Math 510 **Fall 2022**
Linear Algebra II **Drew Armstrong**

Contents

1 Basic Concepts

1.1 Abstract Vector Spaces

Let $\mathbb{R}$ be the set of real numbers. A *vector space over* $\mathbb{R}$ consists of a set V (of "vectors"), with two algebraic operations, called *addition* and *scalar multiplication*:

$$\mathbf{u}, \mathbf{v} \in V \rightsquigarrow \mathbf{u} + \mathbf{v} \in V$$
$$a \in \mathbb{R}, \mathbf{v} \in V \rightsquigarrow a\mathbf{v} \in V.$$

[Remark: We could also write scalar multiplication as $\mathbf{v}a$; the order doesn't matter.] These two operations are required to satisfy the following eight axioms:

(1) Axioms of Addition.

(a) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(b) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

(c) There exists a vector $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.

(d) For every vector $\mathbf{v} \in V$ there exists a vector $\mathbf{u} \in V$ such that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} = \mathbf{0}$.

Remarks: The vector $\mathbf{0}$ in axiom (1c) is unique. Indeed, if $\mathbf{0}$ and $\mathbf{0}'$ are two vectors satisfying (1c) then we must have

$$\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}'.$$

We call the unique element $\mathbf{0} \in V$ satisfying (1c) the *zero vector*. The vector $\mathbf{u}$ in axiom (1d) is also unique. Indeed, suppose we have two vectors $\mathbf{u}$ and $\mathbf{u}'$ satisfying (1d). Then from axioms (1abc) we must have

$$\mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u} + (\mathbf{v} + \mathbf{u}') = (\mathbf{u} + \mathbf{v}) + \mathbf{u}' = \mathbf{0} + \mathbf{u}' = \mathbf{u}'.$$

The unique element $\mathbf{u}$ satisfying (1d) is called the *additive inverse of* $\mathbf{v}$. We denote it by $-\mathbf{v}$. In other words, we have

$$\mathbf{v} + \mathbf{u} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{u} = -\mathbf{v}.$$

Based on this, we define the operation of vector subtraction:

$$\mathbf{u} - \mathbf{v} := \mathbf{u} + (-\mathbf{v}).$$

(2) Axioms of Scalar Multiplication.

(a) For the real number $1 \in \mathbb{R}$ we have $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$.

(b) For all real numbers $a, b \in \mathbb{R}$ and vectors $\mathbf{v} \in V$ we have $a(b\mathbf{v}) = (ab)\mathbf{v}$.[1]

(c) For all $a, b \in \mathbb{R}$ and $\mathbf{v} \in V$ we have $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$.[2]

[1] Note that this identity involves two **different** operations: multiplication of real numbers and scalar multiplication in V. This identity is the reason that we use the same notation for both operations.

[2] This identity is the reason that we use the same notation for addition in $\mathbb{R}$ and addition in V.

(d) For all $a \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in V$ we have $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.

Remarks: These eight axioms imply many other basic properties. For example, I claim that the real number $0 \in \mathbb{R}$ satisfies $0\mathbf{v} = \mathbf{0}$ for all vectors $\mathbf{v} \in V$, where $\mathbf{0}$ is the zero vector. Indeed, since $0 + 0 = 0$ as real numbers, we have

$$\begin{aligned} 0 + 0 &= 0 \\ (0 + 0)\mathbf{v} &= 0\mathbf{v} \\ 0\mathbf{v} + 0\mathbf{v} &= 0\mathbf{v} \\ (0\mathbf{v} + 0\mathbf{v}) - 0\mathbf{v} &= 0\mathbf{v} - 0\mathbf{v} \\ 0\mathbf{v} + (0\mathbf{v} - 0\mathbf{v}) &= \mathbf{0} \\ 0\mathbf{v} + \mathbf{0} &= \mathbf{0} \\ 0\mathbf{v} &= \mathbf{0}. \end{aligned}$$

[We could have taken this as another axiom, but we didn't need to.] It follows from this that the additive inverse $-\mathbf{v}$ is the same as $(-1)\mathbf{v}$ for the real number $-1 \in \mathbb{R}$. Indeed, since $1 + (-1) = 0$ as real numbers, we have

$$\begin{aligned} 1 + (-1) &= 0 \\ (1 + (-1))\mathbf{v} &= 0\mathbf{v} \\ 1\mathbf{v} + (-1)\mathbf{v} &= 0\mathbf{v} \\ \mathbf{v} + (-1)\mathbf{v} &= \mathbf{0} \\ -\mathbf{v} + (\mathbf{v} + (-1)\mathbf{v}) &= -\mathbf{v} + \mathbf{0} \\ (-\mathbf{v} + \mathbf{v}) + (-1)\mathbf{v} &= -\mathbf{v} \\ \mathbf{0} + (-1)\mathbf{v} &= -\mathbf{v} \\ (-1)\mathbf{v} &= -\mathbf{v}. \end{aligned}$$

If this is too pedantic for you, feel free to take the properties $0\mathbf{v} = \mathbf{0}$ and $(-1)\mathbf{v} = -\mathbf{v}$ as axioms.

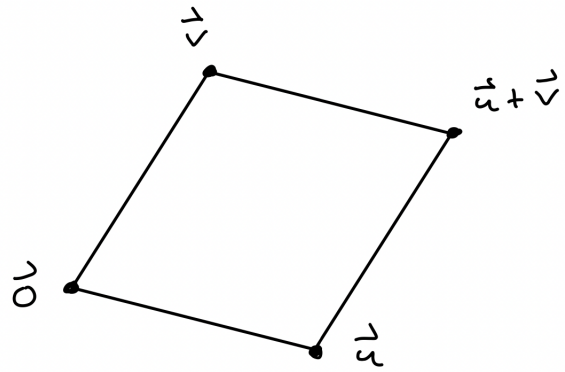
The Prorotype: Euclidean Space. Let $\mathbb{R}^n$ denote the set of ordered n-tupes of real numbers:

$$\mathbb{R}^n = \{\mathbf{v} = (v_1, v_2, \ldots, v_n) : v_1, v_2, \ldots, v_n \in \mathbb{R}\}.$$

It is easy to check that the following operations make $\mathbb{R}^n$ into a vector space over $\mathbb{R}$:

$$\begin{aligned} (u_1, \ldots, u_n) + (v_1, \ldots, v_n) &:= (u_1 + v_1, \ldots, u_n + v_n), \\ a(v_1, \ldots, v_n) &:= (av_1, \ldots, av_n). \end{aligned}$$

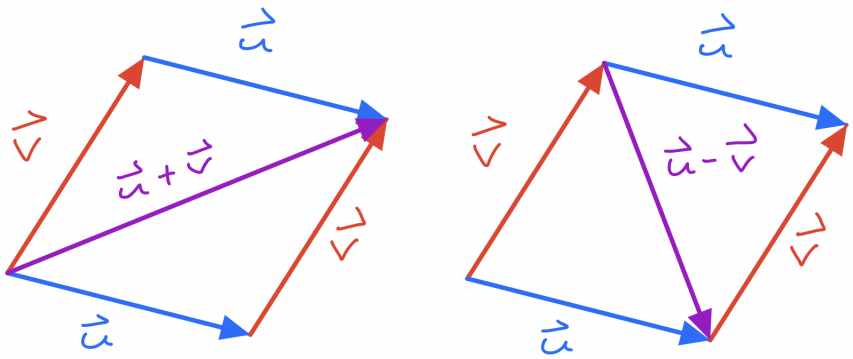
We can think of $\mathbf{v} = (v_1, \ldots, v_n)$ as the coordinates of a point in n-dimensional Euclidean space. In this case, the point $\mathbf{0} = (0, \ldots, 0)$ is called the *origin*. The Parallelogram Law says that for any points $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the four points $\mathbf{0}$, $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u} + \mathbf{v}$ are the vertices of a parallelogram. Picture:



We can also think of an n-tuple $\mathbf{v} = (v_1, \ldots, v_n)$ as a directed line segment (an "arrow") with head at the point $\mathbf{v}$ and tail at the origin $\mathbf{0}$. According to the Pythagorean Theorem, the length $\|\mathbf{v}\|$ of this line segment satisfies

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 \cdots + v_n^2}.$$

Geometrically, arrows add "head-to-tail" and subtract "tail-to-tail":



If we let θ denote the angle between arrows $\mathbf{u}$ and $\mathbf{v}$ then the Law of Cosines tells us that

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\| \cos\theta.$$

On the other hand, the algebraic formula for the length of an arrow tells us that

$$\begin{aligned}
\|\mathbf{u} - \mathbf{v}\|^2 &= \|(u_1 - v_1, \ldots, u_n - v_n)\|^2 \\
&= (u_1 - v_1)^2 + \cdots + (u_n - v_n)^2 \\
&= (u_1^2 - 2u_1v_1 + v_1^2) + \cdots + (u_n^2 - 2u_nv_n + v_n^2) \\
&= (u_1^2 + \cdots + u_n)^2 + (v_1^2 + \cdots + v_n^2) - 2(u_1v_1 + \cdots + u_nv_n)
\end{aligned}$$

$$= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2(u_1v_1 + \cdots + u_nv_n).$$

Then comparing the two equations gives the amazing formula

$$u_1v_1 + \cdots + u_nv_n = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta.$$

This formula allows us to express angles simply in terms of the coordinates. To be precise, we define the *dot product* of two arrows:

$$\mathbf{u} \bullet \mathbf{v} := u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Observe that

$$\mathbf{v} \bullet \mathbf{v} = v_1v_1 + \cdots + v_nv_n = v_1^2 + \cdots + v_n^2 = \|\mathbf{v}\|^2.$$

Hence we have

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{\mathbf{u} \bullet \mathbf{v}}{\sqrt{\mathbf{u} \bullet \mathbf{u}}\sqrt{\mathbf{v} \bullet \mathbf{v}}}.$$

Note that θ is a right angle if and only if $\mathbf{u} \bullet \mathbf{v} = 0$.

1.2 Inner Product Spaces over $\mathbb{R}$

More generally, an *inner product space over* $\mathbb{R}$ consists of a vector space V over $\mathbb{R}$ together with another algebraic operation

$$\mathbf{u}, \mathbf{v} \in V \rightsquigarrow \langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{R},$$

which must satisfy the following axioms:

(3) Axioms of Inner Products.

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$[3]

(c) For all $a \in \mathbb{R}$ and $\mathbf{u}, \mathbf{v} \in V$ we have $\langle a\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, a\mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle$.

(d) For all $\mathbf{v} \in V$ we have $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, with $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

The following important inequality is a direct consequence of the axioms, but its proof is just a little bit tricky. I'll give you a hint and have you prove it on the homework.

Cauchy-Schwarz Inequality. For any vectors $\mathbf{u}, \mathbf{v} \in V$ in an inner product space we have

$$|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle.$$

Why should we bother with this level of abstraction? There are two reasons.

[3] By combining (3ab) we also have $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$.

First of all, there exist important examples of abstract inner product spaces that have nothing to do with arrows or points in Euclidean space.

Example: L^2 Space. Let $L^2[0,1]$ denote the set of real-valued functions $f(x)$ on the interval $[0,1]$ such that the integral of $f(x)$ converges:[4]

$$L^2[0,1] = \{f : [0,1] \to \mathbb{R}, \int_0^1 f(x)^2\,dx < \infty\}.$$

Given functions $f, g \in L^2[0,1]$ and scalar $a \in \mathbb{R}$ we define the new functions $f+g \in L^2[0,1]$ and $af \in L^2[0,1]$ by adding and multiplying their values, as one does in Calculus:

$$\begin{aligned}(f+g)(x) &:= f(x) + g(x),\\ (af)(x) &:= af(x).\end{aligned}$$

One can check that these operations make $L^2[0,1]$ into a vector space over $\mathbb{R}$.[5] Furthermore, one can check that the following operation satisfies the inner product axioms:

$$\langle f(x), g(x)\rangle := \int_0^1 f(x)g(x)\,dx.$$

Such inner product spaces of functions are extremely important in applied mathematics. We will say more below.

Another reason for abstraction in linear algebra has to do with "subspaces".

(4) Axioms of Subspaces. Given a vector space V over $\mathbb{R}$ and a subset $U \subseteq V$, we say that U is a *subspace* when it satisfies the following axioms:

(a) $\mathbf{0} \in U$

(b) If $\mathbf{u}, \mathbf{v} \in U$ then $\mathbf{u} + \mathbf{v} \in U$.

(c) If $a \in \mathbb{R}$ and $\mathbf{v} \in U$ then $a\mathbf{v} \in U$.

For example, any line or plane through the origin in Euclidean space is a subspace.[6] We note that Euclidean spaces comes with a collection of *standard basis vectors*:

$$\begin{aligned}\mathbf{e}_1 &= (1,0,0,\ldots,0,0),\\ \mathbf{e}_2 &= (0,1,0,\ldots,0,0),\end{aligned}$$

[4] Any statement about integrals has some very technical conditions, but we will proceed intuitively, just as a physicist would.

[5] The hardest part of the proof is to show that the sum of square integrable functions is square integrable. This can be shown with the Cauchy-Schwarz inequality.

[6] A line or plane not through the origin is not a subspace because it doesn't satisfy (3a). The concept of "subspace" is not immediately intuitive but it is vital to the theory.

$$\vdots$$
$$\mathbf{e}_n = (0, 0, 0, \ldots, 0, 1).$$

By definition, every vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ has a unique expression as a *linear combination* of these basis vectors:

$$\begin{aligned}\mathbf{v} &= (v_1, \ldots, v_n) \\ &= v_1(1, 0, \ldots, 0, 0) + \cdots + v_n(0, 0, \ldots, 0, 1) \\ &= v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n.\end{aligned}$$

However, **subspaces** of $\mathbb{R}^n$ do not come with standard basis vectors. For example, consider the plane $V \subseteq \mathbb{R}^3$ defined by the equation $x - 2y + z = 0$.[7] I claim that every vector $\mathbf{v} \in V$ this plane has a unique expression of the form

$$\mathbf{v} = a_1(1, 1, 1) + a_2(1, 2, 3).$$

Hence we say that $B = \{\mathbf{b}_1, \mathbf{b}_2\}$ with $\mathbf{b}_1 = (1, 1, 1)$ and $\mathbf{b}_2 = (1, 2, 3)$ is a *basis* for the vector space V, and if $\mathbf{v} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2$ we say that $\mathbf{v} = (a_1, a_2)_B$ are the *coordinates of* $\mathbf{v}$ *in the* B*-basis*. For example, the vector $\mathbf{v} = (1, -1, -3) \in \mathbb{R}^3$ is in the plane V. It has coordinates $(1, -1, -3)$ as an element of $\mathbb{R}^3$ but it has coordinates $(3, -2)_B$ as an element of V, with respect to the B-basis. Here is why we need the concept of an abstract vector space:

> *Subspaces of* $\mathbb{R}^n$ *do not come with a standard basis. Therefore we must study them via the axioms of abstract vector spaces.*

Here is the technical definition of a basis in an abstract vector space.

Definition of Basis. Let V be a vector space over $\mathbb{R}$ and consider a finite subset $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of vectors in V.

- We say that B is a *spanning set* if for all $\mathbf{v} \in V$ **there exists at least one choice** of scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that
$$\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n.$$
- We say that B is an *independent set*[8] if for all $\mathbf{v} \in V$ **there exists at most one choice** of scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that
$$\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n.$$

[7] Check that this is a subspace.

[8] In proofs it is often convenient to use a different form of the definition. Say that B is *independent* if for any scalars $a_1, \ldots, a_n$ we have
$$a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n = \mathbf{0} \text{ implies } a_i = 0 \text{ for all } i.$$
Exercise: Check that the two definitions are equivalent.

- We say that B is a *basis* if it is spanning and independent; that is, if for all $\mathbf{v} \in V$ **there exists a unique choice** of scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that

$$\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n.$$

  In this case we say that $a_1, \ldots, a_n \in \mathbb{R}$ are the *B-coordinates* of $\mathbf{v}$, and we write

$$\mathbf{v} = (a_1, \ldots, a_n)_B.$$

  After we have chosen a basis, we can work with coordinates and pretend that V is $\mathbb{R}^n$.

The following point is fundamental, but its proof is more subtle than you would think.

Definition of Dimension. If a vector space V has a basis with n vectors, then any basis of V must have n vectors. In this case we say that V has *dimension* n, and we write

$$\dim V = n.$$

Proof. This uses a famous trick called "Steinitz Exchange". See the homework.

Example: Euclidean Space. The vector space $\mathbb{R}^n$ has a standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ consisting of n vectors. It follows from Steinitz Exchange that **any** basis for $\mathbb{R}^n$ must have n vectors, and hence $\dim \mathbb{R}^n = n$, as it should be. It is relatively easy to find a basis: any sufficiently random collection in n vectors in $\mathbb{R}^n$ will do. For example:

$(1, 4, 3, 2), (3, -7, 4, 1), (100, 89, -72, 36), (23, 24, 25, 26)$ is almost certainly a basis of $\mathbb{R}^4$.

Not every vector space has a finite basis.

Example: Polynomials. Let $\mathbb{R}[x]$ denote the set of polynomials in x with real coefficients. This set is a vector space over $\mathbb{R}$. It does not have a finite basis, but it does have a fairly obvious infinite basis B consisting of the elements

$$B = \{1(= x^0), x, x^2, \ldots\}.$$

For infinite bases we need to modify slightly the definitions of independence and spanning. In this case, the key fact is that each polynomial $f(x) \in \mathbb{R}[x]$ **has a unique expression**

$$f(x) = \sum_{k \geq 0} a_k x^k,$$

where only finitely many of the coefficients $a_0, a_1, a_2, \ldots$ are nonzero. If we allow infinitely many nonzero coefficients then we obtain *power series*, instead of polynomials.

1.3 Normed Spaces

In order to say anything about convergence of infinite series in a vector space, one needs a way to measure "distance" between vectors.

(5) Axioms of Norms. Let V be a vector space with a function

$$\mathbf{v} \in V \rightsquigarrow \|\mathbf{v}\| \in \mathbb{R}.$$

We call this function a *norm* when it satisfies the following axioms:

(a) $\|\mathbf{v}\| \geq 0$ for all $\mathbf{v} \in V$ with $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(b) For all $a \in \mathbb{R}$ and $\mathbf{v} \in V$ we have $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|$.

(c) For all $\mathbf{u}, \mathbf{v} \in V$ we have $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

(6) Axioms of Metrics. Let V be a vector space with a function

$$\mathbf{u}, \mathbf{v} \in V \rightsquigarrow \text{dist}(\mathbf{u}, \mathbf{v}) \in \mathbb{R}.$$

We call this function a *metric* when it satisfies the following axioms:

(a) $\text{dist}(\mathbf{u}, \mathbf{v}) = \text{dist}(\mathbf{v}, \mathbf{u})$

(b) $\text{dist}(\mathbf{u}, \mathbf{v}) \geq 0$ for all $\mathbf{u}, \mathbf{v} \in V$ with $\text{dist}(\mathbf{u}, \mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$.

(c) $\text{dist}(\mathbf{u}, \mathbf{v}) \leq \text{dist}(\mathbf{u}, \mathbf{w}) + \text{dist}(\mathbf{w}, \mathbf{v})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$.

Every inner product space becomes a normed space[9] by taking $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, and every normed space becomes a metric space by taking $\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Concept of Orthonormal Sets. Let V be an inner product space. A collection of vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots$ is called *orthonormal* if

- $\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0$ for all i, j with $i \neq j$
- $\langle \mathbf{b}_i, \mathbf{b}_i \rangle = 1$ for all i

The first statement says that any two vectors in the set are *orthogonal*,[10] and the second statement says that each vector has length 1:

$$\|\mathbf{b}_i\| = \sqrt{\langle \mathbf{b}_i, \mathbf{b}_i \rangle} = \sqrt{1} = 1.$$

Orthonormal sets are very easy to work with. You will show on the homework that if $\mathbf{b}_i$ are orthonormal and $\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n$ then we must have

$$a_i = \langle \mathbf{v}, \mathbf{b}_i \rangle \quad \text{and} \quad \|\mathbf{v}\|^2 = a_1^2 + \cdots + a_n^2.$$

[9] You will prove this on the homework.

[10] In Euclidean space this corresponds to perpendicular vectors

If the orthonormal set spans V then it is called an *orthonormal basis*. Orthonormal bases are analogous to the standard basis in Euclidean space.

Example: Fourier Series. The inner product space $L^2[0,1]$ of square integrable functions $[0,1] \to \mathbb{R}$ contains a particularly famous orthonormal set of functions. If we define

$$s_n(x) := \sqrt{2}\sin(2\pi n x),$$
$$c_n(x) := \sqrt{2}\cos(2\pi n x),$$

then you will show on the homework that the following set of functions is orthonormal:

$$B = \{1, s_1(x), s_2(x), \ldots, c_1(x), c_2(x), \ldots\}.$$

That is, you will show

- $\langle 1, s_n(x)\rangle = \langle 1, c_n(x)\rangle = 0$ for all $n \geq 1$,
- $\langle s_m(x), c_n(x)\rangle = 0$ for all $m, n \geq 1$,
- $\langle s_m(x), s_n(x)\rangle = 0$ for $m \neq n$ and 1 for $m = n$,
- $\langle c_m(x), c_n(x)\rangle = 0$ for $m \neq n$ and 1 for $m = n$.

It follows from this that the set is independent. Is it also a spanning set? For a given function $f(x) \in L^2[0,1]$, the problem is to find scalars $a_0, a_1, a_2, \ldots, b_1, b_2, \ldots \in \mathbb{R}$ such that

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n s_n(x) + \sum_{n=1}^{\infty} b_n c_n(x). \qquad (*)$$

In Fourier's paper on the analytic theory of heat (1822) he gave a clever formula to find the coefficients. For us this formula is an immediate consequence of the fact that B is orthonormal:

$$a_0 = \langle f(x), 1\rangle = \int_0^1 f(x)\,dx,$$
$$a_n = \langle f(x), s_n(x)\rangle = \sqrt{2}\int_0^1 f(x)\sin(2\pi n x)\,dx,$$
$$b_n = \langle f(x), c_n(x)\rangle = \sqrt{2}\int_0^1 f(x)\cos(2\pi n x)\,dx.$$

So the coefficients are easy to find. The hard question is whether, and in what sense, the infinite series $(*)$ converges. This is an important problem in the history of mathematics; controversies surrounding its solution led to many of the concepts of modern analysis.

I will just state the simplest form of the answer; the proof is well beyond the scope of this course. Consider the distance function induced by the inner product on $L^2[0,1]$. That is, for any functions $f(x), g(x) \in L^2[0,1]$ we define the "distance" between then by

$$\operatorname{dist}(f(x), g(x))^2 = \|f(x) - g(x)\|^2 = \langle f(x) - g(x), f(x) - g(x)\rangle = \int_0^1 (f(x) - g(x))^2\,dx.$$

Now consider any function $f(x) \in L^2[0,1]$ and let a_n, b_n be the corresponding Fourier coefficients. Then we have the following theorems.

- **Convergence of Fourier Series.** The series $(*)$ converges in L^2. That is, we have

$$\operatorname{dist}\left(f(x), a_0 + \sum_{n=1}^{N} a_n s_n(x) + \sum_{n=1}^{N} b_n c_n(x)\right) \to 0 \quad \text{as } N \to \infty.$$

- **Parseval's Identity.** Computing the "length" of each side of $(*)$ gives a convergent series of real numbers:

$$\int_0^1 f(x)^2\,dx = \langle f(x), f(x)\rangle = a_0^2 + a_1^2 + b_1^2 + a_2^2 + b_2^2 + \cdots.$$

For example, consider the square wave function

$$f(x) = \begin{cases} 1 & 0 \le x < 1/2, \\ 0 & 1/2 \le x \le 1. \end{cases}$$

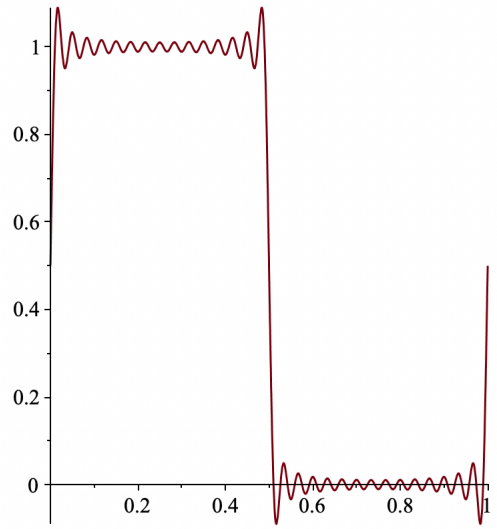
It is easy to check that $a_0 = \langle f(x), 1\rangle = 1/2$ and $b_n = \langle f(x), c_n(x)\rangle = 0$ for all $n \ge 1$. Next, we compute

$$\begin{aligned} a_n &= \langle f(x), s_n(x)\rangle \\ &= \sqrt{2}\int_0^1 f(x)\sin(2\pi n x)\,dx \\ &= \sqrt{2}\int_0^{1/2} \sin(2\pi n x)\,dx \\ &= \frac{\sqrt{2}}{2\pi n}\left[-\cos(2\pi n x)\right]_0^{1/2} \\ &= \frac{\sqrt{2}}{2\pi n}\left[-\cos(\pi n) + 1\right] \\ &= \frac{\sqrt{2}}{2\pi n}\left[-(-1)^n + 1\right] \\ &= \begin{cases} 0 & n \text{ even}, \\ \frac{\sqrt{2}}{\pi n} & n \text{ odd}. \end{cases} \end{aligned}$$

It follows that

$$f(x) = \frac{1}{2} + \frac{\sqrt{2}}{\pi}\sin(2\pi x) + \frac{\sqrt{2}}{3\pi}\sin(6\pi x) + \frac{\sqrt{2}}{5\pi}\sin(10\pi x) + \cdots.$$

Here is a picture of the first 30 terms of this sequence:



Finally, Parseval's Identity gives the following interesting identity:

$$\begin{aligned}
\int_0^1 f(x)^2\,dx &= \left(\frac{1}{2}\right)^2 + \left(\frac{\sqrt{2}}{\pi}\right)^2 + \left(\frac{\sqrt{2}}{3\pi}\right)^2 + \left(\frac{\sqrt{2}}{5\pi}\right)^2 + \cdots \\
\frac{1}{2} &= \frac{1}{4} + \frac{2}{\pi^2} + \frac{2}{3^2\pi^2} + \frac{2}{5^2\pi^2} + \cdots \\
\frac{1}{4} &= \frac{2}{\pi^2} + \frac{2}{3^2\pi^2} + \frac{2}{5^2\pi^2} + \cdots \\
\frac{1}{4} &= \frac{2}{\pi^2}\left(\frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots\right) \\
\frac{\pi^2}{8} &= \frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \cdots
\end{aligned}$$

That's weird.[11]

1.4 Inner Product Spaces over $\mathbb{C}$

Now seems like a good time to bring in complex numbers. I have a beef with the American educational system, in that there is no course that reliably introduces complex numbers. The system is able to sleep at night because complex numbers are in the pre-Calculus curriculum, but the treatment is inadequate, and most math majors don't take pre-Calculus. Indeed, I believe it possible for a student to graduate with a math major having never seen a good

[11] This series is related to the famous *Basel problem*. It is easy to see that the infinite series $1/1^2 + 1/2^2 + 1/3^2 + \cdots$ converges, but is not at all clear how to find a formula for the sum. This problem was posed by Pietro Mengoli 1650 and finally solved by Leonhard Euler in 1734, who showed that the limit is exactly $\pi^2/6$. The appearance of π in the answer was a big surprise.

introduction to complex numbers. As is traditional, I will give a quick review and pretend that you have seen this before, even if you haven't.

Complex Numbers. The complex numbers are defined as

$$\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\},$$

where i is an abstract symbol satisfying $i^2 = -1$. Given a complex number $\alpha = a + ib$, we define its *absolute value* and *complex conjugate*:[12]

$$|\alpha| := \sqrt{a^2 + b^2},$$
$$\alpha^* := a - ib.$$

These satisfying the following properties.

(7) Properties of Complex Numbers. For all $a, b \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{C}$ we have

(a) $(a\alpha + b\beta)^* = a\alpha^* + b\beta^*$

(b) $(\alpha\beta)^* = \alpha^*\beta^*$

(c) $\alpha = \alpha^* \iff \alpha \in \mathbb{R}$

(d) $|\alpha| \geq 0$ with $|\alpha| = 0$ if and only if $\alpha = 0$.

(d) $|\alpha| = \alpha^*\alpha$

(e) $|\alpha\beta| = |\alpha||\beta|$.

(f) If $\alpha \neq 0$ then $\alpha^{-1} = \alpha^*/|\alpha|^2$.

Many applications of linear algebra use complex instead of real scalars. Almost all of the axioms are the same, but there is a key change in the definition of inner product.

(8) Axioms of Hermitian Inner Products. Let V be a vector space over $\mathbb{C}$, together with an algebraic operation

$$\mathbf{u}, \mathbf{v} \in V \rightsquigarrow \langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{C}.$$

We call this a *Hermitian inner product* if it satisfies the following axioms:

(a) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle^*$

(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$[13]

(c) For all $\alpha \in \mathbb{C}$ we have $\langle \mathbf{u}, \alpha\mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle$.[14]

[12] I will use α^* instead of the traditional $\overline{\alpha}$ to avoid conflict with the whiteboard notation for vectors: $\vec{v}$.
[13] By combining (8ab) we also have $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v} + \mathbf{w} \rangle$.
[14] By combining (8ac) we also have $\langle \alpha\mathbf{u}, \mathbf{v} \rangle = \alpha^* \langle \mathbf{u}, \mathbf{v} \rangle$.

(d) For all $\mathbf{v} \in V$, part (a) tells us that $\langle \mathbf{v}, \mathbf{v} \rangle \in \mathbb{R}$. Furthermore, we must have $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ with $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Jargon: A Hermitian inner product is sometimes called *sesquilinear* (one and a half times linear) because it is linear in the second coordinate:

$$\begin{aligned} \langle \mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w} \rangle &= \alpha\langle \mathbf{u}, \mathbf{v} \rangle + \beta\langle \mathbf{u}, \mathbf{w} \rangle, \\ \langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle &= \alpha^*\langle \mathbf{u}, \mathbf{w} \rangle + \beta^*\langle \mathbf{v}, \mathbf{w} \rangle. \end{aligned}$$

Beware, some books switch these.

Example: The standard Hermitian product on $\mathbb{C}^n$. Consider the set

$$\mathbb{C}^n = \{\mathbf{v} = (v_1, \ldots, v_n) : v_1, \ldots, v_n \in \mathbb{C}\}.$$

This is naturally a vector space over $\mathbb{C}$ with the usual operations of addition and scalar multiplication. We can still define the usual dot product

$$\mathbf{u} \bullet \mathbf{v} = u_1 v_1 + \cdots + u_n v_n,$$

but this turns out to have bad properties. For example, we might have $\mathbf{v} \bullet \mathbf{v} < 0$, as with the vector $\mathbf{v} = (i, i)$. To fix this, we instead consider the following operation:

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1^* v_1 + \cdots + u_n^* v_n.$$

One can check that this satisfies the axioms of a Hermitian inner product. Most importantly, we have $\langle \mathbf{v}, \mathbf{v} \rangle = v_1^* v_1 + \cdots + v_n^* v_n = |v_1|^2 + \cdots + |v_n|^2 \geq 0$, with $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$, which allows us to define a norm and a metric:

$$\begin{aligned} \|\mathbf{v}\| &= \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}, \\ \operatorname{dist}(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\|. \end{aligned}$$

Quantum mechanics is the big reason for using complex Hermitian spaces, but the complex numbers also allow us to simplify some classical problems.

Example: Complex Fourier Series. Recall Euler's identities:

$$\begin{aligned} e^{i\theta} &= \cos\theta + i\sin\theta, \\ e^{-i\theta} &= \cos\theta - i\sin\theta, \\ \cos\theta &= (e^{i\theta} + e^{-i\theta})/2, \\ \sin\theta &= (e^{i\theta} - e^{-i\theta})/(2i). \end{aligned}$$

Suppose that we have a *real Fourier series*[15]

$$f(x) = a_0 + \sum_{n \geq 1} a_n \sin(2\pi n x) + \sum_{n \geq 1} b_n \cos(2\pi n x).$$

[15] I'll absorb the $\sqrt{2}$ factors into the coefficients this time.

We can write this as a *complex Fourier series*

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi nx}$$

by defining the complex coefficients:

$$c_n := \begin{cases} a_0 & n = 0, \\ (b_n - ia_n)/2 & n \geq 1, \\ (b_{-n} + ia_{-n})/2 & n \leq -1. \end{cases}$$

Why would we do this? Because the functions $e^{i2\imath nx}$ are easier to work with! Let's define the *complex L^2 space* $L^2[0,1]$ as the set of functions $[0,1] \to \mathbb{C}$ satisfying

$$\int_0^1 |f(x)|^2 \, dx < \infty.$$

This space has a standard Hermitian inner product:

$$\langle f(x), g(x) \rangle = \int_0^1 f(x)^* g(x) \, dx.$$

And the functions $e^{i2\pi nx}$ for $n \in \mathbb{Z}$ are an orthonormal set:

$$\begin{aligned} \langle e^{i2\pi mx}, e^{i2\pi nx} \rangle &= \int_0^1 (e^{i2\pi mx})^* e^{2\pi inx} \, dx \\ &= \int_0^1 e^{-i2\pi mx} e^{2\pi inx} \, dx \\ &= \int_0^1 e^{i2\pi(n-m)x} \, dx. \end{aligned}$$

If $m = n$ then this gives

$$\langle e^{i2\pi nx}, e^{i2\pi nx} \rangle = \int_0^1 1 \, dx = 1,$$

and if $m \neq n$ then we get

$$\begin{aligned} \langle e^{i2\pi mx}, e^{i2\pi nx} \rangle &= \int_0^1 e^{i2\pi(n-m)x} \, dx \\ &= \frac{1}{i2\pi(n-m)} \left[e^{i2\pi(n-m)x} \right]_0^1 \\ &= \frac{1}{i2\pi(n-m)} \left[1 - 1 \right] \\ &= 0. \end{aligned}$$

Note: That was much easier than messing around with trigonometric identities. It also means that we have a single formula for the complex Fourier coefficients:[16]

$$c_n = \langle e^{i2\pi nx}, f(x) \rangle = \int_0^1 e^{-i2\pi nx} f(x)\, dx.$$

Then we can convert back to real coefficients if desired.

Fourier Transform. For the physicists among you, I should mention what happens for functions on the whole real line. Let $L^2(\mathbb{R})$ denote the set of functions $f : \mathbb{R} \to \mathbb{C}$ that are square integrable:

$$\int_{-\infty}^{\infty} |f(x)|^2\, dx < \infty.$$

As with $L^2[0,1]$, this is a Hermitian space with Hermitian product

$$\langle f(x), g(x) \rangle = \int_{-\infty}^{\infty} f(x)^* g(x)\, dx.$$

This space is more complicated than $L^2[0,1]$ because it does not have a countable basis.[17] However, the situation is not hopeless because we can generalize the *Fourier series* to the *Fourier transform*:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi nx} \quad \rightsquigarrow \quad f(x) = \int_{-\infty}^{\infty} c(\omega) e^{i2\pi\omega x}\, d\omega.$$

We can view the function $c : \mathbb{R} \to \mathbb{C}$ as a generalization of the sequence of coefficients c_n for $n \in \mathbb{Z}$. This function $c(\omega)$ is called the *Fourier transform* of $f(x)$ and it is sometimes denoted $\hat{f}(\omega)$. In some sense we can view the set

$$\{e^{i2\pi\omega x} : \omega \in \mathbb{R}\}$$

as an uncountably infinite basis for the space $L^2(\mathbb{R})$. There is just one issue; the functions $e^{i2\pi\omega x}$ are not square integrable:

$$\int_{-\infty}^{\infty} |e^{i2\pi\omega x}|^2\, dx = \int_{-\infty}^{\infty} 1\, dx = \infty.$$

This is a typical problem in physics. It can be surmounted by generalizing the concept of function to that of "distribution", but the rigorous mathematical definitions make the subject less understandable. Dirac showed that the intuitive point of view is a powerful tool for studying quantum mechanics.

[16] Taking the inner product in the other direction gives $\langle f(x), e^{i2\pi nx} \rangle = c_n^*$.

[17] The issue is that $[0,1]$ is a compact infinite set, while $\mathbb{R}$ is not compact.

2 Matrices

2.1 Matrix Multiplication

In the last section we talked about individual vector spaces such as $\mathbb{R}^n$ and $L^2[0,1]$. Each of these has an inner product, hence it also has a vector norm and a metric. Now we discuss linear functions between different vector spaces. In the finite dimensional case we can encode such functions as matrices. But matrix arithmetic does more than just encode linear functions; it is an extremely powerful language that gives out much more than we put in.

I assume you already know the definition of matrix multiplication. Here is a reminder.

Definition of Matrix Multiplication. Consider two matrices

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{\ell 1} & \cdots & a_{\ell m} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}.$$

We say that A has shape $\ell \times m$ and B has shape $m \times n$. (The number of rows comes first.) Since the number of columns of A equals the number of rows of B (they both equal m), we can define the product matrix AB, which has shape $\ell \times n$:

$$AB = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{\ell 1} & \cdots & c_{\ell n} \end{pmatrix}.$$

The entries of A, B and AB are related as follows:

$$\boxed{c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.}$$

I could have postponed this gory definition until it emerged naturally from the theory. But, as I said, the mechanics of matrix arithmetic is more than the sum of its parts, so I wanted to explore the mechanics first.

Row Times Column = Dot Product. Suppose that $\ell = n = 1$, so that

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} \\ \vdots \\ b_{m1} \end{pmatrix}.$$

Then the matrix product AB has shape 1×1 (it is just a scalar) and corresponds to the dot product of vectors:

$$\begin{pmatrix} a_{11} & \cdots & a_{1m} \end{pmatrix} \begin{pmatrix} b_{11} \\ \vdots \\ b_{m1} \end{pmatrix} = a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1m}b_{m1}.$$

From now on we will identify vectors $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ with **column vectors**:

$$\mathbf{v} = (v_1, \ldots, v_n) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

To talk about **row vectors** we will use the operation of *transposition*:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{\ell 1} & \cdots & a_{\ell m} \end{pmatrix} \quad \rightsquigarrow \quad A^T := \begin{pmatrix} a_{11} & \cdots & a_{\ell 1} \\ \vdots & & \vdots \\ a_{1m} & \cdots & a_{\ell m} \end{pmatrix}.$$

Thus the transpose of a column vector is a row vector:

$$\mathbf{v}^T = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}^T = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}.$$

Finally, we can express the dot product of any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ in terms of matrix multiplication:

$$\mathbf{u}^T\mathbf{v} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} u_1v_1 + \cdots + u_nv_n = \mathbf{u} \bullet \mathbf{v}.$$

Column Times Row = Something Else. Warning. A column times a row is not a scalar; it is a matrix of any shape that we want. That is, for any $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$ we obtain an $m \times n$ matrix[18]

$$\mathbf{u}\mathbf{v}^T = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1v_1 & \cdots & u_1v_n \\ \vdots & & \vdots \\ u_mv_1 & \cdots & u_mv_n \end{pmatrix}.$$

Row times column and column times row are the two basic examples. In between there are many different ways to think about matrix multiplication. For example:

$$\begin{aligned} (ij \text{ entry of } AB) &= (i\text{th row of } A)(j\text{th col of } B), \\ (i\text{th row of } AB) &= (i\text{th row of } A)B, \\ (j\text{th col of } AB) &= A(j\text{th col of } B). \end{aligned}$$

If A has shape $\ell \times m$ and B has shape $m \times n$ then we also have

$$AB = \sum_{k=1}^{m} (k\text{th col of } A)(k\text{th row of } B),$$

[18] Later we will call these *rank one matrices*.

where the right hand side is a sum of m matrices, each of shape $\ell \times n$.

All of these rules are examples of a very general recursive property of matrix multiplication.

Theorem: Block Multiplication. Suppose that we partition two matrices into submatrices by inserting vertical and horizontal lines:

$$A = \left(\begin{array}{c|c|c} A_{11} & \cdots & A_{1m} \\ \hline \vdots & & \vdots \\ \hline A_{\ell 1} & \cdots & A_{\ell m} \end{array}\right) \quad \text{and} \quad B = \left(\begin{array}{c|c|c} B_{11} & \cdots & B_{1n} \\ \hline \vdots & & \vdots \\ \hline B_{m1} & \cdots & B_{mn}. \end{array}\right)$$

Let's say that each submatrix A_{ij} has shape $\lambda_i \times \mu_j$ and each B_{ij} has shape $\mu_i \times \nu_j$, so

$$\begin{aligned} \#(\text{rows of } A) &= \lambda_1 + \cdots + \lambda_\ell, \\ \#(\text{cols of } A) &= \mu_1 + \cdots + \mu_m, \\ \#(\text{rows of } B) &= \mu_1 + \cdots + \mu_m, \\ \#(\text{cols of } B) &= \nu_1 + \cdots + \nu_n. \end{aligned}$$

Then I claim that that the product matrix AB can be partitioned as

$$AB = \left(\begin{array}{c|c|c} C_{11} & \cdots & C_{1n} \\ \hline \vdots & & \vdots \\ \hline C_{\ell 1} & \cdots & C_{\ell n} \end{array}\right),$$

where the submatrix C_{ij} is given by

$$\boxed{C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}.}$$

Note that $\#(\text{cols of } A_{ik}) = \mu_k = \#(\text{rows of } B_{kj})$ so that each matrix product $A_{ik}B_{kj}$ is defined and has shape $\ell_i \times \nu_j$. Thus C_{ij} is a sum of m matrices, each of shape $\lambda_i \times \nu_j$. In particular, C_{ij} has shape $\lambda_i \times \mu_j$. Note that the standard formula for unpartitioned matrices corresponds to the case when each submatrix A_{ij} and B_{ij} has size 1×1.

I won't prove right this now because the notation is too hairy.[19] Instead let's see some examples illustrating the few rules that we stated above. Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

[19] Later it will follow easily from properties of linear functions between direct sums of vector spaces.

Multiplying rows of A by columns of B gives

$$\left(\begin{array}{ccc} 1 & 1 & 1 \\ \hline 1 & 2 & 3 \end{array}\right)\left(\begin{array}{c|c} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{array}\right) = \left(\begin{array}{c|c} \begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix} & \begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix} \\ \hline \begin{pmatrix}1&2&3\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix} & \begin{pmatrix}1&2&3\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix} \end{array}\right) = \left(\begin{array}{c|c} 2 & 2 \\ \hline 3 & 5 \end{array}\right).$$

Multiplying rows of A by B gives

$$\left(\begin{array}{ccc} 1 & 1 & 1 \\ \hline 1 & 2 & 3 \end{array}\right)\begin{pmatrix}1&0\\1&1\\0&1\end{pmatrix} = \left(\begin{array}{c} \begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}1&0\\1&1\\0&1\end{pmatrix} \\ \hline \begin{pmatrix}1&2&3\end{pmatrix}\begin{pmatrix}1&0\\1&1\\0&1\end{pmatrix} \end{array}\right) = \left(\begin{array}{cc} 2 & 2 \\ \hline 3 & 5 \end{array}\right).$$

Multiplying A by columns of B gives

$$\begin{pmatrix}1&1&1\\1&2&3\end{pmatrix}\left(\begin{array}{c|c} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{array}\right) = \left(\begin{array}{c|c} \begin{pmatrix}1&1&1\\1&2&3\end{pmatrix}\begin{pmatrix}1\\1\\0\end{pmatrix} & \begin{pmatrix}1&1&1\\1&2&3\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix} \end{array}\right) = \left(\begin{array}{c|c} 2 & 2 \\ 3 & 5 \end{array}\right).$$

Finally, multiplying columns of A by rows of B gives

$$\begin{aligned}\left(\begin{array}{c|c|c} 1 & 1 & 1 \\ 1 & 2 & 3 \end{array}\right)\left(\begin{array}{cc} 1 & 0 \\ \hline 1 & 1 \\ \hline 0 & 1 \end{array}\right) &= \begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}1&0\end{pmatrix} + \begin{pmatrix}1\\2\end{pmatrix}\begin{pmatrix}1&1\end{pmatrix} + \begin{pmatrix}1\\3\end{pmatrix}\begin{pmatrix}0&1\end{pmatrix} \\ &= \begin{pmatrix}1&0\\1&0\end{pmatrix} + \begin{pmatrix}1&1\\2&2\end{pmatrix} + \begin{pmatrix}0&1\\0&3\end{pmatrix} \\ &= \begin{pmatrix}2&2\\3&5\end{pmatrix}.\end{aligned}$$

Each of these kinds of multiplication is useful for a different purpose. It is important to know them all.

2.2 Linear Functions

The ultimate goal of matrices is to hide all of the details of matrix arithmetic behind uppercase Roman letters. This lets us ignore irrelevant details to focus on higher level structure. The magic property that makes this work is the *associative property of matrix multiplication.*

Magic: Associativity of Matrix Multiplication. Consider matrices A, B, C of sizes $\ell \times m$, $m \times n$ and $n \times p$, respectively. Then the matrices AB, BC, $A(BC)$ and $(AB)C$ are defined, and we have

$$A(BC) = (AB)C.$$

This is not at all obvious from the definitions given above. A brute force proof is possible, but not enlightening. There is a much more conceptual explanation.

Definition of Linear Functions. Consider vector spaces V and W over $\mathbb{R}$ (or $\mathbb{C}$). A function $T : V \to W$ is called *linear* when it satisfies the following three properties:

- $T(\mathbf{0}) = \mathbf{0}$,
- $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$,
- $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$.

In other words, a linear function preserves the vector space operations of addition and scalar multiplication. We can also summarize these properties in one step by saying that T preserves *linear combinations*:

$$T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_n T(\mathbf{v}_n).$$

Why? Many natural operations are linear:

- Differention between suitable spaces of functions is linear.
- Integration from a suitable space of functions to $\mathbb{R}$ is linear.
- An inner product $\langle -, - \rangle$ on V over $\mathbb{R}$ is *bilinear*. That is, for any $\mathbf{v} \in V$, each of the following two functions is linear:

$$\langle \mathbf{v}, - \rangle : V \to \mathbb{R} \quad \text{and} \quad \langle -, \mathbf{v} \rangle : V \to \mathbb{R}$$

- A Hermitian inner product $\langle -, - \rangle$ on V over $\mathbb{C}$ is *sesquilinear* (one and a half times linear). This means that for each fixed $\mathbf{v} \in V$, the function $V \to \mathbb{C}$ defined by $\mathbf{u} \mapsto \langle \mathbf{v}, \mathbf{u} \rangle$ is linear, but the function $V \to \mathbb{C}$ defined by $\mathbf{u} \mapsto \langle \mathbf{u}, \mathbf{v} \rangle$ is *conjugate linear*:

$$\langle \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n, \mathbf{v} \rangle = \alpha_1^* \langle \mathbf{u}_1, \mathbf{v} \rangle + \cdots + \alpha_n^* \langle \mathbf{u}_n, \mathbf{v} \rangle.$$

If V and W are finite dimensional with $\dim V = n$ and $\dim W = m$, then choosing bases turns linear transformations $T : V \to W$ into $m \times n$ matrices. To keep things simple, for now we will work with Euclidean space and standard bases. Here is the big idea:

$$\begin{array}{c}\text{linear functions} \\ T : \mathbb{R}^n \to \mathbb{R}^m\end{array} \qquad \leftrightsquigarrow \qquad m \times n \text{ matrices}$$

The correspondence is easy to describe. First of all, let A be an $m \times n$ matrix over $\mathbb{R}$. This defines a function $\mathbb{R}^n \to \mathbb{R}^m$ by multiplying column vectors on the left:

$$\mathbf{v} \in \mathbb{R}^n \quad \rightsquigarrow \quad A\mathbf{v} \in \mathbb{R}^m.$$

Indeed, if $\mathbf{v}$ has shape $n \times 1$ then the product matrix $A\mathbf{v}$ is defined and has shape $m \times 1$. It is straightforward to check that this function is linear:

$$A(\alpha_1 \mathbf{v}_+ \cdots + \alpha_n \mathbf{v}_1) = \alpha_1 A\mathbf{v}_1 + \cdots + \alpha_n A\mathbf{v}_n.$$

Conversely, let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear function. In order to create an $m \times n$ matrix from T we consider the n standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$. Following our convention we will think of these as column vectors:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Now each basis vector $\mathbf{e}_i \in \mathbb{R}^n$ gets sent by T to a column vector $T(\mathbf{e}_i)$ in $\mathbb{R}^m$. We will record the n column vectors $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n) \in \mathbb{R}^m$ as the columns of an $m \times n$ matrix:

$$[T] := \begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix}.$$

Thus the linear function $T : \mathbb{R}^n \to \mathbb{R}^m$ becomes an $m \times n$ matrix $[T]$. Furthermore, the linear function defined by the matrix $[T]$ is the same as the linear function T. To see this, we consider any vector $\mathbf{v} \in \mathbb{R}^n$:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + v_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_n \mathbf{e}_n.$$

Then from the definition of $[T]$ and the linearity of T we have

$$\begin{aligned} T(\mathbf{v}) &= T(v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \cdots + v_n \mathbf{e}_n) \\ &= v_1 T(\mathbf{e}_1) + v_2 T(\mathbf{e}_2) + \cdots + v_n T(\mathbf{e}_n) \\ &= \sum_j v_j T(\mathbf{e}_j) \\ &= \sum_j v_j (j\text{th col of } [T]) \\ &= [T]\mathbf{v}, \end{aligned}$$

where the last expression $[T]\mathbf{v}$ is a matrix product. To summarize: To each linear function $T : \mathbb{R}^n \to \mathbb{R}^n$ we associate an $m \times n$ matrix $[T]$ with the property that

$$\underbrace{T(\mathbf{v})}_{\text{apply the function}} \quad = \quad \underbrace{[T]\mathbf{v}}_{\text{matrix multiplication}}$$

So far this is slightly interesting. It becomes very interesting when we consider functional composition. Suppose we have linear functions $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^\ell$:

Observe that the composite function $S \circ T : \mathbb{R}^n \to \mathbb{R}^\ell$ is also linear:

$$\begin{aligned}(S \circ T)\left(\sum a_i \mathbf{v}_i\right) &= S\left(T\left(\sum a_i \mathbf{v}_i\right)\right) \\ &= S\left(\sum a_i T(\mathbf{v}_i)\right) \\ &= \sum a_i S\left(T(\mathbf{v}_i)\right) \\ &= \sum a_i (S \circ T)(\mathbf{v}_i).\end{aligned}$$

Hence the function $S \circ T : \mathbb{R}^n \to \mathbb{R}^\ell$ corresponds to an $\ell \times n$ matrix $[S \circ T]$. Now we have three matrices:

$$\begin{aligned}[S] &\text{ has shape } \ell \times m, \\ [T] &\text{ has shape } m \times n, \\ [S \circ T] &\text{ has shape } \ell \times n.\end{aligned}$$

The following theorem is the ultimate reason for the concept of matrix multiplication. This theorem could also be taken as the **definition** of matrix multiplication.

Matrix Multiplication = Composition of Linear Functions. For any linear functions $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^\ell$, the composite $S \circ T : \mathbb{R}^n \to \mathbb{R}^\ell$ is also linear, and we have

$$[S \circ T] = [S][T].$$

Proof. The proof will use the following rule of matrix multiplication:

$$(j\text{th col of } AB) = A(j\text{th col of } B).$$

Our goal is to show that $[S \circ T]$ and $[S][T]$ have the same columns. From the definition of the matrix $[S \circ T]$ we have

$$(j\text{th col of } [S \circ T]) = (S \circ T)(\mathbf{e}_j) = S(T(\mathbf{e}_j)).$$

On the other hand, from the above property of matrix multiplication we have

$$(j\text{th col of } [S][T]) = [S](j\text{th col of } [T]) = [S]T(\mathbf{e}_j) = S(T(\mathbf{e}_j)).$$

□

Remark: It is worth meditating on this proof. When you understand it then you can say that you really understand the concept of matrix multiplication.

Before moving to some examples, we pause to give the *correct* (conceptual) proof that matrix multiplication is associative.

Proof of Associativity. Consider matrices A, B, C of shapes $\ell \times m$, $m \times n$ and $n \times p$, respectively. We can use these to define linear functions $R : \mathbb{R}^m \to \mathbb{R}^\ell$, $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^p \to \mathbb{R}^n$ by matrix multiplication:

$$\begin{aligned} R(\mathbf{v}) &:= A\mathbf{v} \text{ for } \mathbf{v} \in \mathbb{R}^\ell, \\ S(\mathbf{v}) &:= B\mathbf{v} \text{ for } \mathbf{v} \in \mathbb{R}^n, \\ T(\mathbf{v}) &:= C\mathbf{v} \text{ for } \mathbf{v} \in \mathbb{R}^m. \end{aligned}$$

Then, of course, the corresponding matrices are $[R] = A$, $[S] = B$ and $[T] = C$. Here is a picture of the functions:

Recall that composition of functions is naturally associative. That is, for any $\mathbf{v} \in \mathbb{R}^p$ we have

$$(R \circ (S \circ T))(\mathbf{v}) = R(S(T(\mathbf{v}))) = ((R \circ S) \circ T)(\mathbf{v}),$$

which means that $R \circ (S \circ T) = (R \circ S) \circ T$ as functions $\mathbb{R}^p \to \mathbb{R}^\ell$. Then the previous theorem tells us that

$$\begin{aligned} A(BC) &= [R]([S][T]) \\ &= [R][S \circ T] \\ &= [R \circ (S \circ T)] \\ &= [(R \circ S) \circ T] \\ &= [R \circ S][T] \\ &= ([R][S])[T] \\ &= (AB)C. \end{aligned}$$

Note that we never had to mention the entries of the matrices. Magic! □

2.3 Matrix Arithmetic

Let's zoom out again. One of the strengths of matrix notation is that we can sometimes solve a problem purely symbolically, without mentioning the entries of the matrices. In fact, by hiding the appropriate details we can sometimes turn a difficult problem into an almost trivial matrix computation.

Here is the context for matrix arithmetic.

Vector Spaces of Matrices. Let $\mathbb{R}^{m\times n}$ denote the set of $m\times n$ with real entries. (We define $\mathbb{C}^{m\times n}$ similarly.) By convention we will write

$$\mathbb{R}^n = \mathbb{R}^{n\times 1} = \text{ the set of } n\times 1 \text{ column vectors.}$$

Matrices can be added and multiplied by scalars in an obvious way. That is, given $m\times n$ matrices $A, B \in \mathbb{R}^{m\times n}$ and a scalar $\alpha\in\mathbb{R}$ we define $m\times n$ matrices $A+B$ and αA such that

$$\begin{aligned}(ij \text{ entry of } A+B) &= (ij \text{ entry of } A) + (ij \text{ entry of } B),\\ (ij \text{ entry of } \alpha A) &= \alpha(ij \text{ entry of } A).\end{aligned}$$

It is easy to check that these operations make $\mathbb{R}^{m\times n}$ into a vector space over $\mathbb{R}$. Furthermore, there is a *standard basis* of matrices E_{ij} with $1\le i\le m$ and $1\le j\le n$, with the entry 1 in the ij position and all other entries equal to zero:

$$B_{ij} = \begin{array}{c} \\ i \\ \\ \end{array}\!\!\begin{array}{c} \begin{array}{c} j \end{array} \\ \begin{pmatrix} \vdots \\ \cdots\ 1 \\ \\ \end{pmatrix} \end{array}$$

(When a matrix contains many zero entries we will simply leave them blank.) Since there mn such basis matrices it follows that

$$\dim \mathbb{R}^{m\times n} = mn.$$

In addition to the vector space structure, we have two additional operations on matrices. First we have *transposition* and *conjugate transposition*:

$$\begin{array}{ccc} \mathbb{R}^{m\times n} & \to & \mathbb{R}^{n\times m} \\ A & \mapsto & A^T \end{array}$$

and

$$\begin{array}{ccc} \mathbb{C}^{m\times n} & \to & \mathbb{C}^{n\times m} \\ A & \mapsto & A^* \end{array}$$

Second we have the all-important operation of matrix multiplication:

$$\begin{array}{ccc} \mathbb{R}^{\ell\times m}\times\mathbb{R}^{m\times n} & \to & \mathbb{R}^{\ell\times n} \\ (A,B) & \mapsto & AB. \end{array}$$

Finally, we have two special classes of matrices. For any shape $m \times n$ we have a *zero matrix*:

$$O_{m\times n} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}.$$

[Note: I use the letter O for zero matrices.] And for any n we have a square *identity matrix*:

$$I_n = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}.$$

This identity matrix corresponds to the *identity function* $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$, which sends each vector to itself. Indeed, for any linear function $T : \mathbb{R}^n \to \mathbb{R}^n$ recall that the ith column of the corresponding matrix $[T]$ is $T(\mathbf{e}_i)$. Since the ith column of $[\mathrm{id}]$ is $\mathrm{id}(\mathbf{e}_i) = \mathbf{e}_i$ we have $[\mathrm{id}] = I_n$.

Rules of Matrix Arithmetic. The operations of matrix arithmetic satisfy the following abstract rules. Here uppercase Roman letters represent matrices and lowercase Greek letters are scalars. Assume that the matrices have appropriate shape so the indicated matrix sums and products exist.

- **Vector Space Rules.**

$$\begin{aligned} A + B &= B + A, \\ A + (B + C) &= (A + B) + C, \\ A + O &= O + A = A, \\ 1A &= A, \\ 0A &= O, \\ \alpha(\beta A) &= (\alpha\beta)A, \\ (\alpha + \beta)A &= \alpha A + \beta A, \\ \alpha(A + B) &= \alpha A + \alpha B. \end{aligned}$$

- **Multiplication is not Commutative.** In general we have

$$AB \neq BA,$$

  even when both matrices are defined and have the same shape.

- **Multiplication is Bilinear.**

$$\begin{aligned} A(\beta B + \gamma C) &= \beta AB + \gamma AC, \\ (\alpha A + \beta B)C &= \alpha AC + \beta BC. \end{aligned}$$

- **Multiplication by O and I.**

$$\begin{aligned} AO &= O, \\ OA &= O, \\ AI &= A, \\ IA &= A. \end{aligned}$$

- **Properties of Transpose and Conjugate Transpose.**

$$\begin{array}{rclcrcl} (A^T)^T & = & A, & & (A^*)^* & = & A, \\ (A+B)^T & = & A^T + B^T, & & (A+B)^* & = & A^* + B^* \\ (\alpha A)^T & = & \alpha A^T, & & (\alpha A)^* & = & \alpha^* A^*, \\ (AB)^T & = & B^T A^T, & & (AB)^* & = & B^* A^*. \end{array}$$

  Remark: If A is $\ell \times m$ and B is $m \times n$ then A^T is $m \times \ell$ and B^T is $n \times m$. The matrix $B^T A^T$ always exists and is equal to AB. In general, the matrix $A^T B^T$ does not exist.

In addition to arithmetic operations, we also need a way to measure the "size" of a matrix.

Axioms of Matrix Norms. Let $\| - \|$ be a function that assigns to each matrix A a real number $\|A\|$. We call this a *matrix norm* when it satisfies the following axioms:

(a) $\|A\| \geq 0$ for all A, with $\|A\| = 0$ if and only if $A = O$.

(b) $\|\alpha A\| = |\alpha| \|A\|$

(c) $\|A + B\| \leq \|A\| + \|B\|$

(d) $\|AB\| \leq \|A\| \|B\|$.

Here are the two main examples.

The Frobenius Norm. We define this by analogy with the standard vector norm:

$$\|A\|_F := \begin{cases} \sqrt{\sum_{i,j} a_{ij}^2} & \text{over } \mathbb{R}, \\ \sqrt{\sum_{i,j} |a_{ij}|^2} & \text{over } \mathbb{C}. \end{cases}$$

We observe that $\|\mathbf{v}\|_F = \|\mathbf{v}\|$ for all column vectors $\mathbf{v}$. The fact that $\| - \|_F$ satisfies (abc) follows from this vector case. You will prove that $\|AB\|_F \leq \|A\|_F \|B\|_F$ on the homework.

The L^2 Norm (Also Called the Operator Norm). The Frobenius norm only applies to matrices. The operator norm also applies to linear functions on infinite dimensional normed vector spaces:

$$\|A\|_2 := \max\{\|A\mathbf{u}\| : \text{over all unit vectors } \|\mathbf{u}\| = 1\}.$$

Since $\|A\mathbf{u}\| \geq 0$ for all $\mathbf{u}$ we have $\|A\|_2 \geq 0$. And if $\|A\|_2 = 0$ then we must have $\|A\mathbf{u}\| = 0$ (and hence $A\mathbf{u} = \mathbf{0}$) for all unit vectors $\mathbf{u}$. In particular, letting $\mathbf{u}$ range of the standard basis vectors we find that each column is A is a zero vector, hence $A = O$. This proves property (a). For property (b) we observe that $\|\alpha A\mathbf{u}\| = |\alpha|\|A\mathbf{u}\|$, hence the maximum value of $\|\alpha A\mathbf{u}\|$ is $|\alpha|$ times the maximum value of $\|A\mathbf{u}\|$. For part (c) we use the triangle inequality for vector norms to observe that[20]

$$\|(A+B)\mathbf{u}\| = \|A\mathbf{u} + B\mathbf{u}\| \leq \|A\mathbf{u}\| + \|B\mathbf{u}\| \text{ for all matrices } A, B \text{ and unit vectors } \mathbf{u}.$$

To prove part (d) we first show that $\|A\mathbf{v}\|_2 \leq \|A\|_2\|\mathbf{v}\|$ for any nonzero vector $\mathbf{v}$. Indeed if $\mathbf{v}$ is nonzero then $\mathbf{v}/\|\mathbf{v}\|$ is a unit vector and hence

$$\|A\|_2 = \max\{\|A\mathbf{u}\| : \text{all unit vectors } \mathbf{u}\} \geq \|A(\mathbf{v}/\|\mathbf{v}\|)\| = \|A\mathbf{v}\|/\|\mathbf{v}\|.$$

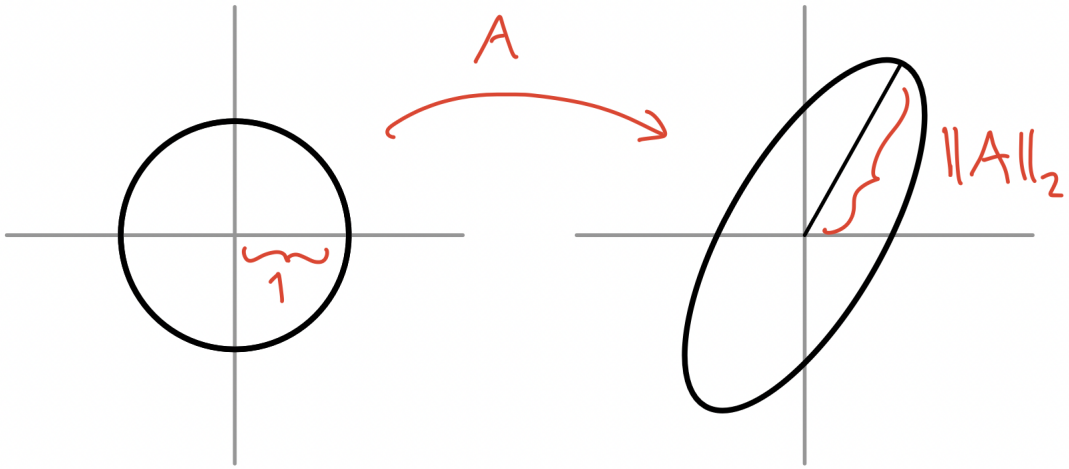
Finally, to show that $\|AB\|_2 \leq \|A\|_2\|B\|_2$, consider any unit vector $\mathbf{u}$. Note that $B\mathbf{u}$ is not necessarily a unit vector, but from the previous remark with $\mathbf{v} = B\mathbf{u}$ we still have[21]

$$\|(AB)\mathbf{u}\| = \|A(B\mathbf{u})\| \leq \|A\|_2\|B\mathbf{u}\| \leq \|A\|_2\|B\|_2.$$

It follows that

$$\|AB\|_2 = \max\{\|AB\mathbf{u}\| : \text{all unit vectors } \mathbf{u}\} \leq \|A\|_2\|B\|_2.$$

Here is a picture of the L^2 norm of a 2×2 matrix:



The matrix A sends the unit circle to an ellipse. The operator norm $\|A\|_2$ is the longest axis of the ellipse. More generally, the longest axis is called the first *singular value* σ_1 and the smaller axis is the second singular value σ_2. We will discuss the SVD (singular value decomposition) in a later section.

The Frobenius norm is harder to visualize.

[20] Details: The maximum value of $\|(A+B)\mathbf{u}\|$ is $\leq$ the maximum value of $\|A\mathbf{u}\| + \|B\mathbf{u}\|$ which is $\leq$ the sum of the maximum values of $\|A\mathbf{u}\|$ and $\|B\mathbf{u}\|$.

[21] If $B\mathbf{u} = \mathbf{0}$ then we have $\|AB\mathbf{u}\| = 0$ and there is nothing to show.

2.4 Inverse Matrices

We have seen how to multiply matrices, but can we also divide? If we can then this will be extremely useful for solving matrix equations. For example, suppose we have an equation

$$AX = B,$$

where A and B are given matrices and X is an unknown matrix. If we can find a matrix C such that $CA = I$ then multiplying both sides on the left by C gives

$$\begin{aligned} AX &= B \\ C(AX) &= CB \\ (CA)X &= CB \\ IX &= CB \\ X &= CB. \end{aligned}$$

Definition of Inverse Matrices. Let A be an $m \times n$ matrix. Any $n \times m$ matrix B satisfying

$$AB = I_m$$

is called a *right inverse* of A. Any $n \times m$ matrix C satisfying

$$CA = I_n$$

is called a *left inverse* of A. Left and right inverses, if they exist, need not be unique. However, suppose that A has **both** a right inverse B **and** a left inverse C. Then we must have

$$B = I_n B = (CA)B = C(AB) = CI_m = C.$$

In this case $B = C$ is the unique *two-sided inverse* of A, and we write

$$A^{-1} = B = C.$$

When A has a two-sided inverse we say that A is *invertible*. It will follow from the Fundamental Theorem below that **an invertible matrix must be square** (i.e., have $m = n$) but this theorem is surprisingly difficult to prove.

For example, consider the following non-square matrix:

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

If B is a right inverse of A then it must have two columns $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^3$ and it must satisfy the block matrix equation

$$I_2 = AB$$

$$\left(\begin{array}{c|c} 1 & 0 \\ 0 & 1 \end{array}\right) = A\left(\begin{array}{c|c} \mathbf{b}_1 & \mathbf{b}_2 \end{array}\right)$$
$$\left(\begin{array}{c|c} \mathbf{e}_1 & \mathbf{e}_2 \end{array}\right) = \left(\begin{array}{c|c} A\mathbf{b}_1 & A\mathbf{b}_2 \end{array}\right)$$

From Linear Algebra I you know how to solve the systems $A\mathbf{b}_1 = \mathbf{e}_1$ and $A\mathbf{b}_2 = \mathbf{e}_2$ to obtain all possible column vectors $\mathbf{b}_1, \mathbf{b}_2$. In turns out $\mathbf{b}_1 = (2+s, -1-2s, s)$ and $\mathbf{b}_2 = (-1+t, 1-2t, t)$ for any parameters s, t. Thus we obtain a two-dimensional family of right inverses:[22]

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 2+s & -1+t \\ -1-2s & 1-2t \\ s & t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This already tells us that A has **no left inverse**, since if it did then any two right inverses would be equal. Indeed, let B, B' be any two right inverses of A and suppose that A has a left inverse C. Then we get

$$\begin{aligned} I_2 &= I_2 \\ AB &= AB' \\ C(AB) &= C(AB') \\ (CA)B &= (CA)B' \\ I_3 B &= I_3 B' \\ B &= B'. \end{aligned}$$

Since our matrix A has many different right inverses, no left inverse can exist.

I mentioned above that any invertible matrix (i.e., any matrix with a two-sided inverse) must be square. It is also true that any left inverse of a given square matrix must also be a right inverse, and vice versa. I will state these theorems now, but the proofs are surprisingly subtle and are postponed until the next section.

Two Subtle Theorems.

- Any invertible matrix must be square.
- For any square matrices A and B of the same size, we have

$$AB = I \quad \Longleftrightarrow \quad BA = I.$$

To be concrete, consider the matrices

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and} \quad A' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}.$$

[22]Note: The family of right inverses of A is not a vector subspace of $\mathbb{R}^{3\times 2}$ because it does not contain the zero matrix. However, it is an *affine subspace* of $\mathbb{R}^{3\times 2}$, i.e., a translation of a linear subspace.

The matrix equation $AA' = I$ is equivalent to the following four equations:

$$\begin{cases} aa' + bc' &= 1, \\ ab' + bd' &= 0, \\ ca' + dc' &= 0, \\ cb' + dd' &= 1. \end{cases}$$

And the matrix equation $A'A = I$ is equivalent to the system

$$\begin{cases} a'a + b'c &= 1, \\ a'b + b'd &= 0, \\ c'a + d'c &= 0, \\ c'b + d'd &= 1. \end{cases}$$

The second theorem above tells us that these two systems of equations have the same solutions for the eight unknowns $a, b, c, d, a', b', c', d'$. It is tempting to look for a direct algebraic proof of this but you won't be able to find one because this is the wrong approach. The correct approach requires us to consider the dimensions of certain vector spaces associated to the matrices. See the Fundamental Theorem in the next section.

For now we will prove some easy and purely symbolic properties of inverse matrices.

Algebraic Properties of Inverse Matrices.

(a) Suppose that A^{-1} exists. Then $(A^*)^{-1}$ exists and is equal to $(A^{-1})^*$.

(b) Suppose that A^{-1}, B^{-1} and AB exist. Then $(AB)^{-1}$ exists and is equal to $B^{-1}A^{-1}$.

Proof. (a): We only need to show that $A^*(A^{-1})^* = I$ and $(A^{-1})^*A^* = I$. For the first identity we have[23]

$$A^*(A^{-1})^* = (A^{-1}A)^* = I^* = I.$$

The other direction is similar. (b): We only need to show that $(AB)(B^{-1}A^{-1}) = I$ and $(B^{-1}A^{-1})AB = I$. This follows easily from the associativity of matrix multiplication:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I.$$

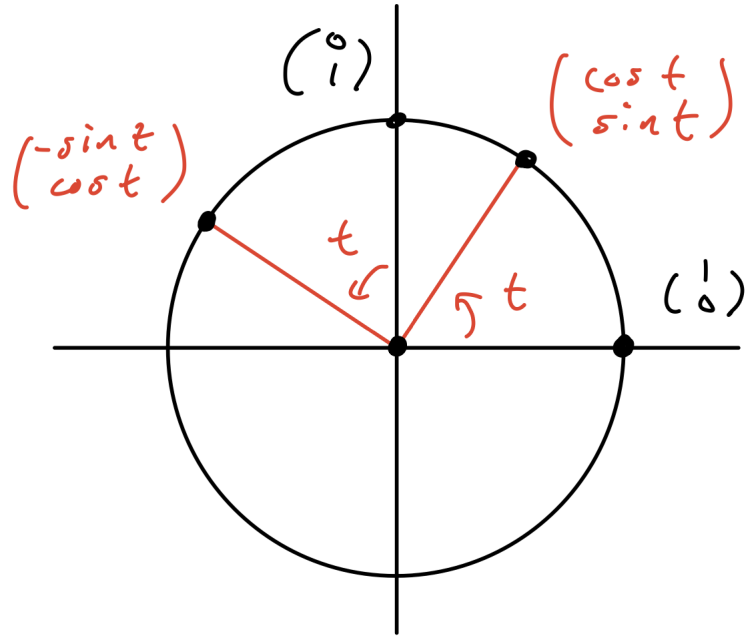
The other direction is similar. □

2.5 Important Kinds of Matrices

It is high time for some examples.

Rotations. Consider the function $R_t : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates each point by angle t, counter-clockwise around the origin. This function is linear because it sends the origin to itself and it sends parallelograms to parallelograms. To determine the corresponding matrix we only need to rotate the standard basis vectors:

[23] Recall that $A^*B^* = (BA)^*$.



Since no confusion will result, I will use the notation R_t for the function and for the corresponding matrix. Thus we have

$$R_t = \begin{pmatrix} R_t \begin{pmatrix} 1 \\ 0 \end{pmatrix} & R_t \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

Once we have the matrix we can use this to rotate a general point:

$$R_t \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos t - y\sin t \\ x\sin t + y\cos t \end{pmatrix}.$$

It would be much harder to solve this problem without the theory of matrices. Next we consider the composition of two rotations. Thinking in terms of functions, it is clear that $R_s R_t = R_{s+t} = R_t R_s$, since rotating first by one angle and then by the other angle is the same as rotating once by the sum of the two angles. On the other hand, since matrix multiplication is the same as functional composition, we obtain the following matrix identity, which is equivalent to the angle sum trigonometric identities:

$$\begin{pmatrix} \cos s & -\sin s \\ \sin s & \cos s \end{pmatrix} \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} = \begin{pmatrix} \cos(s+t) & -\sin(s+t) \\ \sin(s+t) & \cos(s+t) \end{pmatrix}.$$

Note that rotation clockwise by angle t is the same as rotation counterclockwise by angle $-t$. Thus the functions R_t and R_{-t} are inverses:

$$R_t R_{-t} = R_{-t} R_t = R_0 = I.$$

Note that rotation by angle zero is just the identity function. It is interesting to observe that

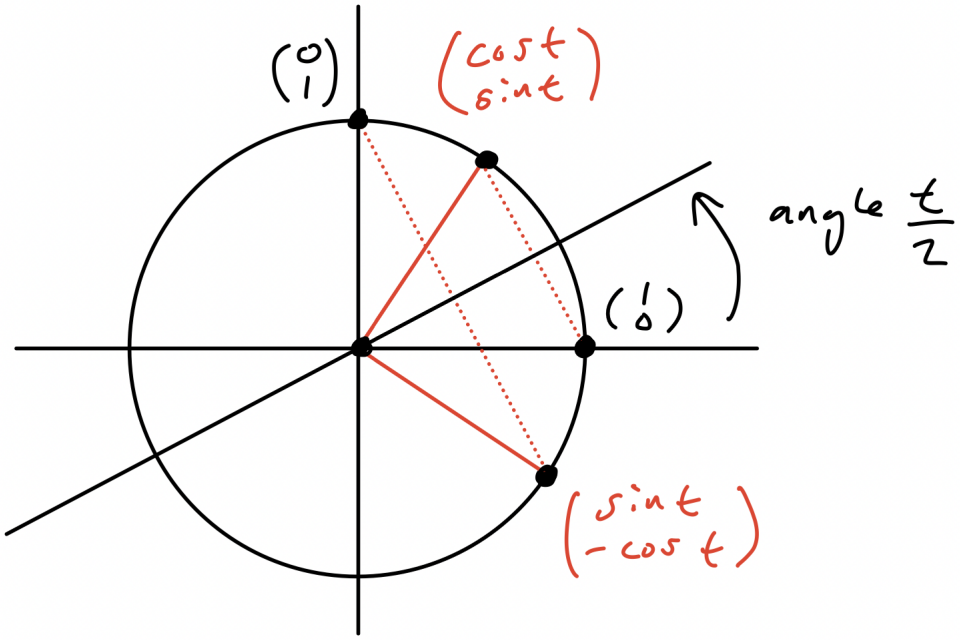
$$\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}^{-1} = (R_t)^{-1} = R_{-t} = \begin{pmatrix} \cos(-t) & -\sin(-t) \\ \sin(-t) & \cos(-t) \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} = (R_t)^T.$$

We will see below the matrices satisfying $A^{-1} = A^T$ are called *orthogonal matrices*. Finally, let me remark that the determinant of a rotation matrix is always 1:

$$\det(R_t) = \det \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} = \cos^2 t + \sin^2 t = 1.$$

We will discuss the general theory of determinants later.

Reflections. Let $F_t : \mathbb{R}^2 \to \mathbb{R}^2$ be the function that reflects each point across the line that makes angle $t/2$ from the positive x-axis. Again, this is a linear function because it sends the origin to itself and sends parallelograms to parallelograms. To determine the corresponding matrix we reflect the standard basis vectors:



Thus we obtain the matrix

$$F_t = \begin{pmatrix} F_t \begin{pmatrix} 1 \\ 0 \end{pmatrix} & F_t \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix}.$$

The composition of two reflections in two different lines turns out to be a rotation:

$$F_s F_t = \begin{pmatrix} \cos s & \sin s \\ \sin s & -\cos s \end{pmatrix} \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix}$$

$$
\begin{aligned}
&= \begin{pmatrix} \cos s \cos t + \sin s \sin t & \cos s \sin t - \sin s \cos t \\ \sin s \cos t - \cos s \sin t & \sin s \sin t + \cos s \cos t \end{pmatrix} \\
&= \begin{pmatrix} \cos(s-t) & -\sin(s-t) \\ \sin(s-t) & \cos(s-t) \end{pmatrix} \\
&= R_{s-t}.
\end{aligned}
$$

This would be more difficult to see geometrically. In particular, we find that reflection matrices do not commute in general:

$$F_s F_t = R_{s-t} \neq R_{t-s} = F_t F_s \quad \text{unless angles } s-t \text{ and } t-s \text{ are equal.}$$

Taking $s = t$ shows that the composition of a reflection with itself is the identity matrix:

$$F_t^2 = F_t F_t = R_{t-t} = R_0 = I.$$

In other words, reflecting in the same line twice is the same thing as doing nothing. This implies that each reflection matrix F_t is equal to its own inverse:

$$(F_t)^{-1} = F_t.$$

It also happens that $(F_t)^T = F_t$, so F_t is another example of an orthogonal matrix. Finally, let me remark that the determinant of any reflection matrix is -1:

$$\det(F_t) = \det \begin{pmatrix} \cos t & \sin t \\ \sin t & -\cos t \end{pmatrix} = -\cos^2 t - \sin^2 t = -1.$$

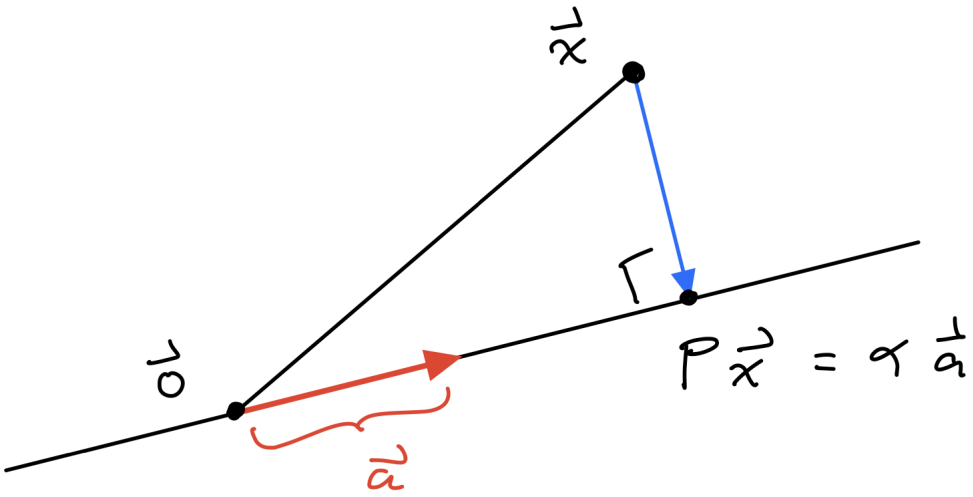
Projections. Consider the following matrix:

$$P_t = \begin{pmatrix} \cos^2 t & \cos t \sin t \\ \cos t \sin t & \sin^2 t \end{pmatrix}.$$

As with any 2×2 matrix, this defines a linear function $\mathbb{R}^2 \to \mathbb{R}^2$. What is the geometric description of this function? It is convenient to solve this problem in greater generality.

Suppose that we want to project[24] a point $\mathbf{x} \in \mathbb{R}^n$ onto the line in $\mathbb{R}^n$ generated by a vector $\mathbf{a}$:

[24]Here we are talking about *orthogonal projection*, i.e., projection at right angles. Later we will talk about more general kinds of projection.



Since projection is a linear function there will be some $n \times n$ matrix P that achieves this projection. We know exactly two things about this situation:

(1) Since the projection $P\mathbf{x}$ is on the line generated by $\mathbf{a}$ we must have $P\mathbf{x} = \alpha\mathbf{x}$ for some scalar α. This scalar will change depending on the point $\mathbf{x}$.

(2) Since the projection is orthogonal we know that the blue vector $P\mathbf{x} - \mathbf{x}$ is orthogonal to the red vector $\mathbf{a}$.

Putting these two facts together gives[25]

$$\begin{aligned} \mathbf{a}^T(P\mathbf{x} - \mathbf{x}) &= 0 && (2)\\ \mathbf{a}^T(\alpha\mathbf{a} - \mathbf{x}) &= 0 && (1)\\ \alpha\mathbf{a}^T\mathbf{a} - \mathbf{a}^T\mathbf{x} &= 0 \\ \alpha &= \mathbf{a}^T\mathbf{x}/\mathbf{a}^T\mathbf{a} \\ \alpha &= \mathbf{a}^T\mathbf{x}/\|\mathbf{a}\|^2. \end{aligned}$$

Hence the projection of $\mathbf{x}$ is given by

$$P\mathbf{x} = \alpha\mathbf{a} = \underbrace{\left(\frac{\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\right)}_{\text{scalar}} \underbrace{\mathbf{a}}_{\text{vector}}.$$

To find a formula for the $n \times n$ projection matrix P we simply rearrange using the fact that matrix multiplication is associative:[26]

$$P\mathbf{x} = \left(\frac{\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\right)\mathbf{a}$$

[25] I will express this using inner products because the ideas generalize beyond Euclidean space.

[26] The associativity of matrix multiplication is behind many clever proofs like this.

$$
\begin{aligned}
&= \mathbf{a}\left(\frac{\mathbf{a}^T\mathbf{x}}{\|\mathbf{a}\|^2}\right) && \text{scalars commute with matrices}\\
&= \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}(\mathbf{a}^T\mathbf{x})\\
&= \underbrace{\frac{1}{\|\mathbf{a}\|^2}(\mathbf{a}\mathbf{a}^T)}_{n\times n \text{ matrix}}\underbrace{\mathbf{x}}_{\text{vector}}.
\end{aligned}
$$

Since this identity holds for any vector $\mathbf{x}$[27] we conclude that the projection matrix is given by

$$P = \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}\mathbf{a}^T.$$

If $\mathbf{a} = \mathbf{u}$ is a unit vector then the formula is particularly simple:

$$P = \mathbf{u}\mathbf{u}^T = \text{the projection onto the line in } \mathbb{R}^n \text{ spanned by unit vector } \mathbf{u}.$$

Now we go back to two dimensions. Consider the line in $\mathbb{R}^2$ that makes angle t counterclockwise from the positive x-axis. This line is generated by the unit vector $\mathbf{u} = (\cos t, \sin t)$. Hence the matrix that projects onto this line is

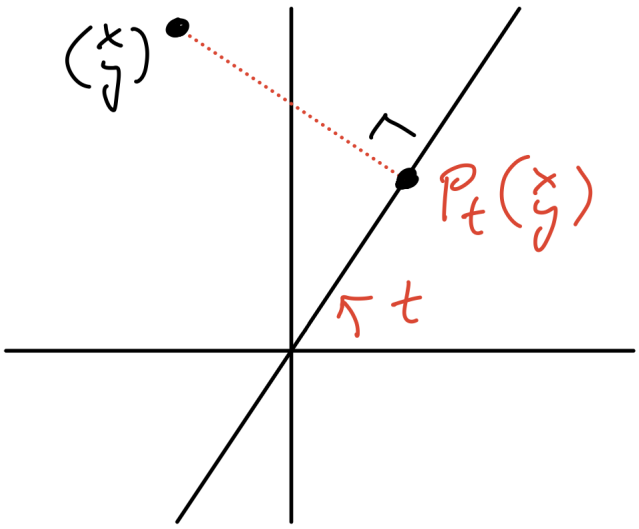
$$P_t = \mathbf{u}\mathbf{u}^T = \begin{pmatrix}\cos t\\ \sin t\end{pmatrix}\begin{pmatrix}\cos t & \sin t\end{pmatrix} = \begin{pmatrix}\cos^2 t & \cos t\sin t\\ \cos t\sin t & \sin^2 t\end{pmatrix}.$$

The image of a general point $\mathbf{x} = (x, y)$ under this projection is

$$P_t\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}\cos^2 t & \cos t\sin t\\ \cos t\sin t & \sin^2 t\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}x\cos^2 t + y\cos t\sin t\\ x\cos t\sin t + y\sin^2 t\end{pmatrix}.$$

Here is a picture:

[27] Let A, B be two $n\times n$ matrices such that $A\mathbf{x} = B\mathbf{x}$ for all $\mathbf{x}\in\mathbb{R}^n$. If $\mathbf{x} = \mathbf{e}_j$ then the identity $A\mathbf{e}_j = B\mathbf{e}_j$ tells us that the jth columns of A and B are the same. Since this holds for any j we conclude that A and B are the same matrix.



Note that this projection is **not invertible**. To see this, let's consider the point $(-\sin t, \cos t)$. This point gets projected to the origin:

$$\begin{aligned} P_t \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= \begin{pmatrix} \cos^2 t & \cos t \sin t \\ \cos t \sin t & \sin^2 t \end{pmatrix} \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} \\ &= \begin{pmatrix} -\cos^2 t \sin t + \cos^2 t \sin t \\ -\cos t \sin^2 t + \cos t \sin^2 t \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

But the origin gets projected to itself: $P_t\mathbf{0} = \mathbf{0}$. If P_t had an inverse matrix $(P_t)^{-1}$ then this would imply that

$$\begin{aligned} P_t \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= P_t \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ (P_t)^{-1} P_t \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= (P_t)^{-1} P_t \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Contradiction.[28] More generally, let $P = \mathbf{u}\mathbf{u}^T$ be the matrix that projects onto the line in $\mathbb{R}^n$ generated by some unit vector $\mathbf{u}$ and let $\mathbf{v} \in \mathbb{R}^n$ be any vector that is perpendicular to $\mathbf{u}$, so that $\mathbf{u}^T\mathbf{v} = 0$. Then we have

$$P\mathbf{v} = (\mathbf{u}\mathbf{u}^T)\mathbf{v} = \mathbf{u}(\mathbf{u}^T\mathbf{v}) = \mathbf{u}(0) = \mathbf{0}.$$

[28] More generally, a linear function that is not injective cannot have a left inverse.

This shows that the projection onto a line in $\mathbb{R}^n$ is never invertible. Finally, let me note that the matrix P_t has determinant zero:

$$\det P_t = \det \begin{pmatrix} \cos^2 t & \cos t \sin t \\ \cos t \sin t & \sin^2 t \end{pmatrix} = \cos^2 t \sin^2 t - \cos^2 t \sin^2 t = 0.$$

Later we will see that a square matrix A is invertible if and only if $\det A \neq 0$.

The Group of Orthogonal Matrices. I mentioned above that a square matrix A satisfying $A^{-1} = A^T$ is called an *orthogonal matrix*. We denote the set of all such matrices by

$$\mathrm{O}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n \times n} : A^T A = I \text{ and } AA^T = I\}.$$

Sometimes the set $\mathrm{O}_n(\mathbb{R})$ is called the *orthogonal group*, because it satisfies the three group axioms from abstract algebra:

- The identity matrix is in $\mathrm{O}_n(\mathbb{R})$. Indeed, we have $I^T = I$ and $II = I$, so that $I^T I = II = I$ and $II^T = II = I$.
- If A is in $\mathrm{O}_n(\mathbb{R})$ then A is invertible and A^{-1} is also in $\mathrm{O}_n(\mathbb{R})$. Indeed, the conditions $A^T A = I$ and $AA^T = I$ just tell us that A is invertible with $A^{-1} = A^T$. But then we also have
$$(A^{-1})^{-1} = A = (A^T)^T = (A^{-1})^T,$$
which tells us that A^{-1} is in $\mathrm{O}_n(\mathbb{R})$.
- If A and B are in $\mathrm{O}_n(\mathbb{R})$ (i.e., if $A^{-1} = A^T$ and $B^{-1} = B^T$) then so is their product AB. Indeed, we have
$$(AB)^{-1} = B^{-1}A^{-1} = B^T A^T = (AB)^T.$$

Remark: Particle physicists are particularly interested in matrix groups but they prefer the complex version of orthogonal matrices, which are called *unitary matrices*:

$$\mathrm{U}_n(\mathbb{C}) = \{A \in \mathbb{C}^{n \times n} : A^* A = I \text{ and } AA^* = I\}.$$

It is worth mentioning a geometric interpretation of orthogonal matrices:[29]

$$A^T A = I \quad \Longleftrightarrow \quad \text{the columns of } A \text{ are orthonormal.}$$

Indeed, suppose that $A \in \mathbb{R}^{n \times n}$ has column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$, so that A^T has row vectors $\mathbf{a}_1^T, \ldots, \mathbf{a}_n^T$. Then the i, j entry of the matrix $A^T A$ is the dot product of $\mathbf{a}_i$ and $\mathbf{a}_j$:

$$\begin{aligned}(i,j \text{ entry of } A^T A) &= (i\text{th row of } A^T)(j\text{th col of } A) \\ &= \mathbf{a}_i^T \mathbf{a}_j\end{aligned}$$

[29]The same result holds for unitary matrices, with respect to the Hermitian inner product.

$$= \mathbf{a}_i \bullet \mathbf{a}_j.$$

On the other hand, the i, j entry of the identity matrix is the Kronecker delta δ_{ij}. Hence we have $A^T A = I$ if and only if

$$\mathbf{a}_i \bullet \mathbf{a}_j = \delta_{ij},$$

i.e., if and only if the column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ are orthonormal. This is one reason for the term "orthogonal matrix". On the homework you used this fact to prove that every 2×2 orthogonal matrix is either a rotation or a rotation.

The Fundamental Theorem, which we will prove in the next section, tells us that the equations $A^T A = I$ and $AA^T = I$ are equivalent when A is square, which means that the columns of a square matrix are orthonormal if and only if the rows are orthonormal. I find this mysterious.

3 Subspaces Associated to Linear Functions

In the last section we discussed purely symbolic properties of matrix inversion. Recall: Let A be an $m \times n$ matrix. An $n \times m$ matrix B is called a *right inverse of* A when $AB = I_m$ and an $m \times n$ matrix C is called a *left inverse of* A when $CA = I_n$. If A has both a right inverse B and a left inverse C then the two must be equal because

$$B = I_n B = (CA)B = C(AB) = CI_m = C.$$

In this case we say that $A^{-1} = B = C$ is the unique *two-sided inverse of* A. Any matrix having a two-sided inverse is called *invertible*. We also proved the following basic facts: If A^{-1} exists then $(A^*)^{-1}$ exists and is equal to $(A^{-1})^*$. If A^{-1}, B^{-1} and AB exist then $(AB)^{-1}$ exists and is equal to $B^{-1}A^{-1}$.

Precisely when do inverse matrices exist? This question is surprisingly subtle. In order to answer it we must ascend to a higher level of abstraction. To each linear function between vector spaces $V \to W$ we associate certain subspaces of V and W.

3.1 Kernel and Image of a Linear Function.

Consider a linear function $f : V \to W$ between vector spaces.[30] We define the *kernel* and the *image* of f as follows:

$$\begin{aligned} ker(f) &:= \{\text{the set of } \mathbf{v} \in V \text{ such that } f(\mathbf{v}) = \mathbf{0}\}, \\ im(f) &:= \{\text{the set of } \mathbf{w} \in W \text{ such that } \mathbf{w} = f(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}. \end{aligned}$$

Remark: The kernel and image of f are sometimes called the *nullspace* and *range*.[31]

[30] Over $\mathbb{R}$ or $\mathbb{C}$; it doesn't matter. Indeed, the same theory applies to vector spaces over arbitrary fields.

[31] Kernel and image are standard terminology in abstract algebra. Nullspace and range are more common in applied linear algebra. For matrices, the image/range is often called the *column space*. (Too many words; I know.) See the next section.

We observe that $ker(f) \subseteq V$ **is a subspace**. Indeed, given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in ker(f)$ in the kernel and scalars $a_1, \ldots, a_n$, the linearity of f implies

$$f(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1 f(\mathbf{v}_1) + \cdots + a_n f(\mathbf{v}_n) = a_1\mathbf{0} + \cdots + a_n\mathbf{0} = \mathbf{0},$$

so the linear combination $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ is also in the kernel. Furthermore, we observe that $im(f) \subseteq W$ **is a subspace**. Indeed, consider any vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n \in im(f)$ in the image and any scalars $a_1, \ldots, a_n$. By definition we can write $\mathbf{w}_i = f(\mathbf{v}_i)$ for some vectors $\mathbf{v}_i$, hence from the linearity of f we have

$$a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n = a_1 f(\mathbf{v}_1) + \cdots + a_n f(\mathbf{v}_n) = f(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n).$$

Since $a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n = f(\mathbf{v}')$ for some vector $\mathbf{v}'$ we conclude that the linear combination $a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n$ is also in the image.

The invertibility of a linear function is closely related to its kernel and image. The first observation is true by definition of the words *image* and *surjective*:[32]

$$f : V \to W \text{ is surjective if and only if } im(f) = W.$$

The next observation requires a short proof:

$$f : V \to W \text{ is injective if and only if } ker(f) = \{\mathbf{0}\}.$$

Proof. Recall that any linear function satisfies $f(\mathbf{0}) = \mathbf{0}$. If f is injective then $f(\mathbf{v}) = \mathbf{0} = f(\mathbf{0})$ implies $\mathbf{v} = \mathbf{0}$, and hence $ker(f) = \{\mathbf{0}\}$. Conversely, suppose that $ker(f) = \{\mathbf{0}\}$. To show that f is injective, let $f(\mathbf{v}_1) = f(\mathbf{v}_2)$ for some vectors $\mathbf{v}_1, \mathbf{v}_2$. Then we have

$$\begin{aligned}
f(\mathbf{v}_1) &= f(\mathbf{v}_2) && \\
f(\mathbf{v}_1) - f(\mathbf{v}_2) &= \mathbf{0} && \\
f(\mathbf{v}_1 - \mathbf{v}_2) &= \mathbf{0} && \text{linearity of } f \\
\mathbf{v}_1 - \mathbf{v}_2 &= \mathbf{0} && ker(f) = \{\mathbf{0}\} \\
\mathbf{v}_1 &= \mathbf{v}_2. &&
\end{aligned}$$

Hence f is injective. □

[32] The words *surjective* and *injective* were introduced by Bourbaki in the 1940s. The older equivalent terms are *onto* and *one-to-one*.

3.2 Isomorphism of Vector Spaces.

Let $f : V \to W$ be a function between vector spaces. We say that f is an *isomorphism*[33] when the following properties are satisfied:

(a) f is linear,

(b) f is surjective,

(c) f is injective.

Properties (b) and (c) say that f is a *bijection*,[34] which is equivalent to being invertible. Furthermore, one can check that the inverse function $f^{-1} : W \to V$ is also linear. If there exists an isomorphism between vector spaces V and W then we will write

$$V \cong W.$$

When V and W are finite dimensional we have the following important fact:

Isomorphism of Finite Dimensional Vector Spaces.

$$\boxed{V \cong W \quad \iff \quad \dim(V) = \dim(W).}$$

Proof. $\Longrightarrow$: Suppose that $V \cong W$ and let $f : V \to W$ be a specific isomorphism. Suppose that $\dim(V) = n$ and let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for V. Then I claim that $\{f(\mathbf{b}_1), \ldots, f(\mathbf{b}_n)\}$ is a basis for W, from which it will follow that $\dim(W) = n$. There are two things to show:

- **Independent.** Suppose that $a_1 f(\mathbf{b}_1) + \cdots + a_n f(\mathbf{b}_n) = \mathbf{0}$ for some scalars $a_1, \ldots, a_n$. Linearity of f implies that

  $$\mathbf{0} = a_1 f(\mathbf{b}_1) + \cdots + a_n f(\mathbf{b}_n) = f(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n),$$

  and then the fact that f is injective implies that

  $$\mathbf{0} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n.$$

  Finally, the fact that $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is independent implies that $a_1 = \cdots = a_n = 0$.

- **Spanning.** Consider any vector $\mathbf{w} \in W$. Since f is surjective we have $\mathbf{w} = f(\mathbf{v})$ for some $\mathbf{v} \in V$, and since $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ spans $\mathbf{v}$ we can write

  $$\mathbf{v} = a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n$$

  for some scalars $a_1, \ldots, a_n$. Finally, by linearity of f we have

  $$\mathbf{w} = f(\mathbf{v}) = f(a_1\mathbf{b}_1 + \cdots + a_n\mathbf{b}_n) = a_1 f(\mathbf{b}_1) + \cdots + a_n f(\mathbf{b}_n),$$

  which shows that $\{f(\mathbf{b}_1), \ldots, f(\mathbf{b}_n)\}$ spans W.

[33] Also called a linear isomorphism, or an isomorphism of vector spaces.
[34] Another Bourbaki term. The older word is *one-to-one correspondence*.

$\Longleftarrow$: Suppose that $\dim(V) = \dim(W) = n$. Choose bases $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ and $\mathbf{w}_1, \ldots, \mathbf{w}_n \in W$ and define a linear function $f : V \to W$ by sending $\mathbf{v}_i \mapsto \mathbf{w}_i$ for all i. Then for any vector $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \in V$ we have

$$f(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1 f(\mathbf{v}_1) + \cdots + a_n f(\mathbf{v}_n) = a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n.$$

Furthermore, the function $f :^{-1}: W \to V$ defined by sending $\mathbf{w}_i \mapsto \mathbf{v}_i$ is the inverse of f:

$$f^{-1}(a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n) = a_1 f^{-1}(\mathbf{w}_1) + \cdots + a_n f^{-1}(\mathbf{w}_n) = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n.$$

□

As a consequence of this theorem, any n-dimensional vector space over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$. Indeed, let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ be a basis. Then the following function $V \to \mathbb{R}^n$ is an isomorphism:

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \mapsto \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

We will apply these ideas in the next section.

4 Subspaces Associated to Matrices

Recall that an $m \times n$ matrix over $\mathbb{R}$ is the same thing as a linear function $\mathbb{R}^n \to \mathbb{R}^m$.[35] In this case the kernel and image have a special interpretation.

4.1 The Nullspace of a Matrix.

Given an $m \times n$ matrix A we define the *nullspace*:

$$\mathcal{N}(A) := \{\text{the set of } \mathbf{x} \in \mathbb{R}^n \text{ such that } A\mathbf{x} = \mathbf{0}\}.$$

It is easy to check that $\mathcal{N}(A) \subseteq \mathbb{R}^m$ is a subspace. Indeed, $\mathcal{N}(A)$ is just the kernel of the linear function $A : \mathbb{R}^n \to \mathbb{R}^m$. More interestingly, we can use the concept of the nullspace to express the fact that a given vector is simultaneously orthogonal to a given set of vectors:

$$\boxed{\mathbf{x} \in \mathcal{N}(A) \quad \Longleftrightarrow \quad A\mathbf{x} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{x} \text{ is orthogonal to every row of } A.}$$

Indeed, let $\mathbf{a}_i^T$ be the ith row vector of A. If $A\mathbf{x} = \mathbf{0}$ then we have

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0} = A\mathbf{x} = \begin{pmatrix} - & \mathbf{a}_1^T & - \\ & \vdots & \\ - & \mathbf{a}_m^T & - \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}_1^T\mathbf{x} \\ \vdots \\ \mathbf{a}_m^T\mathbf{x} \end{pmatrix}.$$

[35] When I write $\mathbb{R}^n$ I always assume that we are working with the standard basis.

Comparing entries on the left and right gives $\mathbf{a}_i^T\mathbf{x} = 0$ for all i. In other words, the vector $\mathbf{x}$ is orthogonal to each row vector of A. Equivalently, we have

$$\boxed{A^T\mathbf{x} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{x} \text{ is orthogonal to every column of } A.}$$

It is important to get comfortable with this idea because it is the foundation of least squares.[36]

4.2 The Column Space of a Matrix.

We can think of an $m \times n$ matrix A as a linear function $A : \mathbb{R}^n \to \mathbb{R}^m$. In this case the image of A is called the *column space*:

$$\mathcal{C}(A) = \{\text{the set of } A\mathbf{x} \in \mathbb{R}^m \text{ for all } \mathbf{x} \in \mathbb{R}^n\}.$$

But *why* is it called the column space? Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ be the column vectors of A. Then for any vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we have

$$A\mathbf{x} = \left(\begin{array}{c|c|c} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{array} \right) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n,$$

which is a linear combination of the columns of A. So we can also write

$$\mathcal{C}(A) = \{\text{all linear combinations of the columns of } A\}.$$

Similarly, we can define the *row space* of A:

$$\mathcal{R}(A) := \mathcal{C}(A^T) = \{\text{all linear combinations of the rows of } A\}.$$

Note that $\mathcal{C}(A)$ is a subspace of $\mathbb{R}^m$ because each column of an $m \times n$ matrix lives in $\mathbb{R}^m$, while $\mathcal{R}(A)$ is a subspace of $\mathbb{R}^n$. So the row space and column space cannot be directly compared.

4.3 Orthogonality of the Subspaces.

We observed above that $\mathbf{x} \in \mathcal{N}(A)$ if and only if $\mathbf{x}$ is orthogonal to every row of A. We can express this as follows:

$$\boxed{\mathcal{N}(A) = \mathcal{R}(A)^\perp.}$$

In general, given a subspace $U \subseteq V$ of an inner product space V we let $U^\perp \subseteq V$ denote the set of vectors that are orthogonal to every vector in U:[37]

$$U^\perp = \{\text{the set of } \mathbf{v} \in V \text{ such that } \langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{u} \in U\}.$$

[36] For the impatient: Let $P\mathbf{x}$ be the orthogonal projection of a point $\mathbf{x}$ onto the column space of a matrix A. Then the vector $P\mathbf{x} - \mathbf{x}$ must be orthogonal to every column of A, hence $A^T(P\mathbf{x} - \mathbf{x}) = \mathbf{0}$.

[37] We read $U^\perp$ as "U perp".

You will check on the homework that $U^\perp \subseteq V$ is also a subspace. Furthermore, if V finite dimensional then you will prove the following dimension formula:

$$\dim U + \dim U^\perp = \dim V.$$

In the case of the rowspace and nullspace of a matrix A we obtain the following theorem.

The Rank-Nullity Theorem. For any matrix A we have

$$\boxed{\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = \text{the number of columns of } A.}$$

Indeed, if A is $m \times n$ then $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are orthogonal subspaces of $\mathbb{R}^n$, so that

$$\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = n.$$

This is often called the *rank-nullity theorem* because $\dim \mathcal{R}(A)$ is called the *rank* and $\dim \mathcal{N}(A)$ is called the *nullity* of the matrix A.[38] By replacing A with A^T we obtain the equivalent formula

$$\dim \mathcal{C}(A) + \dim \mathcal{N}(A^T) = m,$$

which does not have a nice name.

5 The Fundamental Theorem

In this section we will prove the most important theorem about matrices. Following Gilbert Strang, I will call this "The Fundamental Theorem".

The Fundamental Theorem of Linear Algebra. For any $m \times n$ matrix A we have

$$\dim \mathcal{R}(A) = \dim \mathcal{C}(A).$$

This common dimension is called the *rank* of A, sometimes written $\text{rank}(A)$.

This result is a bit surprising because the row space $\mathcal{R}(A)$ lives in $\mathbb{R}^n$, while the column space $\mathcal{C}(A)$ lives in $\mathbb{R}^m$, so there is no direct way to compare them. Evidently there is some subtle form of communication between the rows and columns of a matrix. We will see in the next section that the Fundamental Theorem implies the following facts:

- Invertible matrices are square.
- If A and B are square of the same size, then $AB = I$ if and only if $BA = I$.
- If A is square then A has orthonormal columns if and only if it has orthonormal rows.

The proof is more difficult than you might expect, but it is worth going through the details because the ideas in the proof quite useful. There are two main steps:

[38] The dimension of $\mathcal{C}(A)$ is also called the rank of A. The fact that $\mathcal{R}(A)$ and $\mathcal{C}(A)$ have the same dimension is a deep fact called the Fundamental Theorem. See the next section.

(1) Let E and F be any matrices such that E has a left inverse $E'E = I$ and F has a right inverse $FF' = I$.[39] Then we will show that

$$\dim \mathcal{R}(EAF) = \dim \mathcal{R}(A) \quad \text{and} \quad \dim \mathcal{C}(EAF) = \dim \mathcal{C}(A).$$

(2) For any matrix A, we will find matrices E and F, as in (1), so that EAF has the following simple form:

$$EAF = \left(\begin{array}{c|c} I_r & O_{r,n-r} \\ \hline O_{m-r,r} & O_{m-r,n-r} \end{array}\right),$$

where I_r is the square $r \times r$ identity matrix. Since the matrix on the right clearly has row space and column space of dimension r,[40] it will follow that

$$\dim \mathcal{R}(A) = \dim \mathcal{R}(EAF) = r = \dim \mathcal{C}(EAF) = \dim \mathcal{C}(A).$$

Aside from these two main steps, we will further organize the proof into substeps, labeled by (a), (b), etc., since there are many details.

Proof of Step (1).

(a) For any matrix E such that EA exists, we have

$$\mathcal{R}(EA) \subseteq \mathcal{R}(A).$$

Indeed, I claim that each row of EA is a linear combination of the rows of A. To see this, let E have ith row $(e_{i1}, \ldots, e_{im})$ and let A have ith row $\mathbf{a}_i^T$. Then

$$\begin{aligned} (i\text{th row of } EA) &= (i\text{th row of } E)A \\ &= \begin{pmatrix} e_{i1} & \cdots & e_{im} \end{pmatrix} A \\ &= \left(\begin{array}{c|c|c} e_{i1} & \cdots & e_{im} \end{array}\right) \left(\begin{array}{c} \mathbf{a}_1^T \\ \hline \vdots \\ \hline \mathbf{a}_m^T \end{array}\right) \\ &= e_{i1}\mathbf{a}_1^T + \cdots + e_{im}\mathbf{a}_m^T. \end{aligned}$$

In the last step we used block multiplication. Since every row of EA is in the rowspace $\mathcal{R}(A)$ it follows that any linear combination of rows of EA is in $\mathcal{R}(A)$. In other words, $\mathcal{R}(EA) \subseteq \mathcal{R}(A)$.

(b) If E has a left inverse $E'E = I$ then we also have

$$\mathcal{R}(A) \subseteq \mathcal{R}(EA).$$

[39] These one-sided inverses need not be unique.

[40] The first r rows are a basis for the row space, while the first r columns are a basis for the column space.

Indeed, applying step (a) to the matricx $B = EA$ and E' shows that

$$\mathcal{R}(A) = \mathcal{R}(E'EA) = \mathcal{R}(E'B) \subseteq \mathcal{R}(B) = \mathcal{R}(EA).$$

Then combining (a) and (b) shows that $\mathcal{R}(EA) = \mathcal{R}(A)$, hence

$$\dim \mathcal{R}(EA) = \dim \mathcal{R}(A).$$

(c) For any matrix F such that AF exists, we have

$$\mathcal{C}(A) \subseteq \mathcal{C}(AF).$$

Indeed, I claim that any column of AF is a linear combination of the columns of A. The proof is similar to part (a). Let $(f_{1j}, \ldots, f_{nj})$ be the jth column of F and let $\mathbf{a}_j$ be the jth column of A. Then we have

$$\begin{aligned}(j\text{th column of } AF) &= A(j\text{th column of } F) \\ &= \begin{pmatrix} \mathbf{a}_1 \,\big|\, \cdots \,\big|\, \mathbf{a}_n \end{pmatrix} \begin{pmatrix} f_{1j} \\ \vdots \\ f_{nj} \end{pmatrix} \\ &= f_{1j}\mathbf{a}_1 + \cdots + f_{nj}\mathbf{a}_n.\end{aligned}$$

(d) If F has a right inverse $FF' = I$, then applying (c) to the matrix $B = AF$ and F' gives

$$\mathcal{C}(AF) = \mathcal{C}(B) \subseteq \mathcal{C}(BF') = \mathcal{C}(AFF') = \mathcal{C}(A),$$

hence $\mathcal{C}(AF) = \mathcal{C}(A)$. It follows that

$$\dim \mathcal{C}(AF) = \dim \mathcal{C}(A).$$

Next we will show that $\dim \mathcal{R}(A) = \dim \mathcal{R}(AF)$ and $\dim \mathcal{C}(EA) = \dim \mathcal{C}(A)$. This time the corresponding spaces are **not equal**, but they are still **isomorphic**.

(e) For any matrix A with rows $\mathbf{a}_i^T$ and any matrix F of appropriate shape, note that

$$(i\text{th row of } AF) = (i\text{th row of } A)F = \mathbf{a}_i^T F.$$

Consider the function $\varphi : \mathcal{R}(A) \to \mathcal{R}(AF)$ defined by multiplying on the right by F. That is, for any vector[41] $\mathbf{b}^T = b_1\mathbf{a}_1^T + \cdots + b_m\mathbf{a}_m^T \in \mathcal{R}(A)$ we define

$$\begin{aligned}\varphi(\mathbf{b}^T) &:= \mathbf{b}^T F \\ &= \varphi(b_1\mathbf{a}_1^T + \cdots + b_m\mathbf{a}_m^T)F\end{aligned}$$

[41] Usually we think of $\mathcal{R}(A)$ as space of column vectors, but for the purpose of this proof it is more convenient to think of $\mathcal{R}(A)$ as space of row vectors.

$$= b_1(\mathbf{a}_1^T F) + \cdots + b_m(\mathbf{a}_m^T F) \in \mathcal{R}(AF).$$

Matrix multiplication is linear, so φ is a linear function. Next, for any vector

$$\mathbf{c}^T := c_1(\mathbf{a}_1^T F) + \cdots + c_m(\mathbf{a}_m^T F) \in \mathcal{R}(AF)$$

we have

$$\mathbf{c}^T = \varphi(c_1\mathbf{a}_1^T + \cdots + c_n\mathbf{a}_m^T),$$

so that φ is surjective. Finally, since F has a right inverse $FF' = I$ we see that φ is injective. Indeed, if $\varphi(\mathbf{b}^T) = \varphi(\mathbf{c}^T)$ then

$$\begin{aligned}
\varphi(\mathbf{b}^T) &= \varphi(\mathbf{c}^T) \\
\mathbf{b}^T F &= \mathbf{c}^T F \\
(\mathbf{b}^T F)F' &= (\mathbf{c}^T F)F' \\
\mathbf{b}^T(FF') &= \mathbf{c}^T(FF') \\
\mathbf{b}^T &= \mathbf{c}^T.
\end{aligned}$$

Hence φ is an isomorphism $\mathcal{R}(A) \cong \mathcal{R}(AF)$, and it follows from the previous section that

$$\dim \mathcal{R}(AF) = \dim \mathcal{R}(A).$$

(f) Similarly, if E has a left inverse $E'E = I$ then we will show that $\mathcal{C}(EA) \cong \mathcal{C}(A)$. To do this we consider the function $\psi : \mathcal{C}(A) \to \mathcal{C}(EA)$ defined by multiplying on the left by E. To be explicit, let $\mathbf{a}_j$ be the jth column of A,[42] so that

$$(j\text{th column of } EA) = E(j\text{th column of } A) = E\mathbf{a}_j.$$

Consider the function $\psi : \mathcal{C}(A) \to \mathcal{C}(EA)$ defined by multiplying on the left by E. That is, for any vector $\mathbf{b} = b_1\mathbf{a}_1 + \cdots + b_n\mathbf{a}_n \in \mathcal{C}(A)$ we define

$$\begin{aligned}
\psi(\mathbf{b}) &:= E\mathbf{b} \\
&= E(b_1\mathbf{a}_1 + \cdots + b_n\mathbf{a}_n) \\
&= b_1(E\mathbf{a}_1) + \cdots + b_n(E\mathbf{a}_n) \in \mathcal{C}(EA).
\end{aligned}$$

Following an argument similar to (e), we see that ψ is a vector space isomorphism, and hence

$$\dim \mathcal{C}(EA) = \dim \mathcal{C}(A).$$

Proof of Step (2). The proof of this step is an algorithm. For this purpose we introduce the important new idea of *elementary matrices*.

(g) Elementary Matrices. We define three families of **square** matrices.[43]

[42] In part (e) we used $\mathbf{a}_i^T$ for the ith row of A. Hopefully you don't mind that I'm recycling the notation $\mathbf{a}_j$ for a different purpose. Gilbert Strang uses $\mathbf{a}_i^*$ to denote rows of a matrix, but I don't like this because I use $*$ for conjugate transpose.

[43] It is always a struggle to find a notation for elementary matrices. Here I use the Wikipedia notation. I guess that D is for Diagonal, T is for Transposition and L is for Lower triangular, since many algorithms only use lower triangular $L_{ij}(\lambda)$ (i.e., with $i > j$). I prefer to think of D for Dilation and L for eLimination.

- For any index i and **nonzero** scalar λ we define

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

  The main diagonal entries are 1, except of the ii entry, which is λ. The off-diagonal entries are all zero.

- For any indices $i \neq j$ and any scalar λ we define

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

  The main diagonal entries are 1. The only other nonzero entry is λ in the ij position.

- For any indices $i \neq j$ we define

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

  The main diagonal entries are 1 except for zeros in the ii and jj positions. The off-diagonal entries are zero except for 1 in the ij and ji positions.

We observe that each of these (square) *elementary matrices* is invertible. That is, we have

$$\begin{aligned} D_i(\lambda)^{-1} &= D_i(1/\lambda) \\ L_{ij}(\lambda)^{-1} &= L_{ij}(-\lambda) \\ T_{ij}^{-1} &= T_{ij}. \end{aligned}$$

But what are these matrices **for**?

(h) Row and Column Operations. Let A be an $m \times n$ matrix. For any matrix E we have seen that each row of EA is a linear combination of the rows of A. To be precise, if $(e_{i1}, \ldots, e_{im})$ is the ith row of E and $\mathbf{a}_i^T$ is the ith row of A, then

$$(i\text{th row of } EA) = e_{i1}\mathbf{a}_1^T + \cdots + e_{im}\mathbf{a}_m^T.$$

When E is an $m \times m$ elementary matrix then we have the following *elementary row operations.*

- The function $A \rightsquigarrow D_i(\lambda)A$ multiplies the ith row of A by λ.
- The function $A \rightsquigarrow L_{ij}(\lambda)A$ replaces the ith row of A by itself plus λ times the jth row of A. Indeed, when $k \neq i$, the kth row of $L_{ij}(\lambda)$ is just a standard basis vector, whereas the ith row of $L_{ij}(\lambda)$ is $(0, \ldots, 0, 1, 0, \ldots, 0, \lambda, 0, \ldots, 0)$ with 1 in the ith position and λ in the jth position. Hence the ith row of EA is

$$0\mathbf{a}_1^T + \cdots + 0\mathbf{a}_{i-1}^T + 1\mathbf{a}_i^T + 0\mathbf{a}_{i+1}^T + \cdots + 0\mathbf{a}_{j-1}^T + \lambda\mathbf{a}_j^T + 0\mathbf{a}_{j+1}^T + \cdots + 0\mathbf{a}_m^T.$$

- The function $A \rightsquigarrow T_{ij}A$ swaps the ith and jth rows of A.

Similarly, if F has jth column $(f_{1j}, \ldots, f_{nj})$ and A has jth column $\mathbf{a}_j$, then

$$(j\text{th column of } AF) = f_{1j}\mathbf{a}_1 + \cdots + f_{nj}\mathbf{a}_n.$$

When F is an elementary matrix then we have the following *elementary column operations.*

- The function $A \rightsquigarrow AD_i(\lambda)$ multiplies the ith column of A by λ.
- The function $A \rightsquigarrow AL_{ij}(\lambda)$ replaces the jth column of A by itself plus λ times the ith column of A. The proof is the same as for rows.
- The function $A \rightsquigarrow AT_{ij}$ swaps the ith and jth columns of A.

(i) The Algorithm. Finally, we can use elementary matrices to put the $m \times n$ matrix A into a particularly nice form. If $E_1, \ldots, E_k$ are elementary $m \times m$ matrices and if $F_1, \ldots, F_\ell$ are elementary $n \times n$ matrices then by performing row and column operations we will obtain

$$E_k \cdots E_1 E_1 A F_1 F_2 \cdots F_\ell = EAF.$$

Since elementary matrices are invertible, the products $E = E_k \cdots E_2 E_1$ and $F = F_1 F_2 \cdots F_\ell$ are also invertible. In particular, E has a left inverse and F has a right inverse, so we can apply the results from step (1).

Now we explain how to choose the operations.[44] If the top left entry of A is zero, swap rows or columns until it is not zero. Then scale the first row or column so the top left entry is equal to 1. Next apply elimination matrices $L_{ij}(\lambda)$ on both sides to eliminate the other entries in the first row and column. The result is a matrix of the form

$$\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right),$$

[44] We are concerned here with clarity of exposition, not with efficiency of implementation.

where A' has size $(m-1)\times(n-1)$. If A' is the zero matrix then we are done. Otherwise we repeat the previous steps on the smaller matrix to obtain

$$\left(\begin{array}{c|c|ccc} 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & & & \\ \vdots & \vdots & & A'' & \\ 0 & 0 & & & \end{array}\right),$$

where A'' has size $(m-2)\times(n-2)$. We repeat this process until the bottom right corner is a zero matrix. If the process terminates after r steps then the bottom right corner is the zero matrix of size $(m-r)\times(n-r)$. □

This completes our proof of the Fundamental Theorem. This is not the shortest proof, but it is the clearest proof that I know. And it has the added benefit of introducing important ideas (such as elementary matrices) that we will use in the future.

Remark: There is a variant of this algorithm that works over the integers. The difference when working over $\mathbb{Z}$ is that we cannot divide, so we cannot scale the top left entry to equal 1. However, we can arrange that the top left entry is as small as possible, and that each diagonal entry divides the next. We omit the proof because it requires a bit of number theory.[45]

Theorem (Smith Normal Form). Let A be an $n\times m$ matrix of rank r with integer entries. Then there exist invertible matrices E and F with integer entries, whose inverses E^{-1} and F^{-1} and also have integer entries, such that

$$EAF = \left(\begin{array}{cccc|c} d_1 & & & & \\ & d_2 & & & \\ & & \ddots & & O_{r,n-r} \\ & & & d_r & \\ \hline & O_{m-r,r} & & & O_{m-r,n-r} \end{array}\right).$$

The diagonal integers $0 \le d_1 \le \ldots \le d_r$ have the property that d_{i+1} is an integer multiple of d_i for all i. These diagonal entries are called the *elementary divisors* of the matrix A. The Smith Normal Form is useful in cryptography and in algebraic topology, but we will have no use for it in this course.

6 Existence of Inverse Matrices

As promised, we now apply the Fundamental Theorem to the existence of inverse matrices. Before doing so we make a basic observation. For any $m\times n$ matrix A and $n\times 1$ column $\mathbf{b}$,

the matrix equation $A\mathbf{x}=\mathbf{b}$ has a solution $\mathbf{x}\in\mathbb{R}^n$ if and only if $\mathbf{b}\in\mathcal{C}(A)$.

[45] In general the algorithms for linear algebra over $\mathbb{Z}$ are much more expensive than for linear algebra over a field such as $\mathbb{R}$ or $\mathbb{C}$. The complexity of the algorithms makes the subject useful for cryptography.

Indeed, this is just a way of rephrasing the definition of the column space, since every linear combination of the columns of A has the form $A\mathbf{x}$ for some vector $\mathbf{x}$.

First we state conditions for the existence of left and right inverse matrices.

6.1 Existence of Right Inverses.

Given an $m \times n$ matrix A, recall that a right inverse is any $n \times m$ matrix X satisfying $AX = I_m$. In order to find such a matrix, let $\mathbf{x}_j \in \mathbb{R}^n$ be the unknown jth column of X. Then using block multiplication gives

$$\left(\begin{array}{c|c|c} A\mathbf{x}_1 & \cdots & A\mathbf{x}_m \end{array} \right) = A \left(\begin{array}{c|c|c} \mathbf{x}_1 & \cdots & \mathbf{x}_m \end{array} \right) = AX = I_m = \left(\begin{array}{c|c|c} \mathbf{e}_1 & \cdots & \mathbf{e}_m \end{array} \right).$$

In other words, we have $AX = I_m$ if and only if we have $A\mathbf{x}_j = \mathbf{e}_j$ for each column vector $\mathbf{x}_j$, where $\mathbf{e}_j$ is the jth column of the identity matrix I_m, i.e., the jth standard basis vector in $\mathbb{R}^m$. By the previous remark, such vectors $\mathbf{x}_j$ exist if and only if each basis vector $\mathbf{e}_j \in \mathbb{R}^m$ is in the column space $\mathcal{C}(A)$. Finally, since $\mathcal{C}(A)$ is a subspace of $\mathbb{R}^m$, this happens if and only if $\mathcal{C}(A)$ fills up all of $\mathbb{R}^m$.[46] Here is a summary:

$$\begin{array}{rcl} A \text{ has a right inverse} & \iff & AX = I_m \text{ for some matrix } X \\ & \iff & A\mathbf{x}_j = \mathbf{e}_j \text{ for some vectors } \mathbf{x}_1, \ldots, \mathbf{x}_m \\ & \iff & \mathbf{e}_j \in \mathcal{C}(A) \text{ for the standard basis vectors } \mathbf{e}_1, \ldots, \mathbf{e}_m \\ & \iff & \mathcal{C}(A) = \mathbb{R}^m \\ & \iff & \dim \mathcal{C}(A) = m. \end{array}$$

Furthermore, the Rank-Nullity Theorem tells us that $\dim \mathcal{C}(A) + \dim \mathcal{N}(A^T) = m$, hence

$$\begin{array}{rcl} A \text{ has a right inverse} & \iff & \dim \mathcal{C}(A) = m \\ & \iff & \dim \mathcal{N}(A^T) = 0 \\ & \iff & \mathcal{N}(A^T) = \{\mathbf{0}\} \\ & \iff & A^T\mathbf{x} = \mathbf{0} \text{ implies } \mathbf{x} = \mathbf{0} \\ & \iff & \text{the columns of } A^T \text{ are independent} \\ & \iff & \text{the rows of } A \text{ are independent.} \end{array}$$

6.2 Existence of Left Inverses.

We could do this from scratch, or we could observe that A has a right inverse if and only if A^T has a left inverse. Indeed, if X is a right inverse of A then $AX = I_m$ implies $X^T A^T = I_m$,

[46] Indeed, if $\mathcal{C}(A)$ contains every basis vector $\mathbf{e}_1, \ldots, \mathbf{e}_n$ then since $\mathcal{C}(A)$ is a subspace, it contains every linear combination of the basis vectors, i.e., it contains every vector in $\mathbb{R}^m$.

so that X^T is a left inverse of A^T. Conversely, if Y is a left inverse of A^T then $YA^T = I_n$ implies $AY^T = I_m$, so that Y^T is a right inverse of A. Hence

$$\begin{aligned}
A \text{ has a left inverse} &\iff A^T \text{ has a right inverse} \\
&\iff \mathcal{C}(A^T) = \mathbb{R}^n \\
&\iff \mathcal{R}(A) = \mathbb{R}^n \\
&\iff \dim \mathcal{R}(A) = n \\
&\iff \dim \mathcal{N}(A) = 0 \qquad \text{Rank-Nullity} \\
&\iff \mathcal{N}(A) = \{\mathbf{0}\} \\
&\iff A\mathbf{x} = \mathbf{0} \text{ implies } \mathbf{x} = \mathbf{0} \\
&\iff \text{the columns of } A \text{ are independent.}
\end{aligned}$$

6.3 Existence of Two-Sided Inverses.

Now we will use the Fundamental Theorem, which says that $\dim \mathcal{R}(A) = \dim \mathcal{C}(A)$. First we observe that

$$A \text{ has a two-sided inverse} \iff A \text{ has a right inverse and a left inverse.}$$

Indeed, any two sided inverse is by definition a right inverse and a left inverse. Conversely, suppose that A has a right inverse $AB = I_m$ and a left inverse $CA = I_n$. Then (as we have seen before) we must have

$$B = I_n B = (CA)B = C(AB) = CI_m = C,$$

so that $A^{-1} = B = C$ is the unique two-sided inverse of A. Finally, let r be the rank of A so that $r = \dim \mathcal{R}(A) = \dim \mathcal{C}(A)$ and observe that[47]

$$\begin{aligned}
A \text{ has a two-sided inverse} &\iff A \text{ has a right inverse and a left inverse} \\
&\iff \dim \mathcal{C}(A) = m \text{ and } \dim \mathcal{R}(A) = n \\
&\iff r = m \text{ and } r = n \\
&\iff m = n = r.
\end{aligned}$$

In particular, A must be square.

These ideas lead to some subtle properties of square matrices. Apparently the columns know what the rows are doing, and vice versa.

[47] There are many more equivalent conditions for invertibility. Wolfram MathWorld lists twenty three: `https://mathworld.wolfram.com/InvertibleMatrixTheorem.html`. Twenty of these follow easily from the results in this section. The remaining three refer to determinants, eigenvalues and singular values, which we haven't discussed yet.

6.4 Proof that $AB = I \iff BA = I$ for Square Matrices.

Let A and B be square matrices with $r = \text{rank}(A)$. Then

$$\begin{aligned} AB = I &\implies A \text{ has a right inverse} \\ &\implies r = \text{ the number of columns of } A \\ &\implies r = \text{ the number of rows of } A \\ &\implies A \text{ has a left inverse, say } CA = I. \end{aligned}$$

But then from the above computation we must have $B = C$, so $BA = I$. Switching the roles of A and B shows that $BA = I$ implies $AB = I$.

Here is an interesting application.

6.5 For Square Matrices, Orthonormal Columns $\iff$ Orthonormal Rows.

Let A be a square matrix. Then we have

$$\begin{aligned} A \text{ has orthonormal columns} &\iff A^T A = I \\ &\iff AA^T = I \\ &\iff A \text{ has orthonormal rows.} \end{aligned}$$

I think this theorem is a small miracle.

Now we know when inverse matrices exist. In the next section we will describe methods to compute inverse matrices.

7 Linear Systems

I assume you have some familiarity with the solution of linear systems, which is the main topic of Linear Algebra I. In this section we will go deeper into the topic.

Recall that a system of m linear equations in n unknowns has the form

$$\begin{cases} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & b_1, \\ \vdots & & & & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & b_m, \end{cases}$$

which can be expressed as a single matrix equation:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

At a higher level of abstraction we just write $A\mathbf{x} = \mathbf{b}$. Given a matrix of coefficients $A \in \mathbb{R}^{m\times n}$ and a vector of constants $\mathbf{b} \in \mathbb{R}^m$, the goal is to solve for the vector of unknowns $\mathbf{x} \in \mathbb{R}^n$. Recall from the previous section that

the system $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$ if and only if $\mathbf{b}$ is in the column space $\mathcal{C}(A)$.

If this is the case, then we can view the solution of $A\mathbf{x} = \mathbf{b}$ as an $(n - r)$-dimensional affine subspace of $\mathbb{R}^n$, which is parallel to the nullspace $\mathcal{N}(A)$. To be precise, we have the following.

7.1 Shape of the Solution

Let A be an $m\times n$ matrix and consider any vector $\mathbf{b} \in \mathcal{C}(A)$ in the column space. By definition this means we can write $\mathbf{b} = A\mathbf{x}'$ for some vector $\mathbf{x}' \in \mathbb{R}^m$, which might not be unique. Then every solution $A\mathbf{x} = \mathbf{b}$ has the form

$$\mathbf{x} = \mathbf{x}' + \mathbf{x}_0$$

for some *homogeneous solution* $A\mathbf{x}_0 = \mathbf{0}$, i.e., for some element of the nullspace $\mathbf{x}_0 \in \mathcal{N}(A)$. In more colloquial terms:

$$\text{(general solution)} = \text{(one particular solution)} + \text{(general } \textit{homogeneous} \text{ solution)}.$$

Proof. Fix a particular solution $A\mathbf{x}' = \mathbf{b}$. Then for any $\mathbf{x}_0 \in \mathcal{N}(A)$ we have

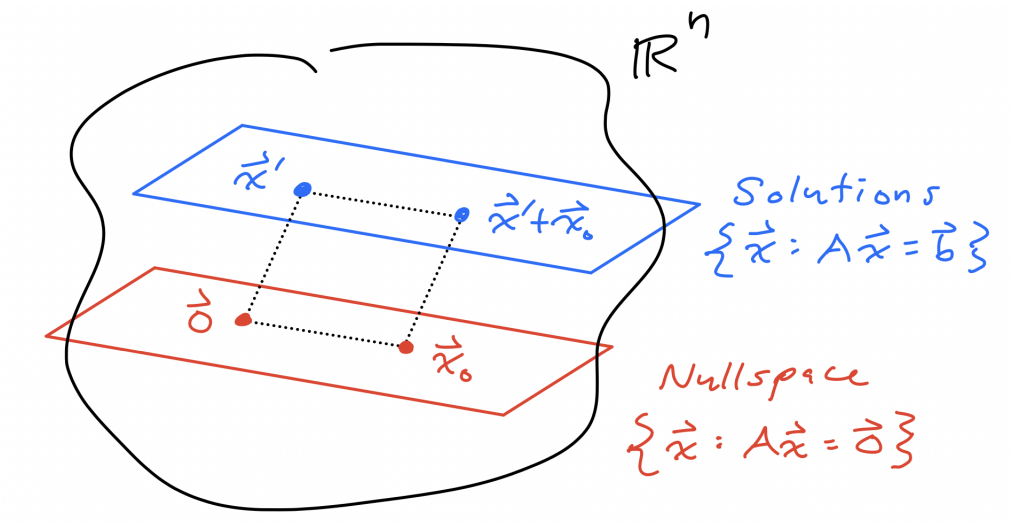
$$A(\mathbf{x}' + \mathbf{x}_0) = A\mathbf{x}' + A\mathbf{x}_0 = A\mathbf{x}' + \mathbf{0} = A\mathbf{x}' = \mathbf{b},$$

so that $\mathbf{x} = \mathbf{x}' + \mathbf{x}_0$ is also a solution. Conversely, let $\mathbf{x}$ be any solution $A\mathbf{x} = \mathbf{b}$. Then

$$\begin{aligned} \mathbf{b} &= \mathbf{b} \\ A\mathbf{x} &= A\mathbf{x}' \\ A\mathbf{x} - A\mathbf{x}' &= \mathbf{0} \\ A(\mathbf{x} - \mathbf{x}') &= \mathbf{0}, \end{aligned}$$

so that $\mathbf{x} - \mathbf{x}'$ is an element of the nullspace, say $\mathbf{x} - \mathbf{x}' = \mathbf{x}_0 \in \mathcal{N}(A)$. Hence every solution has the form $\mathbf{x} = \mathbf{x}' + \mathbf{x}_0$ for some $\mathbf{x}_0$. □

Here is a picture where the nullspace is a 2-dimensional plane living in $\mathbb{R}^n$, so the general solution is also a 2-dimensional plane:



7.2 Uniqueness of the Solution

Suppose that $\mathbf{b} \in \mathcal{C}(A)$, so the system $A\mathbf{x} = \mathbf{b}$ has a solution. In the previous section we saw that this solution has the same shape as the nullspace. Hence the solution is unique if and only if the nullspace is a single point. If A has shape[48] $m \times n$ and rank r, recall from the Rank-Nullity theorem that $\dim \mathcal{N}(A) = n - \dim \mathcal{R}(A) = n - r$. Hence

$$\begin{aligned} \text{the solution to } A\mathbf{x} = \mathbf{b} \text{ is unique} &\iff \mathcal{N}(A) = \{\mathbf{0}\}, \\ &\iff \dim N(A) = 0 \\ &\iff r = n \\ &\iff A \text{ has independent rows} \\ &\iff A \text{ has a left inverse.} \end{aligned}$$

Indeed, suppose that $CA = I$ and $A\mathbf{x} = \mathbf{b}$. Then we must have

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ CA\mathbf{x} &= C\mathbf{b} \\ I\mathbf{x} &= C\mathbf{b} \\ \mathbf{x} &= C\mathbf{b}, \end{aligned}$$

so that $C\mathbf{b}$ is the **unique** solution.

[48] Here I am using the word "shape" for matrices and for subspaces. Don't take it too literally in either case.

7.3 How to Compute the Solution

Linear systems are solved using *row reduction*, also called *Gaussian elimination*. Gauss developed this method together with the method of least squares when he was 24, in order to determine the orbit of the dwarf planet Ceres. A similar method for solving linear systems was used in China since at least the 5th century AD.[49]

We will perform row reduction using elimination matrices, which were defined in the previous section. The goal is to put the system in a standardized simple form. Given a general matrix A, we first multiply on the left by **lower triangular** elimination matrices $L_{ij}(\lambda)$ (i.e., with $i > j$) until we obtain a matrix in "staircase form":

$$L_k \cdots L_2 L_1 A = \begin{pmatrix} * & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & * & \cdot & \cdot & \cdot \\ & & & & * & \cdot \\ & & & & & \end{pmatrix}.$$

Here the blank entries are zero. The entries labeled $*$ are **nonzero**; these are called the *pivots*. And the entries marked $\cdot$ are arbitrary. Next we multiply by dilation matrices $D_i(\lambda)$ to turn the pivot entries into 1s:

$$D_\ell \cdots D_2 D_1 L_k \cdots L_2 L_1 A = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & 1 & \cdot & \cdot & \cdot \\ & & & & 1 & \cdot \\ & & & & & \end{pmatrix}.$$

Finally, we multiply by **upper triangular** elimination matrices $L_{ij}(\lambda)$ (i.e., with $i < j$) to eliminate the entries above the pivots:

$$U_m \cdots U_2 U_1 D_\ell \cdots D_2 D_1 L_k \cdots L_2 L_1 A = \begin{pmatrix} 1 & \cdot & 0 & \cdot & 0 & \cdot \\ & & 1 & \cdot & 0 & \cdot \\ & & & & 1 & \cdot \\ & & & & & \end{pmatrix}.$$

Finally, this is called the *reduced row echelon form* (or RREF) of A. It has the virtue of being **unique**, i.e., independent of the particular order of row operations.[50]

We can summarize this process as follows. Multiply the elementary matrices together to obtain

$$L := L_k \cdots L_2 L_1, \quad D := D_\ell \cdots D_2 D_1 \quad \text{and} \quad U := U_m \cdots U_2 U_1.$$

The names indicate that L is lower trianglular (i.e., has zeros above the diagonal), D is diagonal (i.e., has zeros away from the diagonal) and U is upper triangular (i.e., has zeros

[49] The Chinese method was concerned with **integer solutions**, and is the precursor of the Chinese Remainder Theorem in abstract algebra.

[50] We will not prove this uniqueness because it is a bit tricky, and we will never need it.

below the diagonal). Furthermore, let's define $E = UDL$, which is invertible because it is a product of invertible matrices. Let R denote the RREF of A, so that

$$EA = R.$$

Since E is invertible, it follows from the section on the Fundamental Theorem that R has the same row space and nullspace as A:

$$\mathcal{R}(R) = \mathcal{R}(A) \quad \text{and} \quad \mathcal{N}(R) = \mathcal{N}(A).$$

In other words, the **homogeneous** system equation $A\mathbf{x} = \mathbf{0}$ is equivalent to $R\mathbf{x} = \mathbf{0}$, and the solution of this second system is particularly easy to read off. To solve the **non-homogeneous** system $A\mathbf{x} = \mathbf{b}$ we simply multiply both sides on the left by E to obtain

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ EA\mathbf{x} &= E\mathbf{b} \\ R\mathbf{x} &= E\mathbf{b}, \end{aligned}$$

and the solution is again easy to read off.

Example. Solve the linear system

$$\begin{cases} x + 3y + 8z = 2, \\ x + 2y + 6z = 1, \\ 0 + y + 2z = 1, \end{cases}$$

which can be expressed in matrix notation as

$$\begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix},$$
$$A\mathbf{x} = \mathbf{b}.$$

First we perform down elimination on A:

$$\begin{pmatrix} 1 & & \\ -1 & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{pmatrix}$$
$$\begin{pmatrix} 1 & & \\ & 1 & \\ & +1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$$

Next we scale the pivots:

$$\begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 8 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then we perform up elimination:[51]

$$\begin{pmatrix} 1 & -3 & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

The single matrix that performs the elimination is

$$\begin{aligned} E &= UDL \\ &= \left[\begin{pmatrix} 1 & -3 & \\ & 1 & \\ & & 1 \end{pmatrix} \right] \left[\begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \right] \left[\begin{pmatrix} 1 & & \\ & 1 & \\ & +1 & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ -1 & 1 & \\ & & 1 \end{pmatrix} \right] \\ &= \begin{pmatrix} 1 & -3 & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ -1 & 1 & \\ -1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix}. \end{aligned}$$

Check:

$$EA = R,$$
$$\begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

To solve the homogeneous system $A\mathbf{x} = \mathbf{0}$ we multiply both sides by E:

$$\begin{gathered} A\mathbf{x} = \mathbf{0} \\ EA\mathbf{x} = E\mathbf{0} \\ R\mathbf{x} = \mathbf{0} \\ \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \end{gathered}$$

This is equivalent to the linear system

$$\begin{cases} x + 0 + 2z = 0, \\ 0 + y + 2z = 0, \\ 0 + 0 + 0 = 0. \end{cases}$$

Note that the third equation is redundant, which shows that our original system of three equations really only contains two equations. The solution, which is also called the nullspace

[51] In class I circled pivots and drew arrows, which is extremely difficult to do in LaTeX.

of A, is a line:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2z \\ -2z \\ z \end{pmatrix} = z \begin{pmatrix} -2 \\ -2 \\ 1 \end{pmatrix}.$$

To solve the **non-homogeneous system** $A\mathbf{x} = \mathbf{b}$ we again multiply both sides by E:

$$\begin{gathered} A\mathbf{x} = \mathbf{b} \\ EA\mathbf{x} = E\mathbf{b} \\ R\mathbf{x} = E\mathbf{b} \\ \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}. \end{gathered}$$

This is equivalent to the linear system

$$\begin{cases} x + 0 + 2z = -1, \\ 0 + y + 2z = 1, \\ 0 + 0 + 0 = 0, \end{cases}$$

whose solution is a line parallel to the null space:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 - 2z \\ 1 - 2z \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} -2 \\ -2 \\ 1 \end{pmatrix}.$$

In the language of 5.1, $\mathbf{x}_0 = z(-2, -2, 1)$ is the general homogeneous solution and $\mathbf{x}' = (-1, 1, 0)$ is one particular solution. Note that there are infinitely many equivalent ways to describe this solution. For example, we can also write

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} + t \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}.$$

On the other hand, the following system has **no solution** because $(1, 0, 0)$ is not in the column space of A:

$$\begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

If we **try** to solve the system then we obtain

$$\begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix},$$

which is equivalent to the system

$$\begin{cases} x + 0y + 2z = -2, \\ 0x + y + 2z = 1, \\ 0x + 0y + 0z = -1. \end{cases}$$

This system has no solution because the third equation $0x + 0y + 0z = -1$ has no solution.

7.4 How to Compute the Inverse of a Square Matrix

We have seen a method for solving linear systems. Now we apply this method to compute the inverse of a square matrix. Let A be an invertible $n \times n$ matrix, and let E be the product of elementary matrices that puts A in reduced row echelon form: $EA = R$. Since A is invertible it has independent rows, and, since $\mathcal{R}(A) = \mathcal{R}(R)$, this implies that R has independent rows. In particular, R has no zero rows, which finally implies that R is the identity matrix. Summary:

$$\boxed{\text{The RREF of an invertible matrix } A \text{ is the identity matrix } I.}$$

This idea gives an algorithm to compute the inverse. Begin with the *augmented matrix*

$$\left(\begin{array}{c|c} A & I \end{array}\right).$$

Then apply elementary matrices on the left to put A in RREF:

$$\begin{aligned} &\left(\begin{array}{c|c} A & I \end{array}\right) \\ \rightsquigarrow &\left(\begin{array}{c|c} E_1A & E_1I \end{array}\right) \\ \rightsquigarrow &\left(\begin{array}{c|c} E_2E_1A & E_2E_1I \end{array}\right) \\ &\vdots \\ \rightsquigarrow &\left(\begin{array}{c|c} E_k\cdots E_2E_1A & E_k\cdots E_2E_1I \end{array}\right) \\ = &\left(\begin{array}{c|c} EA & E \end{array}\right) \\ = &\left(\begin{array}{c|c} R & E \end{array}\right). \end{aligned}$$

If A is invertible, so that $R = I$ and $E = A^{-1}$ then the process gives

$$\left(\begin{array}{c|c} A & I \end{array}\right) \overset{\text{RREF}}{\rightsquigarrow} \left(\begin{array}{c|c} I & A^{-1} \end{array}\right).$$

We don't even need to keep track of the elementary matrices.

Example.

$$\left(\begin{array}{c|c} A & I \end{array}\right)$$

$$
\begin{aligned}
&= \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & -1 & -1 & -1 & 0 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & -2 & -2 & 1 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right) \\
&= \left(\begin{array}{c|c} I & A^{-1} \end{array}\right).
\end{aligned}
$$

Check:

$$
\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1/2 & -1/2 \\ 1 & -1/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
$$

Recall that for square matrices A and B we have $AB = I$ if and only if $BA = I$ so we only need to check one.

What happens if we try to invert a non-invertible matrix? Consider the matrix A from Section 5.3. We perform elimination until we reach the RREF:

$$
\begin{aligned}
&\left(\begin{array}{ccc|ccc} 1 & 3 & 8 & 1 & 0 & 0 \\ 1 & 2 & 6 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right) \\
&= \left(\begin{array}{c|c} A & I \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{c|c} EA & E \end{array}\right)
\end{aligned}
$$

$$= \left(\begin{array}{c|c} R & E \end{array} \right)$$
$$= \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & -2 & 3 & 0 \\ 0 & 1 & 2 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{array} \right)$$

And then we're stuck.

8 Least Squares Approximation

8.1 The Four Fundamental Subspaces

Let me summarize our results so far. To each $m \times n$ matrix A we associate four subspaces; two of $\mathbb{R}^m$ and two of $\mathbb{R}^n$:

$$\mathcal{R}(A), \mathcal{N}(A) \subseteq \mathbb{R}^n \quad \text{and} \quad \mathcal{C}(A), \mathcal{N}(A^T) \subseteq \mathbb{R}^m.$$

The subspaces $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are orthogonal complements in $\mathbb{R}^n$, while $\mathcal{C}(A)$ and $\mathcal{N}(A^T)$ are orthogonal complements in $\mathbb{R}^m$.[52] It follows from the general theorem on dimensions of orthogonal complements[53] that

$$\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = n \quad \text{and} \quad \dim \mathcal{C}(A) + \dim N(A^T) = m.$$

These results are called the Rank-Nullity Theorem. The Fundamental Theorem says that the rank of A is well-defined:

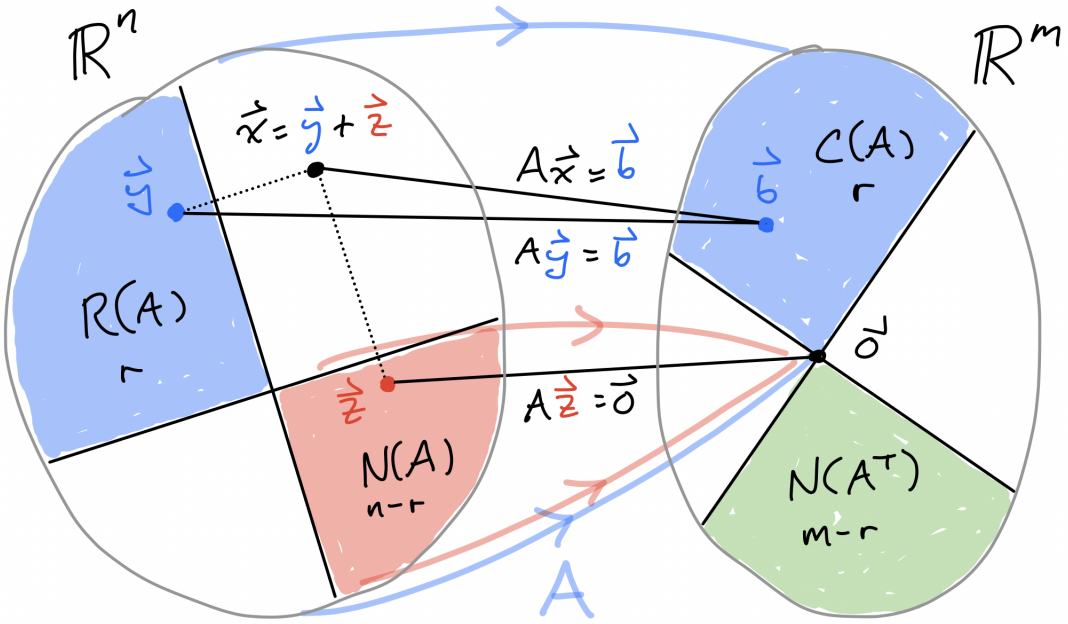
$$r = \text{rank}(A) := \dim \mathcal{R}(A) = \dim \mathcal{C}(A).$$

Hence we also have

$$\dim \mathcal{N}(A) = n - r \quad \text{and} \quad \dim \mathcal{N}(A^T) = m - r.$$

Here is "the big picture" in the style of Gilbert Strang:[54]

[52] Remind yourself right now why this is true.
[53] See the homework.
[54] A similar picture appears on the cover of his 4th edition of *Introduction to Linear Algebra*.



The matrix A maps the space $\mathbb{R}^n$ on the left to the space $\mathbb{R}^m$ on the right. Actually, A maps all of $\mathbb{R}^n$ onto the blue column space $\mathcal{C}(A)$. The red nullspace $\mathcal{N}(A)$ gets squashed onto the origin $\mathbf{0} \in \mathbb{R}^m$. Any vector $\mathbf{x} \in \mathbb{R}^n$ can be expressed uniquely as $\mathbf{x} = \mathbf{y} + \mathbf{z}$ with $\mathbf{y} \in \mathcal{R}(A)$ and $\mathbf{z} \in \mathcal{N}(A)$. If $A\mathbf{y} = \mathbf{b}$ then we also have $A\mathbf{x} = \mathbf{b}$ because

$$A\mathbf{x} = A(\mathbf{y} + \mathbf{z}) = A\mathbf{y} + A\mathbf{z} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

This picture is rather impressionistic but it does a good job of showing a lot of information. One thing it doesn't show is the set of all solutions to the equation $A\mathbf{x} = \mathbf{b}$, which is an affine subspace of $\mathbb{R}^n$ that is parallel to $N(A)$ and passes through $\mathbf{x}$ and $\mathbf{y}$. I guess that would make the picture unreadable.

Next we work through an explicit example. Consider the rank 2 matrix

$$A = \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}.$$

In Section 5.3 we already computed the nullspace:

$$\mathcal{N}(A) = \text{ the line in } \mathbb{R}^3 \text{ spanned by } (2, 2, -1).$$

The rowspace is the orthogonal complement of the nullspace, which is a plane:

$$\mathcal{R}(A) = \mathcal{N}(A)^\perp = \text{the plane in } \mathbb{R}^3 \text{ defined by } 2x + 2y - z = 0.$$

Since no two rows of A are parallel, any two rows will form a basis for $\mathcal{R}(A)$. More systematically, we can look at the RREF:

$$EA = R,$$
$$\begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Since the product of elementary matrices E is invertible, we know that $\mathcal{R}(A) = \mathcal{R}(EA) = \mathcal{R}(R)$, and it is very easy to read a basis from R:

$$\mathcal{R}(A) = \mathcal{R}(R) = \text{Span}\{(1,0,2),(0,1,2)\}.$$

To compute the column space $\mathcal{C}(A)$ and left nullspace $\mathcal{N}(A^T)$ we can apply the same methods to the transposed matrix A^T. That is, we should compute $RREF(A^T)$:[55]

$$\begin{pmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \\ 8 & 6 & 2 \end{pmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

From this we see that

$$\mathcal{C}(A) = \mathcal{R}(A^T) = \text{Span}\{(1,0,1),(0,1,-1)\}.$$

Finally, the left nullspace is the solution to the homogeneous system $A^T\mathbf{x} = \mathbf{0}$, which from the RREF of A^T is equivalent to

$$\begin{cases} x + 0 + +z = 0, \\ 0 + y + -z = 0, \\ 0 + 0 + 0 = 0. \end{cases}$$

The solution is the line spanned by $(1,-1,-1)$:

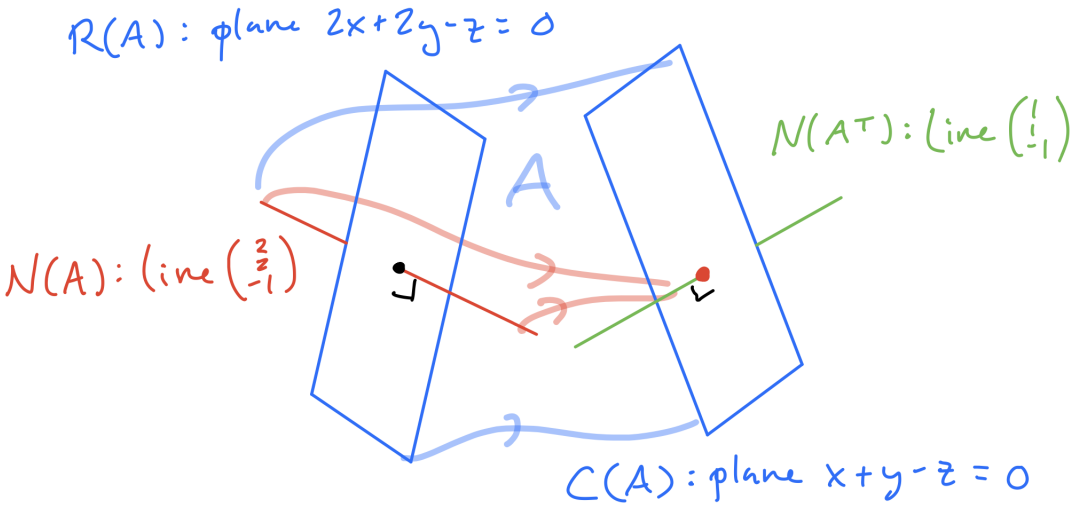
$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -z \\ z \\ z \end{pmatrix} = z \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = \text{Span}\{(1,-1,-1)\}.$$

As expected, this line is the orthogonal complement of the column space:

$$\text{Span}\{(1,-1,-1)\}^\perp = (\text{plane } x - y - z = 0) = \text{Span}\{(1,0,1),(0,1,-1)\} = \mathcal{C}(A).$$

Here is a picture:

[55] This is equivalent to applying elementary matrices on the right of A to compute *reduced column echelon form* (RCEF), but nobody uses this terminology.



Note that the line $\mathcal{N}(A)$ gets squashed onto the origin, while all of $\mathbb{R}^3$ gets squashed onto the plane $\mathcal{C}(A)$. Since this matrix is square, we could have drawn all four subspaces in the same copy of $\mathbb{R}^3$, but that would just be a mess.

In summary:

The four fundamental subspaces can be read off from RREF(A) and RREF(A^T).

8.2 The Matrices A^TA and AA^T

We have seen that a non-square matrix A cannot have an inverse. To fix this we sometimes consider the square matrices A^TA and AA^T. To be precise, suppose that A has shape $m \times n$, so that A^TA is square of shape $n \times n$ and AA^T is square of shape $m \times m$. We also observe that these matrices are *symmetric* because

$$(A^TA)^T = A^T(A^T)^T = A^TA$$

and

$$(AA^T)^T = (A^T)^TA^T = AA^T.$$

The matrices A^TA and AA^T show up surprisingly often in applied mathematics. We will see our first glimpse of this in the next section when we discuss least squares approximation. To prepare for this we develop some basic properties. The key observation is that A and A^TA have **the same nullspace**:

$$\boxed{\mathcal{N}(A^TA) = \mathcal{N}(A).}$$

This would be easy to prove if A^T had a left inverse. Indeed, if E is a matrix with a left inverse $E'E = I$ then we recall from Section 3 that

$$\mathcal{R}(EA) = \mathcal{R}(A),$$

and hence

$$\mathcal{N}(EA) = \mathcal{R}(EA)^\perp = \mathcal{R}(A)^\perp = \mathcal{N}(A).$$

But the matrix A^T might **not** have a left inverse, so we cannot use this fact. Instead we use a clever trick:[56]

$$\boxed{\text{For any } \mathbf{x} \in \mathbb{R}^n \text{ we have } \mathbf{x}^T A^T A\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = (A\mathbf{x}) \bullet (A\mathbf{x}) = \|A\mathbf{x}\|^2.}$$

Proof that $\mathcal{N}(A^TA) = \mathcal{N}(A)$. First we note that $\mathcal{N}(A) \subseteq \mathcal{N}(A^TA)$ because

$$A\mathbf{x} = \mathbf{0} \quad \Longrightarrow \quad (A^TA)\mathbf{x} = A^T(A\mathbf{x}) = A^T\mathbf{0} = \mathbf{0}.$$

On the other hand, suppose that $(A^TA)\mathbf{x} = \mathbf{0}$. Then from the trick we have

$$\|A\mathbf{x}\|^2 = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T(A^TA\mathbf{x}) = \mathbf{x}^T\mathbf{0} = 0,$$

and hence $\|A\mathbf{x}\| = 0$. But recall that the standard norm $\| - \|$ satisfies $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$. Hence we must have $A\mathbf{x} = \mathbf{0}$ as desired. □

We obtain a similar identity by replacing A with A^T. To be precise, let $B = A^T$, so that

$$\boxed{\mathcal{N}(AA^T) = \mathcal{N}(B^TB) = \mathcal{N}(B) = \mathcal{N}(A^T).}$$

And it follows from these identities that

$$\boxed{\text{rank}(A^TA) = \text{rank}(A) = \text{rank}(A^T) = \text{rank}(AA^T).}$$

Indeed, the first and third equations follow by applying dimension to the identities $\mathcal{N}(A^TA) = \mathcal{N}(A)$ and $\mathcal{N}(AA^T) = \mathcal{N}(A^T)$, while the middle equation is just the Fundamental Theorem. This is quite interesting since the four matrices A, A^T, A^TA and AA^T have different shapes.

We combine these results to prove the main result of this section.

Theorem (Invertibility of A^TA and AA^T). For any matrix A, the matrices A^TA and AA^T are square, hence they might be invertible. I claim that

$$\begin{aligned} (A^TA)^{-1} \text{ exists} \quad &\Longleftrightarrow \quad A \text{ has independent columns,} \\ (AA^T)^{-1} \text{ exists} \quad &\Longleftrightarrow \quad A \text{ has independent rows.} \end{aligned}$$

Proof. Let A have shape $m \times n$ and rank r. To prove the first statement, note that A^TA has shape $n \times n$, hence

$$(A^TA)^{-1} \text{ exists} \quad \Longleftrightarrow \quad \text{rank}(A^TA) = n$$

[56] The idea lurking in the background is that matrices of the form A^TA are related to inner products. See Problem 5 on Homework 3.

$$\begin{aligned} &\iff && \operatorname{rank}(A) = n && \text{previous result} \\ &\iff && \dim \mathcal{C}(A) = n \\ &\iff && A \text{ as independent columns.} \end{aligned}$$

Similarly, since AA^T has shape $m \times m$, we have

$$\begin{aligned} (AA^T)^{-1} \text{ exists} &\iff && \operatorname{rank}(AA^T) = m \\ &\iff && \operatorname{rank}(A) = m && \text{previous result} \\ &\iff && \dim \mathcal{R}(A) = m \\ &\iff && A \text{ as independent rows.} \end{aligned}$$

□

To end this section we give two theoretical applications.[57]

Explicit formulas for left and right inverses. For any matrix A we recall from 4.1 that

$$\begin{aligned} A \text{ has a left inverse} &\iff A \text{ has independent columns,} \\ A \text{ has a right inverse} &\iff A \text{ has independent rows.} \end{aligned}$$

Such left and right inverses are **not unique**, but we can use the previous theorem to give a formula for **specific** left and right inverse. If A has independent columns then $(A^TA)^{-1}$ exists and $(A^TA)^{-1}A^T$ is a left inverse:

$$[(A^TA)^{-1}A^T]A = (A^TA)^{-1}(A^TA) = I.$$

If A has independent rows then $(AA^T)^{-1}$ exists and $A^T(AA^T)^{-1}$ is a right inverse:

$$A[A^T(AA^T)^{-1}] = (AA^T)(AA^T)^{-1} = I.$$

CMR Factorization. Applied linear algebra is often expressed in terms of matrix factorizations. Here we will show that any $m \times n$ matrix A of rank r can be factored as $A = CMR$, where the matrices C, M and R have shapes $m \times r$, $r \times r$ and $r \times n$. The matrices C and R are defined as follows:

- Choose any r independent columns of A and let these be the columns of C.
- Choose any r independent rows of A and let these be the rows of R.

By construction, C has independent columns and R has independent rows, so the matrices $(C^TC)^{-1}$ and $(RR^T)^{-1}$ exist. In this case we will show that **there exists a unique** $r \times r$ **matrix** M **satisfying** $A = CMR$. This matrix is invertible and is determined by the formula

$$M = (C^TC)^{-1}(C^TAR^T)(RR^T)^{-1}.$$

It is difficult to see that the matrix defined by this formula has the desired properties, so we will proceed in two steps:

[57] The section on Least Squares below gives some practical applications.

(1) There exists an invertible matrix M satisfying $A = CMR$.

(2) The matrix from part (1) must satisfy the desired formula.

The proof of (1) is tricky and algorithmic.[58] Feel free to skip it.

(1): First let T be an invertible product of column transpositions so that the first r columns of AT are equal to C; let's say

$$AT = \left(\begin{array}{c|c} C & F \end{array}\right),$$

for some $m \times (n-r)$ matrix F. Next we consider the reduced row echelon form of AT. Let E be an invertible product of elementary row operations satisfying $EAT = \mathrm{RREF}(AT)$. Since the first r columns of AT (i.e., the columns of C) are independent, so will be the first r columns of the RREF, and it follows that

$$EAT = \mathrm{RREF}(AT) = \left(\begin{array}{c|c} I_r & G \\ \hline O_{m-r,r} & O_{m-r,n-r} \end{array}\right),$$

for some $r \times (n-r)$ matrix G. I claim that $AT = C\left(\begin{array}{c|c} I & G \end{array}\right)$. Indeed, if we write $E^{-1} = \left(\begin{array}{c|c} X & Y \end{array}\right)$ where X is $m \times r$ and Y is $m \times (m-r)$ then we find

$$\left(\begin{array}{c|c} C & F \end{array}\right) = AT = E^{-1}\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} X & Y \end{array}\right)\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} X & YG \end{array}\right),$$

which implies that $X = C$.[59] It follows that

$$AT = E^{-1}\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} C & Y \end{array}\right)\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} C & CG \end{array}\right) = C\left(\begin{array}{c|c} I & G \end{array}\right).$$

At this point we have

$$A = C\left(\begin{array}{c|c} I & G \end{array}\right)T^{-1} = CR',$$

where we have defined $R' := \left(\begin{array}{c|c} I & G \end{array}\right)T^{-1}$. Our final goal is to prove that $R' = MR$ for some invertible M. Since C has independent columns (and hence has a left inverse) we see from Section 3 that A and R' have the same row space:

$$\mathcal{R}(A) = \mathcal{R}(CR') = \mathcal{R}(R'),$$

Since this row space is r-dimensional, and since R' has r rows, it follows that the rows of R' are a basis for $\mathcal{R}(A)$. In particular, each row of R can be expressed as a linear combination of the rows of R', which gives a matrix equation $R = MR'$. Similarly, since the rows of R are a basis for $\mathcal{R}(A)$ we can write $R' = NR$ for some matrix N. Putting these together gives

[58] I apologize that I assigned this as homework; I didn't realize how tricky it is. Gilbert Strang fooled me. Maybe there is a more direct proof but I couldn't find it.

[59] We also have $YG = F$, but we don't care about this.

$R = MNR$. Finally, since R has a right inverse this implies $MN = I$, which implies that M is invertible.[60] □

(2): Once we know that M exists, it is not difficult to prove that satisfies the desired formula. Indeed, suppose that $A = CMR$. Then since $(C^TC)^{-1}$ and $(RR^T)^{-1}$ exist we must have

$$\begin{aligned} CMR &= A \\ C^T(CMR)R^T &= C^TAR^T \\ (C^TC)M(RR^T) &= C^TAR^T \\ M &= (C^TC)^{-1}C^TAR^T(RR^T)^{-1}. \end{aligned}$$

□

For example, let's consider our favorite matrix

$$A = \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}.$$

This matrix has rank 2, so we should choose two independent columns and two independent rows. Choosing the first two columns and the first two rows gives

$$A = \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \end{pmatrix}.$$

Choosing columns $1, 3$ and rows $2, 3$ gives

$$A = \begin{pmatrix} 1 & 8 \\ 1 & 6 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & -3 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}.$$

Remark: There is another interesting description of the matrix M. In the paper *LU and CR Elimination* by Strang and Moler,[61] they prove that M^{-1} is the matrix obtained from A by intersecting the columns of C with the rows of R. We observe that this is true for the two examples just given:

$$\begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -3 \\ 0 & 1/2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 6 \\ 0 & 2 \end{pmatrix}.$$

Pretty cool.

[60] Recall that $MN = I$ implies $NM = I$ for square matrices.

[61] I think there's a better proof in Hamm and Huang, `https://arxiv.org/abs/1907.12668`. I need to look into it.

8.3 Least Squares Approximation

We have seen that a linear system $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$ if and only if $\mathbf{b}$ is in the column space of A. In fact, this statement is just the definition of the column space:

$$\begin{aligned}\mathcal{C}(A) &= \{\text{all linear combinations of the columns of } A\},\\ &= \{\text{all vectors of the form } A\mathbf{x} \text{ for some } \mathbf{x}\}.\end{aligned}$$

What happens when $\mathbf{b}$ is not in the column space?

The Problem of Least Squares. Given an $m \times n$ matrix A and an $m \times 1$ column vector $\mathbf{b}$, find an $n \times 1$ column vector $\mathbf{x}$ such that the distance $\|A\mathbf{x} - \mathbf{b}\|$ is minimized.

Obviously a true solution $A\mathbf{x} = \mathbf{b}$ makes $\|A\mathbf{x} - \mathbf{b}\| = 0$. When $\mathbf{b} \notin \mathcal{C}(A)$, the minimum value of $\|A\mathbf{x} - \mathbf{b}\|$ will be strictly positive. The problem is called *least squares approximation* since the length $\|A\mathbf{x} - \mathbf{b}\|$ is minimized if and only if the squared length $\|A\mathbf{x} - \mathbf{b}\|^2$ is minimized, and the squared length is a sum of squares:[62]

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \left\| \begin{pmatrix} \mathbf{a}_1^T\mathbf{x} - b_1 \\ \vdots \\ \mathbf{a}_m^T\mathbf{x} - b_m \end{pmatrix} \right\|^2 = (\mathbf{a}_1^T\mathbf{x} - b_1)^2 + \cdots + (\mathbf{a}_m^T\mathbf{x} - b_m)^2,$$

where $\mathbf{a}_i^T$ is the ith row of A and $\mathbf{b} = (b_1, \ldots, b_m)$. There are two ways to solve this problem:

(1) Calculus

(2) Linear Algebra

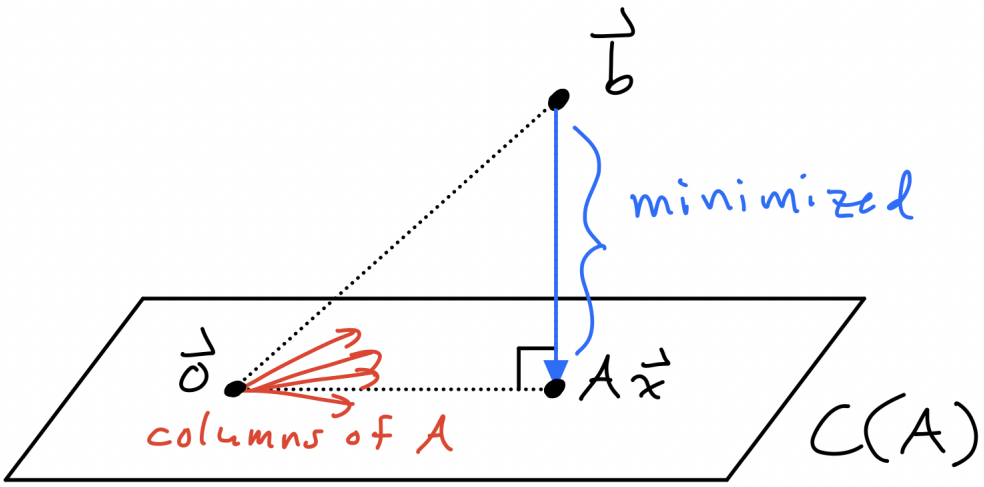
The calculus solution uses the typical method of Lagrange multipliers. This solution is more common in textbooks because every student knows calculus, whereas not every student knows linear algebra. However, the linear algebra solution is conceptually much simpler and is easier to generalize.

The key idea is to view $\|A\mathbf{x} - \mathbf{b}\|$ as the distance between two points in $\mathbb{R}^n$. The expression $A\mathbf{x}$ represents a general point of the column space, while $\mathbf{b}$ is a point that is not in the column space. Here is a picture:

[62] There are certainly other ways to define a "best approximate solution". For example, one could try to minimize the sum of absolute values:

$$|\mathbf{a}_1^T\mathbf{x} - b_1| + \cdots + |\mathbf{a}_m^T\mathbf{x} - b_m|.$$

This is a reasonable idea, but the mathematics is much more difficult. We will see some other methods of approximation after we discuss the singular value decomposition.



For geometric reasons[63] we see that the length of the blue vector $A\mathbf{x} - \mathbf{b}$ is minimized when it is perpendicular to the column space. Since the orthogonal complement of $\mathcal{C}(A)$ is $\mathcal{N}(A^T)$, this happens precisely when

$$\boxed{A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}.} \qquad (*)$$

Since you might not remember the details, I will repeat them one more time.[64] Let $\mathbf{a}_i$ be the ith column of A, so that $\mathbf{a}_i^T$ is the ith row of A^T. To say that $A\mathbf{x} - \mathbf{b}$ is perpendicular to the column space means that $A\mathbf{x} - \mathbf{b}$ is perpendicular to each column. That is, we must have

$$\mathbf{a}_i^T(A\mathbf{x} - \mathbf{b}) = \mathbf{a}_i \bullet (A\mathbf{x} - \mathbf{b}) = 0 \quad \text{for all } i.$$

But this is equivalent to saying that $A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$ because

$$A^T(A\mathbf{x} - \mathbf{b}) = \begin{pmatrix} - & \mathbf{a}_1^T & - \\ & \vdots & \\ - & \mathbf{a}_m^T & - \end{pmatrix} (A\mathbf{x} - \mathbf{b}) = \begin{pmatrix} \mathbf{a}_1^T(A\mathbf{x} - \mathbf{b}) \\ \vdots \\ \mathbf{a}_m^T(A\mathbf{x} - \mathbf{b}) \end{pmatrix},$$

which is the zero vector if and only if each component $\mathbf{a}_i^T(A\mathbf{x} - \mathbf{b})$ is zero.

We may proceed to solve equation $(*)$ which is called the *normal equation*:[65]

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

[63] Ultimately, this follows from the triangle inequality.

[64] David Hilbert said that every idea must be repeated five times before the students will remember it. See the very interesting biography of Hilbert by Constance Reid.

[65] The word normal here indicates that $A\mathbf{x} - \mathbf{b}$ is perpendicular to the columns of A.

Whereas the equation $A\mathbf{x} = \mathbf{b}$ did not have a solution, it is worth noting that the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$ always has a solution. To see this, we only need to check that $A^T\mathbf{b}$ is in the column space $\mathcal{C}(A^TA)$. In the previous section on the matrices A^TA and AA^T we proved the key fact that $\mathcal{N}(A^TA) = \mathcal{N}(A)$, which implies that

$$\mathcal{R}(A^TA) = \mathcal{N}(A^TA)^\perp = \mathcal{N}(A)^\perp = \mathcal{R}(A).$$

But then we must have

$$\mathcal{C}(A^TA) = \mathcal{R}((A^TA)^T) = \mathcal{R}(A^TA) = \mathcal{R}(A) = \mathcal{C}(A^T).$$

This implies that any vector in the column space of A^T, for example $A^T\mathbf{b}$, is in the column space of A^TA, so can be expressed in the form $A^TA\mathbf{x}$.

In general, suppose that A has shape $m \times n$ and rank r. Then the solution of the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$ is an affine subspace of $\mathbb{R}^n$ that is parallel to the nullspace $\mathcal{N}(A^TA) = \mathcal{N}(A)$, and so has dimension $n - r$. This solution will be unique if and only if $r = n$, i.e., if and only if A has independent columns. In this case we know from the previous section that $(A^TA)^{-1}$ exists, and hence the unique least squares solution has a symbolic form:

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

Here is a summary:

- If $\mathbf{b} \in \mathcal{C}(A)$ then the system $A\mathbf{x} = \mathbf{b}$ has an exact solution.
- If $\mathbf{b} \notin \mathcal{C}(A)$ then the system $A\mathbf{x} = \mathbf{b}$ does not have an exact solution.
- The length $\|A\mathbf{x} - \mathbf{b}\|$ is minimized when $A\mathbf{x} - \mathbf{b}$ is perpendicular to $\mathcal{C}(A)$.
- This happens if and only if $A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$, or $A^TA\mathbf{x} = A^T\mathbf{b}$.
- The *normal equation* $A^TA\mathbf{x} = A^T\mathbf{b}$ always has a solution.
- If A has independent columns then A^TA is invertible, so the solution is unique:

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}.$$

We often use a different notation such as $\hat{\mathbf{x}}$ to denote the least squares solution $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$, to distinguish it from an exact solution $A\mathbf{x} = \mathbf{b}$. However, if there exists an exact solution $A\mathbf{x} = \mathbf{b}$, then we note that $\hat{\mathbf{x}} = \mathbf{x}$ since multiplying both sides on the left gives

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A^TA\mathbf{x} &= A^T\mathbf{b}. \end{aligned}$$

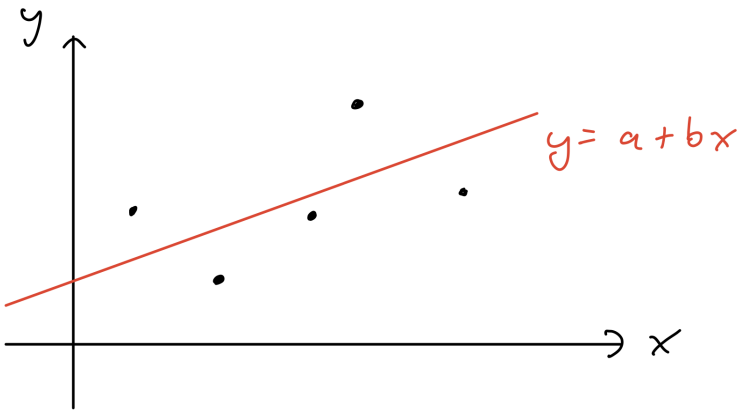
8.4 Examples of Least Squares

The classical application of least squares is to curve fitting. Indeed, this is the purpose for which Gauss invented the method.[66]

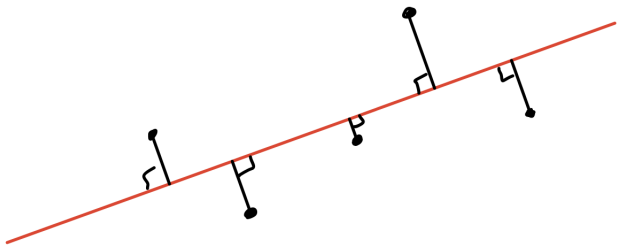
Curve Fitting. Suppose that we have a collection of n data points in the x, y-plane:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

We would like to find the line of the form $y = a + bx$ that is the "best fit" for these points:

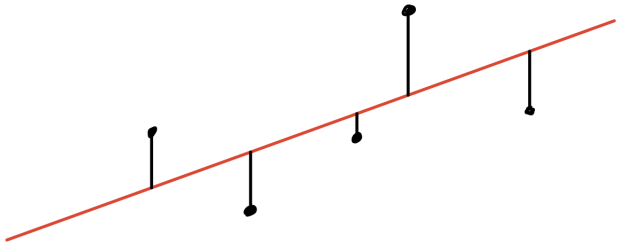


There are different ways one might interpret the word "best". The most obvious definition might be to minimize the orthogonal distances[67] from the points to the line:



This idea is called *total least squares*, or *orthogonal least squares*. It is a hard non-linear problem, which we will solve after discussing the singular value decomposition. In statistics this problem is called *principal component analysis*. It is much easier to minimize the sum of squares of the vertical distances:

[66] He used it to fit the elliptical orbit of the dwarf Planet Ceres to a collection of observed data points.
[67] Typically we want to minimize the sum of squared distances.



This problem is called *ordinary least squares*, or just least squares regression.

Here's how we solve it. We start by being optimistic and assuming that all of the data points fit perfectly on the line, which leads to a system of n linear equations in the two unknowns a and b:

$$\begin{cases} a + bx_1 = y_1, \\ a + bx_2 = y_2, \\ \vdots \\ a + bx_n = y_n, \end{cases}$$

It is an unfortunate feature of curve fitting problems that the roles of variables and constants get switched around, so instead of a system looking like $A\mathbf{x} = \mathbf{b}$ we get a system looking like $X\mathbf{a} = \mathbf{y}$. In our case we have

$$X\mathbf{a} = \mathbf{y}$$
$$\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

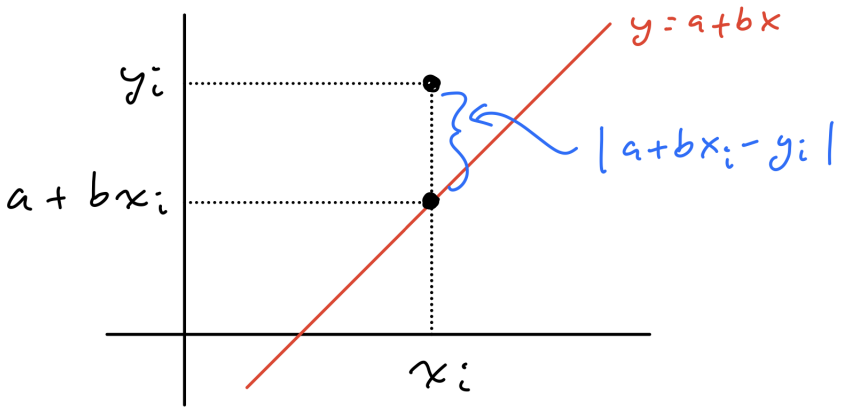
However, this system almost certainly does not have a solution, since any three or more points almost certainly do not fit perfectly on a straight line. Hence we will apply the method of least squares. If the data points do not all have the same x value, then the two columns of X are are independent and we get a unique solution:

$$\begin{aligned} X\mathbf{a} &= \mathbf{y} \\ X^T X\mathbf{a} &= X^T\mathbf{y} \\ \mathbf{a} &= (X^T X)^{-1} X\mathbf{y}. \end{aligned}$$

Recall that this "least squares solution" minimizes the length $\|X\mathbf{a}-\mathbf{y}\|$, hence it also minimizes the squared length $\|X\mathbf{a} - \mathbf{y}\|^2$. In terms of the data points, this becomes

$$\|X\mathbf{a} - \mathbf{y}\|^2 = \left\| \begin{pmatrix} a + bx_i - y_i \\ \vdots \\ a + bx_n - y_n \end{pmatrix} \right\|^2 = \sum (a + bx_i - y_i)^2,$$

which is, indeed, the sum of the squared vertical errors:



To be explicit, the normal equation has the following form, which you might recognize:

$$X^T X\mathbf{a} = X^T\mathbf{y}$$

$$\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix},$$

which is equivalent to the linear system

$$\begin{cases} an & + & b\sum x_i & = & \sum y_i, \\ a\sum x_i & + & b\sum x_i^2 & = & \sum x_i y_i. \end{cases}$$

This is the form usually presented in introductory statistics courses, when the students don't know linear algebra.

However, the linear algebra formulation is much more powerful because it generalizes easily. For example, we can fit our data to polynomial curve of degree d:

$$y = a_0 + a_1 x + \cdots + \cdots + a_d x^d.$$

Assuming optimistically that all n data points lie on this curve gives a system of n linear equations in the $d+1$ unknown coefficients $a_0, \ldots, a_d$:

$$X\mathbf{a} = \mathbf{y}$$

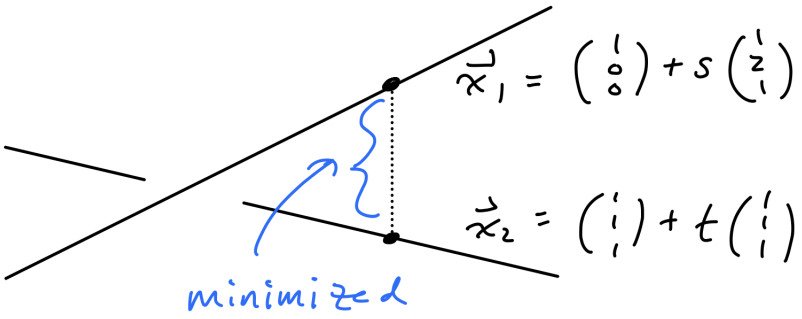
$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_d \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then the least squares solution (which minimizes the sum of squares of the vertical errors) is given by the normal equation $X^T X\mathbf{a} = X^T\mathbf{y}$. This equation is much harder to obtain using calculus, and the explicit formulas for the entries of the matrix $X^T X$ are not so nice.

Distance Between Subspaces. Consider the following parametrized lines in $\mathbb{R}^3$:

$$\begin{aligned} L_1: &\quad (1,0,0) + s(1,2,1), \\ L_2: &\quad (1,1,1) + t(1,1,1). \end{aligned}$$

These lines (probably) do not intersect. We would like to find points $\mathbf{x}_1 \in L_1$ and $\mathbf{x}_2 \in L_2$ such that the distance $\|\mathbf{x}_1 - \mathbf{x}_2\|$ is minimized:



We could solve this problem from scratch, but instead we will apply the general theory of least squares. First we assume, optimistically, that the lines intersect, so that

$$\begin{gathered} \mathbf{x}_1 = \mathbf{x}_2 \\ \begin{pmatrix}1\\0\\0\end{pmatrix} + s\begin{pmatrix}1\\2\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} + t\begin{pmatrix}1\\1\\1\end{pmatrix} \\ s\begin{pmatrix}1\\2\\1\end{pmatrix} - t\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix} \\ \begin{pmatrix}1 & -1\\2 & -1\\1 & -1\end{pmatrix}\begin{pmatrix}s\\t\end{pmatrix} = \begin{pmatrix}0\\1\\1\end{pmatrix}. \end{gathered}$$

Whether this system has an exact solution or not,[68] we can proceed by multiplying on the left by the transpose of the coefficient matrix:

$$\begin{pmatrix}1 & -1\\2 & -1\\1 & -1\end{pmatrix}\begin{pmatrix}s\\t\end{pmatrix} = \begin{pmatrix}0\\1\\1\end{pmatrix}$$

[68] If the system did, unexpectedly, have an exact solution, we would see this at the end.

$$\begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$
$$\begin{pmatrix} 6 & -4 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 6 & -4 \\ -4 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$\begin{pmatrix} s \\ t \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$= \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}.$$

The least squares solution $(s, t) = (1/2, 0)$ corresponds to the points

$$\mathbf{x}_1 = (1, 0, 0) + \frac{1}{2}(1, 2, 1) = (3/2, 1, 1/2) \quad \text{and} \quad \mathbf{x}_2 = (1, 1, 1) + 0(1, 1, 1) = (1, 1, 1).$$

But what exactly have we minimized here? Recall that the least squares solution of $A\mathbf{x} = \mathbf{b}$ minimizes the distance $\|A\mathbf{x} - \mathbf{b}\|$. In our case we have minimized the distance

$$\left\| \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - t \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\| = \|\mathbf{x}_1 - \mathbf{x}_2\|,$$

which is exactly what we wanted to do.

More generally, we can use this method to find the distance between any two affine subspaces living in $\mathbb{R}^n$. Recall that an *affine subspace* of $\mathbb{R}^n$ has the form

$$\mathbf{p} + U = \{\text{the set of points } \mathbf{p} + \mathbf{u} \text{ for all } \mathbf{u} \in U\},$$

where $\mathbf{p} \in \mathbb{R}^n$ is a point and $U \subseteq \mathbb{R}^n$ is a linear subspace (i.e., passing through $\mathbf{0}$). For the current discussion, it is convenient to represent a d-dimensional affine subspace as $\mathbf{p} + \mathcal{C}(A)$ for some $n \times d$ matrix A with independent columns. We can also express this as

$$\mathbf{p} + \mathcal{C}(A) = \{\text{the set of points } \mathbf{p} + A\mathbf{x} \text{ for all } \mathbf{x} \in \mathbb{R}^d\}.$$

Now let A and B be matrices of shapes $n \times d$ and $n \times e$, each with independent columns, and consider any two points $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. We want to find the distance between the following two subspaces:

$$\mathbf{a} + \mathcal{C}(A) = \{\text{the set of } \mathbf{a} + A\mathbf{x} \text{ for } \mathbf{x} \in \mathbb{R}^d\},$$
$$\mathbf{b} + \mathcal{C}(B) = \{\text{the set of } \mathbf{b} + B\mathbf{y} \text{ for } \mathbf{y} \in \mathbb{R}^e\}.$$

We begin optimistically, by assuming that $\mathbf{a}+\mathcal{C}(A)$ and $\mathbf{b}+\mathcal{C}(B)$ share a common point:

$$\begin{aligned}
\mathbf{a}+A\mathbf{x} &= \mathbf{b}+B\mathbf{y}\\
A\mathbf{x}-B\mathbf{y} &= \mathbf{b}-\mathbf{a}\\
\left(\begin{array}{c|c} A & -B \end{array}\right)\left(\frac{\mathbf{x}}{\mathbf{y}}\right) &= \mathbf{b}-\mathbf{a}\\
C\mathbf{z} &= \mathbf{c},
\end{aligned}$$

where the matrices C, $\mathbf{z}$ and $\mathbf{c}$ have shapes $n\times(d+e)$, $(d+e)\times 1$ and $n\times 1$, respectively. Next we multiply on the left by C^T to obtain

$$\begin{aligned}
C\mathbf{z} &= \mathbf{c}\\
C^TC\mathbf{z} &= C^T\mathbf{c}\\
\left(\frac{A^T}{-B^T}\right)\left(\begin{array}{c|c} A & -B \end{array}\right)\mathbf{z} &= \left(\frac{A^T}{-B^T}\right)\mathbf{c}\\
\left(\begin{array}{c|c} A^TA & -A^TB \\ \hline -B^TA & B^TB \end{array}\right)\mathbf{z} &= \left(\frac{A^T\mathbf{c}}{-B^T\mathbf{c}}\right).
\end{aligned}$$

The matrix C need not have independent columns. However, if the column spaces $\mathcal{C}(A)$ and $\mathcal{C}(B)$ have trivial intersection (i.e., if $\mathcal{C}(A)\cap\mathcal{C}(B)=\{\mathbf{0}\}$), then C **will** have independent columns.[69] In this case the inverse $(C^TC)^{-1}$ exists and we have a unique least squares solution:

$$\left(\frac{\mathbf{x}}{\mathbf{y}}\right)=\left(\begin{array}{c|c} A^TA & -A^TB \\ \hline -B^TA & B^TB \end{array}\right)^{-1}\left(\frac{A^T(\mathbf{b}-\mathbf{a})}{-B^T(\mathbf{b}-\mathbf{a})}\right).$$

To check that this makes sense, we consider the case when

$$\mathbf{a}=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad A=\begin{pmatrix}1\\2\\1\end{pmatrix},\quad \mathbf{b}=\begin{pmatrix}1\\1\\1\end{pmatrix},\quad B=\begin{pmatrix}1\\1\\1\end{pmatrix}.$$

This is just our previous example with $L_1=\mathbf{a}+\mathcal{C}(A)$ and $L_2=\mathbf{b}+\mathcal{C}(B)$. Then we have

$$\begin{aligned}
\left(\frac{\mathbf{x}}{\mathbf{y}}\right) &= \left(\begin{array}{c|c} A^TA & -A^TB \\ \hline -B^TA & B^TB \end{array}\right)^{-1}\left(\frac{A^T(\mathbf{b}-\mathbf{a})}{-B^T(\mathbf{b}-\mathbf{a})}\right)\\
&= \left(\begin{array}{c|c} \begin{pmatrix}1&2&1\end{pmatrix}\begin{pmatrix}1\\2\\1\end{pmatrix} & -\begin{pmatrix}1&2&1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} \\ \hline -\begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}1\\2\\1\end{pmatrix} & \begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} \end{array}\right)^{-1}\left(\begin{array}{c} \begin{pmatrix}1&2&1\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix} \\ \hline -\begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix} \end{array}\right)
\end{aligned}$$

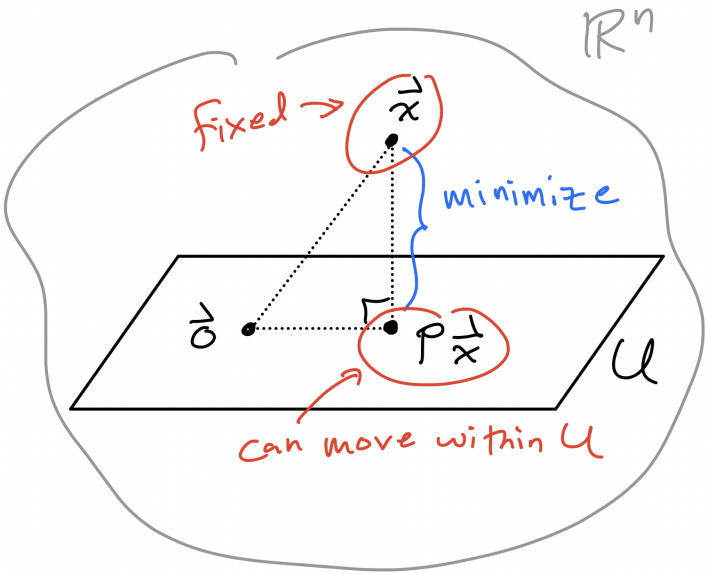
[69] This is a bit tricky so we omit the proof.

$$= \left(\begin{array}{c|c} 6 & -4 \\ \hline -4 & 3 \end{array}\right)^{-1} \left(\begin{array}{c} 3 \\ \hline -2 \end{array}\right),$$
$$= \left(\begin{array}{c} 1/2 \\ \hline 0 \end{array}\right),$$

which is exactly what we had before.

8.5 Projection Matrices

When solving the least squares problem we implicitly solved the problem of projecting onto a (linear) subspace. Given a linear subspace $U \subseteq \mathbb{R}^n$ and a point $\mathbf{x} \in \mathbb{R}^n$ we want to find the point $\mathbf{y} \in U$ that is closest to $\mathbf{x}$. We will denote the point by $\mathbf{y} = P(\mathbf{x})$ and call it the *projection of* $\mathbf{x}$ *onto* U. Here is a picture:



It is not immediately obvious, but we will see that $P : \mathbb{R}^n \to \mathbb{R}^n$ is a linear function, hence it corresponds to an $n \times n$ matrix. The easiest way to find this matrix is to represent U as a column space. Suppose that $\dim U = d$ and let $\mathbf{a}_1, \ldots, \mathbf{a}_d \in U$ be any basis. Then we can form the $n \times d$ matrix

$$A = \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_d \\ | & & | \end{pmatrix} \quad \text{so that} \quad U = \mathcal{C}(A).$$

From geometric considerations (the triangle inequality) we see that the distance $\|P(\mathbf{x}) - \mathbf{x}\|$ is minimized when the vector $P(\mathbf{x}) - \mathbf{x}$ is perpendicular to U. And since $U^\perp = \mathcal{C}(A)^\perp = \mathcal{N}(A^T)$, we see that[70]

$$P(\mathbf{x}) - \mathbf{x} \in U^\perp \quad \iff \quad A^T(P(\mathbf{x}) - \mathbf{x}) = \mathbf{0}.$$

[70] We already saw this argument in 6.3 so I went faster this time.

Furthermore, since $P(\mathbf{x}) \in U$ and since $U = \mathcal{C}(A)$ we can write $P(\mathbf{x}) = A\hat{\mathbf{x}}$ for some vector $\hat{\mathbf{x}} \in \mathbb{R}^d$.[71] Thus we have the following two facts about the projection:

- $A^T(P(\mathbf{x}) - \mathbf{x}) = \mathbf{0}$,
- $P(\mathbf{x}) = A\hat{\mathbf{x}}$.

Combining these facts gives

$$\begin{aligned} A^T(A\hat{\mathbf{x}} - \mathbf{x}) &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} - A^T\mathbf{x} &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} &= A^T\mathbf{x} \\ \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{x} && A \text{ has independent columns} \\ A\hat{\mathbf{x}} &= A(A^TA)^{-1}A^T\mathbf{x} \\ P(\mathbf{x}) &= A(A^TA)^{-1}A^T\mathbf{x}. \end{aligned}$$

Finally, since this equality holds for any vector $\mathbf{x} \in \mathbb{R}^n$ we conclude that P is linear and is represented by the $n \times n$ matrix $A(A^TA)^{-1}A^T$. We have thus proved the following theorem.

Theorem (Projection Onto a Subspace). Let A be an $n \times d$ matrix with independent columns, so the column space $U = \mathcal{C}(A)$ is a d-dimensional subspace of $\mathbb{R}^n$. The function $P : \mathbb{R}^n \to \mathbb{R}^n$ that projects onto U is linear and is represented by the following matrix:

$$\boxed{P = A(A^TA)^{-1}A^T.}$$

If A has **orthonormal columns** then the formula simplifies because $A^TA = I$:

$$\boxed{P = AA^T.}$$

A given subspace is represented by many matrices. For example, consider the 3×1 matrices

$$A = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}.$$

The column spaces $\mathcal{C}(A)$ and $\mathcal{C}(B)$ are the same line in $\mathbb{R}^3$. Thus we expect that the matrices $A(A^TA)^{-1}A^T$ and $B(B^TB)^{-1}B^T$ are equal. Indeed, we have

$$\begin{aligned} A(A^TA)^{-1}A^T &= \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \left(\begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} (3)^{-1} \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \end{aligned}$$

[71] In the least squares problem the vector $\hat{\mathbf{x}}$ is the main event. Here it is only a temporary convenience.

$$= \frac{1}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & -1 & 1 \end{pmatrix}$$
$$= \frac{1}{3}\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

and

$$A(A^TA)^{-1}A^T = \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}\left(\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}\begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}\right)^{-1}\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}$$
$$= \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}(12)^{-1}\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}$$
$$= \frac{1}{12}\begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}$$
$$= \frac{1}{12}\begin{pmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{pmatrix}$$
$$= \frac{1}{3}\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}.$$

More generally, if A is $n \times d$ then for any invertible $d \times d$ matrix C we have

$$\mathcal{C}(AC) = \mathcal{C}(A).$$

If A has independent columns then AC also has independent columns, and we observe that

$$\begin{aligned}(AC)((AC)^T(AC))^{-1}(AC)^T &= AC(C^T(A^TA)C)^{-1}C^TA^T \\ &= ACC^{-1}(A^TA)^{-1}(C^T)^{-1}C^TA^T \\ &= AI(A^TA)^{-1}IA^T \\ &= A(A^TA)^{-1}A^T.\end{aligned}$$

So far we have discussed explicit properties of projection in Euclidean space. Next we discuss some abstract properties of projection that apply also to operators on infinite dimensional spaces.

Definition of Abstract Projection. Let V be a real inner product space and consider a linear function $P : V \to V$. If P satisfies certain mild conditions,[72] then there exists a unique linear function $P^T : V \to V$, called the *adjoint of* P, satisfying

$$\langle P\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, P^T\mathbf{v}\rangle \quad \text{for all } \mathbf{u}, \mathbf{v} \in V.$$

[72]For example, this holds when V is complete and P is continuous.

We say that P is an *abstract projection operator* when

$$P^2 = P \quad \text{and} \quad P^T = P.$$

For example, if V is Euclidean space then the adjoint P^T is just the transpose matrix. In this case we observe that the matrix $P = A(A^TA)^{-1}A^T$ is an abstract projection because

$$\begin{aligned} P^2 &= [A(A^TA)^{-1}A^T][A(A^TA)^{-1}A^T] \\ &= A\cancel{(A^TA)^{-1}(A^TA)}(A^TA)^{-1}A^T \\ &= AI(A^TA)^{-1}A^T \\ &= P \end{aligned}$$

and

$$\begin{aligned} P^T &= [A(A^TA)^{-1}A^T]^T \\ &= (A^T)^T[(A^TA)^{-1}]^T(A)^T \\ &= A[(A^TA)^T]^{-1}A^T \\ &= A[A^T(A^T)^T]^{-1}A^T \\ &= A(A^TA)^{-1}A^T \\ &= P. \end{aligned}$$

Later we will see that any abstract projection matrix satisfying $P^2 = P$ and $P^T = P$ is a "real" (i.e., geometric) projection, hence it can be represented as $P = A(A^TA)^{-1}A^T$. To summarize: For any square matrix P we have

$$\boxed{P^2 = P \text{ and } P^T = P \quad \iff \quad P = A(A^TA)^{-1}A^T \text{ for some } A.}$$

I think that's pretty surprising. In fact, there is a more general version:[73] For any square matrix P we have

$$\boxed{P^2 = P \quad \iff \quad P = A(B^TA)^{-1}B^T \text{ for some } A \text{ and } B.}$$

If P has shape $n \times n$ and rank d then the matrices A and B both have shape $n \times d$ and independent columns. Geometrically, this is a "non-orthogonal projection". It projects all points onto the column space of A, but it does this at a strange angle that is perpendicular to the column space of B.

For example, suppose we want to project onto the line $t(1,1)$ in $\mathbb{R}^2$ in a direction that is perpendicular to $(3,1)$. Then we can take

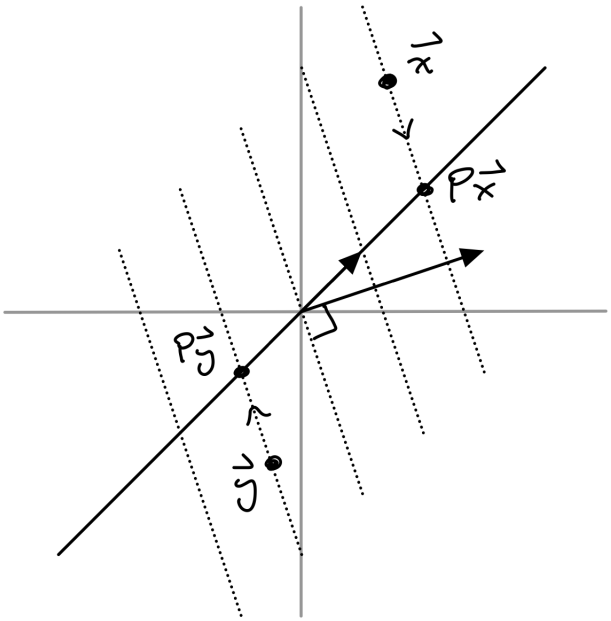
$$A = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

[73] Maybe we'll prove this later; maybe not. Here are some links:
`https://math.stackexchange.com/questions/600745/are-idempotent-matrices-always-diagonalizable`
`https://math.stackexchange.com/questions/2817221/decomposition-of-idempotent-matrix`

to get

$$\begin{aligned} P = A(B^TA)^{-1}B^T &= \begin{pmatrix}1\\1\end{pmatrix}\left(\begin{pmatrix}3 & 1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}\right)^{-1}\begin{pmatrix}3 & 1\end{pmatrix} \\ &= \begin{pmatrix}1\\1\end{pmatrix}(4)^{-1}\begin{pmatrix}3 & 1\end{pmatrix} \\ &= \frac{1}{4}\begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}3 & 1\end{pmatrix} \\ &= \frac{1}{4}\begin{pmatrix}3 & 1\\3 & 1\end{pmatrix}. \end{aligned}$$

Picture:



Projection Matrices Come in Pairs.[74] To end this section, I want to observe that projection matrices come in pairs. Let P be a projection matrix satisfying

$$P^2 = P \quad \text{and} \quad P^T = P.$$

Then the matrix $Q = I - P$ is also a projection since

$$Q^2 = (I-P)^2 = I^2 - 2P + P^2 = I - 2P + P = I - P = Q$$

and

$$Q^T = (I-P)^T = I^T - P^T = I - P = Q.$$

[74]This topic does not apply very well to infinite dimensional vector spaces, since one of the pair will have infinite rank.

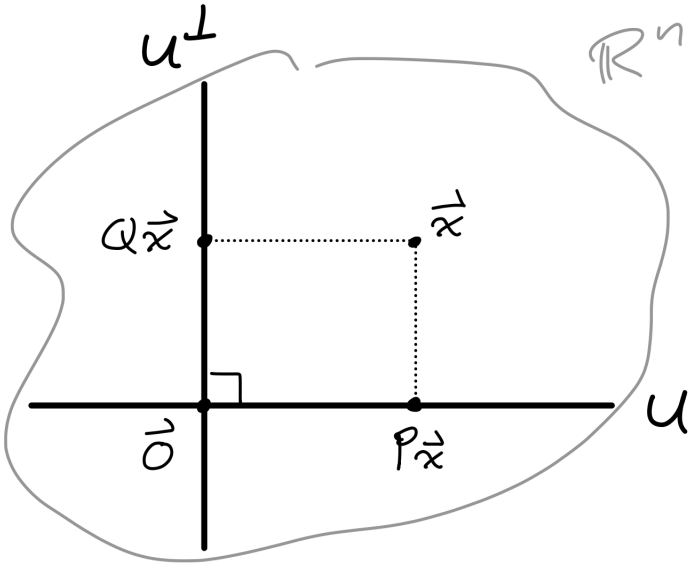
Furthermore, we observe that

$$PQ = QP = P^2 - P = P - P = O.$$

Thus we have the following situation:

- P and Q are projections,
- $P + Q = I$,
- $PQ = O$.

Suppose that P and Q have shape $n \times n$. If P is the projection onto a subspace $U \subseteq \mathbb{R}^n$ then Q is the projection onto the orthogonal complement $U^\perp \subseteq \mathbb{R}^n$ and vice versa. We can see this by looking at $\mathbb{R}^n$ "from the side":



For any point $\mathbf{x} \in \mathbb{R}^n$ we know that the four points $\mathbf{0}, \mathbf{x}, P\mathbf{x}, Q\mathbf{x}$ form a rectangle because

$$P\mathbf{x} + Q\mathbf{x} = (P + Q)\mathbf{x} = I\mathbf{x} = \mathbf{x}$$

and

$$(P\mathbf{x}) \bullet (Q\mathbf{x}) = (P\mathbf{x})^T(Q\mathbf{x}) = \mathbf{x}^T P^T Q\mathbf{x} = \mathbf{x}^T PQ\mathbf{x} = \mathbf{x}^T O\mathbf{x} = 0.$$

This pairing sometimes shortens calculations. For example, suppose that we want to find the 3×3 matrix P that projects onto the plane $x - 2y + z = 0$ in $\mathbb{R}^3$. Then the complementary matrix $Q = I - P$ projects onto the line generated by $(1, -2, 1)$, which is easier to calculate:

$$Q = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \left(\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} (6)^{-1} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$
$$= \frac{1}{6} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$
$$= \frac{1}{6} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}.$$

It follows that

$$\begin{aligned} P &= I - Q \\ &= \frac{1}{6} \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} \\ &= \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}. \end{aligned}$$

Of course, we could also do this the long way, by first finding a basis for the plane $x-2y+z=0$. Let's take $(1,0,-1)$ and $(0,1,2)$ and form the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix}.$$

Then with a bit of work, one can verify that

$$A(A^TA)^{-1}A^T = \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

It seems a bit surprising that these two methods give the same answer. To be more precise, consider any complementary subspaces U and $U^\perp$ in $\mathbb{R}^n$, and choose any matrices A and B with independent columns, such that $U = \mathcal{C}(A)$ and $U^\perp = \mathcal{C}(B)$. Then it must be true that

$$A(A^TA)^{-1}A^T + B(B^TB)^{-1}B^T = I,$$

but this seems mysterious. I'll end by giving an argument to make it feel more natural.

Suppose that $\dim U = d$ so that A has shape $n \times d$ and B has shape $n \times (n-d)$. Form the augmented matrix

$$C = \left(\begin{array}{c|c} A & B \end{array} \right),$$

which has shape $n \times n$. Since the columns of A are a basis for U and the columns of B are a basis for $U^\perp$, the columns of C are a basis for the whole space. In particular, C is invertible, which implies that

$$C(C^TC)^{-1}C^T = CC^{-1}(C^T)^{-1}C^T = I.$$

On the other hand, since every column of A is perpendicular to every column of B we know that $A^TB = O$ and $B^TA = O$, hence

$$C^TC = \left(\begin{array}{c} A^T \\ \hline B^T \end{array}\right)\left(\begin{array}{c|c} A & B \end{array}\right) = \left(\begin{array}{c|c} A^TA & A^TB \\ \hline B^TA & B^TB \end{array}\right) = \left(\begin{array}{c|c} A^TA & O \\ \hline O & B^TB \end{array}\right).$$

And since A and B each have independent columns, we know that A^TA and B^TB are invertible, hence

$$(C^TC)^{-1} = \left(\begin{array}{c|c} A^TA & O \\ \hline O & B^TB \end{array}\right)^{-1} = \left(\begin{array}{c|c} (A^TA)^{-1} & O \\ \hline O & (B^TB)^{-1} \end{array}\right).$$

Finally, we observe that

$$\begin{aligned} C(C^TC)^{-1}C^T &= \left(\begin{array}{c|c} A & B \end{array}\right)\left(\begin{array}{c|c} (A^TA)^{-1} & O \\ \hline O & (B^TB)^{-1} \end{array}\right)\left(\begin{array}{c} A^T \\ \hline B^T \end{array}\right) \\ &= \left(\begin{array}{c|c} A(A^TA)^{-1} & B(B^TB)^{-1} \end{array}\right)\left(\begin{array}{c} A^T \\ \hline B^T \end{array}\right) \\ &= A(A^TA)^{-1}A^T + B(B^TB)^{-1}B^T. \end{aligned}$$

9 Linear and Bilinear Forms

9.1 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[75] 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.

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

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.[76] Let $\varphi : V \to \mathbb{R}$ be a linear functional. Then

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

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:[77]

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

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 $\varphi_{\mathbf{u}} \in V^\vee$:

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

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

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

[76] 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 $N \to \infty$.

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

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 $V \cong V^\vee$.

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:[78]

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

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 $\langle f|g\rangle$.

9.2 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:

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

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

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

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

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

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

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[79]

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

(a) **Symmetric.** We have

$$\begin{array}{rcl} \varphi_B(\mathbf{x}, \mathbf{y}) = \varphi_B(\mathbf{y}, \mathbf{x}) \text{ for all } \mathbf{x}, \mathbf{y} & \iff & B^T = B, \\ \varphi_B(\mathbf{x}, \mathbf{y})^* = \varphi_B(\mathbf{y}, \mathbf{x}) \text{ for all } \mathbf{x}, \mathbf{y} & \iff & B^* = B. \end{array}$$

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

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

In this case the form φ_B and the matrix B are called *positive semi-definite*.[80]

(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

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

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*.

[79] 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.

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

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.[81] But if A has independent

[81] 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.

columns then this implies that $\mathbf{x} = \mathbf{0}$. There are many ways to see this. One method uses the fact that $(A^TA)^{-1}$ exists to get

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

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

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

(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.[82] 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.[83]

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.[84]

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

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

[84] Indeed, we have seen that the "functions" $\delta(x)$ and $e^{2\pi ix}$ are treated as elements of $L^2(\mathbb{C})$, even though

9.3 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:

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

$e^{2\pi ix}$ 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.

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

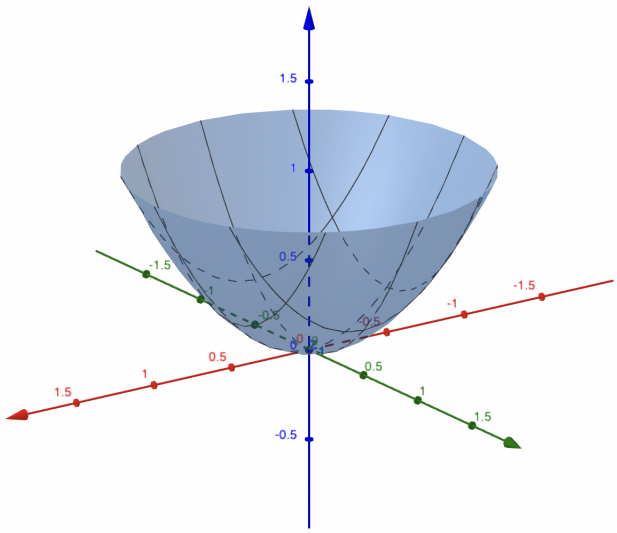
$$\begin{aligned} f(x,y,z) &= b + b_1x + b_2y + b_3z + b_{11}x^2 + b_{22}y^2 + b_{33}z^2 + b_{12}xy + b_{13}xz + b_{23}yz \\ &= b + \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} b_{11} & b_{12}/2 & b_{13}/2 \\ b_{12}/2 & b_{22} & b_{23}/2 \\ b_{13}/2 & b_{23}/2 & b_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= b + \mathbf{b}^T\mathbf{x} + \mathbf{x}^T B\mathbf{x}. \end{aligned}$$

Thus the degree zero terms correspond to a scalar b, the degree 1 terms correspond to a vector $\mathbf{b}^T$,[85] 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.[86]

Here are three simplest examples of quadratic forms. Let

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

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



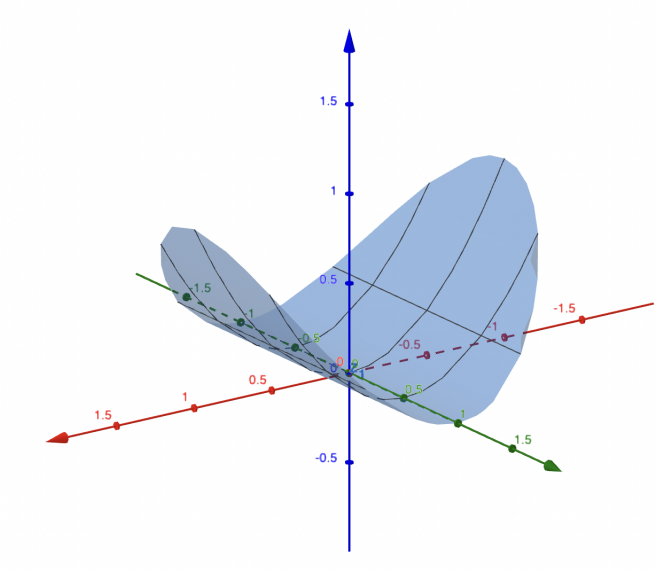
[85] It doesn't matter whether we write the degree 1 terms as $\mathbf{b}^T\mathbf{x}$ or $\mathbf{x}^T\mathbf{b}$. I am simply following the convention from Section 1.1, where linear forms correspond to row vectors.

[86] 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.

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

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

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



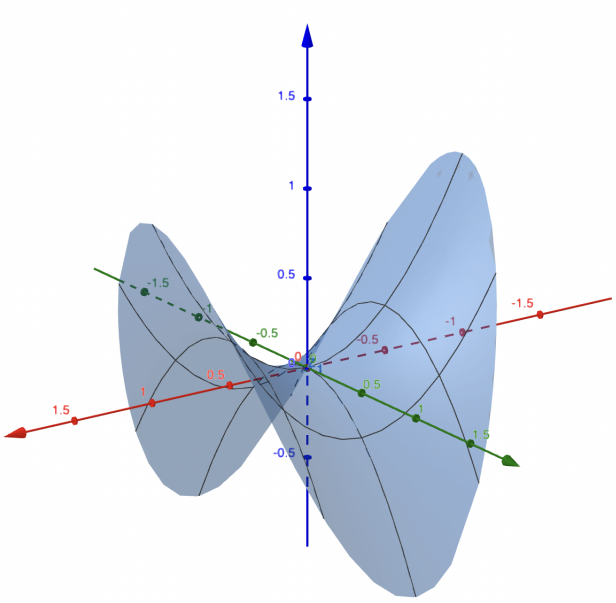
This time the minimum is not unique, since $Q_B(0, y) = 0$ for any value of y. Indeed, this matrix can be factored as

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

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.

9.4 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.[87]

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

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

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

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 $\mathbf{p}$:

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

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

$$\begin{aligned} f(0 + x, 0 + y) &= f(0, 0) + (\nabla f)^T_{(0,0)}\mathbf{x} + \frac{1}{2}\mathbf{x}^T(Hf)_{(0,0)}\mathbf{x} \\ &= 2 + \begin{pmatrix} 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 6 & 2 \\ 2 & 8 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}, \end{aligned}$$

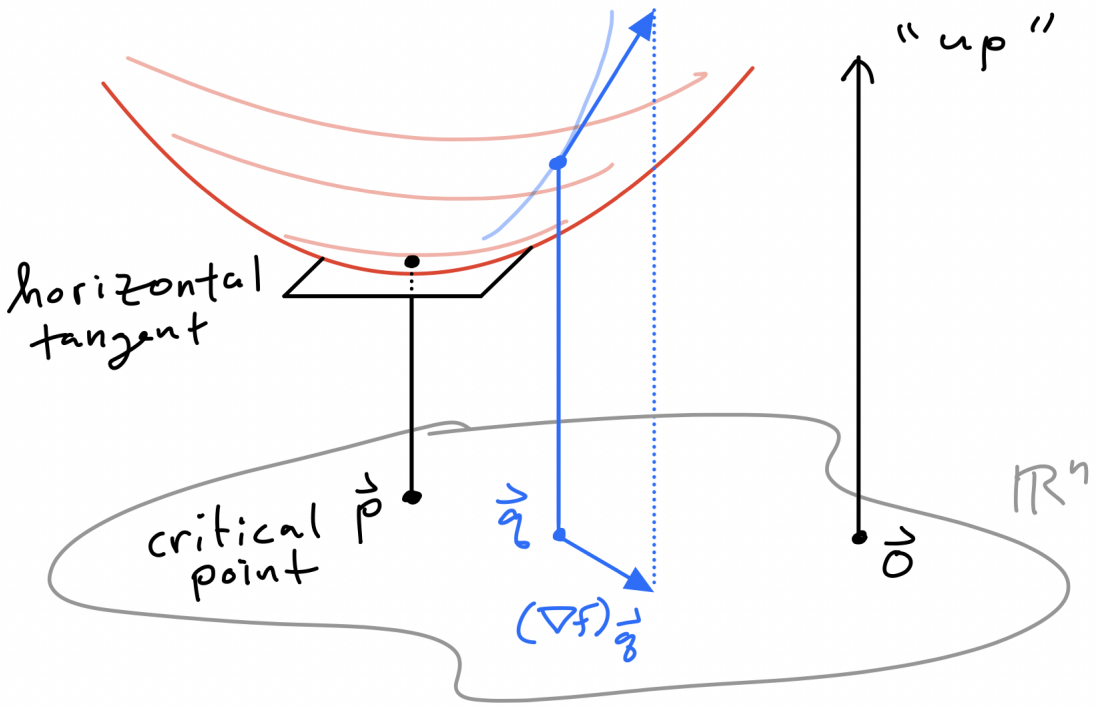
which we already computed in the previous section. The Taylor expansion at $\mathbf{p} = (1, 1)$ is

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

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

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

Note that $\mathbf{p} = \left(\frac{-10}{44}, \frac{8}{44}\right)$ 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[88]

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

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.

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

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:

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

10 Determinants

10.1 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:[89]

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

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

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

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

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[90]

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

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 $1 \le i \le n$, 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

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

[90] 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 $1 \le i, j \le n$ is a basis for $\mathbb{R}^{n\times n}$. More generally, one can show that $\mathbb{R}^{m\times n}$ has dimension mn.

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

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

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

is equal to

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

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[91]

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

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

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

[91] 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.

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

$$\begin{aligned}
\mathcal{S}^1(\mathbb{R}^3) &= \operatorname{Span}\{\varepsilon_1, \varepsilon_2, \varepsilon_3\},\\
\mathcal{S}^2(\mathbb{R}^3) &= \operatorname{Span}\{\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, \varepsilon_{12}+\varepsilon_{21}, \varepsilon_{13}+\varepsilon_{31}, \varepsilon_{23}+\varepsilon_{32}\},\\
\mathcal{S}^3(\mathbb{R}^3) &= \operatorname{Span}\{\varepsilon_{111}, \varepsilon_{222}, \varepsilon_{333},\\
&\qquad \varepsilon_{112}+\varepsilon_{121}+\varepsilon_{211}, \varepsilon_{113}+\varepsilon_{131}+\varepsilon_{311}, \varepsilon_{223}+\varepsilon_{232}+\varepsilon_{322},\\
&\qquad \varepsilon_{221}+\varepsilon_{212}+\varepsilon_{122}, \varepsilon_{331}+\varepsilon_{313}+\varepsilon_{133}, \varepsilon_{332}+\varepsilon_{323}+\varepsilon_{233},\\
&\qquad \varepsilon_{123}+\varepsilon_{132}+\varepsilon_{213}+\varepsilon_{231}+\varepsilon_{312}+\varepsilon_{321}\}.
\end{aligned}$$

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[92] 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:

$$\mathcal{A}^1(\mathbb{R}^2) = \operatorname{Span}\{\varepsilon_1, \varepsilon_2\},$$

[92] 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.

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

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 $0 \leq k \leq n$ I claim that[93]

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

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

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

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

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

[93]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 $k = 1$.

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

$$\begin{aligned}(\varepsilon_{123} + \varepsilon_{231} + \varepsilon_{312} - \varepsilon_{132} - \varepsilon_{213} - \varepsilon_{321})(A) \\ = \varepsilon_{123}(A) + \varepsilon_{231}(A) + \varepsilon_{312}(A) - \varepsilon_{132}(A) - \varepsilon_{231}(A) - \varepsilon_{321}(A), \\ = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_3c_1 - a_3b_2c_1.\end{aligned}$$

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 $\mathbb{R}^3$?

10.2 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

$$A = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \mid \mathbf{a} \end{pmatrix}, \quad B = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \mid \mathbf{b} \end{pmatrix}, \quad C = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \mid \mathbf{a} + \lambda\mathbf{b} \end{pmatrix},$$

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

$$A = \begin{pmatrix} \mathbf{a}_1 \mid \mathbf{a}_2 \end{pmatrix} \text{ and } B = \begin{pmatrix} \mathbf{b}_1 \mid \mathbf{b}_2 \end{pmatrix}, \text{ hence } A + B = \begin{pmatrix} \mathbf{a}_1 + \mathbf{b}_1 \mid \mathbf{a}_2 + \mathbf{b}_2 \end{pmatrix}.$$

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 $\varphi(A) + \varphi(B)$.[94]

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

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:

(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,[95] 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

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

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:

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

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

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

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

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

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

$$\varphi(AT_{ij}) = -\varphi(A).$$

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

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

□

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[96]

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

□

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

$$\varphi(E_1 \cdots E_k) = \varphi(E_1 \cdots E_{k-1})\varphi(E_k)$$

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

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

□

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:

$$AE_1E_2 \cdots E_k = I.$$

Since elementary matrices are invertible, this becomes

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

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

$$E_k \cdots E_2E_1A = I$$

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.

10.3 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

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

$$= \det(A)\det(B).$$

(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.

10.4 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:

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

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.[97] Since we can get from $(1,2,3)$ to $(3,2,1)$ using only odd numbers of swaps, we define

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

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	$\operatorname{sgn}(\sigma)$
$(1,2,3)$	$+1$
$(2,3,1)$	$+1$
$(3,1,2)$	$+1$
$(1,3,2)$	-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:

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

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

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

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

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

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

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:

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

10.5 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[98]

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

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

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

[98] 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.

□

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

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

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 $X = \begin{pmatrix} \mathbf{x}_1 \,\big|\, \cdots \,\big|\, \mathbf{x}_n \end{pmatrix}$ 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

$$\begin{aligned} AX &= I \\ A \begin{pmatrix} \mathbf{x}_1 \,\big|\, \cdots \,\big|\, \mathbf{x}_n \end{pmatrix} &= \begin{pmatrix} \mathbf{e}_1 \,\big|\, \cdots \,\big|\, \mathbf{e}_n \end{pmatrix} \\ \begin{pmatrix} A\mathbf{x}_1 \,\big|\, \cdots \,\big|\, A\mathbf{x}_n \end{pmatrix} &= \begin{pmatrix} \mathbf{e}_1 \,\big|\, \cdots \,\big|\, \mathbf{e}_n \end{pmatrix}, \end{aligned}$$

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![99]

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:

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

10.6 Geometric Interpretation

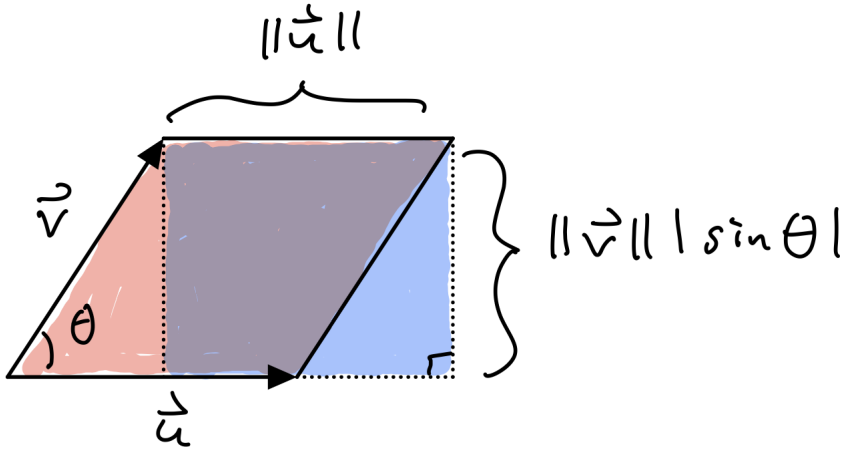
My bias is that algebra is based on geometry,[100] hence for me the "true meaning" of the determinant is its geometric interpretation.

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 $\|\mathbf{u}\|\|\mathbf{v}\||\sin\theta|$.[101] Indeed, in the following picture the red parallelogram and the blue rectangle have the same area:

[99] I have forgotten this many times.

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

[101] 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 $\mathbf{v}$:

$$A = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \end{pmatrix}.$$

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

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

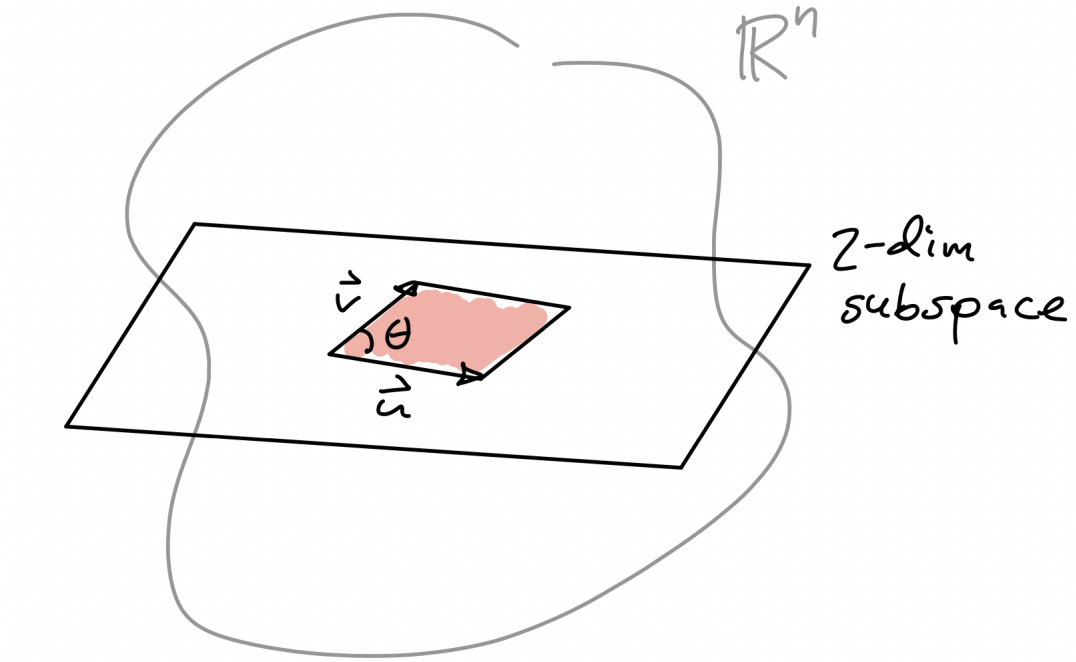
But the determinant of $A^T A$ can also be computed as follows:

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

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:

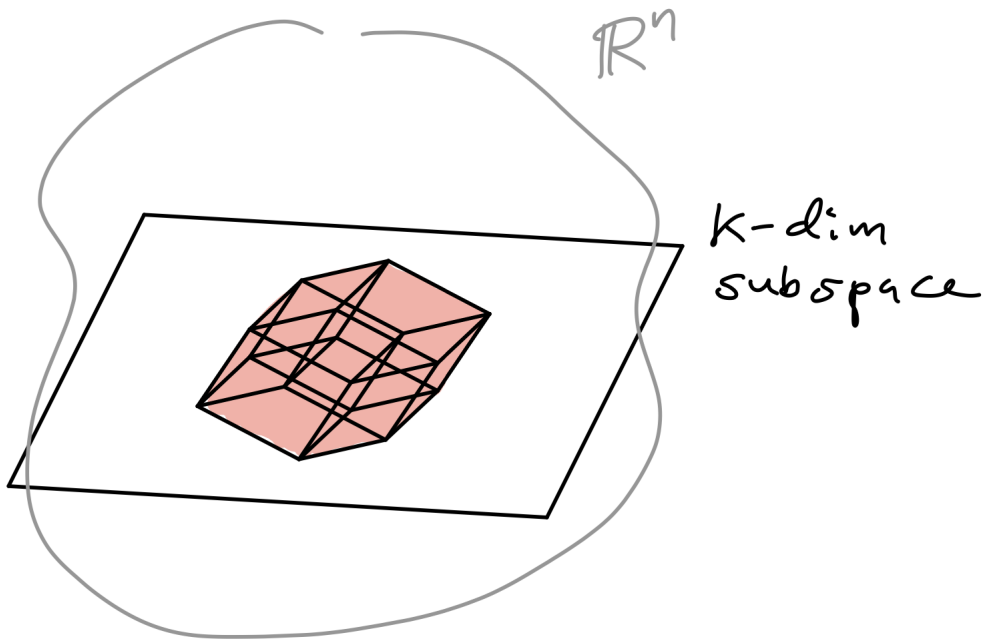


For geometric reasons, the area of the parallelogram is still $\|\mathbf{u}\|\|\mathbf{v}\||\sin\theta|$, but now the $n \times 2$ matrix $A = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \end{pmatrix}$ is **not square**, so $\det(A)$ is not defined. However, the matrix A^TA is still square, so we may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

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

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[102]

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

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 $k \neq n$.

[102] 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). □

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.[103]

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

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

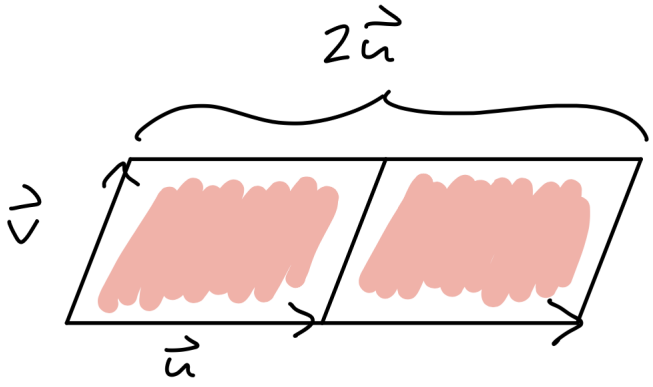
where the sign depends on the ordering of the columns, and is not relevant to the geometry. Thus we will show that the determinant can be interpreted as a "signed volume".[104] According to Section 2.2, we only need to show that the function Vol_n from $(\mathbb{R}^n)^n$ to $\mathbb{R}$ satisfies the 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",[105] 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:

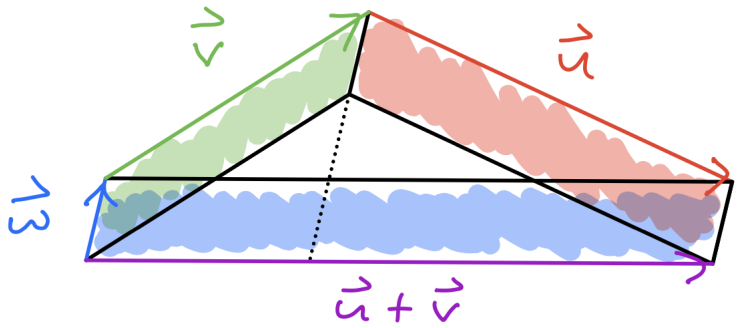


[103] If $n = 3$ then we can also express the volume in terms of the cross product, but doing so breaks the symmetry, and the cross product doesn't generalize to higher dimensions.

[104] 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.

[105] 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.[106]

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

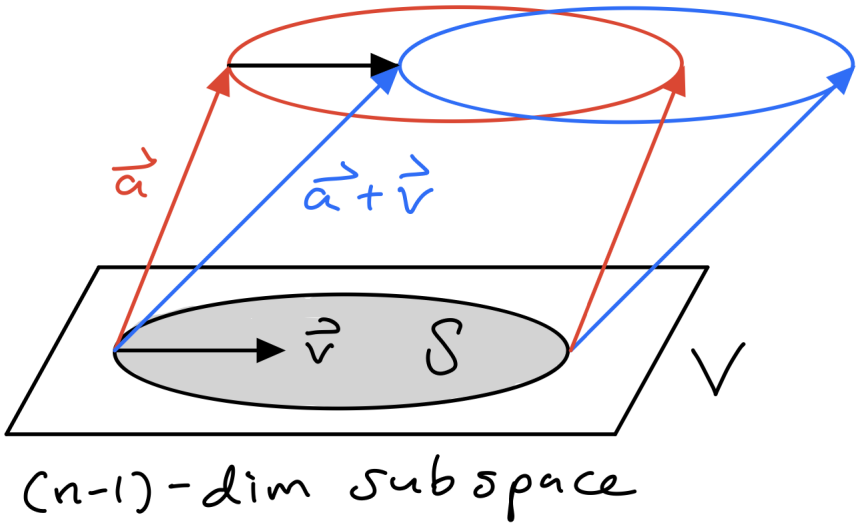
$$\mathrm{Vol}_n(\mathrm{Prism}_S(\mathbf{a})) = \mathrm{Vol}_n(\mathrm{Prism}_S(\mathbf{a}+\mathbf{v})) \quad \text{for any vector } \mathbf{v} \in V.$$

More colloquially:

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

Here is a picture:

[106] 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:

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

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 $\text{Prism}_{S_j}(\mathbf{a}_j)$. 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

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

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.

10.7 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

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

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

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

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[107]

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

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[108] subsets of $\mathbb{R}^n$. To be precise, for any subset $S \subseteq \mathbb{R}^n$ we define the image set

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

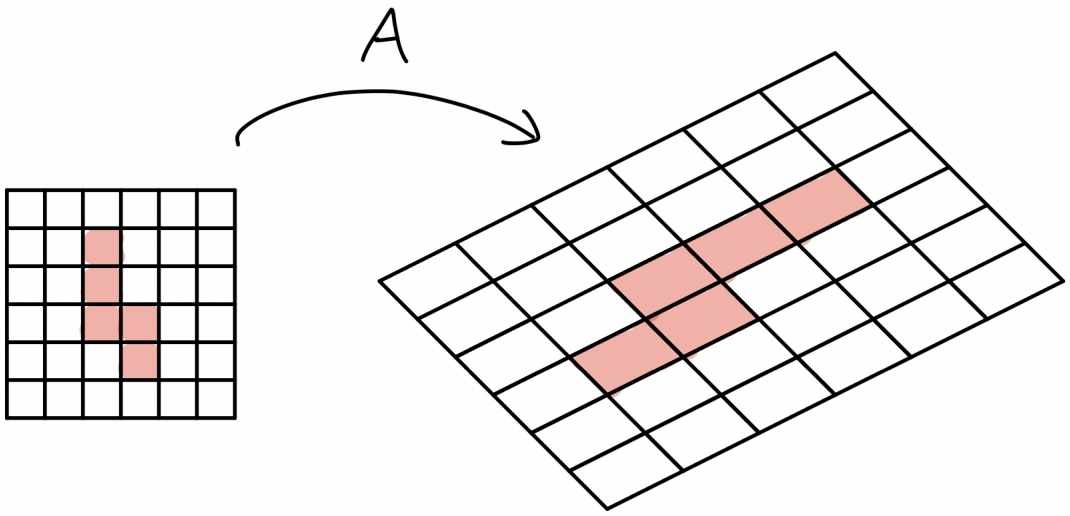
In this case I claim that

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

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:

[107] 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.

[108] 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[109] 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[110]

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

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:

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

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

[110] 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.

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

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

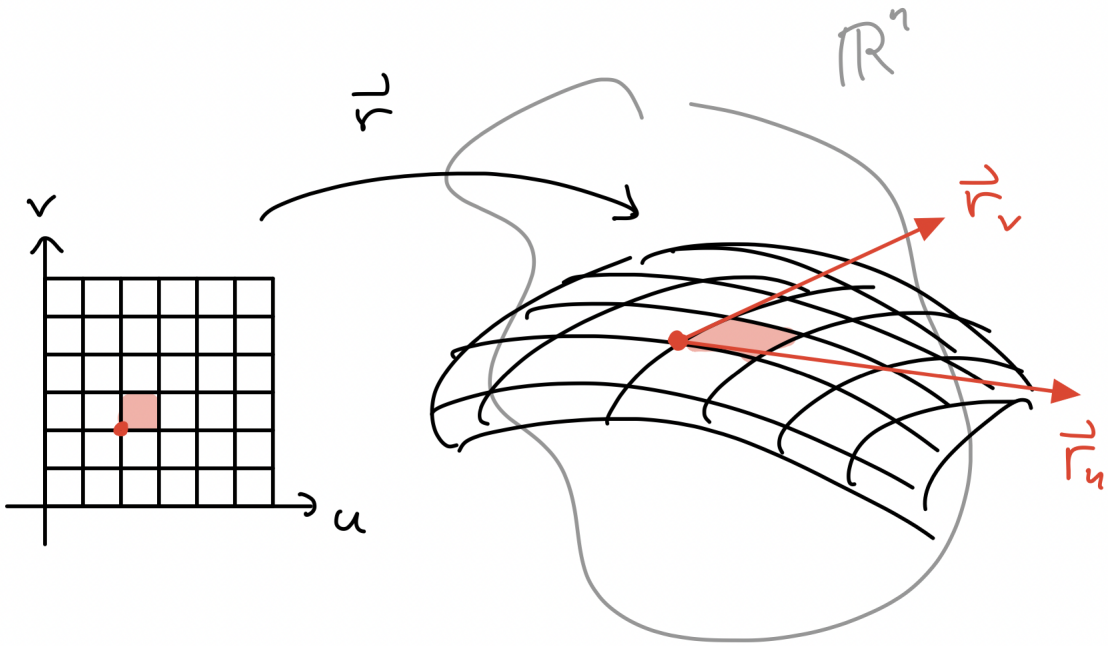
$$(\text{length of the curve } \mathbf{r}(t) \text{ between times } t = a \text{ and } t = b) = \int_a^b \|\mathbf{r}'(t)\|\, dt.$$

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[111]



[111] 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:

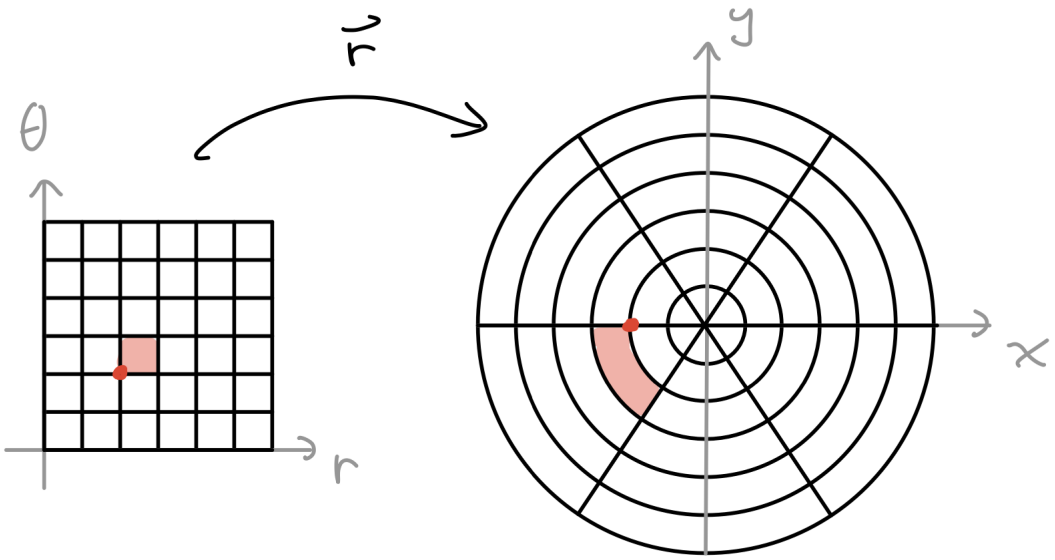
$$\|\mathbf{r}_u \times \mathbf{r}_v\| = \|\mathbf{r}_u\|\|\mathbf{r}_v\||\sin\theta_{uv}|.$$

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[112]

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

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

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

[112] 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.

11 Eigenvalues and Eigenvectors

11.1 A Motivating Example

In order to motivate the concepts of eigenvalues and eigenvectors I will develop one specific example in detail. Some of the steps will seem miraculous, and only make sense later when we discuss the general theory.

Our example comes from the theory of "Markov chains". Consider the following matrix:

$$A = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix}.$$

This matrix has the special property that each of its columns sums to 1. In matrix notation:[113]

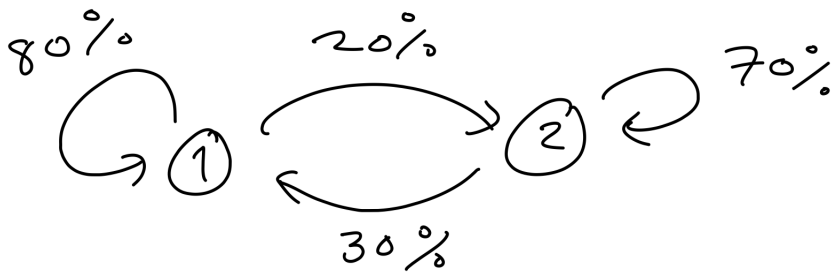
$$\begin{pmatrix} 1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} = \begin{pmatrix} .8+.2 & .3+.7 \end{pmatrix} = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

By induction, this implies that every power of A has columns that sum to 1:

$$\begin{aligned} \begin{pmatrix} 1 & 1 \end{pmatrix} A^n &= \begin{pmatrix} 1 & 1 \end{pmatrix} (AA^{n-1}) \\ &= (\begin{pmatrix} 1 & 1 \end{pmatrix} A)A^{n-1} \\ &= \begin{pmatrix} 1 & 1 \end{pmatrix} A^{n-1} \\ &\vdots \\ &= \begin{pmatrix} 1 & 1 \end{pmatrix}. \end{aligned}$$

Such a matrix is called a *Markov matrix* or a *stochastic matrix*. We can interpret the matrix entries as probabilities. Suppose that a certain particle can be in one of two states. At each discrete time step, the particle can change states, according to the following probabilities:

[113] Jargon: Later we will say that $\begin{pmatrix} 1 & 1 \end{pmatrix}$ is a "left eigenvector" of A with "eigenvalue" 1.



That is, if the particle is currently in state 1 then it has an 80% chance to stay in state 1 and a 20% chance to transition to state 2. If the particle is in state 2 then it has a 70% chance to stay and a 30% chance to stay. This is why the columns of A must sum to 1.

Now let's consider the first few powers of A:

$$\begin{aligned} A &= \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix}, \\ A^2 &= \begin{pmatrix} .7 & .45 \\ .3 & .55 \end{pmatrix}, \\ A^3 &= \begin{pmatrix} .65 & .525 \\ .35 & .475 \end{pmatrix}, \\ &\vdots \\ A^{10} &= \begin{pmatrix} 0.600390625 & 0.5994140625 \\ 0.399609375 & 0.4005859375 \end{pmatrix}. \end{aligned}$$

Do you see any pattern here? It seems likely that

$$A^n \to \begin{pmatrix} .6 & .6 \\ .4 & .4 \end{pmatrix} \text{ as } n \to \infty,$$

but the entries of the matrices look complicated. Nevertheless, at the end of this section we will obtained exact formulas for the entries of each power A^n.

We have shown that each column of each power A^n sums to 1. This fact has a probabilistic interpretation. Let p_k and q_k be the probabilities that the particle is in state 1 or 2, respectively, after k seconds, and let p_0, q_0 denote the initial probabilities. Then I claim that[114]

$$\boxed{\mathbf{p}_n = \begin{pmatrix} p_n \\ q_n \end{pmatrix} = A \begin{pmatrix} p_{n-1} \\ q_{n-1} \end{pmatrix} = A\mathbf{p}_{n-1}.}$$

[114] Alternatively, suppose that we have an ensemble of particles and let x_n, y_n denote the **expected number** of particles in each state. Then the same theory will hold.

To prove this we use the *law of total probability* (which I won't explain here). Given any two events S and T, we have the following identities:

$$\begin{aligned} P(S) &= P(T)P(S|T) + P(T')P(S|T'), \\ P(S') &= P(T)P(S'|T) + P(T')P(S'|T'). \end{aligned}$$

Let S_n be the event that "the particle is in state 1 after n seconds", so that

$$p_n = P(S_n) \text{ and } q_n = P(S'_n).$$

The transition matrix A tells us that

$$\begin{aligned} P(S_n|S_{n-1}) &= .8, \\ P(S'_n|S_{n-1}) &= .2, \\ P(S_n|S'_{n-1}) &= .3, \\ P(S'_n|S'_{n-1}) &= .7, \end{aligned}$$

which are independent of n. Hence we have

$$\begin{aligned} p_n &= P(S_n) \\ &= P(S_{n-1})P(S_n|S_{n-1}) + P(S'_{n-1})P(S_n|S'_{n-1}) \\ &= p_{n-1}(.8) + q_{n-1}(.3) \end{aligned}$$

and

$$\begin{aligned} q_n &= P(S'_n) \\ &= P(S_{n-1})P(S'_n|S_{n-1}) + P(S'_{n-1})P(S'_n|S'_{n-1}) \\ &= p_{n-1}(.2) + q_{n-1}(.7), \end{aligned}$$

as desired. But enough about probability.

Given the initial distribution $\mathbf{p}_0 = (p_0, q_0)$, the distribution after n seconds is given by

$$\begin{aligned} \mathbf{p}_n &= A\mathbf{p}_{n-1}, \\ &= AA\mathbf{p}_{n-2} \\ &\vdots \\ &= AA\cdots A\mathbf{p}_0 \\ &= A^n\mathbf{p}_0. \end{aligned}$$

Our goal is to find explicit formulas for p_n and q_n in terms of p_0 and q_0.

Now comes the key trick. We have the following mysterious identitites:

$$A\begin{pmatrix}3\\2\end{pmatrix} = \begin{pmatrix}3\\2\end{pmatrix} \quad \text{and} \quad A\begin{pmatrix}1\\-1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}. \tag{$*$}$$

Jargon: We say that $(3,2)$ and $(1,-1)$ are "eigenvectors" of A with corresponding "eigenvalues" 1 and $1/2$. More generally, if $A\mathbf{x} = \lambda\mathbf{x}$ for some vector $\mathbf{x}$ and scalar λ, then the action of A^n on $\mathbf{x}$ is easy to compute:

$$\begin{aligned} A^n\mathbf{x} &= (A^{n-1}A)\mathbf{x} \\ &= A^{n-1}(A\mathbf{x}) \\ &= A^{n-1}(\lambda\mathbf{x}) \\ &= \lambda A^{n-1}\mathbf{x} \\ &\vdots \\ &= \lambda^n\mathbf{x}. \end{aligned}$$

Once we know $(*)$, the rest of the solution is straightforward. First we want to express our initial condition $\mathbf{p}_0$ as a linear combination of eigenvectors. In other words, we want to find a and b such that

$$\begin{aligned} \mathbf{p}_0 &= a\begin{pmatrix}3\\2\end{pmatrix} + b\begin{pmatrix}1\\-1\end{pmatrix} \\ &= \begin{pmatrix}3 & 1\\2 & -1\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}. \end{aligned}$$

Since the two eigenvectors are not parallel, the matrix of eigenvectors is invertible, hence

$$\begin{aligned} \begin{pmatrix}a\\b\end{pmatrix} &= \begin{pmatrix}3 & 1\\2 & -1\end{pmatrix}^{-1}\begin{pmatrix}p_0\\q_0\end{pmatrix} \\ &= -\frac{1}{5}\begin{pmatrix}-1 & -1\\-2 & 3\end{pmatrix}\begin{pmatrix}p_0\\q_0\end{pmatrix} \\ &= -\frac{1}{5}\begin{pmatrix}-p_0-q_0\\-2p_0+3q_0\end{pmatrix} \\ &= \begin{pmatrix}1/5\\p_0-3/5\end{pmatrix}. && p_0+q_0=1 \end{aligned}$$

Then we obtain the solution:

$$\begin{aligned} \mathbf{p}_0 &= \frac{1}{5}\begin{pmatrix}3\\2\end{pmatrix} + \left(p_0-\frac{3}{5}\right)\begin{pmatrix}1\\-1\end{pmatrix} \\ A^n\mathbf{p}_0 &= \frac{1}{5}A^n\begin{pmatrix}3\\2\end{pmatrix} + \left(p_0-\frac{3}{5}\right)A^n\begin{pmatrix}1\\-1\end{pmatrix} \\ \mathbf{p}_n &= \frac{1}{5}\begin{pmatrix}3\\2\end{pmatrix} + \left(p_0-\frac{3}{5}\right)\left(\frac{1}{2}\right)^n\begin{pmatrix}1\\-1\end{pmatrix} \\ \begin{pmatrix}p_n\\q_n\end{pmatrix} &= \begin{pmatrix}3/5+(p_0-3/5)/2^n\\2/5-(p_0-3/5)/2^n\end{pmatrix}. \end{aligned}$$

As $n \to \infty$ we observe that $\mathbf{p}_n \to (3/5, 2/5)$, regardless of the initial probabilities p_0 and q_0. The fact that the initial condition is irrelevant is sometimes called the "ergodic property" (or the "mixing property").

But we can do more. Suppose that $A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$ and $A\mathbf{x}_2 = \lambda_2 \mathbf{x}_2$ for some eigenvectors $\mathbf{x}_1, \mathbf{x}_2$ and eigenvalues λ_1, λ_2. We can express both of these conditions simultaneously by forming the matrices

$$X = \left(\begin{array}{c|c} \mathbf{x}_1 & \mathbf{x}_2 \end{array} \right) \quad \text{and} \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Then we have

$$\begin{aligned} AX &= A \left(\begin{array}{c|c} \mathbf{x}_1 & \mathbf{x}_2 \end{array} \right) \\ &= \left(\begin{array}{c|c} A\mathbf{x}_1 & A\mathbf{x}_2 \end{array} \right) \\ &= \left(\begin{array}{c|c} \lambda_1 \mathbf{x}_1 & \lambda_2 \mathbf{x}_2 \end{array} \right) \\ &= \left(\begin{array}{c|c} \mathbf{x}_1 & \mathbf{x}_2 \end{array} \right) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\ &= X\Lambda. \end{aligned}$$

This equation holds even when A is $n \times n$ and X is $n \times 2$. If A is 2×2 and if the vectors $\mathbf{x}_1, \mathbf{x}_2$ are not parallel, then the matrix X is square and invertible, hence

$$\begin{aligned} AX &= X\Lambda \\ A &= X\Lambda X^{-1}. \end{aligned}$$

In this case, we say that we have "diagonalized" the matrix A. In our case, we have

$$\begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix}^{-1}.$$

The powers of A behave well with respect to this factorization. This follows from two key properties. First, the powers of a diagonal matrix are easy to compute:

$$\Lambda^n = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^n = \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}.$$

Second, there is a miraculous cancellation in the powers of $X\Lambda X^{-1}$:

$$\begin{aligned} A^n &= (X\Lambda X^{-1})^n \\ &= (X\Lambda X^{-1})(X\Lambda X^{-1}) \cdots (X\Lambda X^{-1}) \\ &= X\Lambda (X^{-1}X)\Lambda (X^{-1}X) \cdots (X^{-1}X)\Lambda X^{-1} \\ &= X\Lambda\Lambda \cdots \Lambda X^{-1} \\ &= X\Lambda^n X^{-1}. \end{aligned}$$

Putting these together gives us explicit formulas for the entries of A^n:

$$\begin{aligned}
\begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix}^n &= \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}^n \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1^n & 0 \\ 0 & (1/2)^n \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1^n & 0 \\ 0 & (1/2)^n \end{pmatrix} \begin{pmatrix} -\frac{1}{5} \end{pmatrix} \begin{pmatrix} -1 & -1 \\ -2 & 3 \end{pmatrix} \\
&= \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1^n & 0 \\ 0 & (1/2)^n \end{pmatrix} \begin{pmatrix} -\frac{1}{5} \end{pmatrix} \begin{pmatrix} -1 & -1 \\ -2 & 3 \end{pmatrix} \\
&= \frac{1}{5} \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (1/2)^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -3 \end{pmatrix} \\
&= \frac{1}{5} \begin{pmatrix} 3 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2/2^n & -3/2^n \end{pmatrix} \\
&= \frac{1}{5} \begin{pmatrix} 3+2/2^n & 3-3/2^n \\ 2-2/2^n & 2+3/2^n \end{pmatrix}.
\end{aligned}$$

These exact formulas would be very difficult to obtain without the trick of eigenvalues and eigenvectors. By letting n go to infinity, we confirm our experimental observation that

$$A^n \to \frac{1}{5} \begin{pmatrix} 3 & 3 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} .6 & .6 \\ .4 & .4 \end{pmatrix} \text{ as } n \to \infty.$$

Finally, let me mention an alternative expression for A^n that is often more useful. For any vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2$ and scalars λ_1, λ_2, one can check that

$$\left(\begin{array}{c|c} \mathbf{x}_1 & \mathbf{x}_2 \end{array} \right) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \left(\begin{array}{c} \mathbf{y}_1^T \\ \hline \mathbf{y}_2^T \end{array} \right) = \lambda_1 \mathbf{x}_1 \mathbf{y}_1^T + \lambda_2 \mathbf{x}_2 \mathbf{y}_2^T.$$

Thus in our case we have

$$\begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix}^n = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \begin{pmatrix} 1/5 & 1/5 \end{pmatrix} + \left(\frac{1}{2} \right)^n \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 2/5 & -3/5 \end{pmatrix}.$$

This expression emphasizes the fact that A^n converges to a rank 1 matrix:

$$A^n \to \begin{pmatrix} 3 \\ 2 \end{pmatrix} \begin{pmatrix} 1/5 & 1/5 \end{pmatrix} \text{ as } n \to \infty.$$

Remember: This section is just for motivation. I will explain all of the ideas later.

11.2 The Characteristic Polynomial

Let A be a square matrix over $\mathbb{R}$ or $\mathbb{C}$. We say that $\lambda \in \mathbb{C}$ is an *eigenvalue* of A when there exists a nonzero vector $\mathbf{x}$ satisfying $A\mathbf{x} = \lambda\mathbf{x}$. Let me emphasize this:

$$\boxed{\lambda \text{ is an eigenvalue of } A \quad \iff \quad \text{there exists some } \mathbf{x} \neq \mathbf{0} \text{ satisfying } A\mathbf{x} = \lambda\mathbf{x}.}$$

If $A\mathbf{x} = \lambda\mathbf{x}$ then we say that $\mathbf{x}$ is a λ-*eigenvector* of A.

It is not immediately clear that eigenvalues exist. Our first result will show that any matrix has at least one eigenvalue. To do this we will rewrite the definition of eigenvalues in terms of determinants.[115] The following equivalences follow from results in the previous chapter:

$$\begin{aligned}
\lambda \text{ is an eigenvalue of } A &\iff A\mathbf{x} = \lambda\mathbf{x} \text{ for some } x \neq \mathbf{0} \\
&\iff (\lambda\mathbf{x} - A\mathbf{x}) = \mathbf{0} \text{ for some } \mathbf{x} \neq \mathbf{0} \\
&\iff \lambda I\mathbf{x} - A\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \neq \mathbf{0} \\
&\iff (\lambda I - A)\mathbf{x} = \mathbf{0} \text{ for some } \mathbf{x} \neq \mathbf{0} \\
&\iff \text{the nullspace } \mathcal{N}(\lambda I - A) \text{ contains a nonzero vector} \\
&\iff \dim \mathcal{N}(\lambda I - A) \geq 1 \\
&\iff \text{the matrix } \lambda I - A \text{ is not invertible} \\
&\iff \det(\lambda I - A) = 0.
\end{aligned}$$

This last equivalence is the most convenient way to study eigenvalues. In summary,

$$\boxed{\lambda \text{ is an eigenvalue of } A \iff \det(\lambda I - A) = 0.}$$

If λ is an eigenvalue of an $n \times n$ matrix A, we observe that the set of λ-eigenvectors is a subspace of $\mathbb{R}^n$. Indeed, the λ-eigenvectors of A are just the vectors in the nullspace $\mathcal{N}(\lambda I - A)$. We say that

$$\mathcal{N}(\lambda I - A) = \text{the } \lambda\text{-}\textit{eigenspace} \text{ of } A.$$

When λ is **not** an eigenvalue the matrix $\lambda I - A$ is invertible, so in this case the "λ-eigenspace" is trivial: $\mathcal{N}(\lambda I - A) = \{\mathbf{0}\}$. Before going further with the theory, we compute the eigenvalues of a general 2×2 matrix:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We have

$$\begin{aligned}
\det(\lambda I - A) &= \det\left(\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) \\
&= \det \begin{pmatrix} \lambda - a & -b \\ -c & \lambda - d \end{pmatrix} \\
&= (\lambda - a)(\lambda - d) - (-b)(-c) \\
&= \lambda^2 - (a + d)\lambda + (ad - bc).
\end{aligned}$$

Hence the eigenvalues of A are

$$\lambda = \frac{a + d \pm \sqrt{(a + d)^2 - 4(ad - bc)}}{2}.$$

[115] There is a popular textbook called *Linear Algebra Done Right* in which the author goes to great lengths to avoid the use of determinants in the theory of eigenvalues. There is another well-known textbook called *Linear Algebra Done Wrong*, which I greatly prefer.

Let $\Delta = (a+d)^2 - 4(ad-bc)$ denote the discriminant of this quadratic polynomial. If $\Delta = 0$ then the matrix A has only one eigenvalue. Now suppose that A has real entries. If $\Delta > 0$ then A has two distinct real eigenvalues and if $\Delta < 0$ then A has two distinct complex eigenvalues.

After finding the eigenvalues, it is an easy matter to find all of the eigenvectors. Consider our example from the previous section:

$$A = \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix}.$$

The eigenvalues are the roots of the polynomial equation

$$\begin{aligned} \det(\lambda I - A) &= 0 \\ \lambda^2 - (.8+.7)\lambda + (.8)(.7) - (.2)(.3) &= 0 \\ \lambda^2 - 1.5\lambda + 0.5 &= 0, \end{aligned}$$

which are

$$\begin{aligned} \lambda &= \frac{1}{2}(1.5 \pm \sqrt{(1.5)^2 - 4(0.5)}) \\ &= \frac{1}{2}(1.5 \pm \sqrt{0.25}) \\ &= \frac{1}{2}(1.5 \pm 0.5) \\ &= 1 \text{ and } 1/2. \end{aligned}$$

To find the 1-eigenvectors, we use row reduction to compute the nullspace of $1I - A$. First we observe that the matrix $1I - A$ has dependent rows (hence also dependent columns):

$$1I - A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} .8 & .3 \\ .2 & .7 \end{pmatrix} = \begin{pmatrix} .2 & -.3 \\ -.2 & .3 \end{pmatrix}.$$

Indeed, this must be the case because 1 is an eigenvalue. Then we compute the RREF:

$$(1I - A)\mathbf{x} = \mathbf{0} \quad \rightsquigarrow \quad \begin{pmatrix} .2 & -.3 \\ -.2 & .3 \end{pmatrix}\mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \overset{\text{RREF}}{\rightsquigarrow} \quad \begin{pmatrix} 1 & -3/2 \\ 0 & 0 \end{pmatrix}\mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

It follows that there is a line of 1-eigenvectors:[116]

$$\mathbf{x} = t\begin{pmatrix} 3/2 \\ 1 \end{pmatrix}.$$

Next we compute the $(1/2)$-eigenspace:

$$\left(\frac{1}{2}I - A\right)\mathbf{x} = \mathbf{0} \quad \rightsquigarrow \quad \begin{pmatrix} -.3 & -.3 \\ -.2 & -.2 \end{pmatrix}\mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \overset{\text{RREF}}{\rightsquigarrow} \quad \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

[116] In the previous section I chose $\mathbf{x} = (3, 2)$ to avoid fractions.

Thus we have also have a line of (1/2)-eigenvectors:[117]

$$\mathbf{x} = t\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

The procedure is the same for larger matrices. Given the eigenvalues, we can find all of the eigenvectors by row reduction. The hard part is to find the eigenvalues.[118] In general, we define the *characteristic polynomial* of a square matrix A:

$$\boxed{\chi_A(\lambda) := \det(\lambda I - A).}$$

This is, indeed, a polynomial in λ. Furthermore, if A is $n \times n$ then $\chi_A(\lambda)$ is a polynomial of degree n. In general the coefficients are quite complicated, but two of the coefficients have special names. We have

$$\boxed{\chi_A(\lambda) = \lambda^n - \operatorname{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A),}$$

where the *trace* of a square matrix is defined as the sum of its diagonal entries:

$$\boxed{\operatorname{tr}(A) = \operatorname{tr}\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} := a_{11} + a_{22} + \cdots + a_{nn}.}$$

We already know this formula for 2×2 matrices, and the general case is not hard to check.

If matrices A and B satisfy $B = XAX^{-1}$ for some invertible matrix X, then I claim that A and B have the same characteristic polynomial:

$$\boxed{\chi_{XAX^{-1}}(\lambda) = \chi_A(\lambda).}$$

To prove this, we note that

$$\begin{aligned} \chi_B(\lambda) &= \det(\lambda I - B) \\ &= \det(\lambda XX^{-1} - XAX^{-1}) \\ &= \det(X(\lambda I - A)X^{-1}) \\ &= \cancel{\det(X)}\det(\lambda I - A)\cancel{\det(X)^{-1}} \\ &= \det(\lambda I - A) \\ &= \chi_A(\lambda). \end{aligned}$$

[117] In the previous example I chose $\mathbf{x} = (1, -1)$ because I didn't want a negative sign in the first coordinate.

[118] Solving polynomials equations is a non-linear problem. There are no exact algorithms, but there are reasonably good approximation schemes. The state of the art for computing eigenvalues is the *QR algorithm*, which doesn't use the characteristic polynomial at all.

By comparing the coefficients of $\chi_A(\lambda)$ and $\chi_B(\lambda)$, it follows that[119]

$$\boxed{\operatorname{tr}(XAX^{-1}) = \operatorname{tr}(A) \quad \text{and} \quad \det(XAX^{-1}) = \det(A).}$$

The eigenvalues of a square matrix A are the roots of the characteristic polynomial. It follows from the Fundamental Theorem of Algebra that

$$\boxed{\text{Every square matrix has at least one eigenvalue.}}$$

Indeed, since the characteristic polynomial $\chi_A(\lambda)$ has degree n, the FTA says that there exist complex numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that

$$\chi_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n).$$

We can expand this to get

$$\boxed{\chi_A(\lambda) = \lambda^n - (\lambda_1 + \cdots + \lambda_n)\lambda^{n-1} + \cdots + (-1)^n \lambda_1 \cdots \lambda_n.}$$

Then comparing the coefficients with our previous expansion for $\chi_A(\lambda)$ gives

$$\boxed{\operatorname{tr}(A) = \lambda_1 + \cdots + \lambda_n \quad \text{and} \quad \det(A) = \lambda_1 \cdots \lambda_n.}$$

That is, the trace of A equals the sum of the eigenvalues (with multiplicities) and the determinant of A equals the product of the eigenvalues (with multiplicities). This is often useful.

Remarks:

- I guess we could say that every $n \times n$ matrix has n eigenvalues, but they need not be distinct. For example, the identity matrix I_n has characteristic polynomial

$$\det(\lambda I_n - I_n) = \det \begin{pmatrix} \lambda - 1 & & \\ & \ddots & \\ & & \lambda - 1 \end{pmatrix} = (\lambda - 1)^n,$$

  hence 1 is the only eigenvalue. The corresponding eigenspace is all of $\mathbb{R}^n$. Indeed, every vector $\mathbf{x} \in \mathbb{R}^n$ is a 1-eigenvector of the identity matrix: $I_n \mathbf{x} = \mathbf{x} = 1\mathbf{x}$.

- A real matrix has at least one complex eigenvalue, but it need not have any real eigenvalues. For example, consider the matrix that rotates counterclockwise by 90°:

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

  The characteristic polynomial is

$$\det(\lambda I - A) = \begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix} = \lambda^2 + 1,$$

  hence the eigenvalues are $\pm i$. The corresponding eigenspaces are

$$\mathcal{N}(iI - A) = t \begin{pmatrix} 1 \\ -i \end{pmatrix} \quad \text{and} \quad \mathcal{N}(-iI - A) = t \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

  Hence A is a real matrix with no real eigenvalues and no real eigenvectors.

[119] Of course, we already knew this property of the determinant.

11.3 Diagonalization

We say that a square matrix A is *diagonalizable* when it has a basis of eigenvectors. So far we have seen only diagonalizable matrices. Here is the simplest example of a matrix that is **not diagonalizable**:

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

The characteristic polynomial is

$$\det(\lambda I - A) = \det \begin{pmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 1 \end{pmatrix} = (\lambda - 1)^2,$$

hence 1 is the only eigenvalue. But the 1-eigenspace is only one dimensional:

$$(1I - A)\mathbf{x} = \mathbf{0} \quad \rightsquigarrow \quad \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \rightsquigarrow \quad \mathbf{x} = t \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Non-diagonalizable matrices are quite a nuisance. Fortunately, they are rare. The next result shows that any $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Theorem (Distinct Eigenvalues Implies Diagonalizable). Let A be an $n \times n$ matrix. Suppose that the characteristic polynomial factors as

$$\chi_A(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n),$$

where the complex numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ are distinct. Furthermore, let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^n$ be some nonzero vectors satisfying $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$.[120] Then I claim that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is a basis for $\mathbb{C}^n$.

Warning: This theorem is not sharp. If the characteristic polynomial of a matrix has a repeated factor then the matrix may or may not be diagonalizable. For example, the following two matrices both have characteristic polynomial $(\lambda - 1)^2$:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The matrix on the left is not diagonalizable, but the matrix on the right **is** diagonalizable. Indeed, every vector is a 1-eigenvector for the identity matrix. I will give a sharp characterization of diagonalizability in the next section.

Proof. It is enough to show that the set $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is linearly independent. Then the subspace of $\mathbb{C}^n$ spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is n-dimensional, hence it must be the whole space.

First we observe that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct. Indeed, suppose we had $\mathbf{x}_i = \mathbf{x}_j = \mathbf{x}$ for some $i \neq j$. This would imply that

$$A\mathbf{x} = A\mathbf{x}$$

[120] Such vectors exist because $\lambda_1, \ldots, \lambda_n$ are eigenvalues.

$$\lambda_i \mathbf{x} = \lambda_j \mathbf{x}$$
$$(\lambda_i - \lambda_j)\mathbf{x} = \mathbf{0}.$$

But by assumption we have $\lambda_i - \lambda_j \neq 0$ and $\mathbf{x} \neq \mathbf{0}$, which gives a contradiction. This is no big deal; it just says that a given vector can't be an eigenvector for two different eigenvalues.

We will prove by induction on k that the set of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is independent for any $1 \leq k \leq n$, and it will follow that the set $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is independent. The result is trivial for $k = 1$ because any set containing one vector is by convention called independent. It is not logically necessary, but let's also consider the case $k = 2$, to get a feel for the general argument. Suppose that we have

$$b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 = \mathbf{0}$$

for some scalars $b_1, b_2 \in \mathbb{C}$. Our goal is to show that $b_1 = 0$ and $b_2 = 0$. First multiply both sides on the left by A to obtain

$$A(b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2) = A\mathbf{0}$$
$$b_1 A\mathbf{x}_1 + b_2 A\mathbf{x}_2 = \mathbf{0}$$
$$b_1 \lambda_1 \mathbf{x}_1 + b_2 \lambda_2 \mathbf{x}_2 = \mathbf{0}.$$

Subtract λ_2 times the first equation from this equation to obtain

$$(b_1 \lambda_1 \mathbf{x}_1 + b_2 \lambda_2 \mathbf{x}_2) - \lambda_2(b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2) = \mathbf{0}$$
$$b_1(\lambda_1 - \lambda_2)\mathbf{x}_1 + b_2(\lambda_2 - \lambda_2)\mathbf{x}_2 = \mathbf{0}$$
$$b_1(\lambda_1 - \lambda_2)\mathbf{x}_1 = \mathbf{0}.$$

Since $\lambda_1 - \lambda_2 \neq 0$ and $\mathbf{x}_1 \neq \mathbf{0}$ this implies that $b_1 = 0$. But then substituting into the first equation gives

$$b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 = \mathbf{0}$$
$$0\mathbf{x}_1 + b_2 \mathbf{x}_2 = \mathbf{0}$$
$$b_2 \mathbf{x}_2 = \mathbf{0},$$

which implies that $b_2 = 0$ because $\mathbf{x}_2 \neq \mathbf{0}$.

Now we prove the general case. Fix some $k \geq 2$ and suppose that we have

$$b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 + \cdots + b_k \mathbf{x}_k = \mathbf{0} \tag{1}$$

for some scalars $b_1, \ldots, b_k \in \mathbb{C}$. Our goal is to show that $b_1 = b_2 = \cdots = b_k = 0$. To do this we multiply both sides on the left by A to obtain

$$A(b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 + \cdots + b_k \mathbf{x}_k) = A\mathbf{0}$$
$$b_1 A\mathbf{x}_1 + b_2 A\mathbf{x}_2 + \cdots + b_k A\mathbf{x}_k = \mathbf{0}$$
$$b_1 \lambda_1 \mathbf{x}_1 + b_2 \lambda_2 \mathbf{x}_2 + \cdots + b_k \lambda_k \mathbf{x}_k = \mathbf{0}. \tag{2}$$

Then we consider the equation (2) $-\lambda_k(1)$:

$$\begin{aligned}\left(\sum_{i=1}^{k} b_i\lambda_i\mathbf{x}_i\right) - \lambda_k\left(\sum_{i=1}^{k} b_i\mathbf{x}_i\right) &= \mathbf{0}\\ \sum_{i=1}^{k} b_i(\lambda_i-\lambda_k)\mathbf{x}_k &= \mathbf{0}\\ 0\mathbf{x}_k + \sum_{i=1}^{k-1} b_i(\lambda_i-\lambda_k)\mathbf{x}_i &= \mathbf{0}\\ \sum_{i=1}^{k-1} b_i(\lambda_i-\lambda_k)\mathbf{x}_i &= \mathbf{0}.\end{aligned}$$

By induction, the vectors $\mathbf{x}_1,\ldots,\mathbf{x}_{k-1}$ are independent, hence for any $1\le i\le k-1$ we must have $b_i(\lambda_i-\lambda_k)=0$. But by assumption we have $\lambda_i\neq\lambda_k$, and hence $b_i=0$, for any $1\le i\le k-1$. Finally, we substitute back into equation (1) to obtain $0\mathbf{x}_1+\cdots+0\mathbf{x}_{k-1}+b_k\mathbf{x}_k=\mathbf{0}$, which implies that $b_k=0$ because $\mathbf{x}_k\neq\mathbf{0}$. □

This result implies that "almost all" matrices are diagonalizable. To see this, we will use the concept of the *discriminant* of a polynomial. For example, consider the general 2×2 matrix

$$A=\begin{pmatrix}a & b\\ c & d\end{pmatrix}.$$

The characteristic polynomial is

$$\chi_A(\lambda)=\lambda^2-(a+d)\lambda+(ad-bc),$$

and hence the eigenvalues are

$$\lambda=\frac{a+d\pm\sqrt{(a+d)^2-4(ad-bc)}}{2}.$$

The quantity $\Delta(a,b,c,d)=(a+d)^2-4(ad-bc)$ is called the discriminant of the characteristic polynomial. If $\Delta\neq 0$ then we observe that A has two distinct eigenvalues, hence is diagonalizable. If we choose the entries a,b,c,d of the matrix A at random then it would be quite unlikely to have $\Delta(a,b,c,d)=0$. To be more precise, we can view the set of 2×2 matrices as a 4-dimensional vector space:

$$\mathbb{C}^{2\times 2}=\text{the vector space of } 2\times 2 \text{ matrices with complex entries.}$$

Inside this 4-dimensional vector space, the set of matrices satisfying $\Delta(a,b,c,d)=0$ forms a "3-dimensional subset".[121] By analogy, consider a 2-dimensional plane in $\mathbb{R}^3$. A randomly

[121]For a general polynomial $f(x_1,\ldots,x_n)$ in n variables, the set of points $\mathbf{x}\in\mathbb{C}^n$ satisfying $f(\mathbf{x})=0$ forms an $(n-1)$-dimensional subset. I don't want to be too precise about this.

chosen point in $\mathbb{R}^3$ will not lie on this plane. Similarly, a randomly chosen point in a 4-dimensional vector space will not lie in a given 3-dimensional shape.

This discussion generalizes to square matrices of any size. Given an $n \times n$ matrix A with entries a_{ij}, there is a certain polynomial $\Delta(A)$ in the n^2 variables a_{ij} such that $\Delta(A) = 0$ if and only if A has a repeated eigenvalue. Since the equation $\Delta(A) = 0$ defines an $(n^2 - 1)$-dimensional subset of the n^2-dimensional space of $n \times n$ matrices, a randomly chosen matrix will have distinct eigenvalues, and hence will be diagonalizable.

Why do we call it diagonalization? Let A be a diagonalizable $n \times n$ matrix and let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{C}^n$ be a basis of eigenvectors with corresponding eigenvalues $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$. (Here we do not assume that the eigenvalues are distinct.) We can write the n equations $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ simultaneously as a matrix equation:

$$\begin{pmatrix} A\mathbf{x}_1 \mid \cdots \mid A\mathbf{x}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{x}_1 \mid \cdots \mid \lambda_n\mathbf{x}_n \end{pmatrix}$$

$$A \begin{pmatrix} \mathbf{x}_1 \mid \cdots \mid \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \mid \cdots \mid \mathbf{x}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$$AX = X\Lambda,$$

where Λ is a **diagonal** matrix containing the eigenvalues. Since the columns of X are independent by assumption, the matrix X is invertible and we can write

$$\boxed{A = X\Lambda X^{-1}} \quad \text{or} \quad \boxed{X^{-1}AX = \Lambda.}$$

Thus we have used the eigenvector matrix X to convert A into the diagonal matrix Λ. In other words, we have "diagonalized" A. We will see below that this is extremely useful.

11.4 Evaluating a Polynomial at a Matrix

Matrices can be added, multiplied by scalars, and raised to powers. This allows us to consider polynomials of matrices. More precisely, we can "evaluate" polynomials at matrices. Consider a polynomial in one variable, with complex coefficients:

$$f(x) = b_0 + b_1 x + \cdots + b_k x^k.$$

Then for any $n \times n$ matrix A we define the $n \times n$ matrix $f(A)$ by[122]

$$f(A) := b_0 I_n + b_1 A + b_2 A^2 + \cdots + b_k A^k.$$

This evaluation behaves well with respect to eigenvalues and eigenvectors. That is, for any vector $\mathbf{x}$ and scalar λ, we have

$$\boxed{A\mathbf{x} = \lambda\mathbf{x} \quad \Longrightarrow \quad f(A)\mathbf{x} = f(\lambda)\mathbf{x}.}$$

Indeed, if $A\mathbf{x} = \lambda\mathbf{x}$ then we can show by induction that $A^m\mathbf{x} = \lambda^m\mathbf{x}$ for any $m \geq 0$:

[122] Here we use the convention that $A^0 = I_n$.

- **Base Case.** $A^0\mathbf{x} = I_n\mathbf{x} = \mathbf{x} = \lambda^0\mathbf{x}$.
- **Induction Step.** If $A^{m-1}\mathbf{x} = \lambda^{m-1}\mathbf{x}$ then

$$A^m\mathbf{x} = A(A^{m-1}\mathbf{x}) = A^m(\lambda^{m-1}\mathbf{x}) = \lambda^{m-1}(A\mathbf{x}) = \lambda^{m-1}\lambda\mathbf{x} = \lambda^m\mathbf{x}.$$

Then for any polynomial $f(x) = b_0 + b_1x + \cdots + b_kx^k$ we have

$$\begin{aligned} f(A)\mathbf{x} &= (b_0I_n + b_1A + \cdots + b_kA^k)\mathbf{x} \\ &= b_0I_n\mathbf{x} + b_1A\mathbf{x} + \cdots + b_kA^k\mathbf{x} \\ &= b_0\mathbf{x} + b_1\lambda\mathbf{x} + \cdots + b_k\lambda^k\mathbf{x} \\ &= (b_0 + b_1\lambda + \cdots + b_k\lambda^k)\mathbf{x} \\ &= f(\lambda)\mathbf{x}. \end{aligned}$$

Now suppose that the matrix A is a "root" of the polynomial $f(x)$. That is, suppose that $f(A)$ is the zero matrix. Then every eigenvalue of A is also a root of $f(x)$:

$$\boxed{f(A) = O \quad \Longrightarrow \quad \text{every eigenvalue of } A \text{ satisfies } f(\lambda) = 0.}$$

Indeed, if $A\mathbf{x} = \lambda\mathbf{x}$ and $f(A) = O$ then we have

$$f(\lambda)\mathbf{x} = f(A)\mathbf{x} = O\mathbf{x} = \mathbf{0}.$$

And if $\mathbf{x} \neq \mathbf{0}$ then this implies $f(\lambda) = 0$. Here are some examples.

Projections. Any matrix satisfying $P^2 = P$ has eigenvalues in the set $\{0, 1\}$. Indeed, if $P^2 - P = O$ then any eigenvalue λ of P satisfies

$$\begin{aligned} \lambda^2 - \lambda &= 0 \\ \lambda(\lambda - 1) &= 0 \\ \lambda &= 0 \text{ or } 1. \end{aligned}$$

This doesn't mean that both eigenvalues must occur. For example, the zero matrix satisfies $O^2 = O$ and its only eigenvalue is 0, while the identity matrix satisfies $I^2 = I$ and its only eigenvalue is 1.

At the end of this section we will show that any matrix satisfying $P^2 = P$ is diagonalizable. Assuming this for now, we can prove that any $n \times n$ matrix satisfying $P^2 = P$ is a (possibly non-orthogonal) projection matrix. To do this, let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a basis of eigenvectors. Since the only possible eigenvalues are 1 and 0, we can sort the eigenvectors so that $P\mathbf{x}_i = 1\mathbf{x}_i = \mathbf{x}_i$ for $0 \leq i \leq r$ and $P\mathbf{x}_i = 0\mathbf{x}_i = \mathbf{0}$ for $r < i \leq n$.[123] This gives the factorization

$$P = X \left(\begin{array}{c|c} I_r & O_{r,n-r} \\ \hline O_{n-r,r} & O_{n-r}, n-r \end{array} \right) X^{-1}.$$

[123] We allow the possibilities $r = 0$ (all eigenvalues are 0) and $r = n$ (all eigenvalues are 1).

Now let A be the $n \times r$ matrix consisting of the first r columns of X and let B be the $r \times n$ matrix consisting of the first r rows of X^{-1}. Then we have

$$P = \left(\begin{array}{c|c} A & * \end{array} \right) \left(\begin{array}{c|c} I_r & O_{r,n-r} \\ \hline O_{n-r,r} & O_{n-r,n-r} \end{array} \right) \left(\begin{array}{c} B \\ \hline * \end{array} \right) = \left(\begin{array}{c|c} A & * \end{array} \right) \left(\begin{array}{c} B \\ \hline O \end{array} \right) = AB + O = AB.$$

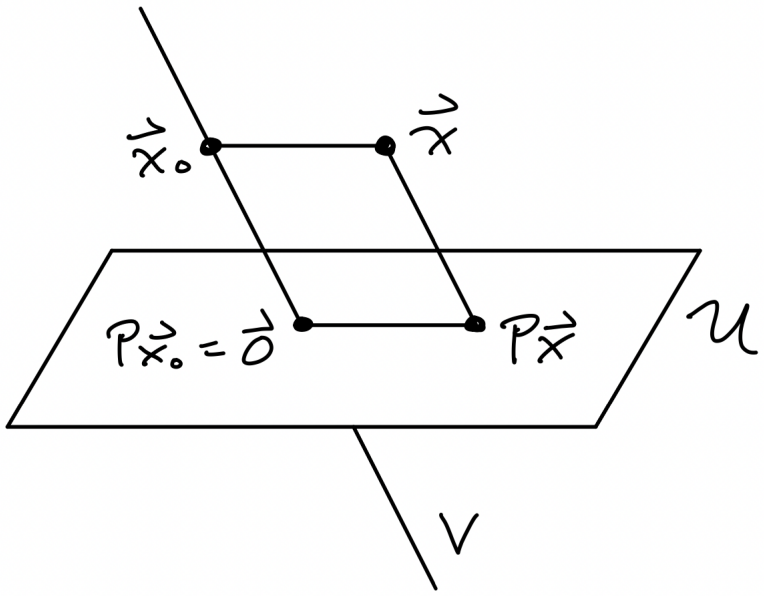
And we also have

$$\left(\begin{array}{c|c} I_r & O_{r,n-r} \\ \hline O_{n-r,r} & I_{n-r,n-r} \end{array} \right) = I_n = X^{-1}X = \left(\begin{array}{c} B \\ \hline * \end{array} \right) \left(\begin{array}{c|c} A & * \end{array} \right) = \left(\begin{array}{c|c} BA & * \\ \hline * & * \end{array} \right),$$

which implies that $BA = I_r$. In summary:

$$\boxed{P^2 = P \quad \Longrightarrow \quad P = AB \text{ for some } A, B \text{ satisfying } BA = I_r, \text{ where } r = \operatorname{rank}(P).}$$

This is the projection onto the column space $U = \mathcal{C}(A)$, in a direction parallel to the null space $V = \mathcal{N}(B)$. Picture:



The projection is orthogonal if and only if $V = U^\perp$. In this case we have $\mathcal{N}(B) = \mathcal{C}(A)^\perp = \mathcal{N}(A^T)$, which impliess that $\mathcal{R}(A) = \mathcal{N}(B)^T = \mathcal{N}(A^T)^\perp = \mathcal{R}(A^T)$. Since $\mathcal{R}(B) = \mathcal{R}(A^T)$ we can find an invertible $r \times r$ matrix S of row operations such that $B = SA^T$. But then $BA = I_r$ implies $SA^TA = I_r$ and hence $S = (A^TA)^{-1}$. Finally, we conclude that

$$P = AB = ASA^T = A(A^TA)^{-1}A^T,$$

which agrees with our previous formula for orthogonal projections.

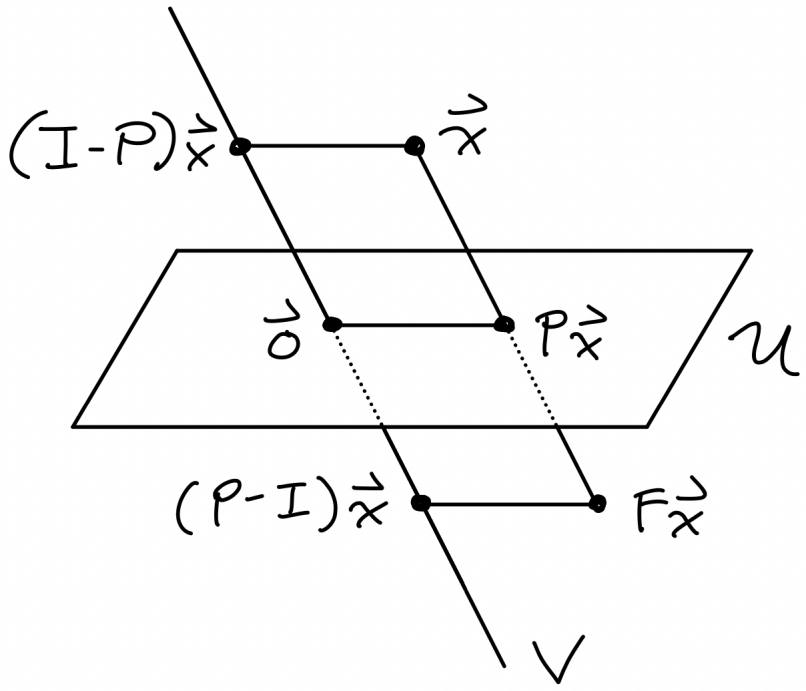
Reflections. Any matrix satisfying $F^2 = I$ has eigenvalues in the set $\{1, -1\}$. Indeed, if $F^2 - I = O$ then any eigenvalue λ of F satisfies

$$\begin{aligned} \lambda^2 - 1 &= 0 \\ \lambda^2 &= 1 \\ \lambda &= 1 \text{ or } -1. \end{aligned}$$

Consider the unique matrix P satisfying $F = 2P - I$ and $P = (F + I)/2$. We observe that

$$P^2 = \frac{1}{4}(F^2 + 2F + I^2) = \frac{1}{2}(I + 2F + I) = \frac{1}{4}(2F + 2I) = \frac{1}{2}(F + I) = P,$$

so that P is a projection. Let U and V be the 1-eigenspace and 0-eigenspace of P as in the previous example, then U is the 1-eigenspace of F and V is the (-1)-eigenspace of F. Geometrically, F is the reflection across the subspace U in the direction of V. Picture:



In terms of matrices, if F is a matrix of rank r satisfying $F^2 = I$ then we can find two $r \times (n-r)$ matrices A, B satisfying $BA = I_r$, such that

$$F = 2P - I = 2AB - I.$$

This is the reflection across the column space $U = \mathcal{C}(A)$, parallel to the nullspace $V = \mathcal{N}(B)$.

Rotations. Any matrix satisfying $R^n = I$ has eigenvalues in the set $\{e^{2\pi ik/n} : k \in \mathbb{Z}\}$. Indeed, since $R^n - I = O$, any eigenvalue λ of R must satisfy $\lambda^n - 1 = 0$, and hence must be an nth

root of unity. Such matrices can be quite complicated. For a simple example, we consider the 2×2 rotation matrix:

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The characteristic polynomial is

$$\chi_{R_\theta}(\lambda) = \lambda^2 - 2\cos\theta\lambda + 1,$$

hence the eigenvalues are[124]

$$\begin{aligned} \lambda &= \frac{2\cos\theta \pm \sqrt{4\cos^2\theta - 4}}{2} \\ &= \frac{2\cos\theta \pm 2\sqrt{\cos^2\theta - 1}}{2} \\ &= \frac{2\cos\theta \pm 2\sqrt{-\sin^2\theta}}{2} \\ &= \frac{2\cos\theta \pm 2i\sin\theta}{2} \\ &= \cos\theta \pm i\sin\theta \\ &= e^{\pm i\theta}. \end{aligned}$$

The case $\theta = 0$ corresponds to the identity matrix, with eigenvalues $(1, 1)$ and the case $\theta = \pi$ corresponds to the negative identity matrix, with eigenvalues $(-1, -1)$. In all other cases, the eigenvalues (and hence also the eigenvectors) are not real. If $\theta = 2\pi/n$ then the eigenvalues $\lambda = e^{\pm 2\pi i/n}$ satisfy $\lambda^n = 1$. This agrees with the fact that

$$(R_{2\pi/n})^n = I.$$

Next, we give an alternative proof for the existence of eigenvalues, which does not use determinants.[125]

Theorem (Existence of Eigenvalues). Let A be any $n \times n$ matrix with real or complex entries and consider an arbitrary nonzero vector $\mathbf{v} \in \mathbb{C}^n$. Since $\mathbb{C}^n$ is n-dimensional, the following $n + 1$ vectors must be linearly dependent:[126]

$$\mathbf{v}, A\mathbf{v}, A^2\mathbf{x}, \ldots, A^n\mathbf{v}.$$

In other words, we can find scalars $b_0, b_1, \ldots, b_n$, not all zero, such that

$$b_0\mathbf{v} + b_1A\mathbf{v} + b_2A^2\mathbf{v} + \cdots + b_nA^n\mathbf{v} = \mathbf{0}.$$

In fact, one of the scalars $b_1, \ldots, b_n$ must be nonzero, otherwise we would have $b_0\mathbf{v} = \mathbf{0}$ and $b_0 \neq 0$, which contradicts the fact that $\mathbf{v} \neq \mathbf{0}$. We can rewrite the previous equation as

$$(b_0I + b_1A + b_2A^2 + \cdots + b_nA^n)\mathbf{v} = \mathbf{0},$$

[124] It's a bit reckless to take square roots in this way, but it gives the correct answer.
[125] This proof is the motivation for Axler's approach in *Linear Algebra Done Right*.
[126] Indeed, **any** collection of $n + 1$ vectors in $\mathbb{C}^n$ is linearly dependent.

$$f(A) = \mathbf{0},$$

for the polynomial $f(x) = b_0 + b_1 x + b_2 x^2 + \cdots b_n x^n$, which has degree between 1 and n because not all of $b_1, \ldots, b_n$ are zero. Let's say $\deg(f) = k$. By the Fundamental Theorem of Algebra we can factor $f(x)$ as

$$f(x) = (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_k),$$

for some complex numbers $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$, not necessarily distinct. Now the equation $f(A)\mathbf{v} = \mathbf{0}$ becomes[127]

$$(A - \alpha_1 I)(A - \alpha_2 I) \cdots (A - \alpha_k I)\mathbf{v} = \mathbf{0}.$$

To save notation, let's write $A_i = A - \alpha_i I$. Thus we have

$$A_1 A_2 \cdots A_k \mathbf{v} = \mathbf{0}.$$

Since $\mathbf{v} \neq \mathbf{0}$, this implies that the matrix $A_1 \cdots A_k$ is not invertible. And since a product of invertible matrices is invertible, this implies that at least one of the factors, say A_i, is not invertible. Finally, since $A_i = A - \alpha_i I$ is not invertible, we conclude that α_i is an eigenvalue of A. In particular, we have shown that A has an eigenvalue. □

Building on this idea, we can give a sharper result about diagonalization. The proof is tricky so you can feel free to skip it.

Theorem (Existence of Diagonalization). A square matrix A is diagonalizable if and only if we have $f(A) = O$ for some polynomial with no repeated roots.

Proof. First suppose that A has a basis of eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with corresponding eigenvalues $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$. Some of these eigenvalues might be repeated. Let $\mu_1, \ldots, \mu_k$ be the list of eigenvalues with repetition removed, and consider the polynomial

$$f(x) = (x - \mu_1) \cdots (x - \mu_k),$$

which has no repeated roots. We will show that $f(A) = O$. To do this, we observe that the matrices $(A - \mu_i I)$ and $(A - \mu_j I)$ commute for any i, j:

$$(A - \mu_i I)(A - \mu_j I) = A^2 - (\mu_i + \mu_j)A + \mu_i \mu_j I = (A - \mu_j I)(A - \mu_i - I).$$

We will use this to show that $f(A)\mathbf{v} = \mathbf{0}$ for any eigenvector $\mathbf{v}$. Then since there exists a basis of eigenvectors, it will follow from this that $f(A)\mathbf{v} = \mathbf{0}$ for **any** vector $\mathbf{v}$, and hence $f(A)$ is the zero matrix. So let $\mathbf{v}$ be an eigenvector with eigenvalue μ_i.[128] Then we have[129]

$$f(A)\mathbf{v} = \left(\prod_j (A - \mu_j I) \right) \mathbf{v}$$

[127] There is a subtle point hiding here. Given a polynomial $f(x)$ and square matrices A, B, it is **not** generally true that $f(AB) = f(A)f(B)$. However, if $AB = BA$ then we do have $f(AB) = f(A)f(B)$. Since the matrices $A - \alpha_1 I, \ldots, A - \alpha_k I$ commute with each other, we are okay in this case.

[128] Recall that every eigenvalue is in the list $\mu_1, \ldots, \mu_k$.

[129] We used commutativity to pull the factor $A - \mu_i I$ to the right.

$$
\begin{aligned}
&= \left(\prod_{j\neq i}(A-\mu_j)\right)(A-\mu_i)\mathbf{v} \\
&= \left(\prod_{j\neq i}(A-\mu_j)\right)(A\mathbf{v}-\mu_{\mathbf{v}}) \\
&= \left(\prod_{j\neq i}(A-\mu_j)\right)\mathbf{0} \\
&= \mathbf{0}.
\end{aligned}
$$

Thus we have shown that a diagonalizable matrix A satisfies an equation $f(A) = O$ for some polynomial $f(x)$ with no repeated roots.

Conversely, suppose that an $n \times n$ matrix A satisfies $f(A) = O$ for some polynomial $f(x)$ with no repeated roots. Suppose that $\deg(f) = k$ and write

$$f(x) = (x-\lambda_1)\cdots(x-\lambda_k)$$

for some distinct complex numbers $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$. We want to show that A has a basis of eigenvectors. First we define the null spaces

$$E_{\lambda_i} = \mathcal{N}(A-\lambda_i I) = \{\mathbf{x} : A\mathbf{x} = \lambda_i \mathbf{x}\}.$$

Note that $E_{\lambda_i} \neq \{\mathbf{0}\}$ if and only if λ_i is an eigenvalue, in which case E_{λ_i} is the corresponding eigenspace. We don't really care if all of the numbers λ_i are eigenvalues. Indeed, some of them might not be. Our goal is merely to show that the spaces $E_{\lambda_1}, \ldots, E_{\lambda_k}$ are big enough to fill up all of $\mathbb{C}^n$. To be precise, we will show that

- $E_{\lambda_i} \cap E_{\lambda_j} = \{\mathbf{0}\}$ for all $i \neq j$,
- $\mathbb{C}^n = \{\mathbf{x}_1 + \cdots + \mathbf{x}_k : \mathbf{x}_i \in E_{\lambda_i} \text{ for all } i\}$.

Then by concatenating bases for $E_{\lambda_1}, \ldots, E_{\lambda_k}$ we will obtain a basis for $\mathbb{C}^n$ that consists of eigenvectors of A. For the first statement, suppose that $\mathbf{x} \in E_{\lambda_i} \cap E_{\lambda_j}$ so that $\lambda_i \mathbf{x} = A\mathbf{x} = \lambda_j \mathbf{x}$. If $\mathbf{x} \neq \mathbf{0}$ then this implies that $\lambda_i = \lambda_j$ and hence $i \neq j$, because the λ_i are distinct. The second statement is trickier. First we consider the partial fraction expansion of $1/f(x)$:

$$\frac{1}{f(x)} = \frac{1}{(x-\lambda_1)\cdots(x-\lambda_k)} = \sum_i \frac{\alpha_i}{x-\lambda_i},$$

for some scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$, not necessarily distinct.[130] Now consider the polynomials

$$p_i(x) = \frac{\alpha_i f(x)}{x-\lambda_i} = \alpha_i \prod_{j\neq i}(x-\lambda_j),$$

[130] I won't prove the existence of the partial fraction expansion. It depends on the theory of greatest common divisors in the ring of polynomials.

and note that

$$\begin{aligned} p_1(x) + \cdots p_k(x) &= \frac{\alpha_1 f(x)}{x - \lambda_1} + \cdots + \frac{\alpha_k f(x)}{x - \lambda_k} \\ &= f(x) \cdot \sum_i \frac{\alpha_i}{x - \lambda_i} \\ &= f(x) \cdot \frac{1}{f(x)} \\ &= 1. \end{aligned}$$

Finally, consider any vector $\mathbf{x} \in \mathbb{C}^n$ and write $\mathbf{x}_i := p_i(A)\mathbf{x}$. On the one hand, by evaluating the polynomial equation $(x - \lambda_i)p_i(A) = \alpha_i f(x)$ at A we have

$$(A - \lambda_i I)\mathbf{x}_i = (A - \lambda_i I)p_i(A)\mathbf{x} = \alpha_i f(A)\mathbf{x} = O\mathbf{x} = \mathbf{0},$$

and hence $\mathbf{x}_i \in E_{\lambda_i}$. On the other hand, by evaluating the polynomial equation $p_1(x) + \cdots + p_k(x) = 1$ at A we have

$$\begin{aligned} \mathbf{x}_1 + \cdots + \mathbf{x}_k &= p_1(A)\mathbf{x} + \cdots + p_k(A)\mathbf{x} \\ &= (p_1(A) + \cdots p_k(A))\mathbf{x} \\ &= I\mathbf{x} \\ &= \mathbf{x}, \end{aligned}$$

as desired. □

That was a tricky proof, but it's a useful theorem. In particular, it implies that any matrix satisfying $P^2 = P$, and hence $P^2 - P = O$, is diagonalizable because the polynomial $x^2 - x = x(x-1)$ has distinct roots. Furthermore, any matrix satisfying $R^n = I$, and hence $R^n - I = O$, is diagonalizable because the polynomial $x^n - 1$ has distinct roots:

$$x^n - 1 = (x - 1)(x - e^{2\pi i/n})(x - e^{4\pi i/n}) \cdots (x - e^{2\pi i(n-1)/n}).$$

Finally, we examine what goes wrong for a specific non-diagonalizable matrix. Consider the following small matrices with repeated eigenvalues:

$$A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

Each of these has characteristic polynomial $(x - \lambda)^2$:

$$(x - \lambda)^2 = \det \begin{pmatrix} x - \lambda & 0 \\ 0 & x - \lambda \end{pmatrix} = \det \begin{pmatrix} x - \lambda & -1 \\ 0 & x - \lambda \end{pmatrix}.$$

In the next section we will show that every matrix satisfies its own characteristic polynomial, which we can easily verify for these two matrices:

$$(A - \lambda I)^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad (B - \lambda I)^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The matrix A also satisfies the polynomial $f(x) = x - \lambda$, which has no repeated roots. This confirms that A is diagonalizable; in fact, it is diagonal. On the other hand, the matrix B is not diagonalizable. This is easy to check directly. Instead, we will prove it indirectly, by showing that B cannot satisfy any polynomial with distinct roots. The basic reason is that

$$(B - \lambda I)^2 = O \quad \text{but} \quad B - \lambda I \neq O.$$

Indeed, consider any polynomial $g(x) = (x - \lambda_1) \cdots (x - \lambda_k)$ with distinct roots $\lambda_1, \ldots, \lambda_k$. If $g(B) = O$ then one of the matrices $B - \lambda_j I$ must be non-invertible, so that λ_j is an eigenvalue and hence $\lambda_j = \lambda$. Since the λ_i are distinct, this implies that the λ_i with $i \neq j$ are **not** eigenvalues. Then since $g(B) = O$ equals $B - \lambda I$ times a product of invertible matrices $B - \lambda_i I$ for $i \neq j$, we conclude that $B - \lambda I = O$. Contradiction.

Remark: In the next chapter we will say more about non-diagonalizable matrices.

11.5 The Functional Calculus

Why are diagonalizable matrices good? As we mentioned in the first section, if we can diagonalize a matrix A then we can find explicit formulas for the entries of its powers A^k. More generally, diagonalizing a matrix allows us to compute any polynomial evaluation of the matrix $f(A)$. We can even compute convergent power series, such as

$$\exp(A) := I + A + \frac{1}{2}A^2 + \cdots + \frac{1}{k!}A^k + \cdots .$$

To begin, suppose that an $n \times n$ matrix A has a basis of eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, with corresponding eigenvalues $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$. (The eigenvalues are not necessarily distinct.) Then, as we showed in the previous section, we can write

$$A = X\Lambda X^{-1} = \begin{pmatrix} \mathbf{x}_1 \,|\, \cdots \,|\, \mathbf{x}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \,|\, \cdots \,|\, \mathbf{x}_n \end{pmatrix}^{-1}.$$

This factorization is compatible with polynomial evaluation. That is, for any polynomial $f(x)$, I claim that

$$\boxed{f(A) = X \cdot f(\Lambda) \cdot X^{-1}.}$$

If A (hence also Λ) is invertible, then we can even allow negative powers of x in the polynomial. Such expressions are called *Laurent polynomials*:

$$f(x) = b_{-\ell}x^{-\ell} + b_{-\ell+1}x^{-\ell+1} + \cdots + b_{k-1}x^{k-1} + b_k x^k \quad \text{for some } k, \ell \geq 0.$$

Actually, we will prove the more general fact that $A = XBX^{-1}$ implies $f(A) = X \cdot f(B) \cdot X^{-1}$ for any polynomial $f(x)$, and for any Laurent polynomial $f(x)$ when A (hence also B) is invertible. The first step is to prove that

$$\boxed{(XBX^{-1})^k = XB^kX^{-1} \text{ for all } k \geq 0, \text{ and also for } k < 0 \text{ when } B \text{ is invertible.}}$$

For this we use induction. When $k = 0$ we have $XB^0X^{-1} = XI_nX^{-1} = XX^{-1} = I_n = (XBX^{-1})^0$. Then for $k \geq 1$ we have

$$\begin{aligned}(XBX^{-1})^k &= (XBX^{-1})(XBX^{-1})^{k-1} \\ &= (XBX^{-1})(XB^{k-1}X^{-1}) && \text{induction} \\ &= XB(X^{-1}X)B^{k-1}X^{-1} \\ &= XBB^{k-1}X^{-1} \\ &= XB^kX^{-1}.\end{aligned}$$

If B is invertible, then for all $k \geq 0$ we also have

$$(XBX^{-1})^{-k} = [(XBX^{-1})^{-1}]^k = (XB^{-1}X^{-1})^k = X(B^{-1})^kX^{-1} = XB^{-k}X^{-1}.$$

Finally, for any polynomial $f(x) = b_0 + b_1x + \cdots + b_kx^k$ we have

$$\begin{aligned}f(A) &= b_0I + b_1A + b_2A^2 + \cdots + b_kA^k \\ &= b_0I + b_1(XBX^{-1}) + b_2(XBX^{-1})^2 + \cdots + b_k(XBX^{-1})^k \\ &= b_0(XX^{-1}) + b_1(XBX^{-1}) + b_2(XB^2X^{-1}) + \cdots + b_k(XB^kX^{-1}) \\ &= X(b_0I + b_1B + b_2B^2 + \cdots + b_kB^k)X^{-1} \\ &= X \cdot f(B) \cdot X^{-1}.\end{aligned}$$

The proof for Laurent polynomials is the same.

So far, this is not very useful. It becomes useful because of the following basic observation.

Multiplication of Diagonal Matrices is Easy. The formula for a general matrix product AB is complicated. However, multiplication of diagonal matrices is easy:

$$\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \begin{pmatrix} b_1 & & \\ & \ddots & \\ & & b_n \end{pmatrix} = \begin{pmatrix} a_1b_1 & & \\ & \ddots & \\ & & a_nb_n \end{pmatrix}.$$

It follows that for any diagonal matrix Λ and any (Laurent) polynomial $f(x)$ we have

$$f(\Lambda) = f\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix}.$$

This is the reason why diagonalization is a powerful technique.

Here is a first application. Recall the following results from the previous two sections:

- If the characteristic polynomial $\chi_A(x)$ has distinct roots then A is diagonalizable.

- If $f(A) = O$ for some polynomial $f(x)$ with distinct roots then A is diagonalizable.

The next theorem ties these results together.

The Cayley-Hamilton Theorem. Let A be a square matrix with characteristic polynomial $\chi_A(x) = \det(xI - A)$. Then we have

$$\boxed{\chi_A(A) = O.}$$

This is a strange idea, so let's first examine the 2×2 case. Consider the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

with characteristic polynomial

$$\chi_A(\lambda) = \lambda^2 - (a+d)\lambda + (ad - bc).$$

Then one can check (as Cayley and Hamilton did) that

$$\begin{aligned} \chi_A(A) &= A^2 - (a+d)A + (ad-bc)I \\ &= \begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \text{some calculations} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Why on earth should this be true? It is because of diagonalization.

Proof of Cayley-Hamilton. Suppose first that A is diagonalizable, with $A = X\Lambda X^{-1}$. For any eigenvalue λ of A, the characteristic polynomial satisfies $\chi_A(\lambda) = 0$ by definition. Hence

$$\chi_A(A) = X \cdot \chi_A(\Lambda) \cdot X^{-1} = X \begin{pmatrix} \chi_A(\lambda_1) & & \\ & \ddots & \\ & & \chi_A(\lambda_n) \end{pmatrix} X^{-1} = XOX^{-1} = O.$$

The result for non-diagonalizable matrices follows by continuity. That is, any non-diagonalizable matrix is a limit of diagonalizable matrices. And the entries of the matrix $\chi_A(A)$ are continuous functions of the entries of A. But each entry of $\chi_A(A)$ is zero for any diagonal matrix. Hence the entries of the limit are zero.[131] □

Remark: The Cayley-Hamilton is actually more general than this. It holds over any commutative ring. As written, our proof only works over the complex numbers.

[131] This is a typical way to deal with non-diagonalizable matrices, i.e., view them as limits of diagonalizable matrices in the space $\mathbb{C}^{n \times n}$ of square matrices.

Next we consider two examples of infinite power series.

The Geometric Series. Consider an $n \times n$ matrix A. Suppose that A is diagonalizable with

$$A = X\Lambda X^{-1} = X \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} X^{-1}.$$

Evaluating A at the polynomial $f(x) = 1 - x$ gives

$$I - A = X(I - \Lambda)X^{-1} = X \begin{pmatrix} 1-\lambda_1 & & \\ & \ddots & \\ & & 1-\lambda_n \end{pmatrix} X^{-1}.$$

If none of the eigenvalues is 1 then $I - \Lambda$ (hence also $I - A$) is invertible, and we obtain

$$(I - A)^{-1} = X(I - \Lambda)^{-1}X^{-1} = X \begin{pmatrix} 1/(1-\lambda_1) & & \\ & \ddots & \\ & & 1/(1-\lambda_n) \end{pmatrix} X^{-1}.$$

On the other hand, for all $k \geq 0$ we can evaluate the polynomial $f(x) = 1 + x + \cdots + x^k$ at A to obtain

$$\begin{aligned} I + A + \cdots + A^k &= X(I + \Lambda + \cdots + \Lambda^k)X^{-1} \\ &= X \begin{pmatrix} 1+\lambda_1+\cdots+\lambda_1^k & & \\ & \ddots & \\ & & 1+\lambda_n+\cdots+\lambda_n^k \end{pmatrix} X^{-1}. \end{aligned}$$

Finally, suppose that the eigenvalues satisfy $0 < |\lambda_i| < 1$ for all i. Then the usual geometric series for scalars implies that

$$I + \Lambda + \cdots + \Lambda^k \to (I - \Lambda)^{-1} \quad \text{as } k \to \infty.$$

The convergence is componentwise in each entry of the matrix. For a fixed invertible matrix X, the function $B \mapsto XBX^{-1}$ is continuous in the matrix entries, hence

$$X(I + \Lambda + \cdots + \Lambda^k)X^{-1} \to X(I - \Lambda)^{-1}X^{-1} \quad \text{as } k \to \infty.$$

In summary, for a diagonalizable matrix A with eigenvalues satisfying $0 < |\lambda| < 1$, we have

$$I + A + \cdots + A^k \to (I - A)^{-1} \text{ componentwise.}$$

And by continuity, the result also holds for non-diagonalizable matrices.[132] On a previous homework you proved a weaker version of this result, using more difficult techniques. Diagonalization makes things easier because it turns matrix arithmetic into scalar arithmetic.

[132]Basically, this is because the eigenvalues depend continuously on the matrix entries. I don't want to get specific about it.

The Matrix Exponential. Given a square matrix A, the functional calculus allows us to define $f(A)$ for any power series $f(x) = a_0 + a_1 x + a_2^2 + \cdots$, as long as this power series converges when evaluated at the eigenvalues of A. For example, consider the power series definition of the exponential function

$$\exp(x) = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots .$$

It is a basic theorem of analysis that $\exp(x)$ converges for any complex number $x \in \mathbb{C}$. In order to define $\exp(A)$ we first suppose that A is diagonalizable:

$$A = X\Lambda X^{-1} = X \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} X^{-1}.$$

For any $k \geq 0$ we have

$$\sum_{i=0}^{k} \frac{1}{i!} A^i = X \left(\sum_{i=0}^{k} \frac{1}{i!} \Lambda^i \right) X^{-1} = X \begin{pmatrix} \sum_{i=0}^{k} \frac{1}{i!}\lambda_1^i & & \\ & \ddots & \\ & & \sum_{i=0}^{k} \frac{1}{i!}\lambda_n^i \end{pmatrix} X^{-1}.$$

Since the power series for $\exp(x)$ converges everywhere, we have

$$\sum_{i=1}^{k} \frac{1}{i!} \cdot \Lambda_i \to \begin{pmatrix} \exp(\lambda_1) & & \\ & \ddots & \\ & & \exp(\lambda_n) \end{pmatrix} \quad \text{as } k \to \infty.$$

Then since conjugation by the fixed matrix X is continuous, we conclude that

$$\sum_{i=1}^{k} \frac{1}{i!} \cdot A^i \to X \begin{pmatrix} \exp(\lambda_1) & & \\ & \ddots & \\ & & \exp(\lambda_n) \end{pmatrix} X^{-1} \quad \text{as } k \to \infty.$$

This establishes the existence of the matrix exponential for any diagonalizable matrix A:[133]

$$\boxed{\exp(A) = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots .}$$

Let me warn you that

$$\exp(A + B) \neq \exp(A)\exp(B) \text{ for general matrices } A, B.$$

[133] We can also prove existence for non-diagonalizable matrices using a continuity argument, though this proof doesn't tell us how to compute $\exp(A)$ in the non-diagonalizable case. The computation of $\exp(A)$ for non-diagonalizable A uses the Jordan canonical form.

The proof of $\exp(x+y) = \exp(x)\exp(y)$ relied on the fact that scalars commute. If $AB = BA$ then this same proof carries over, and we have

$$\boxed{\exp(A+B) = \exp(A)\exp(B) \text{ for matrices satisfying } AB = BA.}$$

Later we will see that the matrix exponential is the key to solving differential equations. In that context we will consider the series

$$\exp(At) = I + At + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \cdots,$$

where t is a real variable representing time.

For now, we present two example computations. First consider the matrix

$$A = \frac{1}{6}\begin{pmatrix} 5 & 4 \\ 2 & -2 \end{pmatrix}.$$

The characteristic polynomial is

$$\begin{aligned} \det(xI - A) &= (x - 5/6)(x + 2/6) - (-4/6)(-2/6) \\ &= x^2 - (1/2)x - (1/2) \\ &= (x-1)(x+1/2), \end{aligned}$$

hence the eigenvalues are 1 and $-1/2$. Since this 2×2 matrix has 2 distinct eigenvalues, we know that it is diagonalizable. After some computation we find the eigenvectors:

$$A\begin{pmatrix} 4 \\ 1 \end{pmatrix} = 1\begin{pmatrix} 4 \\ 1 \end{pmatrix} \quad \text{and} \quad A\begin{pmatrix} 1 \\ -2 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

Hence we obtain the diagonalization:

$$A = \left(\begin{array}{c|c} 4 & 1 \\ 1 & -2 \end{array}\right)\begin{pmatrix} 1 & 0 \\ 0 & -1/2 \end{pmatrix}\left(\begin{array}{c|c} 4 & 1 \\ 1 & -2 \end{array}\right)^{-1} = \frac{1}{9}\begin{pmatrix} 4 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1/2 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -4 \end{pmatrix}.$$

Finally, we obtain the exponential:

$$\exp(A) = \frac{1}{9}\begin{pmatrix} 4 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} \exp(1) & 0 \\ 0 & \exp(-1/2) \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -4 \end{pmatrix}.$$

The last example is more interesting. Consider the matrix that rotates by 90°:

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

For any real number θ, we will show that

$$\exp(R\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos t \end{pmatrix} = \cos\theta \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin\theta \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

which is the matrix that rotates by θ. This is a matrix version of Euler's formula

$$e^{i\theta} = \cos\theta + i\sin\theta,$$

where the 90° rotation matrix R plays the role of the imaginary unit i. We begin by computing the eigenvalues. The characteristic polynomial is

$$\det(xI - R) = \det\begin{pmatrix} x & 1 \\ -1 & x \end{pmatrix} = x^2 + 1,$$

hence there are two distinct eigenvalues: i and $-i$. It is no surprise that these are complex conjugates, since the complex eigenvalues of real matrices come in conjugate pairs. (See the homework.) With a bit of work, one finds the eigenvectors

$$R\begin{pmatrix} 1 \\ -i \end{pmatrix} = i\begin{pmatrix} 1 \\ -i \end{pmatrix} \quad \text{and} \quad R\begin{pmatrix} 1 \\ i \end{pmatrix} = -i\begin{pmatrix} 1 \\ i \end{pmatrix},$$

and hence the exponential:

$$\exp(R\theta) = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}\begin{pmatrix} \exp(i\theta) & 0 \\ 0 & \exp(-i\theta) \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}^{-1}.$$

Then some simplification using Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$ gives the desired result.

But this makes the result look like a miracle. We can gain more insight by looking at the real and imaginary parts of the complex eigenvalues. Let $\mathbf{x} = (1, -i)$, so that $\mathbf{x} = (1, 0) - i(0, 1)$. Since $R\mathbf{x} = i\mathbf{x}$ we must also have $\exp(R\theta)\mathbf{x} = \exp(i\theta)\mathbf{x}$,[134] and hence

$$\begin{aligned}
\exp(R\theta)\begin{pmatrix} 1 \\ 0 \end{pmatrix} - i\exp(R\theta)\begin{pmatrix} 0 \\ 1 \end{pmatrix} &= \exp(R\theta)\begin{pmatrix} 1 \\ -i \end{pmatrix} \\
&= \exp(i\theta)\begin{pmatrix} 1 \\ -i \end{pmatrix} \\
&= (\cos\theta + i\sin\theta)\begin{pmatrix} 1 \\ -i \end{pmatrix} \qquad \text{Euler's formula} \\
&= \begin{pmatrix} \cos\theta + i\sin\theta \\ \sin\theta - i\cos\theta \end{pmatrix} \\
&= \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} - i\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.
\end{aligned}$$

Since the matrix $\exp(R\theta)$ has real entries, comparing real and imaginary parts gives

$$\exp(R\theta)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \quad \text{and} \quad \exp(R\theta)\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix},$$

which is the desired result.

Remark: In general, complex eigenvalues of real matrices lead to rotation. We will examine this in the next section.

[134] The proof that $A\mathbf{x} = \lambda\mathbf{x}$ implies $f(A)\mathbf{x} = f(\lambda)\mathbf{x}$ for polynomials $f(x)$ carries over to power series.

11.6 Complex Eigenvalues and Eigenvectors of Real Matrices

For any complex number $a + ib \in \mathbb{C}$ with $a, b \in \mathbb{R}$ we will denote its complex conjugate by

$$\overline{\alpha} = \overline{a + ib} = a - ib.$$

Recall that complex conjugation satisfies the following properties:

- $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$,
- $\overline{\alpha\beta} = \overline{\alpha} \cdot \overline{\beta}$,
- $\overline{\alpha^k} = (\overline{\alpha})^k$,
- $\overline{\alpha} = \alpha$ if and only if $\alpha \in \mathbb{R}$.

Given a polynomial $f(x) = b_0 + b_1 x + \cdots + b_n x^n$ with **real coefficients** and a complex number $\alpha \in \mathbb{C}$, it follows from these properties that

$$\begin{aligned}
\overline{f(\alpha)} &= \overline{b_0 + b_1\alpha + \cdots + b_n\alpha^n} \\
&= \overline{b_0} + \overline{b_1} \cdot \overline{\alpha} + \cdots + \overline{b_n} \cdot \overline{\alpha^n} \\
&= b_0 + b_1\overline{\alpha} + \cdots + b_n(\overline{\alpha})^n \\
&= f(\overline{\alpha}).
\end{aligned}$$

In particular, we see that α is a root of $f(x)$ if and only if $\overline{\alpha}$ is a root of $f(x)$. Indeed, if $f(\alpha) = 0$ then

$$f(\overline{\alpha}) = \overline{f(\alpha)} = \overline{0} = 0,$$

and if $f(\overline{\alpha}) = 0$ then

$$f(\alpha) = \overline{\overline{f(\alpha)}} = \overline{f(\overline{\alpha})} = \overline{0} = 0.$$

It follows from this that

the non-real roots of a real polynomial come in conjugate pairs.

And, as an interesting consequence,

every real polynomial of odd degree has as least one real root.

We will apply these observations to eigenvalues of real matrices.

Complex Eigenvalues of a Real Matrix. Let A be an $n \times n$ matrix with **real entries**, so the characteristic polynomial $\chi_A(x) = \det(xI - A)$, has **real coefficients**. According to the previous result, the characteristic polynomial can be factored as

$$\chi_A(x) = (x - \lambda_1) \cdots (x - \lambda_{n-2m})(x - \alpha_1)(x - \overline{\alpha_1}) \cdots (x - \alpha_m)(x - \overline{\alpha_m}),$$

for some real numbers $\lambda_i \in \mathbb{R}$ and non-real complex numbers $\alpha_i \in \mathbb{C}$. If n is even, then the matrix A need not have any real eigenvalues. For example, the real matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

has characteristic polynomial $(x-i)(x+i)$. On the other hand, if n is odd then the number $n-2m$ must be ≥ 1, so that A has at least one real eigenvalue.

Complex Eigenvectors. If a real matrix A has a real eigenvalue λ, then the corresponding eigenvectors are real.[135] Indeed, the space of λ-eigenvectors is the null space $\mathcal{N}(\lambda I - A)$, which can be computed by elimination over $\mathbb{R}$. What about complex eigenvalues? Suppose that a real $n \times n$ matrix A has a complex eigenvalue $\lambda \in \mathbb{C}$, and let $\mathbf{x} \in \mathbb{C}^n$ be a corresponding eigenvector:

$$A\mathbf{x} = \lambda\mathbf{x}.$$

If λ is not real then $\mathbf{x}$ cannot have real entries. Indeed, if $\mathbf{x} \in \mathbb{R}^n$ then since A has real entries we would have $\lambda\mathbf{x} = A\mathbf{x} \in \mathbb{R}^n$ which implies that $\lambda \in \mathbb{R}$. Let us suppose that $\lambda = a + ib$ with $a, b \in \mathbb{R}$ and $b \neq 0$. Then we can write

$$\mathbf{x} = \mathbf{u} + i\mathbf{v}$$

for unique real vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}$. By expanding the equation $A\mathbf{x} = \lambda\mathbf{x}$ we obtain

$$\begin{aligned} A\mathbf{u} + iA\mathbf{v} &= A(\mathbf{u} + i\mathbf{v}) \\ &= A\mathbf{x} \\ &= \lambda\mathbf{x} \\ &= (a+ib)(\mathbf{u} + i\mathbf{v}) \\ &= (a\mathbf{u} - b\mathbf{v}) + i(b\mathbf{u} + a\mathbf{v}). \end{aligned}$$

Since the vectors $A\mathbf{u}$, $A\mathbf{v}$, $a\mathbf{u} - \mathbf{b}\mathbf{v}$ and $b\mathbf{u} + a\mathbf{v}$ have real entries, it follows by comparing real and imaginary parts that

$$\begin{cases} A\mathbf{u} &= a\mathbf{u} &- b\mathbf{v}, \\ A\mathbf{v} &= b\mathbf{u} &+ a\mathbf{v}, \end{cases}$$

which can be expressed as a matrix equation:

$$A \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Next, I claim that the vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly independent over $\mathbb{C}$.[136] To see this, we note that the conjugate vector $\overline{\mathbf{x}} = \mathbf{u} - i\mathbf{v}$ is an eigenvector of A corresponding to the conjugate eigenvalue $\overline{\lambda} = a - ib$. Indeed, since A has real entries, conjugating both sides of the equation $A\mathbf{x} = \lambda\mathbf{x}$ gives[137]

$$A\overline{\mathbf{x}} = \overline{A}\overline{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}.$$

Since $\lambda \neq \overline{\lambda}$, the vectors $\mathbf{x}, \overline{\mathbf{x}} \in \mathbb{C}$ correspond to different eigenvalues, hence they are linearly independent over $\mathbb{C}$. But then since

$$\begin{pmatrix} \mathbf{x} \mid \overline{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, \quad \text{where } \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \text{ is invertible,}$$

[135] Technically, every λ-eigenvector is a scalar multiple of a real vector. You could take a real λ-eigenvector $\mathbf{x}$ and scale it to get a complex λ-eigenvector $i\mathbf{x}$, but why would you want to do that?

[136] In this section we will only discuss linear independence over $\mathbb{C}$, which implies linear independence over $\mathbb{R}$.

[137] The equation $\overline{A\mathbf{x}} = \overline{A}\overline{\mathbf{x}}$ needs to be checked. It follows from the standard properties.

we conclude that $\mathbf{u}$ and $\mathbf{v}$ are linearly independent. In particular, we have

$$A = \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} \mathbf{u} \mid \mathbf{v} \end{pmatrix}^{-1}.$$

Furthermore, for any $\mathbf{y} \in \mathbb{C}^n$ we have

$$\begin{pmatrix} \mathbf{y} \mid \mathbf{x} \mid \overline{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \mid \mathbf{u} \mid \mathbf{v} \end{pmatrix} \begin{pmatrix} 1 & & \\ & 1 & 1 \\ & i & -i \end{pmatrix},$$

which implies that the set $\mathbf{y}, \mathbf{x}, \overline{\mathbf{x}}$ is independent if and only if $\mathbf{y}, \mathbf{u}, \mathbf{v}$ is independent.

Real Diagonalizable Matrices. Finally, we discuss diagonalization of real matrices. Let A be a real $n \times n$ matrix. As we saw above, the characteristic polynomial can be factored as

$$\chi_A(x) = (x - \lambda_1) \cdots (x - \lambda_{n-2m})(x - \alpha_1)(x - \overline{\alpha_1}) \cdots (x - \alpha_m)(x - \overline{\alpha_m}),$$

where $\lambda_1, \ldots, \lambda_{n-2m}$ are real and $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$ are non-real.

Suppose that A is diagonalizable over $\mathbb{C}$. This means that we can find nonzero vectors $\mathbf{y}_1, \ldots, \mathbf{y}_{n-2m} \in \mathbb{C}^n$ and $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{C}^n$ such that $A\mathbf{y}_i = \lambda_i \mathbf{y}_i$ and $A\mathbf{x}_i = \alpha_i \mathbf{x}_i$, hence also $A\overline{\mathbf{x}_i} = \overline{\alpha_i \mathbf{x}_i}$, and such that

$$\mathbf{y}_1, \ldots, \mathbf{y}_{n-2m}, \mathbf{x}_1, \overline{\mathbf{x}_1}, \ldots, \mathbf{x}_m, \overline{\mathbf{x}_m}$$

is a basis for $\mathbb{C}^n$. If X is the $n \times n$ (invertible) matrix with these column vectors, then we have

$$A = X \begin{pmatrix} \lambda_1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda_{n-2m} & & & & & & \\ & & & \alpha_1 & & & & & \\ & & & & \overline{\alpha_1} & & & & \\ & & & & & \ddots & & & \\ & & & & & & \alpha_m & & \\ & & & & & & & \overline{\alpha_m} \end{pmatrix} X^{-1}.$$

Now we will eliminate the complex numbers from this factorization. Since the eigenvalues $\lambda_1, \ldots, \lambda_{n-2m}$ are real, we can choose the eignevectors $\mathbf{y}_1, \ldots, \mathbf{y}_{n-2m}$ to be real. Next we write $\mathbf{x}_i = \mathbf{u}_i + i\mathbf{v}_i$ for real vectors $\mathbf{u}_i, \mathbf{v}_i$ with $\mathbf{v}_i \neq \mathbf{0}$. From the previous remarks we see that

$$\mathbf{y}_1, \ldots, \mathbf{y}_{n-2m}, \mathbf{u}_1, \mathbf{v}_1, \ldots, \mathbf{u}_m, \mathbf{v}_m$$

is a basis for $\mathbb{C}^n$ consisting of real vectors. Furthermore, if Y is the (invertible) matrix with these columns, then we have

$$A = Y \begin{pmatrix} \lambda_1 & & & & & & & \\ & \ddots & & & & & & \\ & & \lambda_{n-2m} & & & & & \\ & & & a_1 & b_1 & & & \\ & & & -b_1 & a_1 & & & \\ & & & & & \ddots & & \\ & & & & & & a_m & b_m \\ & & & & & & -b_m & a_m \end{pmatrix} Y^{-1}.$$

This is not quite a "diagonalization", but it has the virtue using only real numbers.

11.7 Normal Operators

In this chapter we have studied the evaluation of polynomials (also power series and Laurent polynomials) at matrices. This discussion has left out one important operation; namely, the transpose and conjugate transpose. In this final section we consider the relationship between eigenvalues and (conjugate) transposition.

The main role is played by *normal matrices*. We say that a matrix A is normal when it commutes with its (conjugate) transpose:

$$A^*A = AA^*.$$

These matrices are extremely common in applications and include the following four families:

- Real symmetric matrices $A^T = A$.
- Complex Hermitian matrices $A^* = A$.
- Real orthogonal matrices $A^T = A^{-1}$.
- Complex unitary matrices $A^* = A^{-1}$.

Of course, these families could be dealt with separately. The reason to combine them under the concept of normal matrices is because of the following fundamental theorem, which we will prove in the next chapter.

The Spectral Theorem. Let A be a square matrix over $\mathbb{R}$ or $\mathbb{C}$. Then

$$A^*A = AA^* \quad \Longleftrightarrow \quad A \text{ has an orthonormal basis of eigenvectors.}$$

Actually, some people think it undignified to call this the Spectral Theorem. They say that the true Spectral Theorem applies to operators on infinite dimensional Hilbert spaces. Recall,

if V is a real or complex Hilbert space and if $A : V \to V$ is a bounded[138] linear operator then there exists a unique bounded linear operator $A^* : V \to V$ satisfying

$$\langle A\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{x}, A^*\mathbf{y}\rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in V.$$

As with many results in functional analysis, the proof is 80% algebra and 20% analysis, which is mostly plausible from geometric intuition.

Anyway, it is convenient to state and prove the results of this section in a language that applies also to Hilbert spaces. Our first theorem was proved by Cauchy in 1829, as part of his extension of the Principal Axes Theorem to higher dimensions. Cauchy's original proof was quite complicated, but today's proof is a one-liner.[139]

Cauchy's Reality Theorem. A real symmetric matrix has real eigenvalues.

Actually, we will prove the following more general statement, since it has the same proof.

Theorem. A self-adjoint operator on a complex inner product space has real eigenvalues.

Proof. Let V be a real or complex inner product space and let $A : V \to V$ be an operator satisfying $A^* = A$. If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ and nonzero vector $\mathbf{x} \neq \mathbf{0}$ then we have

$$\begin{aligned}
\lambda\|\mathbf{x}\|^2 &= \lambda\langle \mathbf{x}, \mathbf{x}\rangle \\
&= \langle \mathbf{x}, \lambda\mathbf{x}\rangle \\
&= \langle \mathbf{x}, A\mathbf{x}\rangle \\
&= \langle A^*\mathbf{x}, \mathbf{x}\rangle \\
&= \langle A\mathbf{x}, \mathbf{x}\rangle && A^* = A \\
&= \langle \lambda\mathbf{x}, \mathbf{x}\rangle \\
&= \overline{\lambda}\langle \mathbf{x}, \mathbf{x}\rangle \\
&= \overline{\lambda}\|\mathbf{x}\|^2.
\end{aligned}$$

Since $\|\mathbf{x}\| \neq 0$ this implies that $\overline{\lambda} = \lambda$, and hence λ is real. □

The next theorem has a similar proof.

Theorem. Unitary (and real orthogonal) operators have eigenvalues of length 1. That is, they have eigenvalues of the form $e^{i\theta}$.

Proof. Let V be a real or complex inner product space and let $A : V \to V$ be an operator satisfying $A^*A = I$. If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ and nonzero vector $\mathbf{x} \neq \mathbf{0}$ then we have

$$\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x}\rangle$$

[138] This means that A sends the unit ball to a bounded set. It is equivalent to A being continuous.

[139] "Dazzled by the brilliance of the new theory of determinants, mathematicians overlooked simple inner product considerations", Hawkins, *The Mathematics of Frobenius in Context*, page 98.

$$
\begin{aligned}
&= \langle \mathbf{x}, I\mathbf{x} \rangle \\
&= \langle \mathbf{x}, A^*A\mathbf{x} \rangle && A^*A = I \\
&= \langle A\mathbf{x}, A\mathbf{x} \rangle \\
&= \langle \lambda\mathbf{x}, \lambda\mathbf{x} \rangle \\
&= \lambda\langle \lambda\mathbf{x}, \mathbf{x} \rangle \\
&= \overline{\lambda}\lambda\langle \mathbf{x}, \mathbf{x} \rangle \\
&= |\lambda|^2 \|\mathbf{x}\|^2.
\end{aligned}
$$

Since $\|\mathbf{x}\| \neq 0$ this implies that $|\lambda| = 1$. □

Though it doesn't involve eigenvalues, we should probably include the following result.

Theorem. Unitary operators preserve lengths and angles.

Proof. This follows from the fact that unitary operators preserve the inner product. If $A^*A = I$ then for all vectors $\mathbf{x}, \mathbf{y}$ we have

$$\langle A\mathbf{x}, A\mathbf{y} \rangle = \langle \mathbf{x}, A^*A\mathbf{y} \rangle = \langle \mathbf{x}, I\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle.$$

□

We have seen that the important families of normal matrices have quite restricted eigenvalues:

- Real symmetric matrices $A^T = A$ and complex Hermitian matrices $A^* = A$ matrices have real eigenvalues.
- Real orthogonal matrices $A^TA = I$ and complex unitary matrices $A^*A = I$ have eigenvalues of the form $e^{i\theta}$.

On the other hand, a general normal matrix can have any eigenvalues you want. Indeed, consider any complex numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ and let Λ be the diagonal matrix with these numbers on the diagonal. Then for any unitary matrix $U^*U = I$, the matrix

$$A = U\Lambda U^*$$

is normal and has eigenvalues $\lambda_1, \ldots, \lambda_n$.

What about eigen**vectors**? In this case, the key property is shared by all normal operators. This result is a precursor to the Spectral Theorem.

Theorem (Normal with Distinct Eigenvalues ⇒ Orthogonal Eigenvectors). Let $A^*A = AA^*$ be a normal operator on an inner product space. Then

$$A\mathbf{x} = \lambda\mathbf{x} \text{ and } A\mathbf{y} = \mu\mathbf{y} \text{ with } \lambda \neq \mu \quad \Longrightarrow \quad \langle \mathbf{x}, \mathbf{y} \rangle = 0.$$

We will work up to the proof by a series of lemmas.

Lemma 1. If $A^*A = AA^*$ then $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle A^*\mathbf{x}, A^*\mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y}$.

Proof. We have $\langle A\mathbf{x}, A\mathbf{y}\rangle = \langle \mathbf{x}, A^*A\mathbf{y}\rangle = \langle \mathbf{x}, AA^*\mathbf{y}\rangle = \langle A^*\mathbf{x}, A^*\mathbf{y}\rangle$. □

Lemma 2. If $A^*A = AA^*$ then we have $A\mathbf{x} = 0$ if and only if $A^*\mathbf{x} = \mathbf{0}$.

Proof. Putting $\mathbf{y} = \mathbf{x}$ in Lemma 1 gives $\|A\mathbf{x}\|^2 = \langle A\mathbf{x}, A\mathbf{x}\rangle = \langle A^*\mathbf{x}, A^*\mathbf{x}\rangle = \|A^*\mathbf{x}\|^2$. Hence

$$A\mathbf{x} = \mathbf{0} \quad \Longleftrightarrow \quad \|A\mathbf{x}\| = 0 \quad \Longleftrightarrow \quad \|A^*\mathbf{x}\| = 0 \quad \Longleftrightarrow \quad A^*\mathbf{x} = \mathbf{0}.$$

□

Lemma 3. Let $A^*A = AA^*$. Then for any vector $\mathbf{x}$ and scalar λ we have

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Longleftrightarrow \quad A^*\mathbf{x} = \overline{\lambda}\mathbf{x}.$$

Proof. Consider the matrix $B = \lambda I - A$, with $B^* = \overline{\lambda}I - A^*$. Then B is normal:

$$\begin{aligned} B^*B &= (\overline{\lambda}I - A^*)(\lambda I - A) \\ &= \overline{\lambda}\lambda I - \overline{\lambda}A - \lambda A^* + A^*A \\ &= \overline{\lambda}\lambda I - \overline{\lambda}A - \lambda A^* + AA^* && A^*A = AA^* \\ &= (\lambda I - A)(\overline{\lambda}I - A^*) \\ &= BB^*. \end{aligned}$$

Hence applying Lemma 2 gives

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Longleftrightarrow \quad B\mathbf{x} = \mathbf{0} \quad \Longleftrightarrow \quad B^*\mathbf{x} = \mathbf{0} \quad \Longleftrightarrow \quad A^*\mathbf{x} = \overline{\lambda}\mathbf{x}.$$

Proof of the Theorem. Let $A^*A = AA^*$ and suppose that $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \mu\mathbf{y}$ with $\lambda \neq \mu$. Then from Lemma 3 we have $A^*\mathbf{x} = \overline{\lambda}\mathbf{x}$, hence

$$\begin{aligned} \lambda\langle \mathbf{x}, \mathbf{y}\rangle &= \langle \overline{\lambda}\mathbf{x}, \mathbf{y}\rangle \\ &= \langle A^*\mathbf{x}, \mathbf{y}\rangle && \text{Lemma 3} \\ &= \langle \mathbf{x}, A\mathbf{y}\rangle \\ &= \langle \mathbf{x}, \mu\mathbf{y}\rangle \\ &= \mu\langle \mathbf{x}, \mathbf{y}\rangle. \end{aligned}$$

Finally, since $(\lambda - \mu)\langle \mathbf{x}, \mathbf{y}\rangle = 0$ and $\lambda \neq \mu$ we have $\langle \mathbf{x}, \mathbf{y}\rangle = 0$. □

In particular, this shows that an $n \times n$ normal matrix $A^*A = AA^*$ with n distinct eigenvalues has an orthogonal basis of eigenvectors. The Spectral Theorem says that this is still true even when A has repeated eigenvalues. The hard part is to show that there are **enough** eigenvectors to fill up the whole space. See the next chapter.

NOTE TO SELF: Maybe I should discuss the "polarization" $A = (A + A^*)/2 + (A - A^*)/2$, i.e., the "real and imaginary parts" of A. This plays a role later in the proof of the Spectral Theorem.

12 Factorization Theorems

12.1 Gram-Schmidt and QR Factorization

Most of the theorems in this chapter deal with orthonormal bases. In this section we lay the groundwork by showing how any basis can be converted into an orthonormal basis. The procedure is quite general. First we consider an infinite dimensional inner product space V over $\mathbb{R}$ or $\mathbb{C}$. Afterwards we will consider finite dimensional spaces and matrices.

Given any linearly independent set $\mathbf{a}_1, \mathbf{a}_2, \ldots \in V$ the Gram-Schmidt procedure produces linearly independent vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots \in V$ with the following properties:[140]

- $\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0$ for $i \neq j$,
- $\text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \text{Span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$.

The definition is recursive:

- First set $\mathbf{b}_1 := \mathbf{a}_1$.
- Then for any $k \geq 1$ set $\mathbf{b}_{k+1} := \mathbf{a}_{k+1} - \text{Proj}_k(\mathbf{a}_{k+1})$, where $\text{Proj}_k : V \to V$ is the orthogonal projection onto the subspace spanned by $\mathbf{b}_1, \ldots, \mathbf{b}_k$. To be precise, we set

$$\mathbf{b}_{k+1} := \mathbf{a}_{k+1} - \frac{\langle \mathbf{a}_{k+1}, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \mathbf{b}_1 - \cdots - \frac{\langle \mathbf{a}_{k+1}, \mathbf{b}_k \rangle}{\langle \mathbf{b}_k, \mathbf{b}_k \rangle} \mathbf{b}_k.$$

You will prove on the homework that this procedure has the desired properties. Afterwards, we can easily turn $\mathbf{b}_1, \mathbf{b}_2, \ldots \in V$ into an ortho**normal** set by dividing each $\mathbf{b}_k$ by its length.

Before applying this to matrices, we give one application to infinite dimensional function spaces. Consider the real Hilbert space $L^2[-1, 1]$ with inner product

$$\langle f(x), g(x) \rangle = \int_{-1}^{1} f(x) g(x)\, dx.$$

And consider the "obvious" basis $1, x, x^2, \ldots \in L^2[-1, 1]$.[141] Note that these functions are not orthogonal. For example,

$$\langle 1, x^2 \rangle = \int_{-1}^{1} x^2\, dx = \frac{1}{3} x^3 \Big|_{-1}^{1} = \frac{1}{3}(1)^3 - \frac{1}{3}(-1)^3 = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \neq 0.$$

Applying the Gram-Schmidt procedure to the non-orthogonal basis $1, x, x^2, \ldots$ produces the orthogonal basis of *Legendre polynomials*: $P_0(x), P_1(x), P_2(x), \ldots$. These are used in physics in the study of spherically symmetric potentials. For example, they determine the "shapes"

[140] If $\mathbf{a}_1, \mathbf{a}_2, \ldots$ is a Hilbert space basis, with appropriate convergence properties, then the vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots$ will also be a Hilbert space basis, though we won't prove this.

[141] It can be shown that this is, indeed, a Hilbert space basis.

of electron orbitals. To be precise, we first define the *associated Legendre function* for integers $\ell, m \in \mathbb{Z}$ with $0 \le m \le \ell$:

$$P_\ell^m(x) = (1-x^2)^{m/2} \cdot \frac{d^m}{dx^m} P_\ell(x).$$

Then the radial equation for the shape of the (ℓ, m)-orbital is[142]

$$\rho = (\text{constant}) \cdot |P_\ell^m(\cos\theta)|.$$

Now we turn to matrices. The matrix form of Gram-Schmidt is called QR factorization. Given an invertible $n \times n$ matrix A, we will produce a unitary matrix $Q^*Q = I$ and an upper-triangular matrix R such that $A = QR$. If A has real entries then Q and R will have real entries. In this case $Q^TQ = I$ is real orthogonal.

Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be a basis for $\mathbb{C}^n$. Then the Gram-Schmidt basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$ satisfies

$$\begin{aligned}
\mathbf{a}_1 &= \mathbf{b}_1, \\
\mathbf{a}_2 &= \mathbf{b}_2 + \frac{\langle \mathbf{a}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \mathbf{b}_1, \\
&\vdots \\
\mathbf{a}_n &= \mathbf{b}_n + \frac{\langle \mathbf{a}_n, \mathbf{b}_{n-1} \rangle}{\langle \mathbf{b}_{n-1}, \mathbf{b}_{n-1} \rangle} \mathbf{b}_{n-1} + \cdots + \frac{\langle \mathbf{a}_n, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \mathbf{b}_1,
\end{aligned}$$

which can be expressed as a matrix equation:

$$A = BU$$

$$\begin{pmatrix} \mathbf{a}_1 \,\middle|\, \cdots \,\middle|\, \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \,\middle|\, \cdots \,\middle|\, \mathbf{b}_n \end{pmatrix} \begin{pmatrix} 1 & \frac{\langle \mathbf{a}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} & \cdots & \cdots & \frac{\langle \mathbf{a}_n, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \\ & 1 & & & \vdots \\ & & \ddots & & \vdots \\ & & & 1 & \frac{\langle \mathbf{a}_n, \mathbf{b}_{n-1} \rangle}{\langle \mathbf{b}_{n-1}, \mathbf{b}_{n-1} \rangle} \\ & & & & 1 \end{pmatrix}$$

By construction, the columns of B are orthogonal. We can make them ortho**normal** by scaling the kth column $\mathbf{b}_k$ by $1/\|\mathbf{b}_k\|$. If S is the diagonal matrix with entries $1/\|\mathbf{b}_k\|$, then the matrix $Q = BS$ has orthonormal columns $\mathbf{q}_k = \mathbf{b}_k/\|\mathbf{b}_k\|$, hence $Q^*Q = I$. To convert $A = BU$ into $A = QR$ we define $R = S^{-1}U$ so that

$$\begin{aligned}
A &= BU \\
&= B(SS^{-1})U \\
&= (BS)(S^{-1}U) \\
&= QR.
\end{aligned}$$

[142] This example is just for fun. See Griffiths, *Introduction to Quantum Mechanics*, Equation 4.32.

It turns out that the matrix $R = S^{-1}U$ has a nice form. To see this, we first observe that

$$\langle \mathbf{a}_k, \mathbf{b}_k \rangle = \langle \mathbf{b}_k + \text{stuff orthogonal to } \mathbf{b}_k, \mathbf{b}_k \rangle = \langle \mathbf{b}_k, \mathbf{b}_k \rangle = \|\mathbf{b}_k\|^2,$$

which implies that

$$\langle \mathbf{a}_k, \mathbf{q}_k \rangle = \left\langle \mathbf{a}_k, \frac{\mathbf{b}_k}{\|\mathbf{b}_k\|} \right\rangle = \frac{1}{\|\mathbf{b}_k\|} \langle \mathbf{a}_k, \mathbf{b}_k \rangle = \frac{1}{\|\mathbf{b}_k\|} \|\mathbf{b}_k\|^2 = \|\mathbf{b}_i\|.$$

Furthermore, for any $1 \leq i < k$ we have

$$\|\mathbf{b}_i\| \cdot \frac{\langle \mathbf{a}_k, \mathbf{b}_i \rangle}{\langle \mathbf{b}_i, \mathbf{b}_i \rangle} = \|\mathbf{b}_i\| \cdot \frac{\langle \mathbf{a}_k, \mathbf{b}_i \rangle}{\|\mathbf{b}_i\|^2} = \frac{1}{\|\mathbf{b}_i\|} \cdot \langle \mathbf{a}_k, \mathbf{b}_i \rangle = \left\langle \mathbf{a}_k, \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|} \right\rangle = \langle \mathbf{a}_k, \mathbf{q}_i \rangle.$$

Putting these together gives

$$\begin{aligned} R &= S^{-1}U \\ &= \begin{pmatrix} \|\mathbf{b}_1\| & & \\ & \ddots & \\ & & \|\mathbf{b}_n\| \end{pmatrix} \begin{pmatrix} 1 & \frac{\langle \mathbf{a}_2, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} & \cdots & \cdots & \frac{\langle \mathbf{a}_n, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle} \\ & 1 & & & \vdots \\ & & \ddots & & \vdots \\ & & & 1 & \frac{\langle \mathbf{a}_n, \mathbf{b}_{n-1} \rangle}{\langle \mathbf{b}_{n-1}, \mathbf{b}_{n-1} \rangle} \\ & & & & 1 \end{pmatrix} \\ &= \begin{pmatrix} \langle \mathbf{a}_1, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{q}_1 \rangle \\ & \ddots & \vdots \\ & & \langle \mathbf{a}_n, \mathbf{q}_n \rangle \end{pmatrix}. \end{aligned}$$

In summary, for any $n \times n$ invertible matrix A with columns $\mathbf{a}_i$, we can find an $n \times n$ unitary matrix $Q^*Q = I$ with columns $\mathbf{q}_i$, such that

$$\begin{pmatrix} \mathbf{a}_1 \mid \cdots \mid \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{q}_1 \mid \cdots \mid \mathbf{q}_n \end{pmatrix} \begin{pmatrix} \langle \mathbf{a}_1, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{a}_n, \mathbf{q}_1 \rangle \\ & \ddots & \vdots \\ & & \langle \mathbf{a}_n, \mathbf{q}_n \rangle \end{pmatrix}$$

$$A = QR.$$

And if A is real then we can choose Q and R with real entries.

Due to rounding errors, the matrix Q computed from Gram-Schmidt is only approximately orthogonal. It is worth mentioning another method, due to Householder, that gives an exactly orthogonal matrix. This method also has an interesting theoretical consequence:

Any real orthogonal matrix $A^TA = I$ is a composition of reflections.

This method uses the *Householder reflection matrices*:

$$H_{\mathbf{v}} = I - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2} \quad \text{for } \mathbf{v} \in \mathbb{R}^n.$$

Recall that

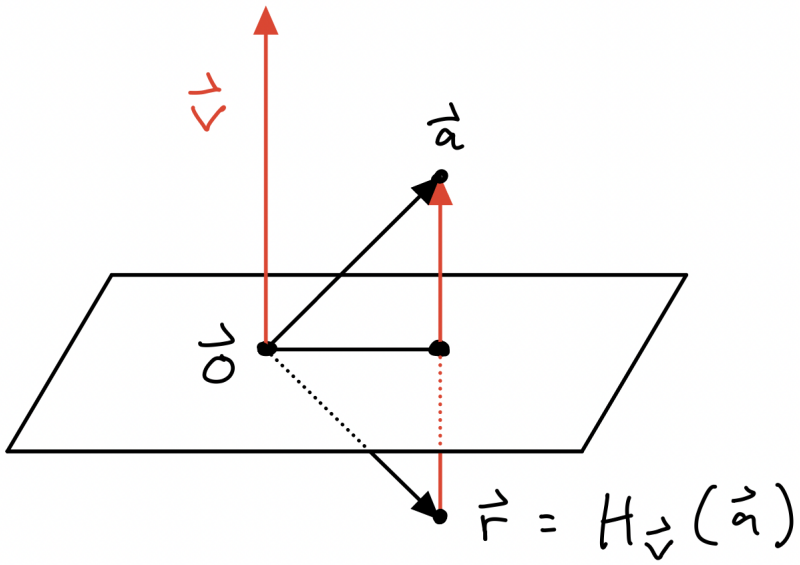
$$P_{\mathbf{v}} = \mathbf{v}(\mathbf{v}\mathbf{v}^T)^{-1}\mathbf{v}^T = \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} = \frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2}$$

is the matrix that projects orthogonally onto the line spanned by $\mathbf{v}$, hence $I - P_{\mathbf{v}}$ is the matrix that projects onto the orthogonal hyperplane $\mathbf{v}^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T\mathbf{v} = 0\}$. From the remarks in the previous chapter, this implies that $2P_{\mathbf{v}} - I$ is the matrix that reflects across the line $\mathbf{v}$ and $2(I - P\mathbf{v}) - I = I - 2P_{\mathbf{v}} = H_{\mathbf{v}}$ is the matrix that reflects across the hyperplane $\mathbf{v}^\perp$. Since $P_{\mathbf{v}}^2 = P_{\mathbf{v}}$ and $P_{\mathbf{v}}^T = P_{\mathbf{v}}$, we find that

$$H_{\mathbf{v}}^{-1} = H_{\mathbf{v}} \quad \text{and } H_{\mathbf{v}}^T = H_{\mathbf{v}}.$$

In particular, $H_{\mathbf{v}}$ is an orthogonal matrix.

The key trick of the Householder algorithm is that $H_{\mathbf{v}}\mathbf{a} = \mathbf{r}$ for any $\mathbf{a}, \mathbf{r}$ satisfying $\|\mathbf{a}\| = \|\mathbf{r}\|$ and $\mathbf{a} - \mathbf{r} = \mathbf{v}$. Picture:



Here is the algorithm.

Householder QR. We are given an invertible matrix A with first column $\mathbf{a} \in \mathbb{R}^n$.

- Let $\mathbf{r} := (\|\mathbf{a}\|, 0, \ldots, 0)$, $\mathbf{v} := \mathbf{a} - \mathbf{r}$ and $H_1 := H_{\mathbf{v}}$, so that $H_1\mathbf{a} = \mathbf{r}$. Then we have

$$H_1 A = \left(\begin{array}{c|c|c|c} H_1\mathbf{a} & * & \cdots & * \end{array} \right) = \left(\begin{array}{c|c|c|c} \mathbf{r} & * & \cdots & * \end{array} \right) = \left(\begin{array}{c|ccc} \|\mathbf{a}\| & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array} \right),$$

  for some matrix A' of size $(n-1) \times (n-1)$.

- Let $\mathbf{a}' \in \mathbb{R}^{n-1}$ be the first column of A', let $\mathbf{r}' = (\|\mathbf{a}'\|, 0, \ldots, 0) \in \mathbb{R}^{n-1}$ and let $\mathbf{v}' = \mathbf{a}' - \mathbf{r}'$, so that $H_{\mathbf{v}'}\mathbf{a}' = \mathbf{r}'$. Then the matrix

$$H_2 := \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & H_{\mathbf{v}'} & \\ 0 & & & \end{array}\right),$$

satisfies

$$\begin{aligned} H_2H_1A &= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & H_{\mathbf{v}'} & \\ 0 & & & \end{array}\right)\left(\begin{array}{c|ccc} \|\mathbf{a}\| & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right) \\ &= \left(\begin{array}{c|ccc} \|\mathbf{a}\| & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & H_2A' & \\ 0 & & & \end{array}\right) \\ &= \left(\begin{array}{c|c|ccc} \|\mathbf{a}\| & * & * & \cdots & * \\ \hline 0 & \|\mathbf{a}'\| & * & \cdots & * \\ \hline 0 & 0 & & & \\ \vdots & \vdots & & A'' & \\ 0 & 0 & & & \end{array}\right), \end{aligned}$$

for some matrix A'' of size $(n-2) \times (n-2)$. We observe that the matrix H_2 is itself a Householder reflection matrix. To see this, let $\mathbf{w} = \left(\begin{array}{c|c} 0 & (\mathbf{v}')^T \end{array}\right)^T$, so that $\|\mathbf{w}\| = \|\mathbf{v}'\|$. Then we have

$$\begin{aligned} H_{\mathbf{w}} &= I - 2\frac{\mathbf{w}\mathbf{w}^T}{\|\mathbf{w}\|} \\ &= I - \frac{2}{\|\mathbf{v}'\|}\left(\begin{array}{c} 0 \\ \hline \mathbf{v}' \end{array}\right)\left(\begin{array}{c|c} 0 & (\mathbf{v}')^T \end{array}\right) \\ &= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & I_{n-1} & \\ 0 & & & \end{array}\right) - \left(\begin{array}{c|ccc} 0 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & -2\frac{\mathbf{v}'(\mathbf{v}')^T}{\|\mathbf{v}'\|} & \\ 0 & & & \end{array}\right) \end{aligned}$$

$$= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & I_{n-1} - 2\frac{\mathbf{v}'(\mathbf{v}')^T}{\|\mathbf{v}'\|} & \\ 0 & & & \end{array}\right)$$

$$= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & H_{\mathbf{v}'} & \\ 0 & & & \end{array}\right)$$

$$= H_2.$$

- Continuing in this way for $n-1$ steps gives an upper triangular matrix:

$$H_{n-1}\cdots H_2 H_1 A = \begin{pmatrix} \|\mathbf{a}\| & * & \cdots & \cdots & * \\ & \|\mathbf{a}'\| & & & \vdots \\ & & \ddots & & \vdots \\ & & & \|\mathbf{a}^{(n-1)}\| & * \\ & & & & b \end{pmatrix} = R,$$

  where each H_i is a Householder matrix $H_{\mathbf{v}_i}$ for some vector $\mathbf{v}_i \in \mathbb{R}$. Note that the diagonal entries of R are nonzero since we have assumed that A is invertible. The real number b can be positive or negative.

- Finally, since each Householder reflection is equal to its own inverse, we obtain

$$\begin{aligned} H_{n-1}\cdots H_2 H_1 A &= R \\ A &= H_1 H_2 \cdots H_{n-1} R \\ A &= QR. \end{aligned}$$

As a consequence, we will prove that every real orthogonal matrix is a composition of reflections. Suppose that $A^T A = I$ and consider the Householder factorization

$$H_{n-1}\cdots H_2 H_1 A = R.$$

Now each matrix on the left is orthogonal. Since a product of orthogonal matrices is orthogonal, we conclude that R is also orthogonal. In particular, the rows of R are orthonormal. Since R is also upper-triangular, this implies that R is diagonal:

$$R = \begin{pmatrix} \|\mathbf{a}\| & & & \\ & \|\mathbf{a}'\| & & \\ & & \ddots & \\ & & & \|\mathbf{a}^{(n-1)}\| \\ & & & & b \end{pmatrix}$$

Finally, since each row of R has length 1, we conclude that

$$R = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & \pm 1 \end{pmatrix}$$

If the last entry is $+1$ then $R = I$ and we obtain

$$A = H_1 H_2 \cdots H_{n-1} I = H_1 H_2 \cdots H_{n-1},$$

which shows that A is a product of $n-1$ reflection matrices. If the last entry of R is -1, then $H_n := R$ equals the Householder matrix $H_{\mathbf{e}_n}$, where $\mathbf{e}_n = (0, \ldots, 0, 1)$. In this case we see that A is a product of n reflection matrices:

$$A = H_1 H_2 \cdots H_{n-1} R = H_1 H_2 \cdots H_{n-1} H_n.$$

Remark: There is some restriction on n. Since each reflection matrix has determinant -1, a product of n reflection matrices has determinant $(-1)^n$. Hence an orthogonal matrix A satisfying $\det(A) = +1$ can only be expressed as an even product of reflections and an orthogonal matrix satisfying $\det(A) = -1$ can only be expressed as an odd product of reflections.

12.2 Schur Triangularization

Given a square matrix A, we always want to find a simpler matrix B that is *similar* to A. That is, we want to find a simpler matrix B and an invertible matrix X such that $A = XBX^{-1}$. Then for any polynomial function $f(x)$ (more generally, for power series or Laurent polynomials) we can compute

$$f(A) = X \cdot f(B) \cdot X^{-1}.$$

The nicest possible situation is when B is diagonal and X is orthogonal or unitary: $X^{-1} = X^T$ or $X^{-1} = X^*$. This is the subject of the Spectral Theorem in the next section. But diagonalization is not always possible. There are three different theorems for dealing with non-diagonalizable matrices:

- Schur triangularization.
- Jordan normal form.
- Singular value decomposition.

We will deal with all three of these in this chapter. We begin with Schur triangularization.

We say that a matrix is upper-triangular if all entries below the main diagonal are zero:[143]

$$T = \begin{pmatrix} t_{11} & * & \cdots & * \\ & t_{22} & & \vdots \\ & & \ddots & * \\ & & & t_{nn} \end{pmatrix}$$

These matrices have some nice properties:

- The eigenvalues of T are the diagonal entries. Indeed, the characteristic polynomial is
$$\chi_T(x) = (x - t_{11})(x - t_{22}) \cdots (x - t_{nn}).$$
- Products and sums of upper-triangular matrices behave as products and sums for the diagonal entries. Thus for any polynomial $f(x)$ we have
$$f(T) = \begin{pmatrix} f(t_{11}) & * & \cdots & * \\ & f(t_{22}) & & \vdots \\ & & \ddots & * \\ & & & f(t_{nn}) \end{pmatrix}$$
Unfortunately, the entries above the diagonal are messy.
- If T is invertible then T^{-1} is also upper-triangular, and the previous formula also applies for Laurent polynomials $f(x)$.
- If the largest eigenvalue satisfies $|\lambda| < 1$ then one can show that $T^k \to O$ as $k \to \infty$, though the proof is a bit tricky.[144]

Here is our main theorem.

Theorem (Schur Triangularization). For **any** square matrix A over $\mathbb{R}$ or $\mathbb{C}$, there exists an upper-triangular matrix T and a unitary matrix $U^{-1} = U^*$ such that

$$A = UTU^{-1}$$
$$A = UTU^*$$
$$\left(\begin{array}{c|c|c|c} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{array} \right) = \left(\begin{array}{c|c|c|c} \mathbf{u}_1 & \ldots & \mathbf{u}_n \end{array} \right) \begin{pmatrix} t_{11} & * & \cdots & * \\ & t_{22} & & \vdots \\ & & \ddots & * \\ & & & t_{nn} \end{pmatrix} \left(\begin{array}{c} \mathbf{u}_1^* \\ \hline \vdots \\ \hline \mathbf{u}_n^* \end{array} \right).$$

Even if A is real, the matrices U and T will generally have complex entries. However, if A is a real matrix with real eigenvalues then we can choose U and T to be real.

[143] Similarly, a lower-triangular matrix has zeros above the main diagonal.
[144] This is easy for diagonalizable matrices.

Proof. We use induction on the size of A. First we note that the theorem is trivially true for 1×1 matrices, i.e., for scalars: $(a) = (1)(a)(1)$. Now let A have shape $n \times n$ for some $n \geq 2$. We have seen that every real matrix has a (possibly complex) eigenvalue. Let $t_{11} \in \mathbb{C}$ be an eigenvalue of A and let $\mathbf{u}_1$ be a corresponding eigenvector of length 1.[145] Now let U_1 be any unitary matrix with first column $\mathbf{u}_1$:

$$U_1 = \left(\begin{array}{c|c|c|c} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{array} \right).$$

To find such a matrix, we first complete $\mathbf{u}_1$ to a basis, $\mathbf{u}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, then apply Gram-Schmidt to convert this into an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.[146] Since these vectors satisfy $\mathbf{u}_1^* \mathbf{u}_1 = 1$ and $\mathbf{u}_i^* \mathbf{u}_1 = 0$ for $i \geq 2$, we observe that

$$\begin{aligned}
U_1^* A U_1 &= \left(\begin{array}{c} \mathbf{u}_1^* \\ \hline \vdots \\ \hline \mathbf{u}_n^* \end{array} \right) \left(\begin{array}{c|c|c|c} A\mathbf{u}_1 & A\mathbf{u}_2 & \cdots & A\mathbf{u}_n \end{array} \right) \\
&= \left(\begin{array}{c} \mathbf{u}_1^* \\ \hline \vdots \\ \hline \mathbf{u}_n^* \end{array} \right) \left(\begin{array}{c|c|c|c} t_{11}\mathbf{u}_1 & A\mathbf{u}_2 & \cdots & A\mathbf{u}_n \end{array} \right) \\
&= \left(\begin{array}{c|c|c|c} t_{11}\mathbf{u}_1^*\mathbf{u}_1 & * & \cdots & * \\ t_{11}\mathbf{u}_2^*\mathbf{u}_1 & & & \\ \vdots & & & \\ t_{11}\mathbf{u}_n^*\mathbf{u}_1 & & & \end{array} \right) \\
&= \left(\begin{array}{c|ccc} t_{11} & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{array} \right),
\end{aligned}$$

for some matrix A_2 of shape $(n-1) \times (n-1)$. By induction there exists an $(n-1) \times (n-1)$ unitary matrix U_2 such that $T_2 := U_2^* A_2 U_2$ is upper-triangular. Now define the matrix

$$U := U_1 \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2 & \\ 0 & & & \end{array} \right).$$

[145] Just take any eigenvector and scale it. If t_{11} is real and if A has real entries then we can choose $\mathbf{u}_1$ to have real entries, in which case we can choose U_1 to have real entries.

[146] Any set of independent vectors can be completed to a basis using Steinitz exchange.

We observe that this matrix is unitary:

$$\begin{aligned}
U^*U &= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2^* & \\ 0 & & & \end{array}\right) U_1^*U_1 \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2 & \\ 0 & & & \end{array}\right) \\
&= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2^* & \\ 0 & & & \end{array}\right) \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2 & \\ 0 & & & \end{array}\right) \\
&= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2^*U_2 & \\ 0 & & & \end{array}\right) \\
&= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & I_{n-1} & \\ 0 & & & \end{array}\right) \\
&= I_n.
\end{aligned}$$

And we observe that the matrix $T := U^*AU$ is upper triangular, as desired:

$$\begin{aligned}
T &= U^*AU \\
&= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2^* & \\ 0 & & & \end{array}\right) U_1^*AU_1 \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2 & \\ 0 & & & \end{array}\right) \\
&= \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2^* & \\ 0 & & & \end{array}\right) \left(\begin{array}{c|ccc} t_{11} & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{array}\right) \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & U_2 & \\ 0 & & & \end{array}\right) \\
&= \left(\begin{array}{c|ccc} t_{11} & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & U_2^*A_2U_2 & \\ 0 & & & \end{array}\right)
\end{aligned}$$

$$= \left(\begin{array}{c|ccc} t_{11} & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & T_2 & \\ 0 & & & \end{array}\right).$$

□

Before moving on, I will mention one application. If $A = XTX^{-1}$ for some (upper or lower) triangular matrix T, then the eigenvalues of A are the diagonal entries of T. One could imagine using this to **compute** the eigenvalues of A. Unfortunately, the proof of Schur triangularization assumes that we already know the eigenvalues of A.

Nevertheless, this is still the good idea, and it is behind the most powerful algorithm for computing eigenvalues. This algorithm uses the QR factorization (which does not assume knowledge of the eigenvalues) in a surprising way to recursively approximate the Schur decomposition, and hence the eigenvalues. It was discovered in the late 1950s by Francis and Kublanovskaya. I will present only the most basic version. The real world version uses extra tricks and optimizations.

The QR Algorithm for Computing Eigenvalues. Given a square matrix A, we recursively define unitary matrices $Q_1, Q_2, \ldots$ and upper-triangular matrices $R_1, R_2, \ldots$ as follows:

- Compute a QR factorization: $A = Q_1 R_1$.
- Next, compute a QR factorization of the matrix $R_1 Q_1$:[147]

$$R_1 Q_1 = Q_2 R_2.$$

- Continue to compute Q_{k+1} and R_{k+1} from the matrix $R_k Q_k$:

$$R_k Q_k = Q_{k+1} R_{k+1}.$$

Let's write $A_1 := A = Q_1 R_1$ and $A_k := Q_k R_k$. Since the Q in the QR factorization is unitary, we have $R_k = Q_k^* A_k$ and hence

$$A_{k+1} = R_k Q_k = Q_k^* A_k Q_k.$$

This implies that the sequence of matrices $A = A_1, A_2, \ldots$ all have the same eigenvalues. The theorem says the following.

Theorem. Suppose that A has eigenvalues with distinct absolute values:[148]

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|.$$

[147] This is a strange idea, but it leads to great results.

[148] There are modified versions of the algorithm that work for all square matrices.

Then the matrix $A_k = R_k Q_k$ in the QR algorithm converges to an upper triangular matrix, whose diagonal entries are the eigenvalues of A.

It is difficult to find a proof of this written down.[149] The only full proof I can find is in Wilkinson, *The Algebraic Eigenvalue Problem* (1965), page 516. Here is a sketch.

Sketch of a Proof. Define $\tilde{Q}_k := Q_1 Q_2 \cdots Q_k$ and $\tilde{R}_k = R_1 R_2 \cdots R_k$. Since the sets of unitary matrices and upper triangular matrices are closed under multiplication,[150] we see that $\tilde{Q}_k$ is unitary and $\tilde{R}_k$ is upper triangular. I claim that $A_{k+1} = \tilde{Q}_k^* A \tilde{Q}_k$ and $A^k = \tilde{Q}_k \tilde{R}_k$. Indeed, we have

$$\begin{aligned} A_{k+1} &= Q_k^* A_k Q_k \\ &= Q_k^* Q_{k-1}^* A_{k-1} Q_{k-1} Q_k \\ &\vdots \\ &= Q_k^* \cdots Q_2^* Q_1^* A Q_1 Q_2 \cdots Q_k \\ &= (Q_1 \cdots Q_k)^* A (Q_1 \cdots Q_k) \\ &= \tilde{Q}_k^* A \tilde{Q}_k \end{aligned}$$

and

$$\begin{aligned} A^k &= (Q_1 R_1) \cdots (Q_1 R_1) \\ &= Q_1 (R_1 Q_1) \cdots (R_1 Q_1) R_1 \\ &= Q_1 (Q_2 R_2) \cdots (Q_2 R_2) R_1 \\ &= Q_1 Q_2 (R_2 Q_2) \cdots (R_2 Q_2) R_2 R_1 \\ &= Q_1 Q_2 (Q_3 R_3) \cdots (Q_3 R_3) R_2 R_1 \\ &\vdots \\ &= (Q_1 Q_2 \cdots Q_k)(R_k \cdots R_2 R_1) \\ &= \tilde{Q}_k \tilde{R}_k. \end{aligned}$$

From our assumption that A has distinct eigenvalues, we can diagonalize A as

$$A = X \Lambda X^{-1} = X \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} X^{-1}.$$

By multiplying X^{-1} on the left by elementary matrices we can write $X^{-1} = L\Gamma$ where Γ is upper triangular and L is lower triangular with 1s on the diagonal. The key to the whole proof is to observe that $\Lambda^k L \Lambda^{-k}$ is lower triangular and converges to the identity matrix as

[149] Pure math books tend not to discuss it and applied math books tend not to prove it.
[150] Jargon: These sets are groups.

$k \to \infty$. Indeed, if ℓ_{ij} is the ij entry of L (so that $\ell_{ii} = 1$ and $\ell_{ij} = 0$ when $i < j$) then the ij entry of $\Lambda^k L \Lambda^{-k}$ is

$$(\Lambda^k L \Lambda^{-k})_{ij} = \begin{cases} 0 & i < j, \\ 1 & i = j, \\ \ell_{ij}(\lambda_i/\lambda_j)^k & i > j. \end{cases}$$

Since we have assumed that $|\lambda_1| > \cdots > |\lambda_n|$, it follows that the entries below the diagonal go to zero as $k \to \infty$.

By combining these ingredients, Wilkinson shows that the sequences $\tilde{Q}_k$ and $\tilde{R}_k$ converge, and that the the sequence Q_k converges to a diagonal matrix, hence $A_k = Q_k R_k$ converges to an upper triangular matrix. Let's say $\tilde{Q}_k \to U$ and $A_k \to T$, for unitary U and upper triangular T. Then in the limit we obtain the Schur triangularization:

$$A = \tilde{Q}_k A_{k+1} \tilde{Q}_k^* \to UTU^*.$$

□

Remark: The proof uses the fact QR factorization is unique up to multiplication with a unitary diagonal matrix D: $QR = (QD)(D^{-1}R)$. This follows from the the fact that any unitary upper triangular matrix must be diagonal.

12.3 The Spectral Theorem

The Spectral Theorem might be viewed as the "fundamental theorem" of spectral theory. It characterizes the best kind of matrices; namely, those that possess an orthonormal basis of eigenvectors. Here is the statement.

The Spectral Theorem. Let A be a square matrix with real or complex entries. Then

$$\boxed{A \text{ has an orthonormal basis of eigenvectors} \quad \Longleftrightarrow \quad A \text{ is normal, i.e., } A^*A = AA^*.}$$

Remarks: This theorem also applies to infinite dimensional Hilbert spaces, with suitable notions of continuity and convergence. The original version of the theorem applies to real symmetric matrices $A^T = A$, in which case we can choose the eigenvectors to be real.

You will find many different proofs in the literature. One proof uses the Schur Triangularization, which is convenient because we proved this in the previous section.

Proof using Schur Triangularization. One direction is easy. If A has an orthonormal basis of eigenvectors then we can find unitary $U^*U = I$ and diagonal Λ such that $A = U\Lambda U^*$. But then A is normal because diagonal matrices are normal:

$$\begin{aligned} A^*A &= (U\Lambda U^*)^*(U\Lambda U^*) \\ &= U\Lambda^* U^* U \Lambda U^* \\ &= U\Lambda^* \Lambda U^* \end{aligned}$$

$$
\begin{aligned}
&= U\Lambda\Lambda^*U^* && \Lambda^*\Lambda = \Lambda^*\Lambda \\
&= (U\Lambda U^*)(U\Lambda^*U^*) \\
&= AA^*.
\end{aligned}
$$

Conversely, suppose that $A^*A = AA^*$. By Schur's Theorem, any square matrix A can be written as $A = UTU^*$ where $U^*U = I$ is unitary and T is upper-triangular. Thus we have

$$T = U^*AU,$$

which implies that $T^*T = TT^*$ by the previous argument. Finally, one can check that any upper-triangular matrix T satisfying $T^*T = TT^*$ must be diagonal. This would make a good homework exercise. The proofs for self-adjoint and unitary matrices are easier. If $A^* = A$ then $T = U^*AU$ implies that $T^* = T$. This says that the above diagonal-entries of T are the conjugates of the below-diagonal entries of T, which are zero, and hence T is diagonal. If $A^*A = I$ then $T = U^*AU$ implies that $T^*T = I$, which says that the rows of T are orthogonal. By working from the bottom row to the top, this implies that the above-diagonal entries of T must be zero, hence T is diagonal. □

However, I think it is useful to give a proof from scratch. I will give the full proof for matrices, and I will sketch out the proof for operators on Hilbert spaces. For full details see John B. Conway, *A Course in Functional Analysis*, Chapter 2.

Proof of the Spectral Theorem, not using Schur Triangularization.

Step 1 for Matrices. If $A^*A = AA^*$ then A has an eigenvalue. Indeed, **any** square matrix has an eigenvalue by the Fundamental Theorem of Algebra.

Step 1 for Operators. If $A^*A = AA^*$ and if the set $\{A\mathbf{u} : \|\mathbf{u}\| = 1\}$ is compact (in which case we say that A is a *compact normal operator*) then A has an eigenvalue. I'll just sketch the proof. Since the set $\{A\mathbf{u} : \|\mathbf{u}\| = 1\}$ is compact,[151] the *operator norm* is finite:

$$\|A\| = \sup\{\|A\mathbf{u}\| : \|\mathbf{u}\| = 1\} < \infty.$$

Since $\|A\|$ exists we can find a sequence of unit vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots$ such that $\langle A\mathbf{v}_n, \mathbf{v}_n\rangle \to \lambda$ as $n \to \infty$ for some $\lambda \in \mathbb{C}$ satisfying $|\lambda| = \|A\|$. Then since the set $\{A\mathbf{u} : \|\mathbf{u}\| = 1\}$ is compact we can find a subsequence $\mathbf{v}_{n_1}, \mathbf{v}_{n_2}, \ldots$ so that $A\mathbf{v}_{n_k} \to \mathbf{v}$ as $k \to \infty$ for some nonzero vector $\mathbf{v}$ satisfying $A\mathbf{v} = \lambda\mathbf{v}$, and hence λ is an eigenvalue.

Step 2 for Matrices. Let $A^*A = AA^*$, with shape $n \times n$. By Step 1 we can find an eigenvalue λ_1. Choose any unit length eigenvector $\mathbf{u}_1$ so that $A\mathbf{u}_1 = \lambda_1\mathbf{u}_1$, and complete this to an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. In particular, we have $\langle \mathbf{u}_1, \mathbf{u}_j\rangle$ for all $j > 1$. Since A is normal, I claim that we also have $\langle \mathbf{u}_1, A\mathbf{u}_j\rangle = 0$ for all $j > 1$. Indeed, if A is normal then

[151] This set is the image of the unit ball under A. In finite dimensions this set is an ellipsoid.

we showed in the previous chapter that $A\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$ implies $A^*\mathbf{u}_1 = \bar{\lambda_1}\mathbf{u}_1$. But then we must have

$$\langle \mathbf{u}_1, A\mathbf{u}_j \rangle = \langle A^*\mathbf{u}_1, \mathbf{u}_j \rangle = \langle \bar{\lambda_1}\mathbf{u}_1, \mathbf{u}_j \rangle = \lambda_1 \langle \mathbf{u}_1, \mathbf{u}_j \rangle = 0.$$

If U is the (unitary) matrix with columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$ then we have

$$A = U \left(\begin{array}{c|ccc} \lambda_1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array} \right) U^*,$$

for some matrix A' of size $(n-1) \times (n-1)$. The displayed matrix containing λ and A' equals U^*AU. This implies that it is normal, and, in particular, the matrix A' is normal. Now we apply induction exactly as in the proof of Schur Triangularization.

Step 2 for Operators. (This is just a sketch; the details fill pages.) Let $A : V \to V$ be a compact normal operator and set $A_1 := A$. By Step 1 there exists an eigenvalue $\lambda_1 \in \mathbb{C}$ such that $|\lambda_1| = \|A_1\|$. If A_1 is not the zero operator then we have $\|A_1\| \neq 0$ and hence $\lambda_1 \neq 0$. Consider the eigenspace $V_1 = ker(\lambda_1 I - A_1) \subseteq V$ and its orthogonal complement $V_1^\perp \subseteq V$. Since A_1 is a compact operator, one can show that the eigenspace V_1 is finite dimensional, say $\dim V_1 = n_1$. Pick an orthonormal basis $\mathbf{u}_{11}, \ldots, \mathbf{u}_{1n_1} \in V_1$. Then since A_1 is normal, the same proof as for matrices shows that A_1 sends $V_1^\perp$ to itself.[152] Let $A_2 : V_1^\perp \to V_1^\perp$ denote the restriction of A_1 to $V_1^\perp$. One can show that A_2 is compact and normal, with $\|A_2\| < \|A_1\|$. If $A_2 \neq 0$ then by Step 1 there exists a nonzero eigenvalue $\lambda_2 \in \mathbb{C}$ with $|\lambda_2| = \|A_2\|$ and with a finite dimensional eigenspace $V_2 = ker(\lambda_2 I - A_2)$. Say $\dim A_2 = n_2$ and pick an orthonormal basis $\mathbf{u}_{21}, \ldots, \mathbf{u}_{2n_2} \in V_2$. Continuing in this way we obtain a sequence of finite dimensional eigenspaces $V_1, V_2, \ldots$ corresponding to eigenvalues with $|\lambda_1| > |\lambda_2| > \cdots$. Finally, one can show that the concatenation of the bases $\mathbf{u}_{k1}, \mathbf{u}_{k2}, \ldots, \mathbf{u}_{kn_k} \in V_k$ is an orthonormal basis for V. The remaining issue is to show that every vector $\mathbf{v} \in V$ can be expressed as a convergent series of eigenvectors $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots$, with $\mathbf{v}_k \in V_k$. □

In applications it is usually safe to ignore convergence issues and just the problem at an algebraic level. For example,

12.4 The Singular Value Decomposition

A form of "generalized diagonalization" that applies to rectangular matrices and non-diagonalizable square matrices.

For any $m \times n$ matrix A and $n \times m$ matrix B, the square matrices AB $(m \times m)$ and BA $(n \times n)$ have the same nonzero eigenvalues. If $m < n$ then the matrix BA has $n - m$ extra zero eigenvalues compared to AB.

For any $m \times n$ matrix A, the eigenvalues of A^TA are real and non-negative: $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. The singular values of A are the non-negative real square roots of the eigenvalues: $\sigma_i = \sqrt{\lambda_i}$. Equivalently, $\sigma_1^2, \ldots, \sigma_n^2$ are the eigenvalues of A^TA.

[152] This is the key point in the proof.

Properties: The largest singular value σ_1 is the operator norm $\|A\|$. The product of the singular values is $\sqrt{\det(A^TA)}$.

Let Σ be the $n \times n$ diagonal matrix of singular values, so $\Lambda = \Sigma^2 = \Sigma^T\Sigma = \Sigma\Sigma^T$ is the diagonal matrix of eigenvalues. From the spectral theorem there exists a unitary (orthogonal) matrix V such that $A^TA = V\Lambda V^T = V\Sigma\Sigma^TV^T = (V\Sigma)(V\Sigma)^T$. The columns $\mathbf{v}_i$ of V are the eigenvectors of A^TA.

Suppose that A^TA has rank r, which is also the rank of A and AA^T, so that there are r nonzero singular values $\sigma_1 \geq \cdots \geq \sigma_r$. From our first remark, A^TA and AA^T have the same non-zero eigenvalues. Define $\mathbf{u}_i = (A\mathbf{v}_i)/\sigma_i$ for $1 \leq i \leq r$. Then $\mathbf{u}_i$ are the eigenvectors of AA^T corresponding to the eigenvalues $\sigma_1^2 \geq \cdots \geq \sigma_r^2$. Complete the $\mathbf{u}_i$ to a basis of $\mathbb{R}^m$ arbitrarily and let U the $m \times m$ unitary (orthogonal) matrix with columns $\mathbf{u}_i$. Then we have $A = U\Sigma V^*$. (I guess we have to pad Σ with some zeros.) This is the singular value decomposition (SVD).

Geometry: A sends the unit ball in $\mathbb{R}^n$ to an ellipsoid in $\mathbb{R}^m$. The singular values of A are the radii of the ellipsoid.

Eckart-Young Theorem. Write $A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. Then $A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is the best rank k approximation to A. That is, for any rank k matrix B we have $\|A - A_k\| \leq \|A - B\|$.

Application: Principal Component Analysis. (Total Least Squares.)

Maybe put all of this in a separate chapter.

12.5 Jordan Canonical Form

If $\chi_A(x) = \prod(x - \lambda_i)^{n_i}$ and $E_{\lambda_i} = \mathcal{N}((\lambda_i I - A)^{n_i}) = n_i$ is the generalized eigenspace then $\dim E_{\lambda_i} = n_i$.

The companion matrix.

13 Applications of Spectral Theory

13.1 The Principal Axes Theorem

The earliest example of the Spectral Theorem goes all the way back to the birth of analytic geometry. It was known to Descartes and Fermat in the early 1600s and was applied by Euler in the 1700s to the mechanics of rotating bodies.[153]

You may have seen this theorem in school: Any polynomial equation of the form

$$f(x, y) = a + bx + cy + dx^2 + exy + fy^2 = 0$$

can be brought into standard form by a translation and a rotation. The standard forms are

$$\begin{aligned} \text{parabola} &: y = ax^2 \text{ or } x = ay^2, \\ \text{ellipse} &: x^2/a^2 + y^2/b^2 = 1, \\ \text{hyperbola} &: \pm(x^2/a^2 - y^2/b^2) = 1. \end{aligned}$$

[153] *Highlights in the History of Spectral Theory*, Steen.

The Principal Axes Theorem generalizes this to higher dimensions.

Theorem (Principal Axes Theorem). Consider a general polynomial $f(x_1, \ldots, x_n)$ of degree 2 in n variables. This can be expressed as

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

for some scalar b, vector $\mathbf{b}$ and **symmetric** matrix B. If B^{-1} exists, then we can find a change of variables $\mathbf{u} = Q\mathbf{x} + \mathbf{t}$, where $Q^TQ = I$ is an orthogonal matrix[154] and $\mathbf{t}$ is a (translation) vector, such that

$$f(u_1, \ldots, u_n) = a + \lambda_1 u_1^2 + \lambda_2 u^2 + \cdots + \lambda_n u_n^2.$$

What if B^{-1} doesn't exist?

Proof. Let $\mathbf{u} = Q\mathbf{x} + \mathbf{t}$ for some invertible matrix Q and vector $\mathbf{t}$. Then we have

$$\begin{aligned}
f(\mathbf{u}) &= f(Q\mathbf{x} + \mathbf{t}) \\
&= b + \mathbf{b}^T(Q\mathbf{x} + \mathbf{t}) + (Q\mathbf{x} + \mathbf{t})^T B(Q\mathbf{x} + \mathbf{t}) \\
&= b + \mathbf{b}^TQ\mathbf{x} + \mathbf{b}^T\mathbf{t} + (\mathbf{x}^TQ^T + \mathbf{t}^T)B(Q\mathbf{x} + \mathbf{t}) \\
&= b + \mathbf{b}^TQ\mathbf{x} + \mathbf{b}^T\mathbf{t} + \mathbf{x}^TQ^TBQ\mathbf{x} + \mathbf{x}^TQ^TB\mathbf{t} + \mathbf{t}^TBQ\mathbf{x} + \mathbf{t}^TB\mathbf{t} \\
&= b + \mathbf{b}^TQ\mathbf{x} + \mathbf{b}^T\mathbf{t} + \mathbf{x}^TQ^TBQ\mathbf{x} + 2\mathbf{t}^TBQ\mathbf{x} + \mathbf{t}^TB\mathbf{t} \qquad (*) \\
&= (b + \mathbf{b}^T\mathbf{t} + \mathbf{t}^TB\mathbf{t}) + (\mathbf{b}^TQ + 2\mathbf{t}BQ)\mathbf{x} + \mathbf{x}^TQ^TBQ\mathbf{x}.
\end{aligned}$$

Step $(*)$ uses the facts that $B^T = B$ and that $\mathbf{t}^TBQ\mathbf{x}$ is a scalar, hence

$$t^TBQ\mathbf{x} = (t^TBQ\mathbf{x})^T = \mathbf{x}^TQ^TB^T\mathbf{t} = \mathbf{x}^TQ^TB\mathbf{x}.$$

If B^{-1} exists, then we can eliminate the linear terms by taking

$$\begin{aligned}
\mathbf{b}^TQ + 2\mathbf{t}BQ &= \mathbf{0}^T \\
2\mathbf{t}BQ &= -\mathbf{b}^TQ \\
\mathbf{t}BQ &= -\frac{1}{2}\mathbf{b}^TQ \\
\mathbf{t} &= -\frac{1}{2}\mathbf{b}^TQQ^{-1}B^{-1} \\
\mathbf{t} &= -\frac{1}{2}\mathbf{b}^TB^{-1}.
\end{aligned}$$

Finally, since B is symmetric, the Spectral Theorem says that we can choose orthogonal $Q^TQ = I$ so that Q^TBQ is diagonal:

Choose $\mathbf{t}$ so that $\mathbf{b}^T + 2\mathbf{t}^TB = \mathbf{0}^T$. Assume B invertible.

[154] If the coefficients of f are complex then $Q^*Q = I$ is unitary, but we are usually interested in the real case.

13.2 Positive Definite Matrices

If $\langle \mathbf{x}, B\mathbf{x} \rangle \geq 0$ (enough to assume $\in \mathbb{R}$) for all $\mathbf{x} \in \mathbb{C}^n$ then $B^* = B$.

Proof: We need to show that $\langle B\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, B\mathbf{y} \rangle$ for all $\mathbf{x}, \mathbf{y}$. First note that $\langle B\mathbf{x}, \mathbf{x} \rangle = \langle \mathbf{x}, B\mathbf{x} \rangle^* = \langle \mathbf{x}, B\mathbf{x} \rangle$, so $\langle \mathbf{x}, (B - B^*)\mathbf{x} \rangle$ for all $\mathbf{x}$. We need to show that $\langle \mathbf{x}, T\mathbf{x} \rangle = 0$ for all $\mathbf{x}$ implies $T = O$. Indeed, we have

$$0 = \langle \mathbf{x} + \mathbf{y}, T(\mathbf{x} + \mathbf{y}) \rangle = \langle \mathbf{x}, T\mathbf{y} \rangle + \langle \mathbf{y}, T\mathbf{x} \rangle + 0 + 0$$

and

$$0 = \langle \mathbf{x} + i\mathbf{y}, T(\mathbf{x} + i\mathbf{y}) \rangle = i\langle \mathbf{x}, T\mathbf{y} \rangle - i\langle \mathbf{y}, T\mathbf{x} \rangle + 0 + 0.$$

Divide the second equation by i and add them to obtain $2\langle \mathbf{x}, T\mathbf{y} \rangle = 0$ and hence $\langle \mathbf{x}, T\mathbf{y} \rangle = 0$ for all $\mathbf{x}, \mathbf{y}$.

In principle, our proof of the Spectral Theorem gives an algorithm to factor a semi-definite matrix $B = A^T A$, but is probably not the most efficient method since it assumes that we already know the eigenvalues. The *Cholesky factorization* is a method to factor $B = A^T A$ that avoids having to compute eigenvalues.

13.3 Differential Equations

The matrix exponential encodes the solution to linear systems of differential equations. To begin, recall the power series definition of the exponential function:

$$\exp(x) := 1 + x + \frac{1}{2}x^2 + \cdots + \frac{1}{k!}x^k + \cdots .$$

It is a basic theorem of analysis that this series converges uniformly for any complex number $x \in \mathbb{C}$. It was invented by Euler because of the following special properties. For any complex numbers $x, y \in \mathbb{C}$ we have

$$\begin{aligned}
\exp(x)\exp(y) &= \left(\sum_{i \geq 0} \frac{1}{i!} \cdot x^i \right) \left(\sum_{j \geq 0} \frac{1}{j!} \cdot x^j \right) \\
&= \sum_{k \geq 0} \left(\sum_{i+j=k} \frac{1}{i!} \cdot x^i \cdot \frac{1}{j!} \cdot x^j \right) \\
&= \sum_{k \geq 0} \frac{1}{k!} \cdot \left(\sum_{i+j=k} \frac{k!}{i!j!} x^i x^j \right) \\
&= \sum_{k \geq 0} \frac{1}{k!} \cdot (x + y)^k \\
&= \exp(x + y).
\end{aligned}$$

This property suggests that $\exp(x) = e^x$ for some number e, which Euler calculated to be ≈ 2.71828. Furthermore, the power series $\exp(x)$ is equal to its own derivative:

$$\frac{d}{dx}\exp(x) = \frac{d}{dx}\left(1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots\right)$$

$$
\begin{aligned}
&= 0 + 1 + \frac{1}{2!} \cdot 2x + \frac{1}{3!} \cdot 3x^2 + \frac{1}{4!} \cdot 4x^3 + \cdots \\
&= 0 + 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots \\
&= \exp(x).
\end{aligned}
$$

Conversely, let $f : \mathbb{C} \to \mathbb{C}$ be any function satisfying $\frac{d}{dx} f(x) = f(x)$. Suppose that $f(x)$ has a convergent power series expansion near $x = 0$:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots .$$

The equation $\frac{d}{dx} f(x) = f(x)$ tells us that

$$a_0 + a_1 x + a_2 x^2 + \cdots = a_1 + 2a_1 x + 3a_2 x^2 + \cdots .$$

Then comparing coefficients tells us that $a_k = (k+1)a_{k-1}$ for all $k \geq 0$, which has the unique solution $a_k = a_0/k!$. Hence we must have $f(x) = a_0 \exp(x)$.

Now consider a vector of functions $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$. We can think of this as a parametrized path in n-dimensional space: $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$.[155] A linear system of ordinary differential equations has the form

$$
\left\{ \begin{array}{rcccccc}
x_1'(t) & = & a_{11}x_1(t) & + & \cdots & + & a_{1n}x_n(t) \\
 & \vdots & & & & & \\
x_n'(t) & = & a_{n1}x_1(t) & + & \cdots & + & a_{nn}x_n(t)
\end{array} \right\} \quad \rightsquigarrow \quad \mathbf{x}'(t) = A\mathbf{x}(t),
$$

for some $n \times n$ matrix A of constants. We can think of $\mathbf{x}'(t)$ as the velocity vector of the path $\mathbf{x}(t)$, and we can think of A as specifying a vector field on $\mathbb{R}^n$, with value $A\mathbf{x}$ at the point $\mathbf{x}$. A solution to the equation $\mathbf{x}'(t) = A\mathbf{x}(t)$ is any path $\mathbf{x}(t)$ in $\mathbb{R}^n$ that flows along the vector field defined by A. For any initial point $\mathbf{x}(0) \in \mathbb{R}^n$

The companion matrix:
`https://math.stackexchange.com/questions/348498/jordan-basis-of-a-when-a-is-the-companion`

13.4 Graph Theory

Powers of the adjacency matrix. Acyclic directed graphs:

Paper: acyclic digraphs and eigenvalues of (0,1)-matrices

13.5 Markov Chains

Perron-Frobenius, Page Rank

[155] I guess we'll work with real numbers to make visualization easier.

13.6 Singular Value Decomposition

13.7 Total Least Squares

Given a matrix of n data points in any dimensional space:

$$X = \left(\begin{array}{c|c|c} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{array} \right)$$

For any vector $\mathbf{a}$ let $P_{\mathbf{a}} = \mathbf{a}\mathbf{a}^T/\|\mathbf{a}\|^2$ be projection onto the line $\mathbf{a}$ and $Q_{\mathbf{a}} = I - P_{\mathbf{a}}$ be projection onto the hyperplane $\mathbf{a}^\perp$. For any i we have

$$\begin{aligned} \mathbf{x}_i &= P_{\mathbf{a}}\mathbf{x}_i + Q_{\mathbf{a}}\mathbf{x}_i \\ \|\mathbf{x}_i\|^2 &= \|P_{\mathbf{a}}\mathbf{x}_i\|^2 + \|Q_{\mathbf{a}}\mathbf{x}_i\|^2 \\ \sum_i \|\mathbf{x}_i\|^2 &= \sum \|P_{\mathbf{a}}\mathbf{x}_i\|^2 + \sum \|Q_{\mathbf{a}}\mathbf{x}_i\|^2. \end{aligned}$$

Goal: Choose $\mathbf{a}$ to minimize $\sum \|Q_{\mathbf{a}}\mathbf{x}_i\|^2$. Since $\sum_i \|\mathbf{x}_i\|^2$ is fixed by the data, this is the same as maximizing $\sum \|P_{\mathbf{a}}\mathbf{x}_i\|^2$. But

$$\|P_{\mathbf{a}}\mathbf{x}_i\|^2 = \frac{1}{\|a\|^2}|\mathbf{a}^T\mathbf{x}_i|^2 = \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}^T\mathbf{x}_i\overline{\mathbf{a}^T\mathbf{x}_i} = \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{a},$$

hence

$$\sum \|P_{\mathbf{a}}\mathbf{x}_i\|^2 = \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}^T XX^T\mathbf{a} = \frac{1}{\|\mathbf{a}\|^2}(X^T\mathbf{a})^T(X^T\mathbf{a}) = \frac{\|X^T\mathbf{a}\|^2}{\|\mathbf{a}\|^2}.$$

This is maximized by letting $\mathbf{a}$ be an eigenvector for the largest (real) eigenvalue of XX^T.

Proof. By S.T., XX^T can be unitarily diagonalized: $XX^T\mathbf{u}_i = \sigma_i^2\mathbf{u}_i$. Let

$$\mathbf{a} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n,$$

so that

$$\frac{1}{\|\mathbf{a}\|^2}\mathbf{a}^T XX^T\mathbf{a} = \frac{\sigma_1^2c_1^2 + \cdots + \sigma_n^2c_n^2}{c_1^2 + \cdots + c_n^2}.$$

Maximum when $c_1 = 1$ and $c_2 = \cdots = c_n = 0$. Maximum under constraint $c_1 = 0$ gives $c_2 = 1$ and $c_3 = \cdots = c_n = 0$, etc.