

Quick Review of Matrix and Linear Algebra

KC Border

Subject to change

Last major revision: September 10, 2008

v. 2012.02.22::18.50

Contents

1	Vector spaces	1
1.1	The vector space \mathbf{R}^n	1
1.2	Other examples of vector spaces	1
1.3	Some elementary consequences	2
1.4	Linear combinations and subspaces	2
1.5	Linear independence	3
1.6	Bases	3
1.7	Linear transformations	4
1.8	The coordinate mapping	4
2	Inner product	5
2.1	Orthogonality	6
2.2	Orthogonal projection	6
3	Linear transformations	9
3.1	Inverse of a linear transformation	10
3.2	Transpose of a transformation	10
3.3	Eigenvalues and eigenvectors	12
4	Matrices	15
4.1	Matrix operations	15
4.2	Systems of linear equations	16
4.3	Matrix representation of a linear transformation	18
4.4	Matrix representation of a composition	20
4.5	Change of basis	21
4.6	Simultaneous diagonalization	23
4.7	Trace	24
5	Determinants	25
5.1	Determinants as multilinear forms	26
5.2	Some simple consequences	28
5.3	Minors and cofactors	29
5.4	Characteristic polynomials	32
5.5	The determinant as an “oriented volume”	33
5.6	Computing inverses and determinants by Gauss’ method	34

1 Vector spaces

Let K be a field of **scalars**—usually either the real numbers \mathbf{R} or the complex numbers \mathbf{C} , or occasionally the rationals \mathbf{Q} . A **vector space** over K is a set V of **vectors** equipped with two operations, vector addition $(x, y) \mapsto x + y$, and scalar multiplication $(\alpha, x) \mapsto \alpha x$, where $x, y \in V$ and $\alpha \in K$. The operations satisfy:

V.1 $x + y = y + x$

V.2 $(x + y) + z = x + (y + z)$

V.3 There is a vector 0 , satisfying $x + 0 = x$ for every vector x .

V.4 $x + (-1)x = 0$

V.5 $\alpha(\beta x) = (\alpha\beta)x$

V.6 $1x = x$

V.7 $\alpha(x + y) = (\alpha x) + (\alpha y)$

V.8 $(\alpha + \beta)x = (\alpha x) + (\beta x)$

1.1 The vector space \mathbf{R}^n

By far the most important example of a vector space, and indeed the model of all vector spaces, is the space \mathbf{R}^n of ordered lists of n real numbers. Given ordered lists $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ the vector sum $x + y$ is the ordered list $(x_1 + y_1, \dots, x_n + y_n)$. The scalar product is given by the ordered list $\alpha x = (\alpha x_1, \dots, \alpha x_n)$. The zero of this vector space is the ordered list $0 = (0, \dots, 0)$. It is a trivial matter to verify that the axioms above are satisfied by these operations.

In \mathbf{R}^n we identify several special vectors, the **unit coordinate vectors**. These are the ordered lists e^i that consist of zeroes except for a single 1 in the i^{th} position. We use the notation e^i to denote this vector regardless of the length n of the list.

1.2 Other examples of vector spaces

The following are also vector spaces. The definition of the operations is usually obvious.

- $\{0\}$ is a vector space, called the **trivial** vector space. A nontrivial vector space contains at least one nonzero vector.
- The field K is a vector space over itself.
- Integrable real-valued functions.

$$\int |f(x)| dx < \infty$$

- Square-integrable random variables.

$$E X^2 < \infty$$

- The set of solutions to $f'' + \alpha f' + \beta f = 0$.

- The ℓ_p spaces.

$$\ell_p = \left\{ x = (x_1, x_2, \dots) : \sum_{n=1}^{\infty} |x_n|^p < \infty \right\}$$

$$\ell_{\infty} = \left\{ x = (x_1, x_2, \dots) : \sup_n |x_n| < \infty \right\}$$

- The set of $m \times n$ matrices.
- The set of linear transformations from one vector space into another.

1.3 Some elementary consequences.

Here are some simple consequences of the axioms that we shall use repeatedly without further comment.

1. 0 is unique.

$$0_2 \underbrace{=}_{\text{by V.3}} 0_2 + 0_1 \underbrace{=}_{\text{by V.1}} 0_1 + 0_2 \underbrace{=}_{\text{by V.3}} 0_1$$

2. $-x = (-1)x$ is the unique vector z satisfying $x + z = 0$. Suppose

$$\begin{aligned} x + y &= 0 \\ \underbrace{-x + x} + y &= -x \\ 0 + y &= -x \\ y &= -x. \end{aligned}$$

3. $0_K x = 0_V$:

$$0_V = 0_K x + (-1)(0_K x) = (0_K - 0_K)x = 0_K x.$$

4. $\underbrace{x + \dots + x}_n = nx$:

$$\begin{aligned} x + x &= 1x + 1x = (1 + 1)x = 2x \\ x + x + x &= (x + x) + x = 2x + x = 2x + 1x = (2 + 1)x = 3x \end{aligned}$$

etc.

1.4 Linear combinations and subspaces

A **linear combination** of x_1, \dots, x_m is any sum of scalar multiples of the form $\sum_{i=1}^m \alpha_i x_i$, $\alpha_i \in K$, $x_i \in V$. A **linear subspace** $M \subset V$ is a subset of V that is closed under linear combinations. A linear subspace of a vector space is a vector space in its own right.

Let $E \subset V$. The **span** of E , denoted $\text{span } E$, is the set of all linear combinations from E . That is,

$$\text{span } E = \left\{ \sum_{i=1}^m \alpha_i x_i : \alpha_i \in K, x_i \in E, m \in \mathbb{N} \right\}.$$

1 Exercise Prove the following.

1. $\{0\}$ is a linear subspace.
2. If M is a linear subspace, then $0 \in M$.
3. The intersection of a family of linear subspaces is a linear subspace.
4. The set $\text{span } E$ is the smallest linear subspace including E .

1.5 Linear independence

A set E of vectors is **linearly dependent** if there are distinct vectors x_1, \dots, x_m belonging to E , and nonzero scalars $\alpha_1, \dots, \alpha_m$, such that $\sum_{i=1}^m \alpha_i x_i = 0$. A set of vectors is **linearly independent** if it is not dependent. That is, E is independent if for every set x_1, \dots, x_m of distinct vectors in E , $\sum_{i=1}^m \alpha_i x_i = 0$ implies $\alpha_1 = \dots = \alpha_m = 0$. We also say that the vectors x_1, \dots, x_m are independent instead of saying that the set $\{x_1, \dots, x_m\}$ is independent.

2 Proposition (Uniqueness of linear combinations) *If E is a linearly independent set of vectors and x belongs to $\text{span } E$, then x is a unique linear combination of elements of E .*

Proof: If x is zero, the conclusion follows by definition of independence. If x is nonzero, suppose

$$x = \sum_{i=1}^m \alpha_i x_i = \sum_{j=1}^n \beta_j y_j,$$

where the x_i s are distinct elements of E , the y_j s are distinct are distinct elements of E (but may overlap with the x_i s), and $\alpha_i, \beta_j > 0$ for $i = 1, \dots, m$ and $j = 1, \dots, n$. Enumerate $A = \{x_i : i = 1, \dots, m\} \cup \{y_j : j = 1, \dots, n\}$ as $A = \{z_k : k = 1, \dots, p\}$. Then we can rewrite $x = \sum_{k=1}^p \hat{\alpha}_k z_k = \sum_{k=1}^p \hat{\beta}_k z_k$, where

$$\hat{\alpha}_k = \begin{cases} \alpha_i & \text{if } z_k = x_i \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{\beta}_k = \begin{cases} \beta_j & \text{if } z_k = y_j \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$0 = x - x = \sum_{k=1}^p (\hat{\alpha}_k - \hat{\beta}_k) z_k \implies \hat{\alpha}_k - \hat{\beta}_k = 0, \quad k = 1 \dots p$$

since E is independent. Therefore $\hat{\alpha}_k = \hat{\beta}_k$, $k = 1 \dots p$, which in turn implies $m = n = p$ and $\{x_i : i = 1, \dots, m\} = \{y_j : j = 1, \dots, n\}$, and the proposition is proved. ■

The coefficients of this linear combination are called the **coordinates of x with respect to the set E** .

3 Exercise *Prove the following.*

1. *The empty set is independent.*
2. *If $0 \in E$, then E is dependent.*

1.6 Bases

4 Definition A **Hamel basis**, or more succinctly, a **basis** for the linear space V is a linearly independent set B such that $\text{span } B = V$. The plural of basis is bases.

The next result is immediate from Proposition 2.

5 Proposition *Every element of the vector space V is a unique linear combination of basis vectors.*

6 Example (The standard basis for \mathbf{R}^m) The set of unit coordinate vectors e^1, \dots, e^m in \mathbf{R}^m is a basis for \mathbf{R}^m , called the **standard basis**.

Observe that the vector $x = (x_1, \dots, x_m)$ can be written uniquely as $\sum_{j=1}^m x_j e^j$. □

The fundamental facts about bases are these. For a proof of the first assertion, see [2, Theorem 1.8, p. 15]

7 Fact *Every nontrivial vector space has a basis. Any two bases have the same cardinality, called the **dimension** of V .*

Prove the second half for finite dimensional case.

Mostly we will deal with finite dimensional vector spaces. The next result summarizes Theorems 1.5, 1.6, and 1.7 in [3, pp. 10–12]

8 Theorem *In an n -dimensional space, every set of more than n vectors is dependent. Consequently, any independent set of n vectors is a basis.*

1.7 Linear transformations

Let V, W be vector spaces. A function $T: V \rightarrow W$ is a **linear transformation** if

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y).$$

If T is a linear transformation from a vector space into the reals \mathbf{R} , then it is customary to call T a **linear functional**.

Note that if T is linear, then $T0 = 0$. We shall use this fact without any special mention. It is traditional to write a linear transformation without parentheses, that is, to write Tx rather than $T(x)$.

The set of linear transformations from the vector space V into the vector space W is denoted $L(V, W)$. It is a vector space under the usual pointwise addition and scalar multiplication of functions.

1.8 The coordinate mapping

In an n -dimensional space V , if we fix a basis in some particular order, we have an **ordered basis** or **frame**. Given this ordered basis the coordinates of a vector comprise an ordered list, and so correspond to an element in \mathbf{R}^n . This mapping from vectors to their coordinates is called the **coordinate mapping** for the ordered basis. The coordinate mapping preserves all the vector operations. The technical term for this is isomorphism. If two vector spaces are isomorphic, then for many purposes we can think of them as being the same space, the only difference being that the names of the points have been changed.

9 Definition *An **isomorphism** between vector spaces is a bijective linear transformation. That is, a function $\varphi: V \rightarrow W$ between vector spaces is an isomorphism if φ is one-to-one and onto, and*

$$\varphi(x + y) = \varphi(x) + \varphi(y) \quad \text{and} \quad \varphi(\alpha x) = \alpha\varphi(x).$$

*In this case we say that V and W are **isomorphic**.*

We have just argued the following.

10 Proposition *Given an ordered basis x_1, \dots, x_n for an n -dimensional vector space V , the coordinate mapping is an isomorphism from V to \mathbf{R}^n .*

2 Inner product

11 Definition A real linear space V has an **inner product** if for each pair of vectors x and y there is a real number, traditionally denoted (x, y) , satisfying the following properties.

IP.1 $(x, y) = (y, x)$.

IP.2 $(x, y + z) = (x, y) + (x, z)$.

IP.3 $\alpha(x, y) = (\alpha x, y) = (x, \alpha y)$.

IP.4 $(x, x) > 0$ if $x \neq 0$.

A vector space V equipped with an inner product is called an **inner product space**.

For a **complex** vector space, the inner product is complex-valued, and property (1) is replaced by $(x, y) = \overline{(y, x)}$, where the bar denotes complex conjugation, and (3) is replaced by $\alpha(x, y) = (\alpha x, y) = (x, \overline{\alpha}y)$.

12 Example The **dot product** of two vectors x and y in \mathbf{R}^n is defined by

$$x \cdot y = \sum_{i=1}^n x_i y_i.$$

The dot product is an inner product, and **Euclidean n -space** is \mathbf{R}^n equipped with this inner product. The dot product of two vectors is zero if they meet at a right angle. \square

13 Lemma If $(x, z) = (y, z)$ for all z , then $x = y$.

Proof: This implies $((x - y), z) = 0$ for all z , so setting $z = x - y$ yields $((x - y), (x - y)) = 0$, which by IP.4 implies $x - y = 0$. \blacksquare

The next result may be found in [3, Theorem 1.8, p. 16].

14 Cauchy–Schwartz Inequality

$$|(x, y)|^2 \leq (x, x)(y, y)$$

with = if and only if x and y are dependent.

Prove this.

15 Definition A **norm** $\|x\|$ is a real function on a vector space that satisfies:

N.1 $\|0\| = 0$

N.2 $\|x\| > 0$ if $x \neq 0$

N.3 $\|\alpha x\| = |\alpha| \|x\|$.

N.4 $\|x + y\| \leq \|x\| + \|y\|$ with equality if and only if $x = 0$ or $y = 0$ or $y = \alpha x$, $\alpha > 0$.

16 Proposition If (\cdot, \cdot) is an inner product, then $\|x\| = (x, x)^{\frac{1}{2}}$ is a norm.

2.1 Orthogonality

17 Definition Vectors x and y are **orthogonal** if $(x, y) = 0$, written

$$x \perp y.$$

A set of vectors $E \subset V$ is **orthogonal** if it is pairwise orthogonal. That is, for all $x, y \in E$ with $x \neq y$ we have $(x, y) = 0$. A set E is **orthonormal** if E is orthogonal and $(x, x) = 1$ for all $x \in E$.

18 Lemma If a set of nonzero vectors is pairwise orthogonal, then the set is independent.

Proof: Suppose $\sum_{i=1}^m \alpha_i x_i = 0$, where the x_i s are pairwise orthogonal. Then for any k , we have $(x_k, \sum_{i=1}^m \alpha_i x_i) = 0$. So

$$\begin{aligned} 0 &= (x_k, 0) = \left(x_k, \sum_{i=1}^m \alpha_i x_i\right) = \sum_{i=1}^m \alpha_i (x_k, x_i) \\ &= \alpha_k (x_k, x_k). \end{aligned}$$

Since x_k is nonzero, $(x_k, x_k) > 0$, so it follows that $\alpha_k = 0$. Since k is arbitrary, $\alpha_1 = \cdots = \alpha_m = 0$, so the vectors x_i are linearly independent. ■

19 Definition The **orthogonal complement** of M in V is the set

$$\{x \in V : (\forall y \in M) [x \perp y]\},$$

denoted M_\perp .

20 Lemma If $x \perp y$ and $x + y = 0$, then $x = y = 0$.

Proof: Now $(x + y, x + y) = (x, x) + 2(x, y) + (y, y)$, so $(x, y) = 0$ implies $(x, x) + (y, y) = 0$. Since $(x, x) \geq 0$ and $(y, y) \geq 0$ we have $x = y = 0$. ■

The next lemma is left as an exercise.

21 Lemma For any set M , the set M_\perp is a linear subspace. If M is a subspace, then

$$M \cap M_\perp = \{0\}.$$

2.2 Orthogonal projection

The next result may be found in [3, Theorem 1.14, p. 24], and the proof is known as the **Gram–Schmidt procedure**.

22 Proposition Let x^1, x^2, \dots be a sequence (finite or infinite) of vectors in an inner product space. Then there exists a sequence y^1, y^2, \dots such that for each m , the span of y^1, \dots, y^m is the same as the span of x^1, \dots, x^m , and the y^i s are orthogonal.

Proof: Set $y^1 = x^1$. For $m > 1$ recursively define

$$y^m = x^m - \frac{(x^m, y^{m-1})}{(y^{m-1}, y^{m-1})} y^{m-1} - \frac{(x^m, y^{m-2})}{(y^{m-2}, y^{m-2})} y^{m-2} - \cdots - \frac{(x^m, y^1)}{(y^1, y^1)} y^1.$$

Use induction on m to prove that the vectors y_1, \dots, y_m are orthogonal and span the same space as the x^i 's. Observe that y_2 is orthogonal to $y^1 = x^1$:

$$(y^2, y^1) = (y^2, x^1) = (x^2, x^1) - \frac{(x^2, x^1)}{(x^1, x^1)}(x^2, x^1) = 0.$$

Furthermore any linear combination of x^1 and x^2 can be replicated with y^1 and y^2 .

For $m > 2$, suppose that y_1, \dots, y_{m-1} are orthogonal and span the same space as x^1, \dots, x^{m-1} . Now compute (y^m, y^k) for $k \leq m$:

$$\begin{aligned} (y^m, y^k) &= (x^m, y^k) - \sum_{i=1}^{m-1} \left(\frac{(x^m, y^i)}{(y^i, y^i)} y^i, y^k \right) \\ &= (x^m, y^k) - \sum_{i=1}^{m-1} \frac{(x^m, y^i)}{(y^i, y^i)} (y^i, y^k) \\ &= (x^m, y^k) - \frac{(x^m, y^k)}{(y^k, y^k)} (y^k, y^k) \\ &= 0, \end{aligned}$$

so y^m is orthogonal to every y^1, \dots, y^{m-1} . As an exercise verify that the span of y^1, \dots, y^m is the same as the span of x^1, \dots, x^m . ■

23 Corollary *Every nontrivial finite dimensional subspace of an inner product space has an orthonormal basis.*

Proof: Apply the Gram–Schmidt procedure to a basis. To obtain an orthonormal basis, simply normalize each y^k by dividing by its norm $(y^k, y^k)^{\frac{1}{2}}$. ■

24 Orthogonal Projection Theorem *Let M be a linear subspace of the real inner product space V . For each $x \in V$ we can write x in a unique way as $x = x_M + x_{\perp}$, where $x_M \in M$ and $x_{\perp} \in M_{\perp}$.*

Proof: Let y^1, \dots, y^m be an orthonormal basis for M . Put $z^i = (x, y^i)y^i$ for $i = 1, \dots, m$. Put $x_M = \sum_{i=1}^m z^i$, and $x_{\perp} = x - x_M$. Let $y \in M$, $y = \sum_{i=1}^m \alpha_i y^i$. Then

$$\begin{aligned} (y, x_{\perp}) &= \left(\sum_{i=1}^m \alpha_i y^i, x - \sum_{j=1}^m (x, y^j) y^j \right) \\ &= \sum_{i=1}^m \alpha_i \left(y^i, x - \sum_{j=1}^m (x, y^j) y^j \right) \\ &= \sum_{i=1}^m \alpha_i \left\{ (y^i, x) - (y^i, \sum_{j=1}^m (x, y^j) y^j) \right\} \\ &= \sum_{i=1}^m \alpha_i \left\{ (y^i, x) - \sum_{j=1}^m (x, y^j) (y^j, y^j) \right\} \\ &= \sum_{i=1}^m \alpha_i \{ (y^i, x) - (x, y^i) \} \\ &= 0. \end{aligned}$$

Uniqueness: Let $x = x_M + x_\perp = z_M + z_\perp$. Then $0 = \underbrace{(x_M - z_M)}_{\in M} + \underbrace{(x_\perp - z_\perp)}_{\in M_\perp}$. Rewriting this yields $x_M - z_M = 0 - (x_\perp - z_\perp)$, so $x_M - z_M$ also lies in M_\perp . Similarly $x_\perp - z_\perp$ also lies in M . Thus $x_M - z_M \in M \cap M_\perp = \{0\}$ and $x_\perp - z_\perp \in M \cap M_\perp = \{0\}$. ■

25 Definition The vector x_M given by the theorem above is called the **orthogonal projection of x on M** .

26 Corollary For any subspace M of V , $\dim M + \dim M_\perp = \dim V$.

There is another important characterization of orthogonal projection.

27 Proposition (Orthogonal projection minimizes the norm of the “residual”) Let M be a linear subspace of the real inner product space V . Let $y \in V$. Then

$$\|y - y_M\| \leq \|y - x\| \quad \text{for all } x \in M.$$

Proof: Let $x \in M$ and note

$$\begin{aligned} \|y - x\|^2 &= \|(y - y_M) + (y_M - x)\|^2 \\ &= \left((y - y_M) + (y_M - x), (y - y_M) + (y_M - x) \right) \\ &= (y - y_M, y - y_M) + 2(y - y_M, y_M - x) + (y_M - x, y_M - x) \end{aligned}$$

but $y - y_M = y_\perp \perp M$, and $y_M - x \in M$, so $(y - y_M, y_M - x) = 0$, so

$$\begin{aligned} &= \|y - y_M\|^2 + \|y_M - x\|^2 \\ &\geq \|y - y_M\|^2. \end{aligned}$$

That is, $x = y_M$ minimizes $\|y - x\|$ over M . ■

28 Proposition (Linearity of Projection) The orthogonal projection satisfies

$$(x + z)_M = x_M + z_M \quad \text{and} \quad (\alpha x)_M = \alpha x_M.$$

Proof: Use $x_M = \sum_{j=1}^k (x, y^j) y^j$ and $z_M = \sum_{j=1}^k (z, y^j) y^j$. Then

$$(x + z)_M = \sum_{j=1}^k (x + z, y^j) y^j.$$

Use linearity of (\cdot, \cdot) . ■

We now present a lemma about linear functions that is true in quite general linear spaces, see [2, Theorem 5.91, p. 212], but we will prove it using some of the special properties of inner products.

29 Lemma Let V be an inner product space. Then y is a linear combination of x^1, \dots, x^m if and only if

$$\bigcap_{i=1}^m \{x^i\}_\perp \subset \{y\}_\perp. \quad (1)$$

Proof: If y is a linear combination of x^1, \dots, x^m , say $y = \sum_{i=1}^m \alpha_i x^i$, then

$$(z, y) = \sum_{i=1}^m \alpha_i (z, x^i),$$

so (1) holds.

For the converse, suppose (1) holds. Let $M = \text{span}\{x^1, \dots, x^m\}$ and orthogonally project y on M to get $y = y_M + y_\perp$, where $y_M \in M$ and $y_\perp \perp M$. In particular, $(x^i, y_\perp) = 0$, $i = 1, \dots, m$. Consequently, by hypothesis, $(y, y_\perp) = 0$ too. But

$$0 = (y, y_\perp) = (y_M, y_\perp) + (y_\perp, y_\perp) = 0 + (y_\perp, y_\perp).$$

Thus $y_\perp = 0$, so $y = y_M \in M$. That is, y is a linear combination of x^1, \dots, x^m . ■

We can rephrase the above result as an alternative.

30 Corollary (Fredholm alternative) *Either there exist real numbers a_1, \dots, a_m such that*

$$y = \sum_{i=1}^m \alpha_i x^i$$

or else there exists a vector z satisfying

$$(z, x^i) = 0, \quad i = 1, \dots, m \quad \text{and} \quad (z, y) = 1.$$

3 Linear transformations

Recall that a linear transformation is a function $T: V \rightarrow W$ between vector spaces satisfying $T(\alpha x + \beta y) = \alpha T x + \beta T y$.

There are three important classes of linear transformations of a space into itself. The first is just a rescaling along various dimensions. The identity function is of this class, where the scale factor is one in each dimension. Another important example is orthogonal projection onto a linear subspace. And another important class is rotation about an axis. These do not cover all the linear transformation, since for instance the composition of two of these kinds of transformations does not belong to any of these groups.

31 Definition *For a linear transformation $T: V \rightarrow W$, the **null space** or **kernel** of T is the inverse image of 0. That is, $\{x \in V : T x = 0\}$.*

32 Proposition *The null space of T , $\{x : T x = 0\}$, is a linear subspace of V , and the range of T , $\{T x : x \in V\}$, is a linear subspace of W .*

33 Proposition *A linear transformation T is one-to-one if and only if its null space is $\{0\}$. In other words, T is one-to-one if and only if $T x = 0$ implies $x = 0$.*

The dimension of the range of T is called the **rank** of T . The dimension of the null space of T is called the **nullity** of T . The next result may be found in [3, Theorem 2.3, p. 34]. You can prove it using Corollary 26 applied to the null space of T .

34 Nullity Plus Rank Theorem *Let $T: V \rightarrow W$, where V and W are finite dimensional inner product spaces. Then*

$$\text{rank } T + \text{nullity } T = \dim V.$$

Proof: Let N denote the null space of T . For x in V , decompose it orthogonally as $x = x_N + x_\perp$, where $x_N \in N$ and $x_\perp \in N_\perp$. Then $Tx = Tx_\perp$, so the range of T is just $T(N_\perp)$. Now let x^1, \dots, x^k be a basis for N_\perp . I claim that Tx^1, \dots, Tx^k are independent and therefore constitute a basis for the range of T . So suppose some linear combination $\sum_{i=1}^k \alpha_i Tx^i$ is zero. By linearity of T we have $T\left(\sum_{i=1}^k \alpha_i x^i\right) = 0$. Which implies $\sum_{i=1}^k \alpha_i x^i$ belongs to the null space N . But this combination also lies in N_\perp , so it must be zero. But since the x^i s are independent, it follows that $\alpha_1 = \dots = \alpha_k = 0$. ■

3.1 Inverse of a linear transformation

35 Corollary *Let $T: V \rightarrow W$ be a linear transformation between m -dimensional spaces. Then T is one-to-one if and only if T is onto.*

Proof: Since T has rank m if and only if its range is all of W . Then by the Nullity Plus Rank Theorem its null space contains only 0. Suppose $Tx = Ty$. Then $T(x - y) = 0$, so $x - y = 0$, which implies T is one-to-one. Therefore T has an inverse. ■

36 Proposition *Let $T: V \rightarrow W$ be linear, one-to-one, and onto, where V and W are finite dimensional inner product spaces. Then $\dim V = \dim W$, T has full rank, the inverse T^{-1} exists and is linear.*

Proof: First we show that T^{-1} is linear: Let $x^i = T^{-1}(u^i)$, $i = 1, 2$. Now

$$T(\alpha x^1 + \beta x^2) = \alpha Tx^1 + \beta Tx^2 = \alpha u^1 + \beta u^2.$$

So taking T^{-1} of both sides gives

$$\alpha \underbrace{T^{-1}(u^1)}_{x^1} + \beta \underbrace{T^{-1}(u^2)}_{x^2} = T^{-1}(\alpha u^1 + \beta u^2)$$

Next we show that $\dim W \leq \dim V$: Let y^1, \dots, y^n be a basis for V . Then Ty^1, \dots, Ty^n span W since T is linear and onto. Therefore $\dim W \leq n = \dim V$.

On the other hand, Ty^1, \dots, Ty^n are linearly independent, so $\dim W \geq \dim V$. To see this, suppose

$$0 = \sum_{i=1}^n \xi_i T e^i = T\left(\sum_{i=1}^n \xi_i e^i\right)$$

But $T0 = 0$, too, and since T is one-to-one, we see $\sum_{i=1}^n \xi_i y^i = 0$. But this implies $\xi_1 = \dots = \xi_n = 0$.

This shows that $\dim V = \dim W$, and since T is onto it has rank $\dim W$. ■

3.2 Transpose of a transformation

Let $T: V \rightarrow W$ be linear where V and W are inner product spaces. The **transpose**, or **adjoint**,¹ of T , denoted T' , is the linear transformation $T': W \rightarrow V$ satisfying

$$(y, Tx) = (T'y, x) \tag{2}$$

for every x in V and every y in W .

A natural question is, how do we know that such a linear transformation T' exists? If T' does exist, then equation (2) implies that $(T'y)_i = (Ty, e^i)$ must be equal to (y, Te^i) . I leave it to you to verify that T' defined by $(T'y)_i = (y, Te^i)$ is the transformation we seek. By the way, Lemma 13 guarantees that T' is uniquely defined by equation (2).

¹Change basis.

37 Proposition Let $T: V \rightarrow W$ and $S: W \rightarrow X$. Then

$$\begin{aligned}(TS)' &= S'T' \\ (S + T)' &= S' + T' \\ (T')' &= T\end{aligned}$$

38 Theorem Let $T: V \rightarrow W$, where V and W are inner product spaces, so $T': W \rightarrow V$. Then $T'y = 0 \iff y \perp \text{range } T$. In other words,

$$\text{null space } T' = (\text{range } T)_{\perp}.$$

Proof: (\implies) If $T'y = 0$, then $(T'y, x) = (0, x) = 0$ for all x in V . But $(T'y, x) = (y, Tx)$, so $(y, Tx) = 0$ for all x in V . That is, $y \perp \text{range } T$.

(\impliedby) If $y \perp \text{range } T$, then $(y, Tx) = 0$ for all x in V . Therefore $(T'y, x) = 0$ for all x in V , so $T'y = 0$. ■

39 Corollary Let $T: V \rightarrow W$, where V and W are inner product spaces. Then $\text{range } T'T = \text{range } T'$. Consequently, $\text{rank } T' = \text{rank } T'T$.

Proof: Clearly $\text{range } T'T \subset \text{range } T'$.

Let x belong to the range of T' , so $x = T'y$ for some y in W . Let M denote the range of T and consider the orthogonal decomposition $y = y_M + y_{\perp}$. Then $T'y = T'y_M + T'y_{\perp}$, but $T'y_{\perp} = 0$ by Theorem 38. Now $y_M = Tz$ some $z \in W$. Then $x = T'Tz$, so x belongs to the range of $T'T$. ■

40 Corollary Let $T: V \rightarrow W$, where V and W are finite dimensional inner product spaces. Then $\text{rank } T' = \text{rank } T$

Proof: By Theorem 38, $\text{null space } T' = (\text{range } T)_{\perp} \subset W$. Therefore

$$\dim W - \text{rank } T' = \text{nullity } T' = \dim(\text{range } T)_{\perp} = \dim W - \text{rank } T,$$

where the first equality follows from the Nullity Plus Rank Theorem 34, the second from Theorem 38, and the third from Corollary 26. Therefore, $\text{rank } T = \text{rank } T'$. ■

41 Corollary Let $T: \mathbf{R}^n \rightarrow \mathbf{R}^m$. Then $\text{null space } T = \text{null space } T'T$.

Proof: Clearly $\text{null space } T \subset \text{null space } T'T$, since $Tx = 0 \implies T'Tx = T'0 = 0$.

Now suppose $T'Tx = 0$. Then $(x, T'Tx) = (x, 0) = 0$. But

$$(x, T'Tx) = (T'Tx, x) = (Tx, Tx),$$

where the first equality is IP.1 and the second is the definition of T' , so $Tx = 0$. ■

42 Proposition (Summary) For a linear transformation T between finite dimensional spaces, $\text{range } T'T = \text{range } T'$ and $\text{range } TT' = \text{range } T$, so

$$\text{rank } T = \text{rank } T' = \text{rank } T'T = \text{rank } TT'.$$

¹When dealing with complex vector spaces, the definition of the adjoint is modified to $(y, Tx) = \overline{(T'y, x)}$.

43 Definition A transformation $T: \mathbf{R}^m \rightarrow \mathbf{R}^m$ is **symmetric** or **self-adjoint** if $T' = T$. A transformation T is **skew-symmetric** if $T' = -T$.

44 Proposition Let $\pi_M: \mathbf{R}^m \rightarrow \mathbf{R}^m$ be the orthogonal projection on M . Then π_M is symmetric.

Proof: To show $(x, \pi_M z) = (\pi_M x, z)$. Observe that

$$(x, z_M) = (x_M + x_\perp, z_M) = (x_M, z_M) + (x_\perp, z_M) = (x_M, z_M) + 0.$$

Similarly,

$$(x_M, z) = (x_M, z_M + z_\perp) = (x_M, z_M) + (x_M, z_\perp) = (x_M, z_M) + 0.$$

Therefore $(x, z_M) = (x_M, z) = (x_M, z_M)$. ■

In light of the following exercise, a linear transformation is an orthogonal projection if and only if it is symmetric and idempotent. A transformation T is **idempotent** if $T^2 x = TT x = T x$ for all x .

45 Exercise Let $P: \mathbf{R}^m \rightarrow \mathbf{R}^m$ be a linear transformation satisfying

$$P' = P \quad \text{and} \quad P^2 x = P x \quad \text{for all } x \in \mathbf{R}^m.$$

That is, P is idempotent and symmetric. Set $M = I - P$ (where I is the identity on \mathbf{R}^m). Prove the following.

1. For any x , $x = Mx + Px$.
2. $M^2 x = Mx$ for all x and $M' = M$.
3. null space $P = (\text{range } P)_\perp = \text{range } M$ and null space $M = (\text{range } M)_\perp = \text{range } P$
4. P is the orthogonal projection onto its range. Likewise for M .

3.3 Eigenvalues and eigenvectors

46 Definition Let $T: \mathbf{R}^m \rightarrow \mathbf{R}^m$ be linear. A scalar λ is an **eigenvalue** of T if there is a nonzero vector x in \mathbf{R}^m such that $Tx = \lambda x$. The vector x is called an **eigenvector** of T associated with λ . Note that the vector 0 is by definition not an eigenvector of T .

If T has an eigenvalue λ with eigenvector x , the transformation “stretches” the space by a factor λ in the direction x .

While the vector 0 is never an eigenvector, the scalar 0 may be an eigenvalue. Indeed 0 is the eigenvalue associated with any nonzero vector in the null space of T .

There are linear transformations with no eigenvalues. For instance, consider the rotation of \mathbf{R}^2 by ninety degrees. This is given by the transformation $(x, y) \mapsto (-y, x)$. In order to satisfy $Tx = \lambda x$ we must have $\lambda x = -y$ and $\lambda y = x$. This cannot happen for any nonzero (x, y) and real λ .

On the other hand, the identity transformation has an eigenvalue 1 , associated with every nonzero vector.

Observe that there is a unique eigenvalue associated with each eigenvector: If $Tx = \lambda x$ and $Tx = \alpha x$, then $\alpha x = \lambda x$, so $\alpha = \lambda$, since by definition x is nonzero.

On the other hand, one eigenvalue must be associated with many eigenvectors, for if x is an eigenvector associated with λ , so is any nonzero scalar multiple of x . More generally, a linear combination of eigenvectors corresponding to an eigenvalue is also an eigenvector corresponding

to the same eigenvalue (provided the linear combination does not equal the zero vector). The span of the set of eigenvectors associated with the eigenvalue λ is called the **eigenspace** of T corresponding to λ . Every nonzero vector in the eigenspace is an eigenvector associated with λ . The dimension of the eigenspace is called the **multiplicity** of λ .

47 Proposition *If $T: V \rightarrow V$ is idempotent, then each of its eigenvalues is either 0 or 1.*

Proof: Suppose $Tx = \lambda x$ with $x \neq 0$. Since T is idempotent, we have $\lambda x = Tx = T^2x = \lambda^2x$. Since $x \neq 0$, this implies $\lambda = \lambda^2$, so $\lambda = 0$ or $\lambda = 1$. ■

For distinct eigenvalues we have the following, taken from [3, Theorem 4.1, p. 100].

48 Theorem *Let x^1, \dots, x^n be eigenvectors associated with distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Then the vectors x^1, \dots, x^n are independent.*

Proof: The proof is by induction n . The case $n = 1$ is trivial, since by definition eigenvectors are nonzero. Now consider $n > 1$ and suppose that the result is true for $n - 1$. Now let

$$\sum_{i=1}^n \alpha_i x^i = 0. \quad (3)$$

Applying the transformation T to both sides gives

$$\sum_{i=1}^n \alpha_i \lambda_i x^i = 0. \quad (4)$$

Let us eliminate x^n from (F4) by multiplying (3) by λ_n and subtracting to get

$$\sum_{i=1}^n \alpha_i (\lambda_i - \lambda_n) x^i = \sum_{i=1}^{n-1} \alpha_i (\lambda_i - \lambda_n) x^i = 0.$$

But x^1, \dots, x^{n-1} are independent so by the induction hypothesis $\alpha_i (\lambda_i - \lambda_n) = 0$ for $i = 1, \dots, n - 1$. Since the eigenvalues are distinct, this implies each $\alpha_i = 0$ for $i = 1, \dots, n - 1$. Thus (3) reduces to $\alpha_n x^n = 0$, which implies $\alpha_n = 0$. This shows that x^1, \dots, x^n are independent. ■

49 Corollary *A linear transformation on an n -dimensional space has at most n distinct eigenvalues. If it has n , then the space has a basis made up of eigenvectors.*

When T is symmetric, not only are eigenvectors associated with distinct eigenvalues independent, they are orthogonal.

50 Proposition *Let T be symmetric. Let x and y be eigenvectors of T corresponding to eigenvalues α and λ with $\alpha \neq \lambda$. Then $x \perp y$.*

Proof: We are given $Tx = \alpha x$ and $Ty = \lambda y$. Thus $(Tx, y) = (\lambda x, y) = \lambda(x, y)$ and $(x, Ty) = (x, \lambda y) = \lambda(x, y)$. Since T is symmetric, $(Tx, y) = (x, Ty)$, so $\alpha(x, y) = \lambda(x, y)$. Since $\lambda \neq \alpha$ we must then have $(x, y) = 0$. ■

Also if T is symmetric, we are guaranteed that it has plenty of eigenvectors.

51 Theorem *Let $T: V \rightarrow V$ be symmetric, where V is an n -dimensional inner product space. Then V has an orthonormal basis consisting of eigenvectors of T .*



Proof: This proof uses some well known results from topology and calculus that are beyond the scope of these notes.

Let $S = \{x \in V : (x, x) = 1\}$ denote the unit sphere in V . Set $M_0 = \{0\}$ and define $S_0 = S \cap M_{0\perp}$. Define the *quadratic form* $Q: V \rightarrow \mathbf{R}$ by

$$Q(x) = (x, Tx).$$

It is easy to see that Q is continuous, so it has a maximizer on S_0 , which is compact. (This maximizer cannot be unique, since $Q(-x) = Q(x)$, and indeed if T is the identity, then Q is constant on S .) Fix a maximizer x^1 of Q over S_0 .

Proceed recursively for $k = 1, \dots, n-1$. Let M_k denote the span of x^1, \dots, x^k , and set $S_k = S \cap M_{k\perp}$. Let x^{k+1} maximize Q over S_k . By construction, $x^{k+1} \in M_{k\perp}$, so the x^k 's are orthogonal, indeed orthonormal.

Since x^1 maximizes Q on $S = S_0$, it maximizes Q subject to the constraint $1 - (x, x) = 0$. Now $Q(x) = (x, Tx)$ is continuously differentiable and $Q'(x) = 2Tx$, and the gradient of the constraint function is $-2x$, which is clearly nonzero (hence linearly independent) on S . It is a nuisance to have these 2s popping up, so let us agree to maximize $\frac{1}{2}(x, Tx)$ subject $\frac{1}{2}(1 - (x, x)) = 0$ instead. Therefore by the well known Lagrange Multiplier Theorem, there exists λ_1 satisfying

$$Tx^1 - \lambda_1 x^1 = 0.$$

This obviously implies that the Lagrange multiplier λ_1 is an eigenvalue of T and x^1 is a corresponding eigenvector. Further, it is the value of the maximum:

$$Q(x^1) = (x^1, Tx^1) = (x^1, \lambda_1 x^1) = \lambda_1,$$

since $(x^1, x^1) = 1$.

We now proceed by induction. Let x^1, \dots, x^n be recursively defined as above and assume that for $i = 1, \dots, k$, each x^i is an eigenvector of A and that $\lambda_i = Q(x^i)$ is its corresponding eigenvalue. We wish to show that x^{k+1} is an eigenvector of T and $\lambda_{k+1} = Q(x^{k+1})$ is its corresponding eigenvalue.

By construction, x^{k+1} maximizes $\frac{1}{2}Q(x)$ subject to the $k + 1$ constraints $\frac{1}{2}(1 - (x, x)) = 0$, $(x, x^1) = 0, \dots, (x, x^k) = 0$. The gradients of these constraint functions are $-x$ and x^1, \dots, x^k respectively. By construction, x^1, \dots, x^{k+1} are orthonormal, so at x^{k+1} the gradients of the constraint are linearly independent. Therefore there exist Lagrange multipliers λ_{k+1} and μ_1, \dots, μ_k satisfying

$$Tx^{k+1} - \lambda_{k+1}x^{k+1} + \mu_1x^1 + \dots + \mu_kx^k = 0. \tag{5}$$

Therefore

$$Q(x^{k+1}) = (x^{k+1}, Tx^{k+1}) = \lambda_{k+1}(x^{k+1}, x^{k+1}) - \mu_1(x^{k+1}, x^1) - \dots - \mu_k(x^{k+1}, x^k) = \lambda_{k+1},$$

since x^1, \dots, x^{k+1} are orthonormal. That is, the multiplier λ_{k+1} is the maximum value of Q over S_k .

By the induction hypothesis, $Tx^i = \lambda_i x^i$ for $i = 1, \dots, k$. Then since T is symmetric,

$$(x^i, Ax^{k+1}) = (x^{k+1}, Tx^i) = (x^{k+1}, \lambda_i x^i) = 0, \quad i = 1, \dots, k.$$

That is, $x^{k+1} \in M_{k\perp}$, so $Tx^{k+1} - \lambda_{k+1}x^{k+1} \in M_{k\perp}$, so by Lemma 20 equation (5) implies

$$Tx^{k+1} - \lambda_{k+1}x^{k+1} = 0 \quad \text{and} \quad \mu_1x^1 + \dots + \mu_kx^k = 0.$$

We conclude therefore that $Tx^{k+1} = \lambda_{k+1}x^{k+1}$, so that x^{k+1} is an eigenvector of T and λ_{k+1} is the corresponding eigenvalue.

Since V is n -dimensional, x^1, \dots, x^n is an orthonormal basis of eigenvectors, as we desire. ■

4 Matrices

A **matrix** is merely a rectangular array of numbers, or equivalently, a doubly indexed ordered list of real numbers. An $m \times n$ matrix has m rows and n columns. The entries in a matrix are doubly indexed, with the first index denoting its row and the second its column. Here is a generic matrix:

$$A = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{m,1} & \cdots & \alpha_{m,n} \end{bmatrix}$$

The generally accepted plural of *matrix* is *matrices*, although *matrixes* is okay too. Matrices are of interest for two separate but hopelessly intertwined reasons. One is their relation to systems of linear equations and inequalities, and the other is their connection to linear transformations between finite dimensional vector spaces.

Given a matrix A , let A_i denote the the i^{th} row of A and let A^j denote the j^{th} column. The i^{th} row and j^{th} column entry is generally denoted by a lower case roman or Greek letter, e.g., $a_{i,j}$ or $\alpha_{i,j}$. We can identify the rows or columns of a matrix with singly indexed lists of real numbers, that is, elements of some \mathbf{R}^k . If A is $m \times n$, the **column space** of A is the subset of \mathbf{R}^m spanned by the n columns of A . Its **row space** is the subspace of \mathbf{R}^n spanned by its m rows.

4.1 Matrix operations

If both A and B are $m \times n$ matrices, the **sum** $A + B$ is the $m \times n$ matrix C defined by

$$c_{i,j} = a_{i,j} + b_{i,j}.$$

The scalar multiple αA of a matrix A by a scalar α is the matrix M defined by $m_{i,j} = \alpha a_{i,j}$. Under these operations the set of $m \times n$ matrices becomes an mn -dimensional vector space.

If A is $m \times p$ and B is $p \times n$, the **product** of A and B is the $m \times n$ matrix whose i^{th} row, j^{th} column entry is the dot product $A_i \cdot B^j$ of the i^{th} row of A with the j^{th} column of B .

$$(AB)_{i,j} = A_i \cdot B^j$$

The reason for this peculiar definition is explained in the next section.

Vectors in \mathbf{R}^n may also be considered matrices as either a column or a row vector. Let x be a row m -vector (a $1 \times m$ matrix), y be a column n -vector (an $n \times 1$ matrix), and let A be an $m \times n$ matrix. Then the matrix product xA considered as a vector in \mathbf{R}^n belongs to the row space of A , and Ay as a vector in \mathbf{R}^m belongs to the column space of A .

$$xA = \sum_{i=1}^m x_i A_i \quad \text{and} \quad Ay = \sum_{j=1}^n y_j A^j$$

Note that the i^{th} row of AB is given by

$$(AB)_i = (A_i)B$$

and the j^{th} column of AB is given by

$$(AB)^j = A(B^j).$$

The **main diagonal** of a square matrix $A = [a_{i,j}]$ is the set of $a_{i,j}$ with $i = j$. A matrix is called a **diagonal matrix** if it is square and all its nonzero entries are on the main diagonal.

A square matrix is **upper triangular** if the only nonzero elements are on or above the main diagonal, that is, if $i > j$ implies $a_{i,j} = 0$. A square matrix is **lower triangular** if $i < j$ implies $a_{i,j} = 0$.

The $n \times n$ diagonal matrix I whose diagonal consists of all 1s has the property that

$$AI = IA = A$$

for any $n \times n$ matrix A . The matrix I is called the **$n \times n$ identity matrix**. The **zero matrix** has all its entries zero, and satisfies $A + 0 = A$.

A **left inverse** for the $n \times n$ matrix A is an $n \times n$ matrix L for which $LA = I$, and a **right inverse** R satisfies $AR = I$.

52 Lemma (Left vs. right inverses) *If A has both a left and right inverse, then they are unique and equal.*

Proof: Let L be a left inverse and R be a right inverse for A , so $LA = I = AR$. Then

$$L = L \underbrace{(AR)}_{=I} = \underbrace{(LA)}_{=I} R = R.$$

Thus every left inverse is a right inverse and vice versa, which by the way shows uniqueness. ■

What the above lemma leaves out is a proof that if A has either a right or left inverse, then it has both. That will have to wait until Theorem 86, where the inverse is computed.

53 Fact (Summary) *Direct computation reveals the following facts.*

$$\begin{aligned} (AB)C &= A(BC) \\ AB &\neq BA \quad (\text{in general}) \\ A(B + C) &= (AB) + (AC) \\ (A + B)C &= AC + BC \end{aligned}$$

54 Exercise *Verify the following.*

1. *The product of upper triangular matrices is upper triangular.*
2. *The product of lower triangular matrices is lower triangular.*
3. *The product of diagonal matrices is diagonal.*
4. *The inverse of an upper triangular matrix is upper triangular (if it exists).*
5. *The inverse of a lower triangular matrix is lower triangular (if it exists).*
6. *The inverse of a diagonal matrix is diagonal (if it exists).*

4.2 Systems of linear equations

One of the main uses of matrices is the simplification of representing a system of linear equations. For instance, consider the following system of equations.

$$\begin{aligned} 3x_1 + 2x_2 &= 8 \\ 2x_1 + 3x_2 &= 7 \end{aligned}$$

It has the unique solution $x_1 = 2$ and $x_2 = 1$. One way to solve this is to take the second equation and solve for $x_1 = \frac{7}{2} - \frac{3}{2}x_2$ and substitute this into the first equation to get $3(\frac{7}{2} - \frac{3}{2}x_2) + 2x_2 = 8$, so $x_2 = 1$ and $x_1 = \frac{7}{2} - \frac{3}{2} = 2$. However Gauss formalized a computationally efficient way to attack these problems using **elementary row operations**. The first step is to write down the so-called augmented coefficient matrix of the system, which is the 2×3 matrix of just the numbers above:

$$\begin{bmatrix} 3 & 2 & 8 \\ 2 & 3 & 7 \end{bmatrix}.$$

There are three elementary row operations, and they correspond to steps used to solve a system of equations. One of these operations is to interchange two rows. We won't use that here. Another is to multiply a row by a nonzero scalar. This does not change the solution. The third operation is to add one row to another. These last two operations can be combined, and we can think of adding a scalar multiple of one row to another as an elementary row operation. We apply these operations until we get a matrix of the form

$$\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \end{bmatrix}$$

which is the augmented matrix of the system

$$\begin{aligned} x_1 &= a \\ x_2 &= b \end{aligned}$$

and the system is solved. There is a simple algorithm for deciding which elementary row operations to apply, namely, the **Gaussian elimination algorithm**. In a section below, we shall go into this algorithm in detail, but let me just give you a hint here. First we multiply the first row by $\frac{1}{3}$, to get a leading 1:

$$\begin{bmatrix} 1 & \frac{2}{3} & \frac{8}{3} \\ 2 & 3 & 7 \end{bmatrix}$$

We want to eliminate x_1 from the second equation, so we add an appropriate multiple of the first row to the second. In this case the multiple is -2 , the result is:

$$\begin{bmatrix} 1 & \frac{2}{3} & \frac{8}{3} \\ 2 - 2 \cdot 1 & 3 - 2 \cdot \frac{2}{3} & 7 - 2 \cdot \frac{8}{3} \end{bmatrix} = \begin{bmatrix} 1 & \frac{2}{3} & \frac{8}{3} \\ 0 & \frac{5}{3} & \frac{5}{3} \end{bmatrix}.$$

Now multiply the second row by $\frac{3}{5}$ to get

$$\begin{bmatrix} 1 & \frac{2}{3} & \frac{8}{3} \\ 0 & 1 & 1 \end{bmatrix}.$$

Finally to eliminate x_2 from the first row we add $-\frac{2}{3}$ times the second row to the first and get

$$\begin{bmatrix} 1 - \frac{2}{3} \cdot 0 & \frac{2}{3} - \frac{2}{3} \cdot 1 & \frac{8}{3} - \frac{2}{3} \cdot 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix},$$

which accords with our earlier result.

4.3 Matrix representation of a linear transformation

Let T be a linear transformation from the n -dimensional space V into the m -dimensional space W . Let x^1, \dots, x^n be an ordered basis for V and y^1, \dots, y^m be an ordered basis for W . Let

$$\begin{aligned} Tx^1 &= \sum_{i=1}^m \alpha_{i,1} y^i \\ Tx^2 &= \sum_{i=1}^m \alpha_{i,2} y^i \\ &\vdots \\ Tx^n &= \sum_{i=1}^m \alpha_{i,n} y^i. \end{aligned}$$

The $m \times n$ array of numbers

$$M(T) = \begin{bmatrix} \alpha_{1,1} & \dots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{m,1} & \dots & \alpha_{m,n} \end{bmatrix}$$

is the **matrix representation of T with respect to the ordered bases $(x^j), (y^i)$** . Note that the j^{th} column of this matrix is the coordinate vector of Tx^j with respect to the ordered basis y^1, \dots, y^m .

Given the coordinate vector of an element x of V , we can use matrix multiplication to compute the coordinates of Tx . Let $x = \sum_{j=1}^n \xi_j x^j$. Then

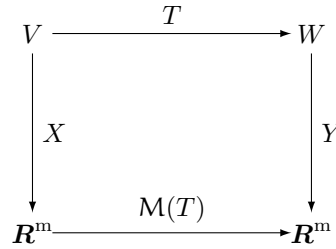
$$\begin{aligned} Tx &= \sum_{j=1}^n \xi_j Tx^j \\ &= \sum_{j=1}^n \xi_j \sum_{i=1}^m \alpha_{i,j} y^i \\ &= \sum_{i=1}^m \left(\sum_{j=1}^n \alpha_{i,j} \xi_j \right) y^i \end{aligned}$$

The coordinates of Tx with respect to (y^i) are given by

$$(Tx)_i = \sum_{j=1}^n \alpha_{i,j} \xi_j.$$

That is, the coordinate vector of Tx is $M(T)$ times the column vector of coordinates of x .

The matrix representation can be thought of in the following terms. Let X denote the coordinate mapping of V onto \mathbf{R}^n with respect to the ordered basis x^1, \dots, x^n . Similarly Y denotes the coordinate mapping from W onto \mathbf{R}^m . In order to compute the coordinates of Tz , we first find the coordinate vector of Xz , and then multiply by the matrix $M(T)$, as the following “commutative diagram” shows.



55 Example (The matrix of the coordinate mapping) Consider the case $V = \mathbf{R}^m$. That is, elements of V are already thought of as ordered lists of real numbers. Let x^1, \dots, x^m be an ordered basis, and let X denote the coordinate mapping from \mathbf{R}^m with this basis to \mathbf{R}^m with the standard basis. Then $M(X)$ is simply the matrix whose j^{th} column is x^j . \square

56 Example (Matrices as linear transformations) Let $A = [a_{i,j}]$ be an $m \times n$ array of numbers and

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

a $n \times 1$ array. Define Ax to be an $m \times 1$ array whose i^{th} row is

$$\sum_{j=1}^n a_{i,j}x_j \quad i = 1, \dots, m.$$

Then $T: x \mapsto Ax$ defines a linear transformation from \mathbf{R}^n to \mathbf{R}^m . The matrix $M(T)$ of this transformation with respect to the standard ordered bases (the bases of unit coordinate vectors) is just A . \square

57 Definition *The rank of a matrix is the largest number of linearly independent columns.*

It follows from Proposition 36 that:

58 Proposition *An $m \times m$ matrix has an inverse if and only if it has rank m .*

59 Example (The identity matrix) What is the matrix for the identity mapping $I: \mathbf{R}^m \rightarrow \mathbf{R}^m$?

$$A_I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix},$$

$m \times m$ identity matrix I . If the transformation T is invertible, so that $TT^{-1} = I$, then $A_T A_{T^{-1}} = I$. The matrix $A_{T^{-1}}$ is naturally referred to as the **inverse of the matrix A** . In general, if A defines an invertible transformation from \mathbf{R}^m onto \mathbf{R}^m , the matrix A^{-1} satisfies

$$AA^{-1} = A^{-1}A = I.$$

\square

60 Example (The zero matrix) The matrix for the zero transformation $0: \mathbf{R}^n \rightarrow \mathbf{R}^m$, is

$$\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix},$$

the $m \times n$ zero matrix. □

The **transpose of a matrix** is the matrix formed by interchanging rows and columns. This definition is justified by the following lemma.

61 Lemma $M(T)_{i,j} = M(T')_{j,i}$.

Proof:

$$M(T)_{i,j} = (e_m^i, T e_n^j) = (T' e_m^i, e_n^j) = (e_n^j, T' e_m^i) = M(T')_{j,i}$$

■

4.4 Matrix representation of a composition

Let S take $\mathbf{R}^p \rightarrow \mathbf{R}^n$ be linear with matrix $M(S) = [b_{j,k}]_{j=1,\dots,n}^{k=1,\dots,p}$. Let T take $\mathbf{R}^n \rightarrow \mathbf{R}^m$ be linear with matrix $M(T) = [a_{i,j}]_{i=1,\dots,m}^{j=1,\dots,n}$. Then $T \circ S: \mathbf{R}^p \rightarrow \mathbf{R}^m$ is linear. What is $M(TS)$?

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$. Then

$$Sx = \sum_{j=1}^n \left(\sum_{k=1}^p b_{j,k} x_k \right) e^j.$$

$$T(Sx) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{i,j} \left(\sum_{k=1}^p b_{j,k} x_k \right) \right) e^i.$$

Set

$$c_{i,k} = \sum_{j=1}^n a_{i,j} b_{j,k}.$$

Then

$$T(Sx) = \sum_{i=1}^m \left(\sum_{k=1}^p c_{i,k} x_k \right) e^i.$$

Thus $M(TS) = [c_{i,k}]_{i=1,\dots,m}^{k=1,\dots,p}$. This proves the following theorem.

62 Theorem $M(TS) = M(T)M(S)$.

Thus multiplication of matrices corresponds to composition of linear transformations.

Let $S, T: \mathbf{R}^n \rightarrow \mathbf{R}^m$ be linear transformations with matrices A, B . Then $S + T$ is linear. The matrix for $S + T$ is $A + B$ where $(A + B)_{i,j} = a_{i,j} + b_{i,j}$.

63 Exercise Let A be an $m \times n$ matrix with $m > n$. Suppose A has rank n . Prove that the matrix $A(A'A)^{-1}A'$ is the orthogonal projection onto the column space of A .

4.5 Change of basis

A linear transformation may have different matrix representations for different bases. Is there some way to tell if two matrices represent the same transformation?

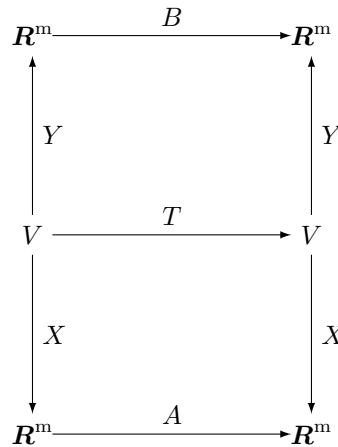
We shall answer this question for the important case where T maps an m -dimensional vector space V into itself, and we use the same basis for V as both the domain and the range. Let $A = [a_{i,j}]$ represent T with respect to the ordered basis x^1, \dots, x^m , and let $B = [b_{i,j}]$ represent T with respect to the ordered basis y^1, \dots, y^m .

That is,

$$Tx^i = \sum_{k=1}^m a_{k,i}x^k \quad \text{and} \quad Ty^i = \sum_{k=1}^m b_{k,i}y^k.$$

Let X be the coordinate map from V into \mathbf{R}^m with respect to the ordered basis x^1, \dots, x^m , and let Y be the coordinate map for the ordered basis y^1, \dots, y^m . Then X and Y have full rank and so have inverses.

Consider the following commutative diagram.



The mapping XY^{-1} from \mathbf{R}^m onto \mathbf{R}^m followed by A from \mathbf{R}^m into \mathbf{R}^m , which maps the upper left \mathbf{R}^m into the lower right \mathbf{R}^m , is the same as B followed by XY^{-1} . Let C be the matrix representation of XY^{-1} with respect to the standard ordered basis of unit coordinate vectors. Then

$$A = CBC^{-1} \quad \text{and} \quad B = C^{-1}AC. \tag{*}$$

64 Definition Two square matrices A and B are called **similar** if there is some nonsingular C such that $(*)$ holds.

So we have already proved half of the following. The second half is left for you. (See [3, Theorem 4.7, p. 110] if you get stuck.)

65 Theorem Two matrices are similar if and only if they represent the same linear transformation.

The following are corollaries, but have simple direct proofs.

66 Proposition If A, B are similar with $A = CBC^{-1}$, then λ is an eigenvalue of A if it is an eigenvalue of B . If x is an eigenvector of A , $C^{-1}x$ is an eigenvector of B .

Proof: Suppose x is an eigenvector of A , $Ax = \lambda x$. Let $y = C^{-1}x$. Since $A = CBC^{-1}$,

$$\lambda x = Ax = CBC^{-1}x = CBy.$$

Premultiplying by C^{-1} ,

$$\lambda y = \lambda C^{-1}x = C^{-1}CBy = By.$$

■

67 Proposition *If A and B are similar, then $\text{rank } A = \text{rank } B$.*

Proof: We prove $\text{rank } B \geq \text{rank } A$. Symmetry completes the argument. Let $z^1 \dots z^k$ be a basis for range A , $z^i = Ay^i$. Put $w^i = C^{-1}y^i$. Then the Bw^i 's are independent. To see this suppose

$$0 = \sum_{i=1}^k \alpha_i (Bw^i) = \sum_{i=1}^k \alpha_i C^{-1}ACC^{-1}y^i = \sum_{i=1}^k \alpha_i C^{-1}z^i = C^{-1} \left(\sum_{i=1}^k \alpha_i z^i \right).$$

Since C^{-1} is nonsingular, this implies $\sum_{i=1}^k \alpha_i z^i = 0$, which in turn implies $\alpha_i = 0$, $i = 1, \dots, k$.

■

The next result describes the diagonalization of a symmetric matrix.

68 Definition *A square matrix X is **orthogonal** if $X'X = I$, or equivalently $X' = X^{-1}$.*

69 Principal Axis Theorem *Let $A: \mathbf{R}^m \rightarrow \mathbf{R}^m$ be a symmetric matrix. Let x^1, \dots, x^m be an orthonormal basis for \mathbf{R}^m made up of eigenvectors of A , with corresponding eigenvalues $\lambda_1, \dots, \lambda_m$. Set*

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix},$$

and set $X = [x^1, \dots, x^m]$.

Then

$$A = X\Lambda X^{-1},$$

$$\Lambda = X^{-1}AX,$$

and X is orthogonal, that is,

$$X^{-1} = X'.$$

Proof: Now $X'X = I$ by orthonormality, so $X^{-1} = X'$. Pick any z and set $y = X^{-1}z$, so $z = Xy = \sum_{j=1}^m y_j x^j$. Then

$$\begin{aligned} Az &= \sum_{j=1}^m y_j Ax^j &= \sum_{j=1}^m y_j (\lambda_j x^j) \\ &= X\Lambda y \\ &= X\Lambda X^{-1}z. \end{aligned}$$

Since z is arbitrary $A = X\Lambda X^{-1}$.

■

4.6 Simultaneous diagonalization

70 Theorem (Simultaneous Diagonalization) Let A, B be symmetric $m \times m$ matrices.

Then $AB = BA$ if and only if there exists an orthonormal basis consisting of vectors that are eigenvectors of both A and B . Then letting X be the orthogonal matrix whose columns are the basis we have

$$\begin{aligned} A &= X\Lambda_A X^{-1} \\ B &= X\Lambda_B X^{-1}, \end{aligned}$$

where Λ_A and Λ_B are diagonal matrices of eigenvalues of A and B respectively.

Partial proof: (\Leftarrow)

$$AB = X\Lambda_A X^{-1} X\Lambda_B X^{-1} = X\Lambda_A \Lambda_B X^{-1} = X\Lambda_B \Lambda_A X^{-1} = X\Lambda_B X^{-1} X\Lambda_A X = BA,$$

since diagonal matrices commute.

(\Rightarrow) We shall prove the result for the special case where A has distinct eigenvalues. In this case, the eigenvectors associated with any eigenvalue are distinct up to scalar multiplication.

Let x be an eigenvector of A corresponding to eigenvalue λ . Suppose A and B commute. Then

$$A(Bx) = BAx = B(\lambda x) = \lambda(Bx).$$

This means that Bx too is an eigenvector of A corresponding to λ , provided $Bx \neq 0$. But as remarked above, this implies that Bx is a scalar multiple of x , so x is an eigenvector of B too. So let X be a matrix whose columns are a basis of orthonormal eigenvectors for both A and B . Then it follows that $A = X\Lambda_A X^{-1}$, where Λ_A is the diagonal matrix of eigenvalues, and similarly $B = X\Lambda_B X^{-1}$.

We now present a sketch of the proof for the general case. The crux of the proof is that in general, the eigenspace M associated with λ may be more than one dimensional, so it is harder to conclude that the eigenvectors of A and B are the same. To get around this problem observe that $B^2A = BBA = BAB = ABB = AB^2$, and in general, $B^nA = AB^n$, so that if $Ax = \lambda x$, then $A(B^n x) = B^n Ax = B^n \lambda x = \lambda(B^n x)$. That is, for every n , the vector $B^n x$ is also an eigenvector of A corresponding to λ . Since the eigenspace M associated with λ is finite dimensional, for some minimal k , the vectors $x, Bx, \dots, B^k x$ are dependent. That is, there are $\alpha_0, \dots, \alpha_k$, not all zero, with

$$\alpha_0 x + \alpha_1 Bx + \alpha_2 B^2 x + \dots + \alpha_k B^k x = 0.$$

Let $\mu_1 \dots \mu_k$ be the roots of the polynomial $\alpha_0 + \alpha_1 y + \alpha_2 y^2 + \dots + \alpha_k y^k$. Then

$$[(B - \mu_1 I)(B - \mu_2 I) \dots (B - \mu_k I)]x = 0.$$

Set $z = [(B - \mu_2 I)(B - \mu_3 I) \dots (B - \mu_k I)]x$

If k is minimal, then $z \neq 0$. (Even if coefficients are complex. Independence over real field implies independence over complex field. Just look at real and imaginary parts.) Therefore $(B - \mu_1 I)z = 0$

Claim: μ_1 and z are real.

Proof of claim: Let $\mu_1 = \alpha + i\beta$ and $z = x + iy$

$$B(x + iy) = (\alpha + i\beta)(x + iy)$$

Therefore $Bx = \alpha x - \beta y$ and $By = \beta x + \alpha y$ (equate real and imaginary parts). Thus $y'Bx = \alpha y'x - \beta y'y$ and by symmetry $y'Bx = x'B'y = \beta x'y + \alpha x'y$.

Now,

$$\alpha(y'x) - \beta(y'y) = \beta(x'x) + \alpha(x'y),$$

so $\beta = 0$ or $z = 0$, i.e., $x = 0$ and $y = 0$. But $z \neq 0$, so $\beta = 0$ and μ_1 is real. Similarly each μ^i is real. \square

Therefore z is a real linear combination of $B^n x$ (all eigenvectors of A) satisfying $(B - \mu_1 I)z = 0$. In other words, $Bz = \mu_1 z$, so z is an eigenvalue of both B and A !

We now consider the orthogonal complement of z in the eigenspace M to recursively construct a basis for M composed of eigenvectors for both A and B .

We must do this for each eigenvalue λ of A . More details are in Rao [10, Result (iii), pp. 41–42]. \blacksquare

4.7 Trace

Let A be an $m \times m$ matrix. The **trace** of A , denoted $\text{tr } A$, is defined by

$$\text{tr } A = \sum_{i=1}^m a_{ii}.$$

Clearly,

$$\begin{aligned} \text{tr}(A + B) &= \text{tr } A + \text{tr } B \\ \text{tr}(\alpha A) &= \alpha(\text{tr } A) \end{aligned}$$

Also, $\text{tr}(AB) = \text{tr}(BA)$:

$$\text{tr } AB = \sum_{i=1}^m \left(\sum_{j=1}^m \alpha_{i,j} \beta_{j,i} \right) = \sum_{j=1}^m \left(\sum_{i=1}^m \beta_{j,i} \alpha_{i,j} \right) = \text{tr } BA$$

Moreover, the trace of a matrix depends only on the linear transformation of \mathbf{R}^m into \mathbf{R}^m that it represents. In light of Theorem 65, it suffices to show that the traces of similar matrices are equal, which is the next proposition.

71 Proposition If $B = C^{-1}AC$, then $\text{tr } B = \text{tr } A$.

Proof: $\text{tr } B = \text{tr } C^{-1}AC = \text{tr } ACC^{-1} = \text{tr } A$ \blacksquare

Remarkably these conditions characterize the trace up to normalization.

72 Proposition There is exactly one linear functional $\ell: \mathbf{L}(\mathbf{R}^m, \mathbf{R}^m) \rightarrow \mathbf{R}$ satisfying

$$\ell(S \circ T) = \ell(T \circ S) \quad \text{for all } T, S \in \mathbf{L}(\mathbf{R}^m, \mathbf{R}^m),$$

and

$$\ell(I) = m.$$

For a proof, see Loomis and Sternberg [8, Theorem 5.1, p. 99].

73 Theorem If A is symmetric and idempotent, then $\text{tr } A = \text{rank } A$.

Proof: Since A is symmetric, $A = XBX^{-1}$ where $X = [x^1, \dots, x^m]$ is an orthogonal matrix whose columns are eigenvectors of A , and B is a diagonal matrix whose diagonal elements are 0 or 1.

Thus $\text{tr } B$ is the number of nonzero eigenvalues of A . Also $\text{rank } B$ is the number of nonzero diagonal elements. Thus $\text{tr } B = \text{rank } B$, but since A and B are similar, $\text{tr } A = \text{tr } B = \text{rank } B = \text{rank } A$. \blacksquare

5 Determinants

The main quick references here are Apostol [3, Chapter 3] and Dieudonné [4, Appendix A.6].

The main things to remember are:

- The determinant assigns a number to each square matrix A , denoted either $\det A$ or $|A|$. A matrix is **singular** if its determinant is zero, otherwise it is **nonsingular**. For an $n \times n$ identity matrix, $\det I = 1$.
- $\det(AB) = \det A \cdot \det B$.
- Thus a matrix has an inverse if and only if its determinant is nonzero.
- Multiplying a row or a column by a scalar multiplies the determinant by the same amount. Consequently for an $n \times n$ matrix A ,

$$\det(-A) = (-1)^n \det A.$$

- Also consequently, the determinant of a diagonal matrix is the product of its diagonal elements.
- Adding a multiple of one row to another does not change the determinant.
- Consequently, the determinant of an upper (or lower) triangular matrix is the product of its diagonal.
- Moreover if a square matrix A is block upper triangular, that is, of the form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

where B and D are square, then $\det A = \det B \cdot \det D$. Likewise for block lower triangular matrices.

- $\det A = \det A'$.
- The determinant can be defined recursively in terms of minors (determinants of submatrices).
- The inverse of a matrix can be computed in terms of these minors. The inverse of A is the transpose of its cofactor matrix divided by $\det A$.
- Cramer's Rule: If $Ax = b$ for a nonsingular matrix A , then

$$x_i = \frac{\det(A^1, \dots, A^{i-1}, b, A^{i+1}, \dots, A^n)}{\det A}.$$

- The determinant is the oriented volume of the "cube" formed by its columns.
- The determinant $\det(\lambda I - A)$, where A and I are $n \times n$ is an n^{th} degree polynomial, called the **characteristic polynomial** of A .
- A root (real or complex) of the characteristic polynomial of A is called a **characteristic root** of A . Characteristic roots that are real are also eigenvalues. If nonzero x belongs to the null space of $\lambda I - A$, then it is an eigenvector corresponding to the eigenvalue λ .
- If A has rank k then every minor of size greater than k has zero determinant and there is at least one minor of order k with nonzero determinant.

5.1 Determinants as multilinear forms

There are several ways to think about determinants. Perhaps the most useful is as an alternating multilinear n -form.²

A function $\varphi: \underbrace{\mathbf{R}^n \times \cdots \times \mathbf{R}^n}_{n \text{ copies}} \rightarrow \mathbf{R}$ is **multilinear** if it is linear in each variable separately.

That is, for each $i = 1, \dots, n$,

$$\begin{aligned} \varphi(x^1, \dots, x^{i-1}, \alpha x^i + \beta y^i, x^{i+1}, \dots, x^n) \\ = \alpha \varphi(x^1, \dots, x^{i-1}, x^i, x^{i+1}, \dots, x^n) + \beta \varphi(x^1, \dots, x^{i-1}, y^i, x^{i+1}, \dots, x^n). \end{aligned}$$

A consequence of this is that if any x^i is zero, then so is φ . That is,

$$\varphi(x^1, \dots, x^{i-1}, 0, x^{i+1}, \dots, x^n) = 0.$$

The multilinear function φ is **alternating** if $x^i = x^j = z$ for distinct i and j , then

$$\varphi(x^1, \dots, z, \dots, z, \dots, x^n) = 0.$$

The reason for this terminology is the following lemma.

74 Lemma *The multilinear function φ is alternating if and only if interchanging x^i and x^j changes the sign of φ , that is,*

$$\varphi(x^1, \dots, x^i, \dots, x^j, \dots, x^n) = -\varphi(x^1, \dots, x^j, \dots, x^i, \dots, x^n).$$

Proof: Suppose first that φ is alternating. Then

$$\begin{aligned} 0 &= \varphi(x^1, \dots, x^i + x^j, \dots, x^j + x^i, \dots, x^n) \\ &= \varphi(x^1, \dots, x^i, \dots, x^j + x^i, \dots, x^n) + \varphi(x^1, \dots, x^j, \dots, x^j + x^i, \dots, x^n) \\ &= \varphi(x^1, \dots, x^i, \dots, x^j, \dots, x^n) + \underbrace{\varphi(x^1, \dots, x^i, \dots, x^i, \dots, x^n)}_{=0} \\ &\quad + \underbrace{\varphi(x^1, \dots, x^j, \dots, x^j, \dots, x^n)}_{=0} + \varphi(x^1, \dots, x^j, \dots, x^i, \dots, x^n) \\ &= \varphi(x^1, \dots, x^i, \dots, x^j, \dots, x^n) + \varphi(x^1, \dots, x^j, \dots, x^i, \dots, x^n). \end{aligned}$$

So $\varphi(x^1, \dots, x^i, \dots, x^j, \dots, x^n) = -\varphi(x^1, \dots, x^j, \dots, x^i, \dots, x^n)$.

Now suppose interchanging x^i and x^j changes the sign of φ . Then if $x^i = x^j = z$,

$$\varphi(x^1, \dots, z, \dots, z, \dots, x^n) = -\varphi(x^1, \dots, z, \dots, z, \dots, x^n),$$

which implies $\varphi(x^1, \dots, z, \dots, z, \dots, x^n) = 0$, so φ is alternating. ■

There is an obvious identification of $n \times n$ square matrices with the elements of $\mathbf{R}^n \times \cdots \times \mathbf{R}^n$. Namely $A \leftrightarrow (A^1, \dots, A^n)$, where you will recall A^j denotes the j^{th} column of A interpreted as a vector in \mathbf{R}^n . (We could have used rows just as well.) Henceforth, for a multilinear form φ and $n \times n$ matrix A , we write $\varphi(A)$ for $\varphi(A^1, \dots, A^n)$. The next fact is rather remarkable, so pay close attention.

²The term *form* refers to a polynomial function in several variables where each term in the polynomial has the same degree. (The *degree* of the term is the sum of the exponents. For example, in the expression $xyz + x^2y + xz + z$, the first two terms have degree three, the third term has degree two and the last one has degree one. It is thus not a form.) This is most often encountered in the phrases *linear form* (each term has degree one) or *quadratic form* (each term has degree two). Tom Apostol once remarked at a cocktail party that mathematicians evidently don't know the difference between form and function.

75 Proposition For every $n \times n$ matrix A , there is a number $\det A$ with the property that for any alternating multilinear n -form φ ,

$$\varphi(A) = \det A \cdot \varphi(I). \tag{*}$$

Proof: Before we demonstrate the proposition in general, let us start with a special case, $n = 2$. Let

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}.$$

Then

$$\begin{aligned} \varphi(A) &= \varphi(A^1, A^2) \\ &= \varphi(a_{1,1}e^1 + a_{2,1}e^2, a_{1,2}e^1 + a_{2,2}e^2) \\ &= a_{1,1}\varphi(e^1, a_{1,2}e^1 + a_{2,2}e^2) + a_{2,1}\varphi(e^2, a_{1,2}e^1 + a_{2,2}e^2) \\ &= a_{1,1}a_{1,2}\varphi(e^1, e^1) + a_{1,1}a_{2,2}\varphi(e^1, e^2) + a_{2,1}a_{1,2}\varphi(e^2, e^1) + a_{2,1}a_{2,2}\varphi(e^2, e^2) \\ &= 0 + a_{1,1}a_{2,2}\varphi(e^1, e^2) + a_{2,1}a_{1,2}\varphi(e^2, e^1) + 0 \\ &= a_{1,1}a_{2,2}\varphi(e^1, e^2) - a_{2,1}a_{1,2}\varphi(e^1, e^2) \\ &= (a_{1,1}a_{2,2} - a_{2,1}a_{1,2})\varphi(I), \end{aligned}$$

so we see that $\det A = a_{1,1}a_{2,2} - a_{2,1}a_{1,2}$. Thus the whole of φ is determined by the single number $\varphi(I)$.

This is true more generally. Write

$$\varphi(A) = \varphi \left(\sum_{i_1=1}^n a_{i_1,1}e^{i_1}, \sum_{i_2=1}^n a_{i_2,2}e^{i_2}, \dots, \sum_{i_j=1}^n a_{i_j,j}e^{i_j}, \dots, \sum_{i_n=1}^n a_{i_n,n}e^{i_n} \right).$$

Now expand this using linearity in the first component:

$$\varphi(A) = \sum_{i_1=1}^n a_{i_1,1}\varphi \left(e^{i_1}, \sum_{i_2=1}^n a_{i_2,2}e^{i_2}, \dots, \sum_{i_j=1}^n a_{i_j,j}e^{i_j}, \dots, \sum_{i_n=1}^n a_{i_n,n}e^{i_n} \right).$$

Repeating this for the other components leads to

$$\varphi(A) = \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_n=1}^n a_{i_1,1}a_{i_2,2} \cdots a_{i_n,n} \varphi(e^{i_1}, e^{i_2}, \dots, e^{i_n}).$$

Now consider $\varphi(e^{i_1}, e^{i_2}, \dots, e^{i_n})$. Since φ is alternating, this term is zero unless i_1, i_2, \dots, i_n are distinct. When these are distinct, then by switching pairs we get to $\pm\varphi(e^1, e^2, \dots, e^n)$ where the sign on whether we need an odd or an even number of switches. It now pays to introduce some new terminology and notation. A **permutation** \mathbf{i} is an ordered list $\mathbf{i} = (i_1, \dots, i_n)$ of the numbers $1, 2, \dots, n$. The **signature** $\text{sgn}(\mathbf{i})$ of \mathbf{i} is 1 if \mathbf{i} can be put in numerical order by switching terms an even number of times and is -1 if it requires an odd number. It follows then that defining

$$\det A = \sum_{\mathbf{i}} \text{sgn}(\mathbf{i}) \cdot a_{i_1,1}a_{i_2,2} \cdots a_{i_n,n}, \tag{6}$$

where the sum runs over all permutations \mathbf{i} , satisfies the conclusion of the proposition. ■

This result still leaves the following question: *Are there any alternating multilinear n -forms at all?* The reason the result above does not settle this is that it would be vacuously true if there were none. Fortunately, it is not hard to verify that \det itself is such an n -form.

How do we know that $\text{sgn}(\mathbf{i})$ is well defined?

76 Proposition *The function $A \mapsto \det A$ as defined in (6) is an alternating multilinear n -form.*

Proof: Observe that in each product $\text{sgn}(\mathbf{i}) \cdot a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n}$ in the sum in (6) there is exactly one element from each row and each column of A . This makes it obvious that $\det(A^1, \dots, \alpha A^j, \dots, A^n) = \alpha \det A$, and it is straightforward to verify that

$$\det(A^1, \dots, x + y, \dots, A^n) = \det(A^1, \dots, x, \dots, A^n) + \det(A^1, \dots, y, \dots, A^n).$$

To see that \det is alternating, suppose $A^i = A^j$ with $i \neq j$. Then for any permutation \mathbf{i} , $i_j \neq i_k$, and there is exactly one permutation \mathbf{i}' satisfying $i_p = i'_p$ for $p \notin \{i, j\}$ and $i'_j = i_k$ and $i'_k = i_j$. Now observe that $\text{sgn}(\mathbf{i}) = -\text{sgn}(\mathbf{i}')$ as it requires an odd number of interchanges to swap two elements in a list (why?). Thus we can rewrite (6) as

$$\det A = \sum_{\mathbf{i}: \text{sgn}(\mathbf{i})=1} a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n} - a_{i'_1,1} a_{i'_2,2} \cdots a_{i'_n,n},$$

but each $a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n} - a_{i'_1,1} a_{i'_2,2} \cdots a_{i'_n,n} = 0$. (Why?) Therefore $\det A = 0$, so \det is alternating. ■

To sum things up we have:

77 Corollary *An alternating multilinear n -form is identically zero if and only if $\varphi(I) = 0$. The determinant is the unique alternating multilinear n -form φ that satisfies $\varphi(I) = 1$. Any other alternating multilinear n -form φ is of the form $\varphi = \varphi(I) \cdot \det$.*

5.2 Some simple consequences

78 Proposition *If A' is the transpose of A , then $\det A = \det A'$.*

Write out a proof.

79 Proposition *Adding a scalar multiple of one column of A to a different column leaves the determinant unchanged. Likewise for rows.*

Proof: By multilinearity, $\det(A^1, \dots, A^j + \alpha A^k, \dots, A^k, \dots, A^n) = \det(A^1, \dots, A^j, \dots, A^k, \dots, A^n) + \alpha \det(A^1, \dots, A^k, \dots, A^k, \dots, A^n)$, but $\det(A^1, \dots, A^k, \dots, A^k, \dots, A^n) = 0$ since \det is alternating. The conclusion for rows follows from that for columns and Proposition 78 on transposes. ■

80 Proposition *The determinant of an upper triangular matrix is the product of the diagonal entries.*

Proof: Recall that an upper triangular matrix is one for which $i > j$ implies $a_{i,j} = 0$. (Diagonal matrices are also upper triangular.) Now examine equation (6). The only summand that is nonzero comes from the permutation $(1, 2, 3, \dots, n)$, since for any other permutation there is some j satisfying $i_j > j$. (Why?) ■

By the way, the result also holds for lower triangular matrices.

81 Lemma *If A is $n \times n$ and φ is an alternating multilinear n -form, so is the form φ_A defined by*

$$\varphi_A(x^1, \dots, x^n) = \varphi(Ax^1, \dots, Ax^n).$$

Furthermore

$$\varphi_A(I) = \varphi(A) = \det A \cdot \varphi(I). \tag{7}$$

Proof: That φ_A is an alternating multilinear n -form is straightforward. Therefore φ_A is proportional to φ . To see that the coefficient of proportionality is $\det A$, consider $\varphi_A(I)$. Direct computation shows that $\varphi_A(I) = \varphi(A)$. ■

As an aside, I mention that we could have defined the determinant directly for linear transformations as follows. For a linear transformation $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$ it follows that φ_T defined by $\varphi_T(x^1, \dots, x^n) = \varphi(Tx^1, \dots, Tx^n)$ is an alternating n -form whenever φ is. It also follows that we could use (7) to define $\det T$ to be the scalar satisfying $\varphi_T(x^1, \dots, x^n) = \det T \cdot \varphi(x^1, \dots, x^n)$. This is precisely Dieudonné's approach. It has the drawback that minors and cofactors are awkward to describe in his framework.

82 Theorem *Let A and B be $n \times n$ matrices. Then*

$$\det AB = \det A \cdot \det B.$$

Proof: Let φ be an alternating multilinear n -form. Applying (7) and (\star), we see

$$\varphi_{AB}(I) = \varphi(AB) = \det AB \cdot \varphi(I).$$

On the other hand

$$\varphi_{AB}(I) = \varphi_A(B) = \det B \cdot \varphi_A(I) = \det B \cdot \det A \cdot \varphi(I).$$

Therefore $\det AB = \det A \cdot \det B$. ■

83 Corollary *If $\det A = 0$, then A has no inverse.*

Proof: Observe that if A has an inverse, then

$$1 = \det I = \det A \cdot \det A^{-1}$$

so $\det A \neq 0$. ■

5.3 Minors and cofactors

Different authors assign different meanings to the term **minor**. Given a square $n \times n$ matrix

$$A = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,n} \end{bmatrix}$$

Apostol [3, p. 87] defines the i, j **minor** of A to be the $(n-1) \times (n-1)$ *submatrix* obtained from A by deleting the i^{th} row and j^{th} column, and denotes it $A_{i,j}$. Gantmacher [6, p. 2] defines a **minor** of a (not necessarily square) matrix A to be the *determinant* of a square submatrix of A , and uses the following notation for minors in terms of the remaining rows and columns:

$$A \left(\begin{array}{c} i_1, \dots, i_p \\ j_1, \dots, j_p \end{array} \right) = \begin{vmatrix} \alpha_{i_1, j_1} & \cdots & \alpha_{i_1, j_p} \\ \vdots & & \vdots \\ \alpha_{i_p, j_1} & \cdots & \alpha_{i_p, j_p} \end{vmatrix}.$$

Here we require that $1 \leq i_1 < i_2 < \cdots < i_p \leq n$ and $1 \leq j_1 < j_2 < \cdots < j_p \leq n$. If the deleted (and hence remaining) rows and columns are the same, that is, if $i_1 = j_1, i_2 = j_2, \dots, i_p = j_p$, then $A \left(\begin{array}{c} i_1, \dots, i_p \\ j_1, \dots, j_p \end{array} \right)$ is called a **principal minor of order p** . I think the following

hybrid terminology is useful: a **minor submatrix** is any square submatrix of A (regardless of whether A is square) and a **minor** of A is the determinant of a minor submatrix (same as Gantmacher). To be on the safe side, I may use the redundant term **minor determinant** to mean minor.

Given a square $n \times n$ matrix

$$A = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,n} \end{bmatrix}$$

the **cofactor** $\text{cof } \alpha_{i,j}$ of $\alpha_{i,j}$ is the determinant obtained by replacing the j^{th} column of A with the i^{th} unit coordinate vector e^i . That is,

$$\text{cof } \alpha_{i,j} = \det(A^1, \dots, A^{j-1}e^i, A^{j+1}, \dots, A^n).$$

By multilinearity we have for any column j ,

$$\det A = \sum_{i=1}^n \alpha_{i,j} \text{cof } \alpha_{i,j}.$$

84 Lemma (Cofactors and minors) For any square matrix A ,

$$\text{cof } \alpha_{i,j} = (-1)^{i+j} \det A_{i,j},$$

where $A_{i,j}$ is the minor submatrix obtained by deleting the i^{th} row and j^{th} column from A . Consequently

$$\det A = \sum_{i=1}^n (-1)^{i+j} \alpha_{i,j} \det A_{i,j}.$$

Similarly (interchanging the roles of rows and columns), for any row i ,

$$\det A = \sum_{j=1}^n (-1)^{i+j} \alpha_{i,j} \det A_{i,j}.$$

Proof: Cf. Apostol [3, Theorem 3.9, p. 87]. By definition

$$\text{cof } \alpha_{i,j} = \begin{vmatrix} \alpha_{1,1} & \cdots & \alpha_{1,j-1} & 0 & \alpha_{1,j+1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{i-1,1} & \cdots & \alpha_{i-1,j-1} & 0 & \alpha_{i-1,j+1} & \cdots & \alpha_{i-1,n} \\ \alpha_{i,1} & \cdots & \alpha_{i,j-1} & 1 & \alpha_{i,j+1} & \cdots & \alpha_{i,n} \\ \alpha_{i+1,1} & \cdots & \alpha_{i+1,j-1} & 0 & \alpha_{i+1,j+1} & \cdots & \alpha_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,j-1} & 0 & \alpha_{n,j+1} & \cdots & \alpha_{n,n} \end{vmatrix}.$$

Adding $-a_{i,k}e^i$ to column k does not change the determinant. Doing this for all $k \neq j$ yields

$$\text{cof } \alpha_{i,j} = \begin{vmatrix} \alpha_{1,1} & \cdots & \alpha_{1,j-1} & 0 & \alpha_{1,j+1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{i-1,1} & \cdots & \alpha_{i-1,j-1} & 0 & \alpha_{i-1,j+1} & \cdots & \alpha_{i-1,n} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \alpha_{i+1,1} & \cdots & \alpha_{i+1,j-1} & 0 & \alpha_{i+1,j+1} & \cdots & \alpha_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,j-1} & 0 & \alpha_{n,j+1} & \cdots & \alpha_{n,n} \end{vmatrix}.$$

Now by repeatedly interchanging columns a total of $j-1$ times we obtain

$$\text{cof } \alpha_{i,j} = (-1)^{j-1} \begin{vmatrix} 0 & \alpha_{1,1} & \cdots & \alpha_{1,j-1} & \alpha_{1,j+1} & \cdots & \alpha_{1,n} \\ \vdots & & & \vdots & & & \vdots \\ 0 & \alpha_{i-1,1} & \cdots & \alpha_{i-1,j-1} & \alpha_{i-1,j+1} & \cdots & \alpha_{i-1,n} \\ 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \alpha_{i+1,1} & \cdots & \alpha_{i+1,j-1} & \alpha_{i+1,j+1} & \cdots & \alpha_{i+1,n} \\ \vdots & & & \vdots & & & \vdots \\ 0 & \alpha_{n,1} & \cdots & \alpha_{n,j-1} & \alpha_{n,j+1} & \cdots & \alpha_{n,n} \end{vmatrix}.$$

Interchanging rows $i-1$ times yields

$$\text{cof } \alpha_{i,j} = (-1)^{i+j} \begin{vmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \alpha_{1,1} & \cdots & \alpha_{1,j-1} & \alpha_{1,j+1} & \cdots & \alpha_{1,n} \\ \vdots & & & \vdots & & & \vdots \\ 0 & \alpha_{i-1,1} & \cdots & \alpha_{i-1,j-1} & \alpha_{i-1,j+1} & \cdots & \alpha_{i-1,n} \\ 0 & \alpha_{i+1,1} & \cdots & \alpha_{i+1,j-1} & \alpha_{i+1,j+1} & \cdots & \alpha_{i+1,n} \\ \vdots & & & \vdots & & & \vdots \\ 0 & \alpha_{n,1} & \cdots & \alpha_{n,j-1} & \alpha_{n,j+1} & \cdots & \alpha_{n,n} \end{vmatrix}.$$

(Recall that $(-1)^{j-1+i-1} = (-1)^{i+j}$.) This last determinant is block diagonal, so we see that it is just $|A_{i,j}|$, which completes the proof.

The conclusion for rows follows from that for columns and Proposition 78 on transposes. ■

By repeatedly applying this result, we can express $\det A$ in terms of 1×1 determinants.

If we take the cofactors from an *alien* column or row, we have:

85 Lemma (Expansion by alien cofactors) *Let A be a square $n \times n$ matrix. For any column j and any $k \neq j$,*

$$\sum_{i=1}^n \alpha_{i,j} \text{cof } \alpha_{i,k} = 0.$$

Likewise for any row i and any $k \neq i$,

$$\sum_{j=1}^n \alpha_{i,j} \text{cof } \alpha_{k,j} = 0.$$

Proof: Consider the matrix $A_1 = [\tilde{\alpha}_{i,j}]$ obtained from A by replacing the k^{th} column with another copy of column j . Then the i, k cofactors, $i = 1, \dots, n$, of A and A_1 are the same. (The cofactors don't depend on the column they belong to, since it is replaced by a unit coordinate vector.) So by Lemma 84 $|A_1| = \sum_{i=1}^n \tilde{\alpha}_{i,k} \text{cof } \alpha_{i,k} = \sum_{i=1}^n \alpha_{i,j} \text{cof } \alpha_{i,k}$. But $|A_1| = 0$ since it has two identical columns.

The conclusion for rows follows from that for columns and Proposition 78 on transposes. ■

The transpose of the cofactor matrix is also called the **adjugate** matrix of A . Combining the previous two lemmas yields the following on the adjugate.

86 Theorem (Cofactors and the inverse matrix) *For a square matrix*

$$A = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,n} \end{bmatrix},$$

we have

$$\begin{bmatrix} \alpha_{1,1} & \dots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{n,1} & \dots & \alpha_{n,n} \end{bmatrix} \begin{bmatrix} \text{cof}\alpha_{1,1} & \dots & \text{cof}\alpha_{n,1} \\ \vdots & & \vdots \\ \text{cof}\alpha_{1,n} & \dots & \text{cof}\alpha_{n,n} \end{bmatrix} = \begin{bmatrix} |A| & & 0 \\ & \ddots & \\ 0 & & |A| \end{bmatrix}.$$

That is, the product of A and its adjugate is $(\det A)I_n$. Consequently, if $\det A \neq 0$, then

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} \text{cof}\alpha_{1,1} & \dots & \text{cof}\alpha_{n,1} \\ \vdots & & \vdots \\ \text{cof}\alpha_{1,n} & \dots & \text{cof}\alpha_{n,n} \end{bmatrix}.$$

Combining this with Corollary 83 yields the following.

87 Corollary (Determinants and invertibility) *A square matrix is invertible if and only if its determinant is nonzero.*

Similar reasoning leads to the following theorem due to Jacobi. It expresses a p^{th} order minor of the adjugate in terms of the corresponding complementary minor of A . The **complement** of the p^{th} -order minor $A \begin{pmatrix} i_1, \dots, i_p \\ j_1, \dots, j_p \end{pmatrix}$ is the $n-p^{\text{th}}$ -order minor obtained by deleting rows i_1, \dots, i_p and columns j_1, \dots, j_p from A .

88 Theorem (Jacobi) *For a square matrix*

$$A = \begin{bmatrix} \alpha_{1,1} & \dots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{n,1} & \dots & \alpha_{n,n} \end{bmatrix},$$

we have for any $1 \leq p \leq n$,

$$|A| \cdot \begin{vmatrix} \text{cof}\alpha_{1,1} & \dots & \text{cof}\alpha_{p,1} \\ \vdots & & \vdots \\ \text{cof}\alpha_{1,p} & \dots & \text{cof}\alpha_{p,p} \end{vmatrix} = |A|^p \cdot \begin{vmatrix} \alpha_{p+1,p+1} & \dots & \alpha_{p+1,n} \\ \vdots & & \vdots \\ \alpha_{n,p+1} & \dots & \alpha_{n,n} \end{vmatrix}.$$

Proof: Observe that

$$\begin{bmatrix} \alpha_{1,1} & \dots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{n,1} & \dots & \alpha_{n,n} \end{bmatrix} \left[\begin{array}{ccc|ccc} \text{cof}\alpha_{1,1} & \dots & \text{cof}\alpha_{p,1} & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \text{cof}\alpha_{1,p} & \dots & \text{cof}\alpha_{p,p} & 0 & \dots & 0 \\ \hline \text{cof}\alpha_{1,p+1} & \dots & \text{cof}\alpha_{p,p+1} & 1 & & 0 \\ \vdots & & \vdots & & \ddots & \\ \text{cof}\alpha_{1,n} & \dots & \text{cof}\alpha_{n,n} & 0 & & 1 \end{array} \right] = \left[\begin{array}{cc|ccc} |A| & 0 & \alpha_{1,p+1} & \dots & \alpha_{1,n} \\ & \ddots & \vdots & & \vdots \\ 0 & & |A| & & \alpha_{p,p+1} & \dots & \alpha_{p,n} \\ \hline 0 & \dots & 0 & \alpha_{p+1,p+1} & \dots & \alpha_{p+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & \alpha_{n,p+1} & \dots & \alpha_{n,n} \end{array} \right]$$

and take determinants on both sides. ■

5.4 Characteristic polynomials

The **characteristic polynomial** f of a square matrix A is defined by $f(\lambda) = \det(\lambda I - A)$. Roots of this polynomial are called **characteristic roots** of A .

89 Lemma *Every eigenvalue of a matrix is a characteristic root, and every real characteristic root is an eigenvalue.*

Proof: To see this note that if λ is an eigenvalue with eigenvector $x \neq 0$, then $(\lambda I - A)x = \lambda x - Ax = 0$, so $(\lambda I - A)$ is singular, so $\det(\lambda I - A) = 0$. That is, λ is a characteristic root of A .

Conversely, if $\det(\lambda I - A) = 0$, then there is some nonzero x with $(\lambda I - A)x = 0$, or $Ax = \lambda x$. ■

90 Lemma *The determinant of a square matrix is the product of its characteristic roots.*

Proof: (cf. [3, p. 106]) Let A be an $n \times n$ square matrix and let f be its characteristic polynomial. Then $f(0) = \det(-A) = (-1)^n \det A$. On the other hand, we can factor f as

$$f(\lambda) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

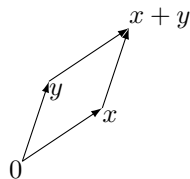
where $\lambda_1, \dots, \lambda_n$ are its characteristic roots. Thus $f(0) = (-1)^n \lambda_1 \cdots \lambda_n$. ■

5.5 The determinant as an “oriented volume”

There is another interpretation of the determinant $\det(x^1, \dots, x^n)$. Namely it the n -dimensional volume of the parallelotope defined by the vectors x^1, \dots, x^n , perhaps multiplied by -1 depending on the “orientation” of the vectors.

I don’t want to go into detail on the notion of orientation or even to prove the assertion above in any generality, but I shall present enough so that you can get a glimmer of what the assertion is.

First, what is a parallelotope. It is a polytope with parallel faces, for example, a cube. The parallelotope generated by the vectors x^1, \dots, x^n is the convex hull of zero and the vectors of the form $x^{i_1} + x^{i_2} + \dots + x^{i_k}$, where $1 \leq k \leq n$ and i_1, i_2, \dots, i_k are distinct. For instance, in R^2 the parallelotope generated by x and y is the plane parallelogram with vertexes $0, x, y$, and $x + y$.



The notion of n -dimensional volume is straightforward. For $n = 1$, it is length, for $n = 2$, it is area, etc. Oriented volume is more complicated. It distinguishes parallelotopes based on the order that the vectors are presented. For instance, in the figure above the angle swept out from x to y counterclockwise is positive, while from y to x is clockwise. The oriented volume is positive in the first case and negative in the second.

The simplest way to convince yourself that the determinant is the oriented volume of the parallelotope, is to realize that oriented volume is an alternating multilinear form. It is clear that multiplying one of the defining vectors by a positive scalar multiplies the volume by the same factor. It is also clear that if two of the vectors are the same, then the parallelotope is degenerate and has zero volume. The additivity in the defining vectors can be appreciated by considering the Figure 1, at least in two dimensions.

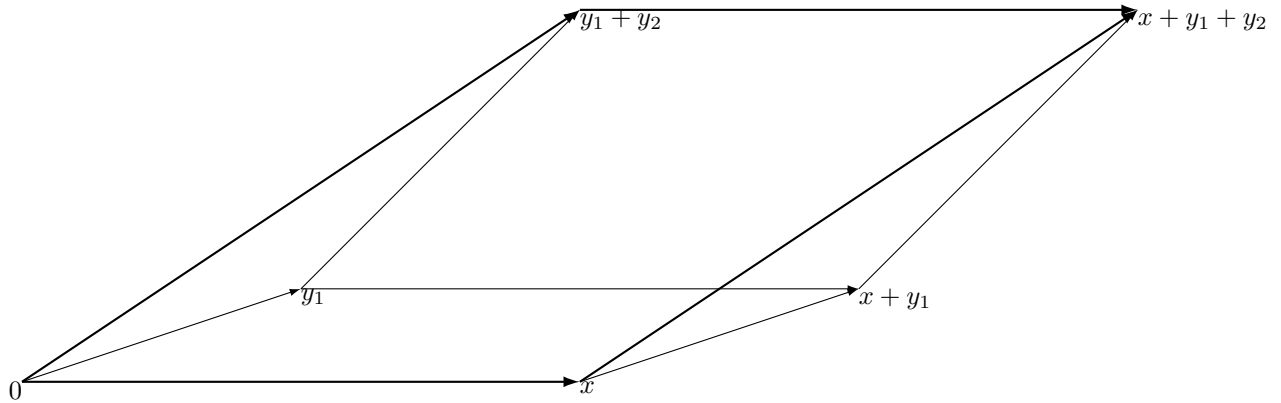


Figure 1. The area of the parallelogram defined by x and y is additive in y .

5.6 Computing inverses and determinants by Gauss' method

This section describes how to use the method of Gaussian elimination to find the inverse of a matrix and to compute its determinant. Let A be an $n \times n$ invertible matrix, so that

$$AA^{-1} = I.$$

To find the j^{th} column $x = A^{-1}I^j$ of A^{-1} recall from the definition of matrix multiplication that x satisfies $Ax = I^j$, where I^j is the j^{th} column of the identity matrix. Indeed if A is invertible, then x is the unique vector with this property. Thus we can form the $n \times (n + 1)$ augmented coefficient matrix $(A|I^j)$ and use elementary row operations of Gaussian elimination to transform it to $(I|x) = (I|A^{-1}I^j)$. The same sequence of row operations is employed to transform A into I regardless of which column I^j we use to augment with. So we can solve for *all* the columns simultaneously by augmenting with all the columns of I . That is, form the $n \times 2n$ augmented matrix $(A|I^1, I^2, \dots, I^n) = (A|I)$. If we can transform this by elementary row operations into $(I|X)$, it must be that $X = A^{-1}$. Furthermore, if we cannot make the transformation, then A is not invertible.

You may ask, how can we tell if we cannot transform A into I ? Perhaps it is possible, but we are not clever enough. To attack this question we must write down a specific algorithm. So here is one version of an algorithm for inverting a matrix by Gaussian elimination, or else showing that it is not invertible.

Let $A^0 = A$. At each stage t we apply an elementary operation to transform A^{t-1} into A^t in such a way that

$$\det A^t \neq 0 \iff \det A \neq 0.$$

The rules for selecting the elementary row operation are described below. By stage $t = n$, either $A^t = I$ or else we shall have shown that $\det A = 0$, so A is not invertible.

Stage t : At stage t , assume that all the columns $j = 1, \dots, t-1$ of A^{t-1} have been transformed into the corresponding columns of I , and that $\det A \neq 0 \iff \det A^{t-1} \neq 0$.

Step 1: Normalize the diagonal to 1.

Case 1: $a_{t,t}^{t-1}$ is nonzero. Divide row t by $a_{t,t}^{t-1}$ to set $a_{t,t}^t = 1$. This has the side effect of setting $\det A^t = a_{t,t}^{t-1} \det A^{t-1}$.

Case 2: $a_{t,t}^{t-1} = 0$, but there is row i with $i > t$ for which $a_{i,t}^{t-1} \neq 0$. Divide row i by $a_{i,t}^{t-1}$ and add it to row t . This sets $a_{t,t}^t = 1$, and leaves $\det A^t = \det A^{t-1}$.

Case 3: If $a_{t,t}^{t-1} = 0$, but there is no row i with $i > t$ for which $a_{i,t}^{t-1} \neq 0$. In this case, the first t columns of A^{t-1} must be dependent. This because there are t column vectors whose only nonzero components are in the first $t - 1$ rows. This implies $\det A^{t-1} = 0$, and hence $\det A = 0$. This shows that A is not invertible.

If Case 3 occurs stop. We already know that A is not invertible. In cases 1 and 2 proceed to:

Step 2: Eliminate the off diagonal elements. Since $a_{t,t}^t = 1$, for $i \neq t$ multiply row i by $a_{i,t}^{t-1}$ and subtract it from row t . This sets $a_{i,t}^t = 0$, for $i \neq t$, and does not change $\det A^t$.

This completes the construction of A^t from A^{t-1} . Proceed to stage $t+1$, and note that all the columns $j = 1, \dots, t$ of A^t have been transformed into the corresponding columns of I , and that $\det A \neq 0 \iff \det A^t \neq 0$.

Now observe how this also can be used to calculate the determinant. Suppose the process runs to completion so that $A^n = I$. Every time a row was divided by its diagonal element $a_{t,t}^{t-1}$ to normalize it, we had $\det A^t = \frac{1}{a_{t,t}^{t-1}} \det A^{t-1}$. Thus we have

$$1 = \det I = \det A^n = \prod_{t:a_{t,t}^{t-1} \neq 0} \frac{1}{a_{t,t}^{t-1}} \det A$$

or

$$\det A = \prod_{t:a_{t,t}^{t-1} \neq 0} a_{t,t}^{t-1}.$$

One of the virtues of this approach is that it is extremely easy to program, and reasonably efficient. Each elementary row operation has at most $2n$ multiplications and $2n$ additions, and at most n^2 elementary row operations are required, so the number of steps grows no faster than $4n^3$.

Here are two illustrative examples, which, by the way, were produced (text and all) by a 475-line program written in `perl` of all things.

91 Example (Matrix inversion by Gaussian elimination) Invert the following 3×3 matrix using Gaussian elimination and compute its determinant as a byproduct.

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

We use only two elementary row operations: dividing a row by a scalar, which also divides the determinant, and adding a multiple of one row to a different row, which leaves the determinant unchanged. Thus the determinant of A is just the product of all the scalars used to divide the rows of A to normalize its diagonal elements to ones. We shall keep track of this product in the variable μ as we go along. At any given stage, μ times the determinant of the left hand block is equal to the determinant of A . Before each step, we put a box around the target entry to be transformed.

$a_{1,1}$ is zero,

$$\mu = 1 \quad \left(\begin{array}{ccc|ccc} \boxed{0} & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

so add row 3 to row 1. To eliminate $a_{3,1} = 1$,

$$\mu = 1 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ \boxed{1} & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

subtract row 1 from row 3. To normalize $a_{3,3} = -1$,

$$\mu = 1 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \boxed{-1} & -1 & 0 & 0 \end{array} \right)$$

divide row 3 (and multiply μ) by -1 . To eliminate $a_{1,3} = 1$,

$$\mu = -1 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & \boxed{1} & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array} \right)$$

subtract row 3 from row 1. This gives us

$$\mu = -1 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{array} \right)$$

To summarize:

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad A^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

And the determinant of A is -1 . We could have gotten this result faster by interchanging rows 1 and 3, but it is hard (for me) to program an algorithm to recognize when to do this. \square

92 Example (Gaussian elimination on a singular matrix) Consider the following 3×3 matrix.

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 4 & 4 \\ 3 & 6 & 6 \end{pmatrix}$$

We shall use the method of Gaussian elimination to attempt transform the augmented block matrix $(A|I)$ into $(I|A^{-1})$. To eliminate $a_{2,1} = 2$,

$$\mu = 1 \quad \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ \boxed{2} & 4 & 4 & 0 & 1 & 0 \\ 3 & 6 & 6 & 0 & 0 & 1 \end{array} \right)$$

multiply row 1 by 2 and subtract it from row 2. To eliminate $a_{3,1} = 3$,

$$\mu = 1 \quad \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & -2 & 1 & 0 \\ \boxed{3} & 6 & 6 & 0 & 0 & 1 \end{array} \right)$$

multiply row 1 by 3 and subtract it from row 3. To normalize $a_{2,2} = 2$,

$$\mu = 1 \quad \left(\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & \boxed{2} & 0 & -2 & 1 & 0 \\ 0 & 3 & 0 & -3 & 0 & 1 \end{array} \right)$$

divide row 2 (and multiply μ) by 2. To eliminate $a_{1,2} = 1$,

$$\mu = 2 \quad \left(\begin{array}{ccc|ccc} 1 & \boxed{1} & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & \frac{1}{2} & 0 \\ 0 & 3 & 0 & -3 & 0 & 1 \end{array} \right)$$

subtract row 2 from row 1. To eliminate $a_{3,2} = 3$,

$$\mu = 2 \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 2 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & -1 & \frac{1}{2} & 0 \\ 0 & \boxed{3} & 0 & -3 & 0 & 1 \end{array} \right)$$

multiply row 2 by 3 and subtract it from row 3. Now notice that the first 3 columns of A are dependent, as each column has at most its first 2 entries nonzero, and any 3 vectors in a 2-dimensional space are dependent.

$$\mu = 2 \quad \left(\begin{array}{ccc|ccc} \boxed{1} & 0 & 2 & 2 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & -1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & -1\frac{1}{2} & 1 \end{array} \right)$$

Therefore A has no inverse, and the determinant of A is zero. (If we had interchanged columns 1 and 3, we would have arrived at this result sooner, but hindsight is also hard to program.) \square

93 Exercise Using the language of your choice, write a program to accept a square matrix and invert it by Gaussian elimination, or stop when you find it is not invertible.

References

- [1] A. C. Aitken. 1954. *Determinants and matrices*, 8th. ed. New York: Interscience.
- [2] C. D. Aliprantis and K. C. Border. 2006. *Infinite dimensional analysis: A hitchhiker's guide*, 3d. ed. Berlin: Springer-Verlag.
- [3] T. M. Apostol. 1969. *Calculus*, 2d. ed., volume 2. Waltham, Massachusetts: Blaisdell.
- [4] J. Dieudonné. 1969. *Foundations of modern analysis*. New York: Academic Press. Volume 1 of Treatise on Analysis.
- [5] J. Franklin. 1968. *Matrix theory*. Englewood Cliffs, N.J.: Prentice-Hall.
- [6] F. R. Gantmacher. 1959. *Matrix theory*, volume 1. New York: Chelsea.
- [7] P. R. Halmos. 1974. *Finite dimensional vector spaces*. New York: Springer-Verlag. Reprint of the edition published by Van Nostrand, 1958.
- [8] L. H. Loomis and S. Sternberg. 1968. *Advanced calculus*. Reading, Massachusetts: Addison-Wesley.
- [9] S. MacLane and G. Birkhoff. 1993. *Algebra*, 3d. ed. New York: Chelsea.
- [10] C. R. Rao. 1973. *Linear statistical inference and its applications*, 2d. ed. New York: Wiley.