mathbf{p}}^T \mathbf{x} \end{pmatrix} \\
&= \mathbf{r}(\mathbf{p}) + \begin{pmatrix} (\nabla \mathbf{r}_1)_{\mathbf{p}}^T \\ \vdots \\ (\nabla \mathbf{r}_m)_{\mathbf{p}}^T \end{pmatrix} \mathbf{x} \\
&= \mathbf{r}(\mathbf{p}) + \begin{pmatrix} \frac{\partial \mathbf{r}_1}{\partial x_1}(\mathbf{p}) & \cdots & \frac{\partial \mathbf{r}_1}{\partial x_n}(\mathbf{p}) \\ \vdots & & \vdots \\ \frac{\partial \mathbf{r}_m}{\partial x_1}(\mathbf{p}) & \cdots & \frac{\partial \mathbf{r}_m}{\partial x_n}(\mathbf{p}) \end{pmatrix} \mathbf{x}.
\end{aligned}$$

The $m \times n$ matrix of partial derivatives of the components of $\mathbf{r}$ is called the *Jacobian matrix*:

$$J\mathbf{r} := \begin{pmatrix} \partial \mathbf{r}_1/\partial x_1 & \cdots & \partial \mathbf{r}_1/\partial x_n \\ \vdots & & \vdots \\ \partial \mathbf{r}_m/\partial x_1 & \cdots & \partial \mathbf{r}_m/\partial x_n \end{pmatrix}.$$

This matrix plays the role of the "linear part" of the multi-multivariable Taylor expansion:

$$\mathbf{r}(\mathbf{p}+\mathbf{x}) = \mathbf{r}(\mathbf{p}) + (J\mathbf{r})_{\mathbf{p}}\mathbf{x} + \text{higher terms}.$$

In summary, suppose that a **possibly non-linear function** $\mathbf{r} : \mathbb{R}^n \to \mathbb{R}^m$ behaves nicely near a point $\mathbf{p} \in \mathbb{R}^n$. Then near this point the function $\mathbf{r}$ is approximately equal to the **linear function** corresponding to the $m \times n$ Jacobian matrix $(J\mathbf{r})_{\mathbf{p}}$. If $\mathbf{r}$ happens to be linear, corresponding to an $m \times n$ matrix A, then one can check that $(J\mathbf{r})_{\mathbf{p}} = A$ for any point $\mathbf{p}$. If $\mathbf{r}$ is non-linear then the matrix $(J\mathbf{r})_{\mathbf{p}}$ changes from point to point.

Application to Integration. In the previous sections we showed the following:

- If a function $\mathbf{r} : \mathbb{R}^m \to \mathbb{R}^n$ has continuous partial derivatives near a point $\mathbf{p} \in \mathbb{R}^n$ then we can approximate $\mathbf{r}$ near $\mathbf{p}$ by an $m \times n$ matrix $(J\mathbf{r})_{\mathbf{p}}$.
- A linear function $A : \mathbb{R}^n \to \mathbb{R}^n$ scales volume by the factor $|\det(A)|$.

Combining these ideas gives us a method to compute the volumes of parametrized shapes in $\mathbb{R}^n$. Before showing some examples, I will state the general theorem.

Theorem (Volume of a k-dimensional submanifold of $\mathbb{R}^n$). We wish to compute the k-volume of a k-dimensional subset $T \subseteq \mathbb{R}^n$. To do this, we look for a *parametrization function* $\mathbf{r} : \mathbb{R}^k \to \mathbb{R}^n$ whose image is T. Suppose that $\mathbf{r}$ sends the subset $S \subseteq \mathbb{R}^k$ to the subset $T \subseteq \mathbb{R}^n$. Then we can compute[35] the k-volume of T by integrating a suitable "volume stretch factor" over the region $S \subseteq \mathbb{R}^k$ using standard Euclidean coordinates:

$$\boxed{\mathrm{Vol}_k(T) = \mathrm{Vol}_k(\mathbf{r}(S)) = \int_{\mathbf{p}\in S} \sqrt{\det((J\mathbf{r})_{\mathbf{p}}^T(J\mathbf{r})_{\mathbf{p}})} \cdot d\mathbf{p}.}$$

Remark: We require the shapes S, T and the function $\mathbf{r}$ to be sufficiently nice. This involves several technical conditions that I am happy to ignore. Basically, S and T should be reasonably smooth, and $\mathbf{r}$ should parametrize T without any overlaps or kinks.

Proof. A tiny cube at the point $\mathbf{p} \in S$ has a tiny volume $d\mathbf{p}$. The function $\mathbf{r}$ is approximately linear at $\mathbf{p}$, given by the $n \times k$ matrix $(J\mathbf{r})_{\mathbf{p}}$. This matrix sends the tiny cube at the point $\mathbf{p}$ to a tiny k-parallelotope at the point $\mathbf{r}(\mathbf{p})$. For any small shape near $\mathbf{p}$, the linear function $(J\mathbf{r})_{\mathbf{p}}$ scales its volume by a factor of[36]

$$\sqrt{\det((J\mathbf{r})_{\mathbf{p}}^T(J_{\mathbf{r}}))}.$$

Hence the volume of the tiny k-parallelotope at the point $\mathbf{r}(\mathbf{p})$ is

$$\sqrt{\det((J\mathbf{r})_{\mathbf{p}}^T(J_{\mathbf{r}}))} \cdot (\text{volume of the tiny cube}) = \sqrt{\det((J\mathbf{r})_{\mathbf{p}}^T(J_{\mathbf{r}}))} \cdot d\mathbf{p}.$$

Then we just add up all these tiny volumes to get the k-volume of T. □

To end this section, I will illustrate how this result unifies several formulas from Calculus III.

Example: Arc Length. Let $\mathbf{r} : \mathbb{R} \to \mathbb{R}^n$ be a parametrized path in $\mathbb{R}^n$. Usually we think of the parameter as time, and we write $\mathbf{r}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$. The Jacobian matrix at time t is just the velocity vector:

$$J\mathbf{r}(t) = \begin{pmatrix} \partial x_1/\partial t \\ \vdots \\ \partial x_n/\partial t \end{pmatrix} = \mathbf{r}'(t).$$

In this case, $(J\mathbf{r})^T(J\mathbf{r})$ is just a scalar, and the 1-volume (i.e., length) scaling factor is just the speed of the parametrization:

$$(J\mathbf{r})^T(J\mathbf{r}) = \mathbf{r}'(t)^T\mathbf{r}(t)$$

[35] In fact, this formula is often used as the **definition** of volume.

[36] We only proved this in the case $k = n$, when $(J_{\mathbf{r}})_{\mathbf{p}}$ is a square matrix and the scaling factor reduces to $|\det((J\mathbf{r})_{\mathbf{p}})|$. The general case follows by the same argument as in 2.6.

$$
\begin{aligned}
(J\mathbf{r})^T(J\mathbf{r}) &= \|\mathbf{r}'(t)\|^2 \\
\det((J\mathbf{r})^T(J\mathbf{r})) &= \|\mathbf{r}'(t)\|^2 \\
\sqrt{\det((J\mathbf{r})^T(J\mathbf{r}))} &= \|\mathbf{r}'(t)\|.
\end{aligned}
$$

Then the theorem tells us that the arc length of the curve is just the integral of the speed:

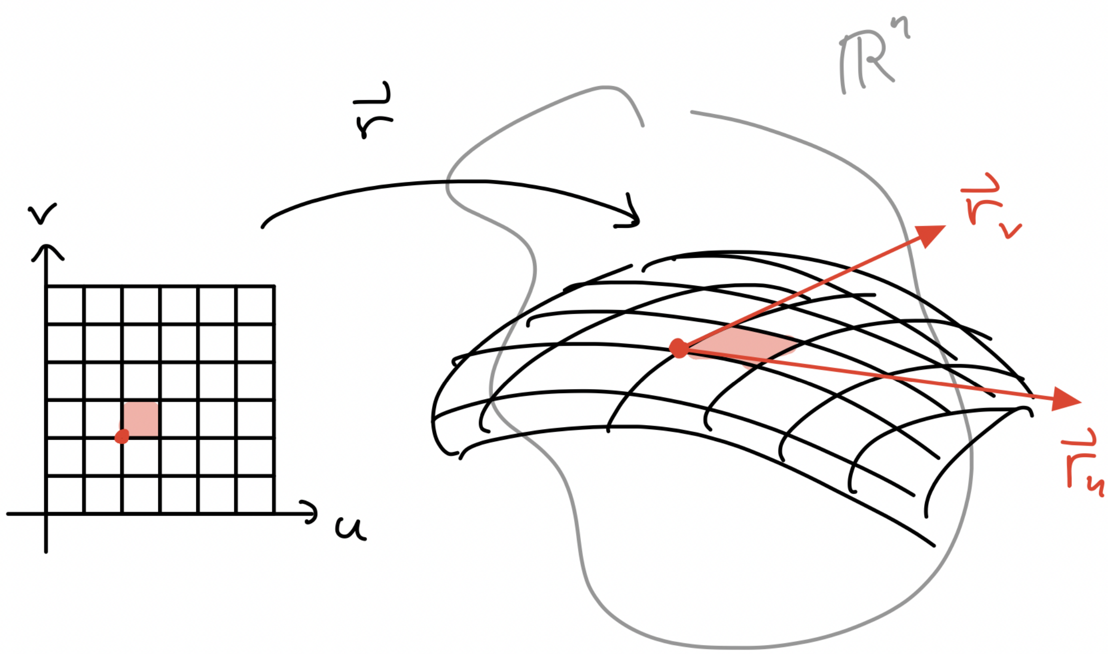
$$
(\text{length of the curve } \mathbf{r}(t) \text{ between times } t = a \text{ and } t = b) = \int_a^b \|\mathbf{r}'(t)\|\, dt.
$$

Of course this makes sense because distance is the time integral of speed.

Example: Surface Area. Let $\mathbf{r} : \mathbb{R}^2 \to \mathbb{R}^n$ be a parametrization for a 2-dimensional surface $T \subseteq \mathbb{R}^n$. It is common to write $\mathbf{r}(u, v) = (x_1(u, v), \ldots, x_n(u, v))$, where each coordinate x_i is a function from $\mathbb{R}^2$ to $\mathbb{R}$. The Jacobian matrix is

$$
J\mathbf{r} = \begin{pmatrix} \partial x_1/\partial u & \partial x_1/\partial v \\ \vdots & \vdots \\ \partial x_n/\partial u & \partial x_n/\partial v \end{pmatrix} = \begin{pmatrix} | & | \\ \mathbf{r}_u & \mathbf{r}_v \\ | & | \end{pmatrix},
$$

where $\mathbf{r}_u$ and $\mathbf{r}_v$ are the "velocity vectors" of $\mathbf{r}$ in the u and v directions[37]



[37] If one of u or v is fixed then you can think of the other as time.

In this case the 2-volume (i.e., area) scaling factor is the area of the parallelogram generated by $\mathbf{r}_u$ and $\mathbf{r}_v$:

$$\begin{aligned}
(J\mathbf{r})^T(J\mathbf{r}) &= \begin{pmatrix} - & \mathbf{r}_u & - \\ - & \mathbf{r}_v & - \end{pmatrix} \begin{pmatrix} | & | \\ \mathbf{r}_u & \mathbf{r}_v \\ | & | \end{pmatrix} \\
&= \begin{pmatrix} \|\mathbf{r}_u\|^2 & \mathbf{r}_u \bullet \mathbf{r}_v \\ \mathbf{r}_u \bullet \mathbf{r}_v & \|\mathbf{r}_v\|^2 \end{pmatrix} \\
\det((J\mathbf{r})^T(J\mathbf{r})) &= \|\mathbf{r}_u\|^2\|\mathbf{r}_v\|^2 - (\mathbf{r}_u \bullet \mathbf{r}_v)^2 \\
&= \|\mathbf{r}_u\|^2\|\mathbf{r}_v\|^2 - (|\mathbf{r}_u\|\|\mathbf{r}_v\| \cos\theta_{uv})^2 \\
&= \|\mathbf{r}_u\|^2\|\mathbf{r}_v\|^2(1 - \cos^2\theta_{uv}) \\
&= \|\mathbf{r}_u\|^2\|\mathbf{r}_v\|^2 \sin^2\theta_{uv} \\
\sqrt{\det((J\mathbf{r})^T(J\mathbf{r}))} &= \|\mathbf{r}_u\|\|\mathbf{r}_v\||\sin\theta_{uv}|,
\end{aligned}$$

where θ_{uv} is the angle between the velocity vectors $\mathbf{r}_u$ and $\mathbf{r}_v$. In the special case of a surface in $\mathbb{R}^3$, we can also describe this area as the length of the cross product vector:

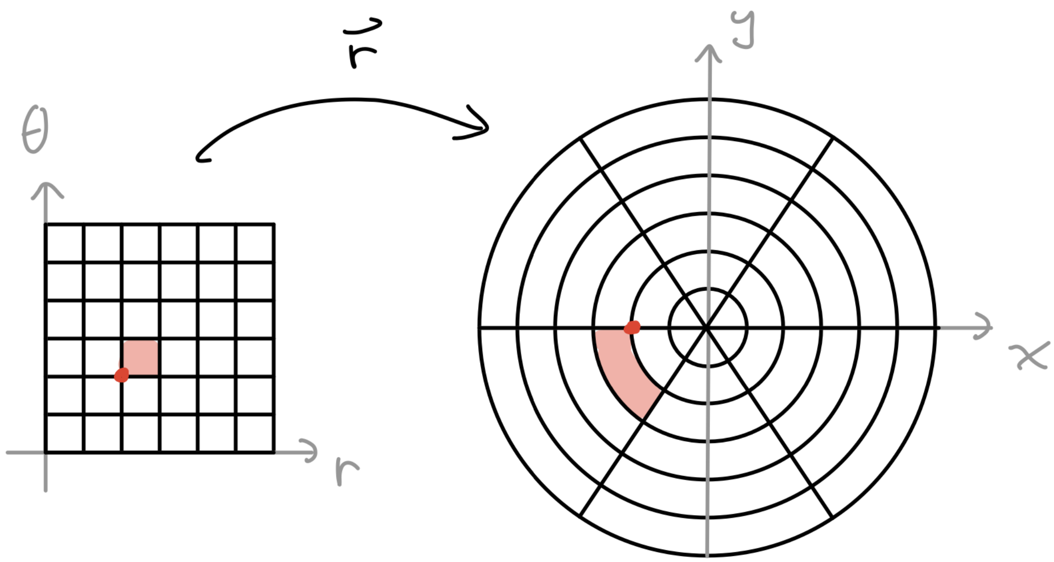
$$\|\mathbf{r}_u \times \mathbf{r}_v\| = \|\mathbf{r}_u\|\|\mathbf{r}_v\||\sin\theta_{uv}|.$$

To compute the area of the surface, we add up all of the areas of tiny parallelograms:

$$(\text{area of the surface } T \subseteq \mathbb{R}^n) = \int \sqrt{\|\mathbf{r}_u\|^2\|\mathbf{r}_v\|^2 - (\mathbf{r}_u \bullet \mathbf{r}_v)^2} \cdot dudv.$$

Example: Change of Coordinates. A parametrization of an n-dimensional shape in n-dimensional space is sometimes viewed as a "change of coordinates" $\mathbf{r} : \mathbb{R}^n \to \mathbb{R}^n$. For example, take the parametrization of $\mathbb{R}^2$ by polar coordinates:

$$\mathbf{r}(r, \theta) = \begin{pmatrix} x(r,\theta) \\ y(r,\theta) \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}.$$



The Jacobian stretch factor at the point (r, θ) is

$$\begin{aligned} J\mathbf{r} &= \begin{pmatrix} \partial x/\partial r & \partial x/\partial \theta \\ \partial y/\partial r & \partial y/\partial \theta \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \\ \sqrt{\det((J\mathbf{r})^T(J\mathbf{r}))} &= |\det(J\mathbf{r})| \\ &= |r\cos^2\theta + r\sin^2\theta| \\ &= |r|. \end{aligned}$$

Hence the area of a region T in the x, y-plane, which is parametrized by a region S in the r, θ-plane is given by[38]

$$\int_S r \cdot dr d\theta.$$

Since a change of coordinates maps a space into itself, changes of coordinates can be composed. Suppose we have functions $\mathbf{r} : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathbf{s} : \mathbb{R}^n \to \mathbb{R}^n$, with composition $\mathbf{r} \circ \mathbf{s} : \mathbb{R}^n \to \mathbb{R}^n$. The multi-multivariable version of the chain rule says that the Jacobian matrix of the composition $\mathbf{r} \circ \mathbf{s}$ is equal to the product of the Jacobian matrices of $\mathbf{r}$ and $\mathbf{s}$. That is, for any point $\mathbf{p} \in \mathbb{R}^n$ we have

$$(J(\mathbf{r} \circ \mathbf{s}))_{\mathbf{p}} = (J\mathbf{r})_{\mathbf{s}(\mathbf{p})} \cdot (J\mathbf{s})_{\mathbf{p}}.$$

[38]In order to ensure the "niceness" of the parametrization, we will take $r \geq 0$ (so that $|r| = r$) and $0 \leq \theta < 2\pi$.

Hence the the Jacobian scaling factors multiply:

$$|\det((J(\mathbf{r} \circ \mathbf{s}))_{\mathbf{p}})| = |\det((J\mathbf{r})_{\mathbf{s}(\mathbf{p})})| \cdot |\det((J\mathbf{s})_{\mathbf{p}})|.$$

Observe that the notation is getting complicated. Indeed, the subject of differential geometry is known for its impenetrable notation. Since no two authors can understand each other, they often invent their own personal notations. Einstein's notation is the most popular among physicists.