Contents

1 Complex numbers

Definition 1.1 *A* field *F is a set of at least two objects together with two binary operations, $+$ and $\cdot$, such that (1)-(5) below hold:*

1. *The field is closed under both operations, i.e. whenever x and y are in F, $x + y$ and xy are in F.*
2. *Both operations are associative and commutative, i.e. for every $x, y, z \in F$, $x + y = y + x$; $xy = yx$; $(x + y) + z = x + (y + z)$; and $x(yz) = (xy)z$.*
3. *There exist identity elements for both operations, i.e. there is a $0 \in F$ such that $x + 0 = x$ for all $x \in F$ and there is a $1 \in F$ such that $1x = x$ for all $x \in F$.*
4. *Every element $x \in F$ has an additive inverse $-x \in F$ such that $x + (-x) = 0$ and every nonzero element $x \in F$ has a reciprocal $x^{-1} \in F$ such that $x \cdot x^{-1} = 1$.*
5. *The distributive property holds: $x(y + z) = xy + xz$ for all $x, y, z \in F$.*

The real numbers form a field under the usual operations; the rational numbers form a field under the usual operations. There are other fields as well, including finite fields. The integers do not form a field because nonzero elements do not have reciprocals in general.

Definition 1.2 *The set of* complex numbers *is defined to be $\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}$. Given a complex number $z = a + ib$, the* real part *of z is $\Re(z) = a$ and the* imaginary part *of z is $\Im(z) = b$. The* complex conjugate *of $z = a + ib$ is $\overline{z} = a - ib$. The* absolute value *or* modulus *of a complex number $z = a + ib$ is $|z| = \sqrt{a^2 + b^2}$.*

We define addition and subtraction in $\mathbb{C}$ by adding and subtracting like terms; and multiplication is defined by setting $i^2 = -1$. These operations make $\mathbb{C}$ into a field; in particular the reciprocal of a nonzero complex number $z = a + ib$ is $z^{-1} = \frac{a}{a^2+b^2} + i\frac{-b}{a^2+b^2}$. We define the *distance* between complex numbers z and w to be $|z - w|$. A complex number $x = a + ib$ is said to be *real* if $b = 0$.

Proposition 1.3 *Let $z, w \in \mathbb{C}$. Then:*

1. *$\Re(z) = \frac{z+\overline{z}}{2}$ and $\Im(z) = \frac{z-\overline{z}}{2i}$;*
2. *$\overline{z + w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z}\,\overline{w}$;*
3. *$\Re(z) \leq |\Re(z)| \leq |z|$ and $\Im(z) \leq |\Im(z)| \leq |z|$;*
4. *$|zw| = |z| \cdot |w|$;*
5. *$|z + w| \leq |z| + |w|$;*
6. *$z \cdot \overline{z} = |z|^2$, so if $z \neq 0$, then $z^{-1} = \frac{\overline{z}}{|z|^2}$ and $\frac{w}{z} = \frac{w\overline{z}}{|z|^2}$;*
7. *$z = \overline{z}$ if and only if z is real.*

Definition 1.4 *The* polar representation *of a complex number $z = a + ib$ is $z = re^{i\theta}$, where (r, θ) are the polar coordinates of the point (a, b). $\theta \in \mathbb{R}$ is called the* argument *of z and is denoted* $\arg(z)$ *(the argument of a complex number is not unique).*

Suppose $f : \mathbb{R} \to \mathbb{R}$ is a function which can be written as a power series which converges for all real numbers. Then the same power series converges for all complex numbers. This allows us to define trigonometric functions:

Definition 1.5 *Given $z \in \mathbb{C}$,*

- $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$.
- $\cos z = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}$.
- $\sin z = \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!}$.

Theorem 1.6 (Euler's Formula) *For all* $\theta \in \mathbb{R}$, $e^{i\theta} = \cos\theta + i\sin\theta$.

By Euler's Formula, we see that if $z = a + ib = re^{i\theta}$, then $a = r\cos\theta$, $b = r\sin\theta$, $r = |z|$ and $\theta = \tan^{-1}(b/a)$.

Corollary 1.7 *Let* $z \in \mathbb{C}$. *Then* $|z| = 1$ *if and only if* $z = e^{it}$ *for some* $t \in \mathbb{R}$. *For all* $z \in \mathbb{C}$, $\cos z = \frac{1}{2}(e^{iz} + e^{-iz})$. *For all* $z \in \mathbb{C}$, $\sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$.

Theorem 1.8 *Let* $z, w \in \mathbb{C}$ *have polar representations* $z = re^{i\theta}$ *and* $w = se^{i\phi}$. *Then*

1. $zw = rse^{i(\theta+\phi)}$.
2. *If* $w \neq 0$, *then* $\frac{z}{w} = \frac{r}{s}e^{i(\theta-\phi)}$.

Theorem 1.9 (DeMoivre's Theorem) *For all* $n \in \mathbb{N}$ *and all* $\theta \in \mathbb{R}$, $(\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$.

2 Vector spaces

A vector space is the most abstract setting in which "linear" objects can be defined.

Definition 2.1 *Given a field* F, *a* vector space V *over* F *is a set of objects called* vectors *together with two operations (under which* V *is closed), namely*
Addition $+ : V \times V \to V$, *and*
Scalar Multiplication $\cdot : F \times V \to V$
such that:

1. *Addition is associative and commutative, i.e.* $\mathbf{v} + (\mathbf{w} + \mathbf{x}) = (\mathbf{v} + \mathbf{w}) + \mathbf{x}$ *and* $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ *for all* $\mathbf{v}, \mathbf{w}, \mathbf{x} \in V$.
2. *Scalar multiplication is associative, i.e.* $(rs)\mathbf{v} = r(s\mathbf{v})$ *for all* $r, s \in F$ *and all* $\mathbf{v} \in V$.
3. *There exists an identity element for addition, i.e.* $\exists\, \mathbf{0} \in V$ *such that* $\mathbf{v} + \mathbf{0} = \mathbf{v}$ *for all* $\mathbf{v} \in V$.
4. *Every vector* $\mathbf{v} \in V$ *has an additive inverse* $-\mathbf{v} \in V$ *such that* $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
5. *Distributive properties hold, i.e.* $(r + s)\mathbf{v} = r\mathbf{v} + s\mathbf{v}$ *and* $r(\mathbf{v} + \mathbf{w}) = r\mathbf{v} + r\mathbf{w}$ *for all* $r, s \in F$ *and all* $\mathbf{v}, \mathbf{w} \in V$.
6. *If* 1 *is the multiplicative identity element of* F, *then* $1\mathbf{v} = \mathbf{v}$ *for all* $\mathbf{v} \in V$.

The field F *is called the* underlying field *of the vector space.*

Many other basic arithmetic properties of vector spaces can be derived from these laws (for example, the uniqueness of the additive identity and additive inverses, the Cancellation Law, the fact that $0\mathbf{v} = \mathbf{0}$ for any $\mathbf{v} \in V$, etc.)

Given any field F, F^n, the set of ordered $n-$tuples of elements of F, is a vector space under the usual (component-wise) addition and scalar multiplication. Every field is a vector space over itself (with the usual addition and multiplication). The set $\{\mathbf{0}\}$ is a vector space over any field, called the *zero space.* The set of infinite sequences of elements of a field forms a vector space over that field. The set of functions from any set X to a field F forms a vector space over that field. The complex numbers are a vector space over the real numbers. $\mathbb{C}$ is also a vector space over itself (but that is a different vector space than $\mathbb{C}$ over $\mathbb{R}$ since the scalars differ).

2.1 Subspaces

Definition 2.2 *Given vector space V over field F, a subset $W \subseteq V$ is a* subspace *of V if:*

1. *W is nonempty.*
2. *W is closed under addition, i.e. given $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W$, it must be that $\mathbf{w}_1 + \mathbf{w}_2 \in W$.*
3. *W is closed under scalar multiplication, i.e. for any $\mathbf{w} \in W$ and any $r \in F$, $r\mathbf{w} \in W$.*

Equivalently, this means W is itself a vector space under the operations it inherits from V.

Proposition 2.3 *Let V be a vector space over field F. If $W \subseteq V$ is a subspace of V, then $\mathbf{0} \in W$.*

The *zero subspace* $\{\mathbf{0}\}$ is a subspace of every vector space, and every vector space is a subspace of itself. The intersection of any collection of subspaces (of the same vector space) is a subspace; however, the union of two subspaces need not be a subspace.

Definition 2.4 *Let $A, B \subseteq V$ where V is a vector space over F. Define $A + B$, the* sum *of A and B, to be the subset of V defined by*

$$A + B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\}.$$

The sum of any two (or any finite number of) subspaces is itself a subspace.

Definition 2.5 *Given a vector space V over a field F, let $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ be any finite list of vectors. The* span *of these vectors is the set of linear combinations of the $\mathbf{v}_j$, i.e.*

$$Span(\mathbf{v}_1, ..., \mathbf{v}_n) = \left\{ \sum_{j=1}^{n} c_j \mathbf{v}_j : c_j \in F \right\}.$$

Given a vector space (or a subspace of another vector space) W, a set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is said to span *W (or* is a spanning set for *W) if $W = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$.*

Equivalently, the span of a set of vectors is the smallest vector space containing all the given vectors.

Proposition 2.6 *Given any vector space V over F, the span of any finite collection of vectors is a subspace of V.*

Proposition 2.7 *Given a vector space V over a field F, let $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ be any finite list of vectors. Suppose $\mathbf{w} = \sum_{j=1}^{n} c_j \mathbf{v}_j$ for scalars $c_1, ..., c_n$. Then*

$$Span(\mathbf{v}_1, ..., \mathbf{v}_n) = Span(\mathbf{v}_1, ..., \mathbf{v}_{n-1}, \mathbf{w}).$$

Proposition 2.8 *Let V be a vector space and suppose $W_1 = Span(\mathbf{v}_1, ..., \mathbf{v}_m)$ and $W_2 = Span(\mathbf{w}_1, ..., \mathbf{w}_n)$. Then $Span(\mathbf{v}_1, ..., \mathbf{v}_m, \mathbf{w}_1, ...\mathbf{w}_n)$ is the subspace $W_1 + W_2$.*

2.2 Affine subspaces

Definition 2.9 *Given a vector space V over F, an* affine subspace *A is a subset of V of the form*

$$A = \mathbf{p} + W = \{\mathbf{p} + \mathbf{w} : \mathbf{w} \in W\}$$

where W is a subspace of V and $\mathbf{p} \in V$.

We remark that all subspaces are affine subspaces, but not all affine subspaces are subspaces. In particular "affine" is not an adjective that describes "subspaces".

Theorem 2.10 *A subset A of a vector space V (over $\mathbb{R}$) is an affine subspace if and only if for every $\mathbf{v}, \mathbf{w} \in A$ and every $t \in \mathbb{R}$, $(t\mathbf{v} + (1-t)\mathbf{w})$ is also in A.*

The previous result implies that a subset A of a vector space is an affine subspace if and only if for any two vectors in A, the line containing those two vectors lies entirely within A.

Proposition 2.11 *If A is an affine subspace of a vector space V, then for any $\mathbf{v} \in A$, $A - \mathbf{v}$ is a subspace of V (and is the same subspace no matter what the choice of $\mathbf{v}$ is).*

Proposition 2.12 *An affine subspace A of a vector space V is a subspace if and only if it contains $\mathbf{0}$.*

Proposition 2.13 *If two affine subspaces have nonempty intersection, then their intersection is an affine subspace.*

Definition 2.14 *Given a vector space V over F, a* line *in V is an affine subspace $A = \mathbf{p} + W$ where $W = Span(\mathbf{v})$ for some nonzero $\mathbf{v} \in V$. Equivalently, a line in V is any set of the form*

$$l = \{\mathbf{p} + t\mathbf{v} : t \in F\}$$

for given vectors $\mathbf{p}$ and $\mathbf{v}$. The vector $\mathbf{v}$ is called a direction vector *for the line.*

Neither the direction vector nor the vector $\mathbf{p}$ are unique for a particular line. Given two distinct vectors $\mathbf{v}$ and $\mathbf{w}$, the line containing those two vectors l can be described as the set of vectors $\mathbf{x}$ satisfying

$$\mathbf{x} = t\mathbf{v} + (1-t)\mathbf{w} \text{ for some } t \in F.$$

Definition 2.15 *Given a vector space V over F, a* plane *in V is an affine subspace $A = \mathbf{p} + W$ where $W = Span(\mathbf{v}, \mathbf{w})$ for some nonparallel vectors $\mathbf{v}, \mathbf{w} \in V$. Equivalently, a plane in V is any set of the form*

$$\mathcal{P} = \{\mathbf{p} + s\mathbf{v} + t\mathbf{w} : s, t \in F\}$$

for given vectors $\mathbf{p}$, $\mathbf{v}$ and $\mathbf{w}$.

Every line in F^n can be specified by writing parametric equations for the line. Every plane in F^n can be specified by writing parametric equations for the plane.

Theorem 2.16 *A subset of $\mathbb{R}^3$ is a plane if and only if it is a set of the form*

$$\{(x, y, z) : ax + by + cz = d\}$$

for some real numbers a, b, c, d.

3 Linear independence, basis and dimension

3.1 Linear independence

Definition 3.1 *Given a vector space V over a field F, two vectors $\mathbf{v}$ and $\mathbf{w}$ in V are called* parallel *if there is a scalar $r \in F$ such that $\mathbf{v} = r\mathbf{w}$ or $\mathbf{w} = r\mathbf{v}$.*

Definition 3.2 *Given a vector space V over $\mathbb{R}$ or $\mathbb{C}$, we say two vectors $\mathbf{v}$ and $\mathbf{w}$ in V are* in the same direction *if there is a scalar $r \geq 0$ (in particular, r must be real, even if $F = \mathbb{C}$) such that $\mathbf{v} = r\mathbf{w}$ or $\mathbf{w} = r\mathbf{v}$.*

Definition 3.3 *Given a vector space V over a field F, a set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is called* linearly dependent *if one of the following equivalent conditions holds:*

1. *There is a nontrivial linear dependence relation among the vectors, i.e.*

$$c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n = \mathbf{0}$$

   for scalars $c_1, ..., c_n \in F$ with at least one $c_j \neq 0$.
2. *There is a $k \leq n$ such that $\mathbf{v}_k = \sum_{j=1}^{k-1} d_j\mathbf{v}_j$ for scalars $d_1, ..., d_{k-1} \in F$ (we describe this by saying that $\mathbf{v}_k$* depends on *$\mathbf{v}_1, ..., \mathbf{v}_{k-1}$).*
3. *There is a $k \leq n$ such that $\mathbf{v}_k \in Span(\mathbf{v}_1, ..., \mathbf{v}_{k-1})$.*

Given a set of vectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$, if any of those vectors is $\mathbf{0}$, then the list is linearly dependent. A set of two vectors is linearly dependent if and only if the two vectors are parallel. More generally, if any two of the vectors in the list are parallel, then the list is linearly dependent. Given any list of linearly dependent vectors, if you add more vectors to the list, the list remains linearly dependent.

Theorem 3.4 *Let* $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *be a set of linearly dependent vectors with* $\mathbf{v}_k = \sum_{j=1}^{k-1} d_j \mathbf{v}_j$. *Then*

$$Span(\mathbf{v}_1, ..., \mathbf{v}_n) = Span(\mathbf{v}_1, ..., \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, ..., \mathbf{v}_n),$$

i.e. vectors which are dependent on the previous vectors in the list can be deleted from the list without changing the span of the list.

Definition 3.5 *Given a vector space* V *over a field* F, *a set of vectors* $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *is called* linearly independent *if they are not linearly dependent, i.e. whenever*

$$c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n = \mathbf{0}$$

for scalars $c_1, ..., c_n \in F$, *it must be that* $c_j = 0$ *for all* j.

Equivalently, this means that there is no $k \leq n$ such that $\mathbf{v}_k \in Span(\mathbf{v}_1, ..., \mathbf{v}_{k-1})$. Roughly speaking, a list of linearly independent vectors does not "repeat" the same direction(s) unnecessarily. Any one nonzero vector forms a linearly independent set by itself; any two nonparallel vectors form a linearly independent set. Given any list of linearly independent vectors, if you remove any number of vectors from the list, the list remains linearly independent.

Proposition 3.6 *Suppose* $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *be a set of linearly dependent vectors. Let* $c_1, ..., c_n \in F$ *with* $c_n \neq 0$. *Then if we let* $\mathbf{w} = \sum_{j=1}^{n} c_j \mathbf{v}_j$, *the set* $\{\mathbf{v}_1, ..., \mathbf{v}_{n-1}, \mathbf{w}\}$ *is linearly independent.*

3.2 Basis and dimension

Definition 3.7 *Given a vector space* V *(or if* V *is a subspace of some other vector space), a set of vectors* $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *is called a* basis *of* V *if*

1. $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *is a linearly independent set, and*
2. $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *spans* V.

The most often used basis of F^n is the *standard basis* $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$, where $\mathbf{e}_k = (0, 0, ..., 0, 1, 0, ..., 0) \in F^n$ has a 1 in the k^{th} position and zeros elsewhere.

Theorem 3.8 (Unique Representation Theorem) *Given a basis* $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$, *every* $\mathbf{v} \in V$ *can be written*

$$\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{v}_j$$

where $c_j \in F$ *are uniquely chosen.*

We remark that if the set $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is assumed only to be linearly independent (but not necessarily a spanning set), then every vector can be written as a linear combination of the $\mathbf{v}_j$ in **at most one way**; if the set $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is assumed to be a spanning set (but possibly linearly dependent), then every vector can be written as a linear combination of the $\mathbf{v}_j$ in **at least one way**.

Theorem 3.9 (Exchange Lemma) *Suppose V is a vector space such that a finite list of vectors spans V. Then the length of any linearly independent list of vectors is less than or equal to the length of any spanning set of vectors.*

Theorem 3.10 (Dimension Theorem) *If $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is a basis of V, then any other basis of $\mathcal{B}$ must also consist of n vectors.*

Definition 3.11 *If V is spanned by a finite set of vectors, we say V is* finite dimensional *and write $\dim V < \infty$. In this case, the* dimension *of V is the number of elements in any basis of V. (We define $\dim\{\mathbf{0}\} = 0$ even though $\{\mathbf{0}\}$ does not have a basis.) If V is not spanned by any finite set of vectors, we say V is* infinite dimensional *and write $\dim V = \infty$.*

Definition 3.12 *Let A be an affine subspace of V. We define the* dimension *of A to be the dimension of the subspace $A - \mathbf{v}$, where $\mathbf{v} \in A$.*

Using this, we can characterize points, lines and planes: a point is an affine subspace of dimension zero, a line is an affine subspace of dimension one, and a plane is an affine subspace of dimension two.

Theorem 3.13 (Spanning Set Theorem) *If $V = Span(\mathbf{v}_1, ..., \mathbf{v}_n)$, then $\dim V \leq n$ (since some subset of $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ forms a basis of V).*

Theorem 3.14 (Linearly Independent Set Theorem) *If $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is a linearly independent set of vectors in V, then $n \leq \dim V$.*

Theorem 3.15 (Basis Extension Theorem) *If $\dim V < \infty$ and $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ is a linearly independent set of vectors in V, then there is a basis of V of the form*

$$\{\mathbf{v}_1, ..., \mathbf{v}_n, \mathbf{w}_1, ..., \mathbf{w}_m\}.$$

Theorem 3.16 (Basis Theorem) *Suppose $n = \dim V < \infty$. Then:*

1. *V has a basis (so long as $V \neq \{\mathbf{0}\}$).*
2. *Any set of n linearly independent vectors in V form a basis of V.*
3. *Any set of n vectors which span V form a basis of V.*

Theorem 3.17 *If W is a subspace of V, then $\dim W \leq \dim V$. If W is a subspace of V and $\dim W = \dim V < \infty$, then $W = V$.*

Subspaces of F^n can therefore be classified by their dimension. In particular:

Proposition 3.18 *The only subspaces of $\mathbb{R}^2$ are $\{\mathbf{0}\}$ (dimension zero), lines passing through the origin (dimension one), and all of $\mathbb{R}^2$ (dimension two).*

Corollary 3.19 *The only affine subspaces of $\mathbb{R}^2$ are points, lines, and all of $\mathbb{R}^2$.*

Proposition 3.20 *The only subspaces of $\mathbb{R}^3$ are $\{\mathbf{0}\}$ (dimension zero), lines passing through the origin (dimension one), planes passing through the origin (dimension two), and all of $\mathbb{R}^3$ (dimension three).*

Corollary 3.21 *The only affine subspaces of* $\mathbb{R}^3$ *are points, lines, planes, and all of* $\mathbb{R}^3$.

Theorem 3.22 *Let* W_1 *and* W_2 *be finite-dimensional subspaces of a vector space* V. *Then* $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \bigcap W_2)$.

4 Matrix theory

4.1 Matrix vocabulary

Definition 4.1 *Given a field* F, *a* matrix A *with entries in* F *is an array of numbers* a_{jk} *where* $1 \leq j \leq m$ *and* $1 \leq k \leq n$ *(the entry* a_{jk} *is in the* j^{th} *row and the* k^{th} *column of* A*). We say that the* size *or* order *of* A *is* $m \times n$ *where* m *is the number of rows and* n *is the number of columns of* A. *The set of* $m \times n$ *matrices with entries in* F *is denoted* $M_{mn}(F)$.

We think of a vector $\mathbf{x} = (x_1, ..., x_n) \in F^n$ as the $n \times 1$ matrix $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

If a matrix has the same number of rows as columns, the matrix is called *square*. The set of square $n \times n$ matrices with entries in F is denoted $M_n(F)$. The *diagonal entries* of a square matrix A are the numbers $a_{11}, a_{22}, ..., a_{nn}$. A matrix is called *diagonal* if all its non-diagonal entries are zero. The $n \times n$ *identity matrix*, denoted I or I_n, is the square matrix with diagonal entries equal to 1 and non-diagonal entries equal to 0. The *trace* of a matrix A, denoted $tr(A)$, is the sum of the diagonal entries in A.

We define addition and scalar multiplication on $M_{mn}(F)$ entry-wise; these operations make $M_{mn}(F)$ into a vector space over F of dimension mn. A commonly used basis of $M_{mn}(F)$ is $\{\Delta_{11}, \Delta_{12}, ..., \Delta_{mn}\}$ where Δ_{jk} has its $j, k-$entry equal to 1 and all its other entries equal to 0. The additive identity for this vector space is called the *zero matrix* (all the entries of this matrix are zero).

Definition 4.2 *Given* $A \in M_{mn}(F)$, *the* transpose *of* A, *denoted* A^T, *is the* $n \times m$ *matrix defined by* $(a^T)_{jk} = a_{kj}$. *If* $F = \mathbb{R}$ *or* $\mathbb{C}$, *the* conjugate *of* $A \in M_{mn}(F)$ *is the* $m \times n$ *matrix* $\overline{A}$ *defined by* $(\overline{a})_{jk} = \overline{(a_{jk})}$, *and the* Hermitian *of* $A \in M_{mn}(F)$ *is the* $n \times m$ *matrix* A^H *defined by* $A^H = \overline{A}^T = \overline{A^T}$.

Definition 4.3 *Let* F *be a field. Given* $A \in M_{mn}(F)$ *and* $B \in M_{pq}(F)$, *if* $n = p$ *then the* product $AB \in M_{mq}(F)$ *is defined by*

$$(ab)_{jk} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

In general, $AB \neq BA$ even if both products are defined and the same size. If it so happens that $AB = BA$, we say A and B *commute*.

Definition 4.4 *Let F be a field. A square matrix $A \in M_n(F)$ is called* invertible *if there is another matrix $A^{-1} \in M_n(F)$ such that $A\,A^{-1} = I$ and $A^{-1}\,A = I$.*

Theorem 4.5 *A 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$, in which case $A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.*

Proposition 4.6 (Properties of Matrix Operations) *Let all matrices in this proposition have entries in the same field F. Then, so long as everything is defined, the following statements always hold:*

1. $A(BC) = (AB)C$;
2. $A(B + C) = AB + AC$;
3. $r(AB) = (rA)B = A(rB)$ *for any scalar* $r \in F$;
4. $IA = A$ *and* $AI = A$;
5. $(A^T)^T = A$; $(A^H)^H = A$; $(A^T)^H = \overline{(A^H)^T} = \overline{A}$;
6. $tr(A^T) = tr(A)$; $tr(\overline{A}) = tr(A^H) = \overline{tr(A)}$;
7. $tr(A + B) = tr(A) + tr(B)$; $tr(AB) = tr(BA)$;
8. $(rA)^T = r(A^T)$ *and* $(rA)^H = \overline{r}(A^H)$ *for any* $r \in F$;
9. $(A_1A_2 \cdots A_n)^T = A_n^T A_{n-1}^T \cdots A_1^T$;
10. $(A_1A_2 \cdots A_n)^H = A_n^H A_{n-1}^H \cdots A_1^H$
11. $\overline{A_1A_2 \cdots A_n} = \overline{A_1}\ \overline{A_2}\ \cdots\ \overline{A_n}$
12. $(A_1 + \cdots + A_n)^T = A_1^T + \cdots + A_n^T$;
13. $\overline{A_1 + A_2 + \cdots + A_n} = \overline{A_1} + \overline{A_2} + \cdots + \overline{A_n}$
14. $(A_1 + \cdots + A_n)^H = A_1^H + \cdots + A_n^H$;
15. *If A is invertible, then $AB = AC$ implies $B = C$ (this does not hold in general if A is not invertible);*
16. $AB = 0$ *does not generally imply* $A = 0$ *or* $B = 0$;
17. *An invertible matrix has only one inverse;*
18. *If A is invertible, then A^{-1} is invertible and $(A^{-1})^{-1} = A$;*
19. *If $A_1, ..., A_n$ are invertible, then $A_1A_2 \cdots A_n$ is invertible and $(A_1 \cdots A_n)^{-1} = A_n^{-1}A_{n-1}^{-1} \cdots A_1^{-1}$;*
20. *If A is invertible, then A^T, $\overline{A}$ and A^H are invertible; in this case $(A^T)^{-1} = (A^{-1})^T$, $(\overline{A})^{-1} = \overline{A^{-1}}$ and $(A^H)^{-1} = (A^{-1})^H$.*

4.2 Special matrices

4.2.1 Triangular and diagonal matrices

Definition 4.7 *Given a matrix A, an entry a_{jk} of A is said to be* below the diagonal *if $j > k$, and is said to be* above the diagonal *if $j < k$. A matrix is called* upper triangular *if all its entries below the diagonal are zero; a matrix is called* lower triangular *if all its entries above the diagonal are zero. A matrix is called* triangular *if it is lower triangular or upper triangular.*

Note that a matrix is diagonal if and only if it is both upper triangular and lower triangular.

Theorem 4.8 *The set of diagonal $n \times n$ matrices is a subspace of $M_n(F)$ of dimension n. The set of upper triangular $n \times n$ matrices and the set of lower triangular $n \times n$ matrices are both subspaces of $M_n(F)$, each of dimension $\frac{1}{2}(n)(n+1)$.*

Proposition 4.9 *Any two diagonal matrices of the same size commute.*

4.2.2 Symmetric, skew-symmetric and Hermitian matrices

Definition 4.10 *A matrix $A \in M_n(F)$ is called* symmetric *if $A = A^T$. A matrix $A \in M_n(F)$ is called* skew-symmetric *if $A^T = -A$. A matrix $A \in M_n(\mathbb{C})$ is called* Hermitian *if $A = A^H$.*

Theorem 4.11 *The set $Sym_n(F)$ of symmetric $n \times n$ matrices is a subspace of $M_n(F)$ of dimension $\frac{n^2+n}{2}$. The set $Skew_n(F)$ of skew-symmetric $n \times n$ matrices is a subspace of $M_n(F)$ of dimension $\frac{n^2-n}{2}$.*

The set of Hermitian matrices do not form a subspace of $M_n(\mathbb{C})$ (taken as a vector space over $\mathbb{C}$) because the diagonal entries of a Hermitian matrix must be real, but they do form a subspace of $M_n(\mathbb{C})$ (taken as a vector space over $\mathbb{R}$).

Theorem 4.12 *Given any $A \in M_{mn}(F)$, both A^TA and AA^T are symmetric matrices. Given any square matrix A, $A + A^T$ is symmetric and $A - A^T$ is skew-symmetric. If A is skew-symmetric, then A^2 is symmetric. The only matrix which is both symmetric and skew-symmetric is the zero matrix.*

Theorem 4.13 *Every $n \times n$ matrix A can be written (uniquely) as the sum of a symmetric matrix and a skew-symmetric matrix (this can be done by writing $A = \frac{A+A^T}{2} + \frac{A-A^T}{2}$; the first matrix in the sum is symmetric and the second matrix is skew-symmetric).*

Theorem 4.14 *Given any $A \in M_{mn}(\mathbb{C})$, A^HA and AA^H are both Hermitian matrices. Given any square matrix A, $A + A^H$ is Hermitian and $A - A^H$ is skew-Hermitian (a matrix B is skew-Hermitian if $B^H = -B$).*

4.2.3 Positive definite matrices

Definition 4.15 *A square matrix $A \in M_n(\mathbb{C})$ is called* positive *if $\mathbf{v}^HA\mathbf{v} \geq 0$ for every $\mathbf{v} \in \mathbb{C}^n$. A square matrix $A \in M_n(\mathbb{C})$ is called* positive definite *if it is positive and if the only vector $\mathbf{v} \in \mathbb{C}^n$ satisfying $\mathbf{v}^HA\mathbf{v} = 0$ is the zero vector.*

Usually, one uses the adjectives “postive” and “positive definite” only to describe symmetric or Hermitian matrices. Indeed, the main applications of positive definite matrices involve only those which are symmetric or Hermitian.

4.2.4 Orthogonal and unitary matrices

Definition 4.16 *A matrix $Q \in M_n(\mathbb{R})$ is called* orthogonal *if Q is invertible and $Q^{-1} = Q^T$. The set of all $n \times n$ orthogonal matrices is called O_n.*

Proposition 4.17 *If Q is orthogonal, then so are Q^T and Q^{-1}. The product of any number of orthogonal matrices is an orthogonal matrix.*

Definition 4.18 *A matrix $U \in M_n(\mathbb{C})$ is called* unitary *if U is invertible and $U^{-1} = U^H$. The set of all $n \times n$ unitary matrices is called U_n.*

Proposition 4.19 *If U is unitary, then so are U^T, U^H, $\overline{U}$ and U^{-1}. The product of any number of unitary matrices is a unitary matrix.*

4.3 Fundamental subspaces of a matrix

Definition 4.20 *Let $A \in M_{mn}(F)$. Define the following subspaces associated to A:*

1. *The* column space *of A, denoted $C(A)$, is the span of the columns of A. This is a subspace of F^m.*
2. *The* row space *of A, denoted $R(A)$, is the span of the rows of A; this is a subspace of F^n.*
3. *The* null space *of A, denoted $N(A)$, is the set of vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$. This is a subspace of F^n.*
4. *The* left nullspace *of A, denoted $N(A^H)$, is the set of vectors $\mathbf{y}$ such that $A^H\mathbf{y} = \mathbf{0}$. This is a subspace of F^m.*

Theorem 4.21 *A vector $\mathbf{z} \in F^m$ is in the column space of A if and only if $\mathbf{z} = A\mathbf{x}$ for some $\mathbf{x} \in F^n$.*

5 Inner products and geometry

5.1 Inner products

Definition 5.1 *Let $F = \mathbb{R}$ or $\mathbb{C}$, and let V be a vector space over F. An* inner product *on V is a function $<,>: V \times V \to F$ satisfying:*

Conjugate symmetry $< \mathbf{v}, \mathbf{w} >= \overline{< \mathbf{w}, \mathbf{v} >}$ *for all $\mathbf{v}, \mathbf{w} \in V$.*

Linearity in first coordinate $< r\mathbf{v}, \mathbf{w} >= r < \mathbf{v}, \mathbf{w} >$ *and* $< \mathbf{v}_1 + \mathbf{v}_2, \mathbf{w} >=< \mathbf{v}_1, \mathbf{w} > + < \mathbf{v}_2, \mathbf{w} >$ *for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V$ and all $r \in F$.*

Positive definiteness $< \mathbf{v}, \mathbf{v} >\geq 0$ *for all $\mathbf{v} \in V$, and $< \mathbf{v}, \mathbf{v} >= 0$ only when $\mathbf{v} = \mathbf{0}$.*

An inner product space *is a vector space V together with an inner product $<,>$.*

Any inner product is also automatically *antilinear in the second coordinate*, i.e.

$$< \mathbf{v}, r\mathbf{w} >= \overline{r} < \mathbf{v}, \mathbf{w} > \text{ and } < \mathbf{v}, \mathbf{w}_1 + \mathbf{w}_2 >=< \mathbf{v}, \mathbf{w}_1 > + < \mathbf{v}, \mathbf{w}_2 >$$

for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V$ and all $r \in F$. More generally:

Theorem 5.2 *Let V be an inner product space. Then for any $c_1, ..., c_m, d_1, ..., d_n \in F$ and any $\mathbf{v}_1, ..., \mathbf{v}_m, \mathbf{w}_1, ..., \mathbf{w}_n \in V$, we have*

$$\left\langle \sum_{j=1}^{m} c_j \mathbf{v}_j, \sum_{k=1}^{n} d_k \mathbf{w}_k \right\rangle = \sum_{j=1}^{m} \sum_{k=1}^{n} c_j \overline{d_k} < \mathbf{v}_j, \mathbf{w}_k > .$$

There are two important examples of inner products:

Example 5.3 *$V = \mathbb{R}^n$; define for $\mathbf{v} = (v_1, ..., v_n)$ and $\mathbf{w} = (w_1, ..., w_n)$ the* dot product *of $\mathbf{v}$ and $\mathbf{w}$ to be*

$$\mathbf{v} \cdot \mathbf{w} =< \mathbf{v}, \mathbf{w} >= \mathbf{v}^T \mathbf{w} = \sum_{j=1}^{n} v_j w_j.$$

Euclidean $n-$dimensional space *is the vector space $\mathbb{R}^n$ endowed with the dot product.*

Example 5.4 *$V = \mathbb{C}^n$; define for $\mathbf{v} = (v_1, ..., v_n)$ and $\mathbf{w} = (w_1, ..., w_n)$ the* Hermitian inner product *of $\mathbf{v}$ and $\mathbf{w}$ to be*

$$< \mathbf{v}, \mathbf{w} >= \mathbf{v}^T \overline{\mathbf{w}} = \mathbf{w}^H \mathbf{v} = \sum_{j=1}^{n} v_j \overline{w_j}.$$

Hermitian $n-$dimensional space *is the vector space $\mathbb{C}^n$ endowed with the Hermitian inner product.*

In fact, every inner product on $\mathbb{R}^n$ is related to dot product, and every inner product on $\mathbb{C}^n$ is related to Hermitian inner product in the following sense:

Theorem 5.5 (Classification of inner products on $\mathbb{R}^n$) . *Let $<,>$ be an inner product on $\mathbb{R}^n$. Then there is a symmetric, positive definite matrix $A \in M_n(\mathbb{R})$ such that*

$$< \mathbf{v}, \mathbf{w} >= \mathbf{w}^T A \mathbf{v}$$

for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. (This matrix is obtained by setting $a_{ij} =< \mathbf{e}_j, \mathbf{e}_i >$.) Conversely, given any symmetric, positive definite matrix $A \in M_n(\mathbb{R})$, the formula

$$< \mathbf{v}, \mathbf{w} >= \mathbf{w}^T A \mathbf{v}$$

defines an inner product on $\mathbb{R}^n$.

Theorem 5.6 (Classification of inner products on $\mathbb{C}^n$) . *Let $<,>$ be an inner product on $\mathbb{C}^n$. Then there is a Hermitian, positive definite matrix $A \in M_n(\mathbb{C})$ such that*

$$< \mathbf{v}, \mathbf{w} >= \mathbf{w}^H A\mathbf{v}$$

for all $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$. (This matrix is obtained by setting $a_{ij} =< \mathbf{e}_j, \mathbf{e}_i >$.) Conversely, given any Hermitian, positive definite matrix $A \in M_n(\mathbb{C})$, the formula

$$< \mathbf{v}, \mathbf{w} >= \mathbf{w}^H A\mathbf{v}$$

defines an inner product on $\mathbb{C}^n$.

The standard inner products correspond to choosing $A = I$ in the above two theorems.

Theorem 5.7 (Dual relations) *Let $F = \mathbb{R}$ or $\mathbb{C}$ and let $<,>$ be the standard (dot or Hermitian) inner product. Then, for any $\mathbf{x} \in F^n$, any $\mathbf{y} \in F^m$ and any $A \in M_{mn}(F)$, we have*

$$< A\mathbf{x}, \mathbf{y} >=< \mathbf{x}, A^H\mathbf{y} >$$

(similarly, for any $B \in M_{nm}(F)$, we have $< \mathbf{x}, B\mathbf{y} >=< B^H\mathbf{x}, \mathbf{y} >$).

5.2 Norms and distances

Definition 5.8 *Given an inner product space V, we define the* length *a.k.a.* norm *a.k.a.* absolute value *a.k.a.* magnitude *of a vector $\mathbf{v}$ (associated to the inner product) to be*

$$||\mathbf{v}|| = \sqrt{< \mathbf{v}, \mathbf{v} >}.$$

The distance *between two vectors $\mathbf{v}$ and $\mathbf{w}$ is $dist(\mathbf{v}, \mathbf{w}) = ||\mathbf{v} - \mathbf{w}||$. A vector is called a* unit vector *if it has norm 1.*

Proposition 5.9 *Let V be an inner product space with associated norm $|| \cdot ||$. Then:*

1. $||\mathbf{v}|| \geq 0$ *for all* $\mathbf{v} \in V$.
2. $||\mathbf{v}|| = 0$ *only when* $\mathbf{v} = \mathbf{0}$.
3. $||k\mathbf{v}|| = |r|\, ||\mathbf{v}||$ *for all* $k \in F$.
4. *Given any nonzero vector $\mathbf{v} \in V$, there is a unit vector in the same direction as $\mathbf{v}$ (namely, $\frac{1}{||\mathbf{v}||}\mathbf{v}$) called a* normalized version *of $\mathbf{v}$.*

Proposition 5.10 *Let V be an inner product space with associated distance function dist. Then for all $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w} \in V$:*

1. $dist(\mathbf{v}, \mathbf{w}) \geq 0$
2. $dist(\mathbf{v}, \mathbf{w}) = 0$ *only when* $\mathbf{v} = \mathbf{w}$
3. $dist(\mathbf{v}, \mathbf{w}) = dist(\mathbf{w}, \mathbf{v})$
4. $dist(\mathbf{u}, \mathbf{w}) \leq dist(\mathbf{u}, \mathbf{v}) + dist(\mathbf{v}, \mathbf{w})$
5. $dist(r\mathbf{u}, r\mathbf{v}) = |r|\, dist(\mathbf{u}, \mathbf{v})$

We see that given any inner product, one obtains a natural notion of length (norm) and distance on the inner product space. In fact, given any natural notion of distance (one that satisfies (1) to (5) in the preceding proposition), one can obtain a norm by setting the norm of a vector equal to its distance from zero. Further still, given a norm, one can show that there is an inner product that generates the norm by applying the following result:

Theorem 5.11 (Polarization Identities) *Let V be an inner product space with associated norm $||\cdot||$. Then for all $\mathbf{v}, \mathbf{w} \in V$:*

1. $\Re(<\mathbf{v}, \mathbf{w}>) = \frac{1}{4}\left(||\mathbf{v}+\mathbf{w}||^2 - ||\mathbf{v}-\mathbf{w}||^2\right)$.
2. $\Im(<\mathbf{v}, \mathbf{w}>) = \frac{1}{4}\left(||\mathbf{v}+i\mathbf{w}||^2 - ||\mathbf{v}-i\mathbf{w}||^2\right)$.

Restated, given a norm $||\cdot||$ on V, the norm is associated to the inner product defined by

$$<\mathbf{v}, \mathbf{w}>= \frac{||\mathbf{v}+\mathbf{w}||^2 - ||\mathbf{v}-\mathbf{w}||^2}{4} + i\left(\frac{||\mathbf{v}+i\mathbf{w}||^2 - ||\mathbf{v}-i\mathbf{w}||^2}{4}\right).$$

Proposition 5.12 *Let $A \in M_n(\mathbb{C})$ and let $<,>$ be any inner product on $\mathbb{C}^n$. If $<A\mathbf{x}, \mathbf{x}>= 0$ for every $\mathbf{x} \in \mathbb{C}^n$, then $A = 0$.*

Theorem 5.13 (Parallelogram Law) *Let V be an inner product space with associated norm $||\cdot||$. Then for all $\mathbf{v}, \mathbf{w} \in V$, $||\mathbf{v}+\mathbf{w}||^2 + ||\mathbf{v}-\mathbf{w}||^2 = 2\left(||\mathbf{v}||^2 + ||\mathbf{w}||^2\right)$.*

5.3 Orthogonality

Definition 5.14 *Let V be an inner product space. We say two vectors $\mathbf{v}, \mathbf{w} \in V$ are* orthogonal *(and write $\mathbf{v} \perp \mathbf{w}$) if $<\mathbf{v}, \mathbf{w}>= 0$.*

The zero vector is orthogonal to every vector in a vector space; it is the only vector orthogonal to every vector in a vector space. The zero vector is the only vector orthogonal to itself.

Theorem 5.15 (Pythagorean Theorem) *Let V be an inner product space and suppose $\mathbf{v} \perp \mathbf{w}$. Then $||\mathbf{v}+\mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2$.*

The converse of the Pythagorean theorem holds if V is a vector space over $\mathbb{R}$, but not in general if V is a complex inner product space.

5.4 Orthogonal complements and projections

Definition 5.16 *Let V be an inner product space. Given a subspace $W \subseteq V$, define $W^\perp$, the* orthogonal complement of *W to be the set of vectors orthogonal to every $\mathbf{w} \in W$.*

If V is an inner product space, then $V^\perp = \{\mathbf{0}\}$ and $\{\mathbf{0}\}^\perp = V$.

Theorem 5.17 *Let V be an inner product space. For any subspace $W \subseteq V$, $W^\perp$ is also a subspace of V.*

As a special case of the preceding theorem, we see that if $\mathbf{v} \perp \mathbf{w}$, then $r\mathbf{v} \perp \mathbf{w}$ for all $r \in F$ and that if $\mathbf{v} \perp \mathbf{w}$ and $\mathbf{x} \perp \mathbf{w}$, then $(\mathbf{v}+\mathbf{x}) \perp \mathbf{w}$.

Proposition 5.18 *Let V be an inner product space and let W be a subspace of V defined by $W = Span(\mathbf{w}_1, ..., \mathbf{w}_n)$. A vector $\mathbf{v}$ is in $W^\perp$ if and only if $<\mathbf{v}, \mathbf{w}_j>= 0$ for all j.*

Theorem 5.19 *Let V be a finite-dimensional inner product space and let W be a subspace of V. Then $W \bigcap W^{\perp} = \{\mathbf{0}\}$.*

Theorem 5.20 *Let V be a finite-dimensional inner product space and let W be a subspace of V. Then $(W^{\perp})^{\perp} = W$.*

If V is infinite-dimensional, we know only that $W \subseteq (W^{\perp})^{\perp}$.

Theorem 5.21 *Let V be a finite-dimensional inner product space. Given any basis $\mathcal{B}$ of W and any basis $B^{\perp}$ of $W^{\perp}$, $\mathcal{B} \cup \mathcal{B}^{\perp}$ is a basis of W.*

Theorem 5.22 *Let V be a finite-dimensional inner product space and let W be a subspace of V. Then*

$$\dim(V) = \dim(W) + \dim(W^{\perp}).$$

Theorem 5.23 (Orthogonal Decomposition Theorem (dimension 1)) *Let V be an inner product space and let $W = Span(\mathbf{w})$ for some nonzero $\mathbf{w} \in V$. Then every vector $\mathbf{v}$ can be written uniquely as $\mathbf{v} = \mathbf{v}^W + \mathbf{v}^{\perp}$ where $\mathbf{v}^W \,||\, \mathbf{w}$ and $\mathbf{v}^{\perp} \perp \mathbf{w}$.*

In particular, the $\mathbf{v}^W$ in the above theorem is the projection of $\mathbf{v}$ onto $\mathbf{w}$ (see below).

Theorem 5.24 (Orthogonal Decomposition Theorem (general case)) *Let V be an inner product space and let W be a subspace of V with $\dim W < \infty$. Then every vector $\mathbf{v}$ can be written uniquely as $\mathbf{v} = \mathbf{v}^W + \mathbf{v}^{\perp}$ where $\mathbf{v}^W \in W$ and $\mathbf{v}^{\perp} \in W^{\perp}$.*

We call the vector $\mathbf{v}^W$ in the above theorem the *projection of* $\mathbf{v}$ *onto* W and the vector $\mathbf{v}^{\perp}$ the *component of* $\mathbf{v}$ *orthogonal to* W.

Definition 5.25 *Let V be an inner product space and let $W \subseteq V$ be a subspace. Given any $\mathbf{v} \in V$, we let the* distance from $\mathbf{v}$ to W, *denoted by $dist(\mathbf{v}, W)$, be the minimum distance from $\mathbf{v}$ to any vector in W.*

Theorem 5.26 *Let V be an inner product space and let W be a finite-dimensional subspace. Then $dist(\mathbf{v}, W) = dist(\mathbf{v}, \mathbf{v}^W) = ||\mathbf{v}^{\perp}||$.*

Definition 5.27 *Let $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$ be two vectors in $\mathbb{R}^3$, endowed with the usual inner product. Define the* cross product *of $\mathbf{a}$ and $\mathbf{b}$, denoted $\mathbf{a} \times \mathbf{b}$, to be*

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

In particular, the cross product of two vectors in $\mathbb{R}^3$ is itself a vector in $\mathbb{R}^3$.

Theorem 5.28 (Properties of cross product) *Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$. Then*

1. *$\mathbf{a}||\mathbf{b}$ if and only if $\mathbf{a} \times \mathbf{b} = \mathbf{0}$;*
2. *$(\mathbf{a} \times \mathbf{b}) \perp \mathbf{a}$ and $(\mathbf{a} \times \mathbf{b}) \perp \mathbf{b}$;*
3. *$\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$;*
4. *$(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \times \mathbf{c}) + (\mathbf{b} \times \mathbf{c})$ and $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c})$;*
5. *$r\mathbf{a} \times \mathbf{b} = r(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times r\mathbf{b}$ for any $r \in \mathbb{R}$;*

6. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are not parallel, then* $(Span(\mathbf{a},\mathbf{b}))^{\perp} = Span(\mathbf{a}\times\mathbf{b})$;
7. $||\mathbf{a}\times\mathbf{b}|| = ||\mathbf{a}||\,||\mathbf{b}||\sin\theta$ *where* θ *is the angle between* $\mathbf{a}$ *and* $\mathbf{b}$ *(see subsection 4.3 below for a definition of angle).*

Definition 5.29 *Let* V *be an inner product space and let* $\mathbf{w}\in V$ *be a nonzero vector. Given any* $\mathbf{v}\in V$, *define the* projection of $\mathbf{v}$ onto $\mathbf{w}$ *to be*

$$proj_{\mathbf{w}}\mathbf{v} = \frac{<\mathbf{v},\mathbf{w}>}{<\mathbf{w},\mathbf{w}>}\mathbf{w}.$$

Proposition 5.30 *Let* V *be an inner product space with associated norm* $||\cdot||$ *and let* $\mathbf{w}\in V$ *be a nonzero vector. Denote by* $\mathbf{u}$ *the normalized version of* $\mathbf{w}$ *(i.e. a unit vector in the direction of* $\mathbf{w}$*). Then for any* $\mathbf{v}\in V$,
1. $proj_{\mathbf{w}}\mathbf{v} = \frac{<\mathbf{v},\mathbf{w}>}{||\mathbf{w}||^2}\mathbf{w} = \frac{<\mathbf{v},\mathbf{w}>}{||\mathbf{w}||}\mathbf{u} =< \mathbf{v},\mathbf{u}> \mathbf{u}$.
2. $\mathbf{v}\perp\mathbf{w}$ *if and only if* $proj_{\mathbf{w}}\mathbf{v} = \mathbf{0}$.
3. $\mathbf{v}\,||\,\mathbf{w}$ *if and only if* $proj_{\mathbf{w}}\mathbf{v} = \mathbf{v}$.
4. $proj_{\mathbf{w}}\mathbf{v}\perp(\mathbf{v}-proj_{\mathbf{w}}\mathbf{v})$.

5.5 Orthonormal bases and the Gram-Schmidt procedure

Definition 5.31 *Let* V *be an inner product space. A set* $\{\mathbf{v}_1,...,\mathbf{v}_n\}$ *of vectors in* V *is called* (pairwise) orthogonal *if* $\mathbf{v}_i\perp\mathbf{v}_j$ *for all* $i\neq j$. *The set is called* orthonormal *if it is orthogonal and* $||\mathbf{v}_j|| = 1$ *for all* j.

Theorem 5.32 (Gram-Schmidt) *Given a basis* $\{\mathbf{v}_1,...,\mathbf{v}_n\}$ *of inner product space* V, *there is a basis* $\{\mathbf{x}_1,...,\mathbf{x}_n\}$ *of* V *such that:*
1. $\{\mathbf{x}_1,...,\mathbf{x}_n\}$ *is an orthonormal basis, and*
2. $Span(\mathbf{v}_1,...,\mathbf{x}_k) = Span(\mathbf{v}_1,...,\mathbf{v}_k)$ *for all* $k\leq n$.

The procedure to produce the $\mathbf{x}$'s from the $\mathbf{v}$'s goes in two steps. First, define

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{x}_1\\
\mathbf{w}_2 &= \mathbf{x}_2 - proj_{\mathbf{w}_1}\mathbf{x}_2\\
\mathbf{w}_3 &= \mathbf{x}_3 - proj_{\mathbf{w}_1}\mathbf{x}_3 - proj_{\mathbf{w}_2}\mathbf{x}_3\\
\vdots &\quad\ \vdots
\end{aligned}$$

to obtain an orthogonal basis $\{\mathbf{w}_1,...,\mathbf{w}_n\}$; then set $\mathbf{x}_j = \frac{\mathbf{w}_j}{||\mathbf{w}_j||}$ for each j to obtain the orthonormal basis.

Corollary 5.33 *Every finite-dimensional inner product space has an orthonormal basis.*

Theorem 5.34 *Let* $\{\mathbf{x}_1,...,\mathbf{x}_n\}$ *be an orthonormal basis of inner product space* V. *Then for every* $\mathbf{v}\in V$, *we have*

$$\mathbf{v} = \sum_{j=1}^{n} < \mathbf{v},\mathbf{x}_j > \mathbf{x}_j.$$

Theorem 5.35 *Let V be an inner product space and let $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ be an orthonormal basis of a finite-dimensional subspace $W \subseteq V$. Then, for every $\mathbf{v} \in V$, the projection of $\mathbf{v}$ onto W is*

$$\mathbf{v}^W = \sum_{j=1}^{n} < \mathbf{v}, \mathbf{x}_j > \mathbf{x}_j.$$

5.6 Angles

Theorem 5.36 (Cauchy-Schwarz Inequality) *Let V be an inner product space with associated norm $|| \cdot ||$. Then for all $\mathbf{v}, \mathbf{w} \in V$,*

$$| < \mathbf{v}, \mathbf{w} > | \leq ||\mathbf{v}|| \, ||\mathbf{w}||.$$

(Equality in the above expression holds only when $\mathbf{v} \, || \, \mathbf{w}$.)

Theorem 5.37 (Triangle Inequality) *Let V be an inner product space with associated norm $|| \cdot ||$. Then for all $\mathbf{v}, \mathbf{w} \in V$, $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$.*

Corollary 5.38 (Generalized Triangle Inequality) *Let V be an inner product space with associated norm $|| \cdot ||$. Then for all $\mathbf{v}_1, ..., \mathbf{v}_n \in V$, $||\mathbf{v}_1 + ... + \mathbf{v}_n|| \leq ||\mathbf{v}_1|| + ... + ||\mathbf{v}_n||$.*

Definition 5.39 *Let V be an inner product space over $\mathbb{R}$. Then for all nonzero $\mathbf{v}, \mathbf{w} \in V$, the* angle between $\mathbf{v}$ and $\mathbf{w}$ is

$$\theta = \cos^{-1} \left(\frac{< \mathbf{v}, \mathbf{w} >}{||\mathbf{v}|| \, ||\mathbf{w}||} \right).$$

Rewritten, this definition says that $< \mathbf{v}, \mathbf{w} >= ||\mathbf{v}|| \, ||\mathbf{w}|| \cos \theta$.

By definition we decree the angle between two vectors to be at least 0 and at most π. The angle between parallel vectors is zero (if they are in the same direction) or π (if they are in opposite directions).

Theorem 5.40 (Law of Cosines) *Let V be an inner product space over $\mathbb{R}$ and let $\mathbf{v}, \mathbf{w} \in V$. Then*

$$||\mathbf{v} - \mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2 - 2||\mathbf{v}|| \, ||\mathbf{w}|| \cos \theta$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$.

5.7 Hyperplanes

Definition 5.41 *Given a finite-dimensional vector space V, a* hyperplane W *is an affine subspace of V satisfying* $\dim W = \dim V - 1$.

If $V = F$, taken as a vector space over itself, hyperplanes in V are points. When $\dim V = 2$, hyperplanes in V are lines. When $\dim V = 3$, hyperplanes in V are planes.

Theorem 5.42 *Let V be a finite-dimensional inner product space. Then given any hyperplane W, there is a vector $\mathbf{n} \in V$, called a* normal vector *to the hyperplane, and a scalar d such that*

$$\mathbf{x} \in W \;\Leftrightarrow < \mathbf{n}, \mathbf{x} >= d.$$

The equation $< \mathbf{n}, \mathbf{x} >= d$ is called the normal equation *of the hyperplane.*

As described earlier, every plane in $\mathbb{R}^3$ has a normal equation $ax + by + cz = d$ for constants a, b, c and d; the normal vector $\mathbf{n} = (a, b, c)$ can be found by taking the cross product of any two nonzero vectors in the plane.

Proposition 5.43 *The distance from a point $\mathbf{v}$ to a hyperplane passing through $\mathbf{0}$ with normal vector $\mathbf{n}$ is $\frac{|<\mathbf{n},\mathbf{v}>|}{||\mathbf{n}||}$.*

5.8 More on orthogonal and unitary matrices

Theorem 5.44 (Characterization of orthogonal matrices) *Let $Q \in M_n(\mathbb{R})$ and let $<,>$ be the standard inner product on $\mathbb{R}^n$. The following are equivalent:*

1. *Q is orthogonal;*
2. *the columns of Q form an orthonormal basis of $\mathbb{R}^n$;*
3. *the rows of Q form an orthonormal basis of $\mathbb{R}^n$;*
4. *Q "preserves dot product", i.e. $< Q\mathbf{x}, Q\mathbf{y} >=< \mathbf{x}, \mathbf{y} >$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$;*
5. *Q "preserves Euclidean angles", i.e. for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the angle between $\mathbf{x}$ and $\mathbf{y}$ is the same as the angle between $Q\mathbf{x}$ and $Q\mathbf{y}$;*
6. *Q "preserves Euclidean norms", i.e. for every $\mathbf{x} \in \mathbb{R}^n$, $||Q\mathbf{x}|| = ||\mathbf{x}||$;*
7. *Q "preserves Euclidean distances", i.e. for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the distance between $\mathbf{x}$ and $\mathbf{y}$ is the same as the distance between $Q\mathbf{x}$ and $Q\mathbf{y}$;*
8. *for any orthonormal set $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ of vectors in $\mathbb{R}^n$, the set $\{Q\mathbf{x}_1, ..., Q\mathbf{x}_k\}$ is also orthonormal;*
9. *for any orthonormal basis $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ of $\mathbb{R}^n$, the set $\{Q\mathbf{x}_1, ..., Q\mathbf{x}_n\}$ is also an orthonormal basis.*

Theorem 5.45 (Classification of 2×2 orthogonal matrices) *$Q \in O_2$ if and only if Q has one of the following two forms:*

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad \textit{or} \qquad \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Matrices in the first class are counterclockwise rotations by θ; matrices in the second class are reflections through a line passing through the origin which sits at an angle $\theta/2$ to the $x-$axis. $n \times n$ orthogonal matrices can be thought of as "$n-$dimensional rotations and reflections" although they are actually more complicated; they can be used to describe a notion of congruence of arbitrary Euclidean objects as follows:

Definition 5.46 *Two subsets A and B of $\mathbb{R}^n$ are* congruent *if there is an orthogonal matrix $Q \in O_n$ and a vector $\mathbf{v} \in \mathbb{R}^n$ such that $QA + \mathbf{v} = B$.*

Theorem 5.47 (Characterization of unitary matrices) *Let $U \in M_n(\mathbb{C})$ and let $<,>$ be the standard Hermitian inner product on $\mathbb{C}^n$. The following are equivalent:*

1. *U is unitary;*
2. *the columns of U form an orthonormal basis of $\mathbb{C}^n$;*
3. *the rows of U form an orthonormal basis of $\mathbb{C}^n$;*
4. *U "preserves Hermitian inner product", i.e. $< U\mathbf{x}, U\mathbf{y} >=< \mathbf{x}, \mathbf{y} >$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$;*
5. *U "preserves norms", i.e. for every $\mathbf{x} \in \mathbb{C}^n$, $||U\mathbf{x}|| = ||\mathbf{x}||$;*
6. *U "preserves distances", i.e. for every $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, the distance between $\mathbf{x}$ and $\mathbf{y}$ is the same as the distance between $U\mathbf{x}$ and $U\mathbf{y}$;*
7. *for any orthonormal set $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ of vectors in $\mathbb{C}^n$, the set $\{U\mathbf{x}_1, ..., U\mathbf{x}_k\}$ is also orthonormal;*
8. *for any orthonormal basis $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ of $\mathbb{C}^n$, the set $\{U\mathbf{x}_1, ..., U\mathbf{x}_n\}$ is also an orthonormal basis.*

6 Systems of linear equations

Definition 6.1 *Let F be a field. A* linear equation *in n variables is any equation of the form*

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

for constants $a_1, .., a_n, b \in F$.

Definition 6.2 *A* system of m linear equations in n variables *is any system which can be expressed in the following three equivalent forms:*
Equation form *As a list of m linear equations, i.e.*

$$\left\{ \begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & + & \cdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \right.$$

Matrix form *As a matrix equation* $A\mathbf{x} = \mathbf{b}$ *where the matrix* $A = (a_{ij})$ *is called the* coefficient matrix *of the system.*

Vector form *As a vector equation, i.e.* $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ *where* $\boldsymbol{a}_j$ *is the* j^{th} *column of* A*.*

A vector $\mathbf{x} \in F^n$ *is called a solution of the system if* $A\mathbf{x} = \mathbf{b}$*. A system is called* consistent *if it has at least one solution and* inconsistent *otherwise; the* solution set *of the system is the set of all solutions of the system. Two systems are called* equivalent *if they have the same solution set.*

Theorem 6.3 *Given any consistent system of linear equations, the solution set forms an affine subspace of* F^n*.*

Systems of linear equations are solved in practice by performing row operations on the *augmented matrix* $(A\,|\,\mathbf{b})$ of the system. The allowable row operations are:

1. Switching two rows of the matrix;
2. Multiplying a row by a nonzero scalar;
3. Adding a nonzero multiple of one row to another.

Each of these operations corresponds to multiplying the matrix by an "elementary matrix"; since all elementary matrices are invertible, all row operations are reversible. In particular:

- if B is obtained by switching rows i and j of A, then $B = E_{i\leftrightarrow j}A$ where $E_{i\leftrightarrow j} =$

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccc} &&&i&&&&j&&& \end{array} \\
\begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} &
\begin{pmatrix}
1 &&&&&&&&&& \\
& \ddots &&&&&&&&& \\
&& 1 &&&&&&&& \\
&&& 0 &&&& 1 &&& \\
&&&& 1 &&&&&& \\
&&&&& \ddots &&&&& \\
&&&&&& 1 &&&& \\
&&& 1 &&&& 0 &&& \\
&&&&&&&& 1 && \\
&&&&&&&&& \ddots & \\
&&&&&&&&&& 1
\end{pmatrix}
\end{array}
$$

- if B is obtained by multiplying row j by the scalar r, then $B = E_{j;r}A$ where $E_{j;r} =$

$$\begin{array}{c@{\quad}c}
\begin{array}{c} \\ \\ \\ j \\ \\ \\ \\ \end{array} &
\left(\begin{array}{ccccccc}
1 &&&&&& \\
& \ddots &&&&& \\
&& 1 &&&& \\
&&& r &&& \\
&&&& 1 && \\
&&&&& \ddots & \\
&&&&&& 1
\end{array}\right)
\end{array}$$

- if B is obtained by adding r times row i to row j of A, then $B = E_{i,j;r}A$ where $E_{i,j;r} =$

$$\begin{array}{c@{\quad}c}
& \begin{array}{ccccccccccc} & & & i & & & & & & & \end{array} \\
\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} &
\left(\begin{array}{ccccccccccc}
1 &&&&&&&&&& \\
& \ddots &&&&&&&&& \\
&& 1 &&&&&&&& \\
&&& 1 &&&&&&& \\
&&&& 1 &&&&&& \\
&&&&& \ddots &&&&& \\
&&&&&& 1 &&&& \\
&&& r &&&& 1 &&& \\
&&&&&&&& 1 && \\
&&&&&&&&& \ddots & \\
&&&&&&&&&& 1
\end{array}\right)
\end{array}$$

- the inverses of each of these classes of matrices are: $E^{-1}_{i\leftrightarrow j} = E_{i\leftrightarrow j}$; $E^{-1}_{j;r} = E_{j;r^{-1}}$; $E^{-1}_{i,j;r} = E_{i,j;-r}$.

Definition 6.4 *Two matrices A and B are called* row equivalent *if one matrix can be transformed into the other by a sequence of row operations.*

Theorem 6.5 *If two systems of equations have row equivalent augmented matrices, then the systems are equivalent.*

Proposition 6.6 *The following are equivalent:*

1. *Matrices A and B are row equivalent;*
2. *There is a list $E_1, E_2, ..., E_k$ of elementary matrices such that $E_k E_{k-1} E_2 \cdots E_1 A = B$;*
3. *There is an invertible matrix E such that $EA = B$.*

Theorem 6.7 *A square matrix A is invertible if and only if A is row equivalent to the identity matrix.*

In this case, the same sequence of row operations which transform A into I transform I into A^{-1} (this method of finding A^{-1} is called the *Gauss-Jordan method*).

Definition 6.8 *A matrix is in* (row-echelon form *if:*
1. *All rows of zeros are at the bottom of the matrix.*
2. *Defining the* leading entry *or* pivot *of a row to be the first nonzero entry of that row, the leading entry of any row is to the right of the leading entry of any above row.*

A matrix is in reduced row-echelon form *if it is in row-echelon form, all pivots are 1, and each pivot is the only nonzero entry in its column.*

Every matrix is row equivalent to one and only one reduced row-echelon form.

Definition 6.9 *Given a matrix A, the* pivot columns *of A are the columns which have a leading entry in any row-echelon form of A. The* free columns *are the columns which are not pivot columns.*

Theorem 6.10 (Rank Theorem) *Let $A \in M_{mn}(F)$. Then the following quantities are all equal to the same number, called the* rank *of A and denoted $r(A)$:*
1. $\dim C(A)$
2. $\dim R(A)$
3. *The number of pivot columns of A*
4. *The number of nonzero rows in any row-echelon form of A.*

Proposition 6.11 *Let $A \in M_{mn}(F)$. Then $r(A) = r(A^T)$ and if $F = \mathbb{C}$, $r(A^H) = r(\overline{A}) = r(A)$.*

In particular, $r(A) \leq m$ and $r(A) \leq n$ if A is $m \times n$. $r(A) = m$ if and only if A has a pivot in every row, and $r(A) = n$ if and only if A has a pivot in every column.

Theorem 6.12 *Let $A \in M_{mn}(F)$. Then a basis for $C(A)$ consists of the pivot columns of A, and a basis for $R(A)$ consists of the nonzero rows of any echelon form of A.*

Theorem 6.13 (Rank-Nullty Theorem) *Let $A \in M_{mn}(F)$. Then:*
1. $\dim C(A) + \dim N(A^H) = m$ *and*
2. $\dim R(A) + \dim N(A) = n$.

In other words, if A has rank $r = r(A)$, then the null space of A has dimension $n - r$ and the left nullspace of A has dimension $m - r$.

Theorem 6.14 (Fundamental Theorem of Linear Algebra) *Let $A \in M_{mn}(F)$, where $F = \mathbb{R}$ or $\mathbb{C}$. Then (with respect to the usual inner product):*

$$[C(A)]^{\perp} = N(A^H) \text{ and } [R(A)]^{\perp} = N(A).$$

Theorem 6.15 (Summary of Theory of Systems of Linear Equations) *Let $A \in M_{mn}(F)$ have rank r. Then:*

1. *The system* $A\mathbf{x} = \mathbf{0}$ *always has at least one solution (namely* $\mathbf{x} = \mathbf{0}$*). There are two possible situations, (a) or (b):*
   (a) $A\mathbf{x} = \mathbf{0}$ *has more than one solution. This is equivalent to all of the following:*
      - $N(A) \neq \{\mathbf{0}\}$
      - $\dim N(A) \geq 1$
      - $r < n$.

      In this case, for any $\mathbf{b} \in F^n$*:*
      - $A\mathbf{x} = \mathbf{b}$ *has no solution if and only if* $\mathbf{b} \notin C(A)$*;*
      - $A\mathbf{x} = \mathbf{b}$ *has infinitely many solutions if and only if* $\mathbf{b} \in C(A)$ *(in this case the solution set of* $A\mathbf{x} = \mathbf{b}$ *is* $\mathbf{x}_p + N(A)$ *where* $\mathbf{x}_p$ *is any particular solution of the system) (this case is assured if* $r = m < n$*);*
      - $A\mathbf{x} = \mathbf{b}$ *never has exactly one solution.*

   (b) $A\mathbf{x} = \mathbf{0}$ *has exactly one solution (only* $\mathbf{x} = \mathbf{0}$*). This is equivalent to all of the following:*
      - $N(A) = \{\mathbf{0}\}$
      - $\dim N(A) = 0$
      - $r = n$.

      In this case, for any $\mathbf{b} \in F^n$*:*
      - $A\mathbf{x} = \mathbf{b}$ *has no solution if and only if* $\mathbf{b} \notin C(A)$*;*
      - $A\mathbf{x} = \mathbf{b}$ *has exactly one solution if and only if* $\mathbf{b} \in C(A)$ *(this is assured if* $r = m = n$*; see below);*
      - $A\mathbf{x} = \mathbf{b}$ *never has infinitely many solutions.*

      Notice that if $m < n$ *(that is, there are fewer variables than equations), case (b) above is impossible because* $r \leq m$ *(so* r *cannot be equal to* n*).*
2. *In the special case where* $r = m = n$ *(i.e. A is square and has* full rank*), then* $C(A) = R(A) = F^n$ *and* $N(A) = N(A^H) = \{\mathbf{0}\}$*. Furthermore, A is invertible and for every* $\mathbf{b} \in F^n$*, the system* $A\mathbf{x} = \mathbf{b}$ *has exactly one solution, namely* $\mathbf{x} = A^{-1}\mathbf{b}$.
3. *The system* $A\mathbf{x} = \mathbf{b}$ *has no solution if and only if* $\mathbf{b} \notin C(A)$ *if and only if an echelon form of the augmented matrix* $(A \,|\, \mathbf{b})$ *contains a* false row *of the form* $(0\ 0\ \cdots\ 0\ z)$ *where* $z \neq 0$.

Theorem 6.16 *Let* $A \in M_{mn}(F)$ *and* $B \in M_{np}(F)$*. Then the rank of AB is at most the rank of B, and the rank of AB is at most the rank of A.*

Theorem 6.17 *Let* $A \in M_{mn}(F)$*. If E is any* $n \times n$ *invertible matrix, then* $r(AE) = r(A)$*. If E is any* $m \times m$ *invertible matrix, then* $r(EA) = r(A)$*.*

7 Linear transformations

Definition 7.1 *Given vector spaces V_1 and V_2 over the same field F, a function $T : V_1 \to V_2$ is called a* linear transformation *if for all $r \in F$ and all $\mathbf{v}, \mathbf{w} \in V_1$, $T(r\mathbf{v}) = rT(\mathbf{v})$ and $T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v})+T(\mathbf{w})$. The set of linear transformations from V_1 to V_2 is called $L(V_1, V_2)$ or $Hom(V_1, V_2)$.*

To describe a linear transformation, it is sufficient to give its values on a basis of V_1. This is because given any basis $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of V_1, we can write any vector $\mathbf{x} \in V_1$ as $\mathbf{x} = \sum_j c_j \mathbf{v}_j$ by the Unique Representation Theorem. Then $T(\mathbf{x}) = \sum_j c_j T(\mathbf{v}_j)$.

Definition 7.2 *Given vector spaces V_1 and V_2 over the same field F, a function $A : V_1 \to V_2$ is called an* affine transformation *if $A(\mathbf{v}) = T(\mathbf{v}) + \mathbf{b}$ for some $T \in L(V_1, V_2)$ and some $\mathbf{b} \in V_2$.*

Classical examples of linear transformations include reflections, rotations, projection, differentiation, integration, evaluation of functions, taking the inner product with a fixed vector in the second input, and taking the transpose of a matrix. The most important example is multiplication by a matrix: every $A \in M_{mn}(F)$ defines a linear transformation $T : F^n \to F^m$ by $T(\mathbf{x}) = A\mathbf{x}$.

Theorem 7.3 *If $T \in L(V_1, V_2)$, then $T(\mathbf{0}) = \mathbf{0}$.*

Theorem 7.4 *$T : V_1 \to V_2$ is linear if and only if $T(r\mathbf{v} + \mathbf{w}) = rT(\mathbf{v}) + T(\mathbf{w})$ for every $r \in F$ and every $\mathbf{v}, \mathbf{w} \in V_1$.*

Theorem 7.5 *Let V_1 and V_2 be vector spaces over the same field F. The set $L(V_1, V_2)$ is itself a vector space over F, with addition and scalar multiplication defined by $(S+T)(\mathbf{x}) = S(\mathbf{x}) + T(\mathbf{x})$ and $(rT)(\mathbf{x}) = r\,T(\mathbf{x})$.*

In particular, the additive identity element of $L(V_1, V_2)$ is the constant function $T(\mathbf{x}) = \mathbf{0}$. If $\dim V_1 = m$ and $\dim V_2 = n$, then $\dim L(V_1, V_2) = mn$.

Theorem 7.6 *Let F be a field and let V_1, V_2 and V_3 be vector spaces over F. If $T_1 \in L(V_1, V_2)$ and $T_2 \in L(V_2, V_3)$, then the composition (a.k.a. product) of the two transformations, denoted $T_2 \circ T_1 = T_2 T_1 : V_1 \to V_3$, is linear.*

Definition 7.7 *Given $T \in L(V_1, V_2)$, the* kernel *of T, denoted $ker(T)$, is the set of vectors $\mathbf{x} \in V_1$ such that $T(\mathbf{x}) = \mathbf{0}$. The* image *of T, denoted $im(T)$ or $T(V_1)$, is the set of vectors $\mathbf{y} \in V_2$ such that $\mathbf{y} = T(\mathbf{x})$ for some $\mathbf{x} \in V_1$.*

Proposition 7.8 *Given a linear transformation $T : V_1 \to V_2$, $ker(T)$ is a subspace of V_1 and $im(T)$ is a subspace of V_2.*

7.1 Injectivity and surjectivity

Definition 7.9 *Let X and Y be any two sets. A function $f : X \to Y$ is called* injective *or* 1-1 *if whenever $f(x) = f(x')$, it must be that $x = x'$. A function $f : X \to Y$ is called* surjective *or* onto *if for every $y \in Y$, there is an $x \in X$ such that $f(x) = y$. A function is called* bijective *if it is both injective and surjective.*

Theorem 7.10 *Let $T \in L(V_1, V_2)$. Then:*

1. *If W is a subspace of V_1, then $T(W)$ is a subspace of V_2 and $\dim T(W) \leq \dim W$.*
2. *T can be injective only if $\dim V_1 \leq \dim V_2$.*
3. *T can be surjective only if $\dim V_1 \geq \dim V_2$.*
4. *T can be bijective only when $\dim V_1 = \dim V_2$.*

Theorem 7.11 *Let $T \in L(V_1, V_2)$. Then the following are equivalent:*

1. *T is surjective;*
2. *$im(T) = V_2$;*
3. *T maps spanning sets to spanning sets, i.e. if the set $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ spans V_1, then $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$ spans V_2.*

Theorem 7.12 *Let $T \in L(V_1, V_2)$. Then the following are equivalent:*

1. *T is injective;*
2. *$ker(T) = \{\mathbf{0}\}$;*
3. *T maps linearly independent sets to linearly independent sets, i.e. if the set $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is linearly independent in V_1, then $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_k)\}$ is a linearly independent set in V_2.*
4. *T preserves dimension, i.e. if W is any subspace of V_1, then $\dim T(W) = \dim W$.*

Proposition 7.13 *Let $T_1 \in L(V_1, V_2)$ and $T_2 \in L(V_2, V_3)$. Then:*

1. *if T_1 and T_2 are both injective, then so is $T_2 T_1$;*
2. *if T_1 and T_2 are both surjective, then so is $T_2 T_1$;*
3. *if $T_2 T_1$ is injective, so is T_1 (but T_2 need not be injective);*
4. *if $T_2 T_1$ is surjective, so is T_2 (but T_1 need not be surjective).*

7.2 Vector space isomorphisms

Definition 7.14 *A function $f : X \to Y$ is called* invertible *if there is another function $f^{-1} : Y \to X$ such that $f \circ f^{-1}(y) = y$ for all $y \in Y$ and $f^{-1} \circ f(x) = x$ for all $x \in X$. f^{-1} is called the* inverse *of f.*

Theorem 7.15 *A function is invertible if and only if it is bijective.*

Theorem 7.16 *The inverse of an invertible linear transformation is itself a linear transformation.*

Definition 7.17 *Let V_1 and V_2 be vector spaces over the same field F. An invertible linear transformation $T : V_1 \to V_2$ is called an* isomorphism *between V_1 and V_2. We say V_1 and V_2 are* isomorphic *vector spaces and write $V_1 \cong V_2$ if there exists an isomorphism $T : V_1 \to V_2$.*

Two isomorphic vector spaces have the same vector space operations up to a "change in language".

Theorem 7.18 *Let V_1, V_2 and V_3 be vector spaces over the same field F. Then*

1. *$V_1 \cong V_1$.*
2. *If $V_1 \cong V_2$ then $V_2 \cong V_1$.*
3. *If $V_1 \cong V_2$ and $V_2 \cong V_3$ then $V_1 \cong V_3$.*

Theorem 7.19 *Let V_1 and V_2 be vector spaces over the same field F.*

1. *If $T \in L(V_1, V_2)$ is invertible, then $\dim V_1 = \dim V_2$.*
2. *If $V_1 \cong V_2$, then $\dim V_1 = \dim V_2$.*
3. *If $\dim V_1 = \dim V_2 < \infty$, then $V_1 \cong V_2$.*
4. *If $\dim V_1 = n < \infty$, then $V_1 \cong F^n$.*

Furthermore, if V is a vector space of dimension n over F, then every isomorphism $V \to F^n$ is a coordinate mapping (see the next section).

Theorem 7.20 *Let V_1 and V_2 be two vector spaces over the same field with $\dim V_1 = \dim V_2 < \infty$. Let $T \in L(V_1, V_2)$. Then the following are equivalent:*

1. *T is injective.*
2. *T is surjective.*
3. *T is an isomorphism.*
4. *If the set $\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ is linearly independent in V_1, then $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_k)\}$ is a linearly independent set in V_2.*
5. *If the set $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ spans V_1, then $\{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$ spans V_2.*
6. *There is a basis $\mathcal{B} = \{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of V_1 such that $T(\mathcal{B}) = \{T(\mathbf{v}_1), ..., T(\mathbf{v}_n)\}$ is a basis of V_2.*
7. *For all bases $\mathcal{B}$ of V_1, $T(\mathcal{B})$ is a basis of V_2.*

7.3 Standard matrices of linear transformations $F^n \to F^m$

Theorem 7.21 *Let F be a field. Given any linear transformation $T : F^n \to F^m$, there is a matrix $A \in M_{mn}(F)$ called the* standard matrix of T *such that $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in F^n$. In particular, the standard matrix is defined by*

$$A = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{pmatrix}.$$

In particular, we see that the map $L(F^n, F^n) \to M_n(F)$, which takes a linear transformation to its standard matrix, is an isomorphism of vector spaces.

Proposition 7.22 *Let F be a field and let $T : F^n \to F^m$ be a linear transformation with standard matrix A. Then:*

1. *$ker(T) = N(A)$.*
2. *$im(T) = C(A)$.*
3. *T is injective if and only if $N(A) = \{\mathbf{0}\}$ if and only if A has as many pivots as columns.*
4. *T is surjective if and only if $C(A) = F^m$ if and only if A has as many pivots as rows.*

Theorem 7.23 *Let F be a field and suppose $T_1 : F^n \to F^m$ and $T_2 : F^m \to F^p$ are linear transformations with standard matrices A_1 and A_2, respectively. Then the composition $T_2T_1 : F^n \to F^p$ has standard matrix A_2A_1.*

The preceding theorem justifies why matrix multiplication is defined the way that it is.

Theorem 7.24 (Equivalent characterizations of invertibility) *Let F be a field and let $T : F^n \to F^n$ (same n) be linear with standard matrix $A \in M_n(F)$. Then, the following are equivalent:*

1. *T is an isomorphism (this is equivalent to many other properties of T by Theorem 7.20).*
2. *A is invertible.*
3. *A is row equivalent to I, i.e. $rref(A) = I$.*
4. *A has n pivots, i.e. A has rank n.*
5. *The columns of A are linearly independent.*
6. *The columns of A span F^n, i.e. $C(A) = F^n$.*
7. *The columns of A form a basis of F^n;*
8. *The rows of A are linearly independent.*
9. *The rows of A span F^n, i.e. $R(A) = F^n$.*
10. *The rows of A form a basis of F^n.*
11. *$N(A) = \{\mathbf{0}\}$, i.e. $A\mathbf{x} = \mathbf{0}$ has only one solution, namely $\mathbf{x} = \mathbf{0}$.*
12. *There is a single $\mathbf{b} \in F^n$ such that $A\mathbf{x} = \mathbf{b}$ has only one solution.*
13. *$A\mathbf{x} = \mathbf{b}$ has only one solution for any $\mathbf{b} \in F^n$ (namely $\mathbf{x} = A^{-1}\mathbf{b}$).*
14. *A^T is invertible (and $\overline{A}$ and A^H are invertible if $F = \mathbb{C}$).*
15. *The corresponding statements (2)-(14) hold if A is replaced with A^T (or $\overline{A}$ or A^H if $F = \mathbb{C}$).*

8 Coordinate systems and changes of basis

8.1 Coordinate mappings

Definition 8.1 *Let V be a finite dimensional vector space over F and let $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ be a basis of V. By the Unique Representation Theorem, for any $\mathbf{x} \in V$, we can write*

$$\mathbf{x} = \sum_{j=1}^{n} c_j \mathbf{b}_j$$

*where the c_j are uniquely chosen scalars. These c_j are called the $\mathcal{B}-$*coordinates of $\mathbf{x}$ *or the* coordinates of $\mathbf{x}$ relative to the basis $\mathcal{B}$. *The vector $[\mathbf{x}]_{\mathcal{B}} = (c_1, ..., c_n)$ is called the* coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$ *or the $\mathcal{B}-$*coordinate vector of $\mathbf{x}$. *The function $\phi_{\mathcal{B}} : V \to F^n$ defined by $\phi_{\mathcal{B}}(\mathbf{x}) = [\mathbf{x}]_{\mathcal{B}}$ is called the $\mathcal{B}-$*coordinate mapping *or* coordinate mapping determined by $\mathcal{B}$.

Theorem 8.2 (Classification of isomorphisms $V \to F^n$) *Let V is a vector space over F of dimension $n < \infty$. Then:*

1. *given any basis $\mathcal{B}$ of V, the $\mathcal{B}-$coordinate mapping $\phi_{\mathcal{B}} : V \to F^n$ is an isomorphism; and*
2. *given any isomorphism $T : V \to F^n$, there is a basis $\mathcal{B}$ of V such that $T = \phi_{\mathcal{B}}$.*

Definition 8.3 *Given any basis $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ of F^n, the matrix $P_{\mathcal{S}\leftarrow\mathcal{B}} \in M_n(F)$ defined by*

$$P_{\mathcal{S}\leftarrow\mathcal{B}} = (\mathbf{b}_1 \ \cdots \ \mathbf{b}_n)$$

is called the change of coordinates matrix for $\mathcal{B}$ *into the standard matrix $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$.*

By Theorem 8.2, every such matrix $P_{\mathcal{S}\leftarrow\mathcal{B}}$ is invertible; denote by $P_{\mathcal{B}\leftarrow\mathcal{S}}$ the matrix $P_{\mathcal{B}\leftarrow\mathcal{S}} = (P_{\mathcal{S}\leftarrow\mathcal{B}})^{-1}$.

Proposition 8.4 *Given any basis $\mathcal{B}$ of F^n, then for all $\mathbf{x} \in F^n$,*

$$\mathbf{x} = P_{\mathcal{S}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} \quad \textit{and} \quad [\mathbf{x}]_{\mathcal{B}} = (P_{\mathcal{S}\leftarrow\mathcal{B}})^{-1}\mathbf{x} = P_{\mathcal{B}\leftarrow\mathcal{S}}\mathbf{x}.$$

Definition 8.5 *Let V be a vector space of dimension $n < \infty$ over a field F. Let $\mathcal{B} = \{\mathbf{b}_1, ..., \boldsymbol{b}_n\}$ and $\mathcal{B}'$ be bases of V. Define $P_{\mathcal{B}'\leftarrow\mathcal{B}} \in M_n(F)$, the* change of basis matrix *from $\mathcal{B}$ to $\mathcal{B}'$, to be*

$$P_{\mathcal{B}'\leftarrow\mathcal{B}} = \left([\boldsymbol{b}_1]_{\mathcal{B}'} \ \cdots \ [\boldsymbol{b}_n]_{\mathcal{B}'}\right).$$

Theorem 8.6 *Let V be a vector space of dimension $n < \infty$ over a field F. Let $\mathcal{B}$, $\mathcal{B}'$ and $\mathcal{B}''$ be bases of V. Then:*

1. *For all $\boldsymbol{x} \in V$, $[\boldsymbol{x}]_{\mathcal{B}'} = P_{\mathcal{B}'\leftarrow\mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}}$.*
2. *$P_{\mathcal{B}'\leftarrow\mathcal{B}}$ is invertible and has inverse $(P_{\mathcal{B}'\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{B}'}$.*
3. *$P_{\mathcal{B}\leftarrow\mathcal{B}} = I$.*
4. *$P_{\mathcal{B}''\leftarrow\mathcal{B}} = P_{\mathcal{B}''\leftarrow\mathcal{B}'}P_{\mathcal{B}'\leftarrow\mathcal{B}}$.*

8.2 Matrices of linear transformations

Definition 8.7 *Let V and W be finite dimensional vector spaces over the same field F with* $\dim V = n$ *and* $\dim W = m$*. Let $T \in L(V, W)$ and let $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, ..., \mathbf{c}_m\}$ be bases of V and W respectively. The* matrix of T relative to $\mathcal{B}$ and $\mathcal{C}$ *is the matrix $A \in M_{mn}(F)$ defined by*

$$A = ([T(\mathbf{b}_1)]_{\mathcal{C}} \ \cdots \ [T(\mathbf{b}_n)]_{\mathcal{C}}) .$$

Fix bases $\mathcal{B}$ of V and $\mathcal{C}$ of W. Then, the map taking T to its matrix A with respect to $\mathcal{B}$ and $\mathcal{C}$ is an isomorphism between the vector spaces $L(V, W)$ and $M_{mn}(F)$ (thus $L(V, W)$ has dimension mn).

Theorem 8.8 *Let V and W be finite dimensional vector spaces over the same field F with* $\dim V = n$ *and* $\dim W = m$*. Let $T \in L(V, W)$ have matrix A relative to bases $\mathcal{B}$ of V and $\mathcal{C}$ of W. Then for all $\mathbf{x} \in V$,*

$$[T(\mathbf{x})]_{\mathcal{C}} = A\,[\mathbf{x}]_{\mathcal{B}}.$$

Theorem 8.9 *Let V be a vector space over F with* $\dim V = n < \infty$*; let $\mathcal{B}$ and $\mathcal{B}'$ be bases of V. Let W be a vector space over F with* $\dim W = m < \infty$ *and let $\mathcal{C}$ and $\mathcal{C}'$ be bases of W. Let $T \in L(V, W)$ have matrix A relative to $\mathcal{B}$ and $\mathcal{C}$ and matrix A' relative to $\mathcal{B}'$ and $\mathcal{C}'$. Then:*

$$A' = P_{\mathcal{C}' \leftarrow \mathcal{C}}\, A\, P_{\mathcal{B} \leftarrow \mathcal{B}'}.$$

Corollary 8.10 *Let V be a vector space over F with* $\dim V = n < \infty$*; let W be a vector space over F with* $\dim W = m < \infty$*; let $T \in L(V, W)$. If A and A' are both matrices for the transformation T, relative to different bases of the domain and range of T, then:*

1. *there exist invertible matrices $P \in M_n(F)$ and $Q \in M_m(F)$ such that $A' = Q^{-1}AP$;*
2. *A and A' have the same rank.*

Corollary 8.11 *Let V and W be finite-dimensional vector spaces over F with* $\dim V = n$ *and* $\dim W = m$*; let $T \in L(V, W)$. Then the following are equivalent:*

1. *T is injective.*
2. *Every matrix A of T (no matter what bases are chosen) satisfies $N(A) = \{\mathbf{0}\}$.*
3. *A single matrix A of T (relative to one particular choice of basis for V and W) satisfies $N(A) = \{\mathbf{0}\}$.*

Corollary 8.12 *Let V and W be finite-dimensional vector spaces over F with* $\dim V = n$ *and* $\dim W = m$*; let $T \in L(V, W)$. Then the following are equivalent:*

1. *T is surjective.*
2. *Every matrix A of T (no matter what bases are chosen) satisfies $C(A) = F^m$.*
3. *A single matrix A of T (relative to one particular choice of basis for V and W) satisfies $C(A) = F^m$.*

In the situation where the domain and codomain of T are the same vector space, the above results can be specialized by asking the same basis to be used for the domain and codomain. In this situation, we have:

Definition 8.13 *Let V be a finite dimensional vector space over field F with* $\dim V = n$. *Let $T \in L(V, V)$ and let $\mathcal{B} = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ be a basis of V. The* matrix of T relative to $\mathcal{B}$ *is the matrix $A \in M_{mn}(F)$ defined by*

$$A = ([T(\mathbf{b}_1)]_{\mathcal{B}} \ \cdots \ [T(\mathbf{b}_n)]_{\mathcal{B}}) .$$

Theorem 8.14 *Let V be a finite dimensional vector space over field F with* $\dim V = n$. *Let $T \in L(V, V)$ have matrix A relative to basis $\mathcal{B}$ of V. Then for all $\mathbf{x} \in V$,*

$$[T(\mathbf{x})]_{\mathcal{B}} = A\, [\mathbf{x}]_{\mathcal{B}}.$$

Theorem 8.15 *Let V be a vector space over F with* $\dim V = n < \infty$*; let $\mathcal{B}$ and $\mathcal{B}'$ be bases of V. Let $T \in L(V, V)$ have matrix A relative to $\mathcal{B}$ and matrix A' relative to $\mathcal{B}'$. Then:*

$$A' = P_{\mathcal{B}' \leftarrow \mathcal{B}}\, A\, P_{\mathcal{B} \leftarrow \mathcal{B}'}.$$

Corollary 8.16 *Let V be a vector space over F with* $\dim V = n < \infty$*; let $T \in L(V, V)$. If A and A' are both matrices for the transformation T, relative to different bases of V (where the same basis is chosen for both the domain and codomain), then:*

1. *There exists an invertible matrix $P \in M_n(F)$ such that $A' = P^{-1}AP$.*
2. *A and A' have the same rank.*

Corollary 8.17 *Let V and W be vector spaces over F with* $\dim V = \dim W = n < \infty$*; let $T \in L(V, V)$. Then the following are equivalent:*

1. *T is invertible;*
2. *Every matrix of T (no matter what bases are chosen) is invertible.*
3. *The matrix of T relative to any one choice of basis of V and W is invertible.*

Corollary 8.18 *Let V be a vector space over F with* $\dim V = n < \infty$*; let $T \in L(V, W)$. Then the following are equivalent:*

1. *T is invertible;*
2. *Every matrix of T (no matter what basis is chosen) is invertible.*
3. *The matrix of T relative to any one choice of basis of V (chosen as the basis for both the domain and codomain) is invertible.*

9 Determinants

9.1 The symmetric group

Definition 9.1 *Let $[n] = \{1, ..., n\}$. Denote by S_n the set of bijections $\sigma : [n] \to [n]$; the set S_n is called the* symmetric group on n letters *and elements of S_n are called* permutations. *A permutation that exchanges two elements of $[n]$ but keeps all other elements of $[n]$ is called a* transposition.

We define a "multiplication" on S_n by composition: if $\sigma \in S_n$ and $\tau \in S_n$, then $\sigma\tau \in S_n$ is defined by $\sigma\tau = \sigma \circ \tau : [n] \to [n]$. Every permutation can be written as a product of transpositions. In particular, a cycle of odd length can only be written as a product of an even number of transpositions, and a cycle of even length can only be written as a product of an odd number of transpositions.

Definition 9.2 *Suppose $\sigma \in S_n$ is the product of k transpositions. Define the* sign *or* signature *of σ to be $sgn(\sigma) = (-1)^k$.*

This is well-defined because a permutation cannot be written as both a product of an even number of transpositions and an odd number of transpositions. Permutations with sign 1 are called *even*; permutations with sign -1 are called *odd*. Exactly half of the $n!$ permutations in S_n are even; the identity permutation is even; every transposition is odd.

Proposition 9.3 *Let $\sigma, \tau \in S_n$. Then $sgn(\sigma^{-1}) = sgn(\sigma)$ and $sgn(\sigma\tau) = sgn(\sigma)sgn(\tau)$.*

9.2 Properties of determinants

Definition 9.4 *The* determinant *is the function* $\det : M_n(F) \to F$ *defined by*

$$\det(A) = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}.$$

In particular, the determinant of an $n \times n$ matrix is the sum of $n!$ terms, half of which are added and half of which are subtracted. Each term is itself the product of n entries of the matrix, where the n entries include exactly one entry from each row and column.

Notice that from the definition, the determinant of a 1×1 matrix (a) is just a; the determinant of a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $ad - bc$; and the determinant of the 3×3 matrix

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

is $aei + bfg + cdh - afh - bdi - ceg$ (this 3×3 determinant can be computed by writing the first two columns of the matrix next to the matrix, then multiplying numbers along the six diagonals, and subtracting the sum of the "upward" products from the sum of the "downward" products. The determinants of larger matrices are most efficiently calculated using row reductions.

Theorem 9.5 (Properties of the determinant) *Let $A, B \in M_n(F)$. Then*

1. *The determinant of a triangular matrix is the product of its diagonal entries;*
2. $\det(A) = \det(A^T)$*;*
3. $\det(\bar{A}) = \det(A^H) = \overline{\det A}$*;*

4. *the determinant is* alternating*, i.e. if* B *is obtained from* A *by swapping two columns of* A*, then* $\det B = -\det A$*;*
5. *if* B *is obtained from* A *by swapping two rows of* A*, then* $\det B = -\det A$*;*
6. *if* A *contains the same row (or column) repeated twice, then* $\det A = 0$*;*
7. *the determinant is* normalized*, i.e.* $\det I = 1$*;*
8. *the determinant is* n-linear *(a.k.a.* column linear*), i.e. if all but one column of a matrix is fixed, then the determinant is a linear transformation of the remaining column;*
9. $\det(AB) = \det A \cdot \det B$*;*
10. *if* B *is obtained from* A *by adding a multiple of one row (column) of* A *to another row (resp. column) of* A*, then* $\det A = \det B$*;*
11. $\det A \neq 0$ *if and only if* A *is invertible (this is equivalent to many other properties by Theorem 7.25).*

In fact, the only function $M_n(F) \to F$ which is alternating, normalized and $n-$linear (i.e. satisfies properties 4,7 and 8 in the above theorem) is the determinant function.

Proposition 9.6 *The volume of the* $n-$*dimensional parallelepiped whose edges are* $\mathbf{v}_1, ..., \mathbf{v}_n \in F^n$ *is* $|\det(\mathbf{v}_1 \ \cdots \ \mathbf{v}_n)|$*.*

The sign of the determinant, loosely speaking, gives the "orientation" of the column vectors of the matrix. If the determinant is positive, the vectors are "positively oriented" which in dimensions 2 and 3 means that they follow a right-hand rule.

Proposition 9.7 *The determinant of any orthogonal matrix is* ± 1*; the determinant of any unitary matrix is a complex number of modulus* 1*.*

10 Eigentheory

10.1 Conjugacy and similarity

Definition 10.1 *Let* V *and* W *be vector spaces over the same field* F*. We say linear transformations* $T : V \to V$ *and* $S : W \to W$ *are* conjugate *a.k.a.* similar *and write* $T \cong S$ *if there exists an isomorphism* $\phi : V \to W$ *such that* $\phi \circ T = S \circ \phi$*.*

Conjugate linear transformations are essentially the same transformation, expressed in different "languages". The isomorphism ϕ is in some sense a translation between the two transformations.

Definition 10.2 *Let* $A, B \in M_n(F)$*. We say* A *and* B *are* conjugate *or* similar*, and write* $A \sim B$*, if there exists an invertible matrix* $P \in M_n(F)$ *such that* $PA = BP$*.*

Proposition 10.3 *Let R, S and T be linear transformations.*

1. $T \cong T$.
2. *If* $T \cong S$, *then* $S \cong T$.
3. *If* $T \cong S$ *and* $S \cong R$, *then* $T \cong R$.

Proposition 10.4 *Let $A, B, C \in M_n(F)$.*

1. $A \sim A$.
2. *If* $A \sim B$, *then* $B \sim A$.
3. *If* $A \sim B$ *and* $B \sim C$, *then* $A \sim C$.
4. *If* $A \sim B$, *then* $\det(A) = \det(B)$ *and* $tr(A) = tr(B)$.

Theorem 10.5 *Let $A, B \in M_n(F)$, and let V be a vector space over F with $\dim V = n$. The following are equivalent:*

1. $A \sim B$.
2. *There exists an invertible matrix $P \in M_n(F)$ such that $B = PAP^{-1}$.*
3. *There exists an invertible matrix $P \in M_n(F)$ such that $B = P^{-1}AP$ (this is not the same P as in statement (2).*
4. *There is a linear transformation $T : V \to V$ such that A and B are both matrices of that transformation.*
5. *Given* ***any*** *linear transformation T whose matrix relative to some basis of V is A, there is a second basis of V such that the matrix of T relative to the second basis is B.*

10.2 Eigenvalues and eigenvectors

Definition 10.6 *Let V be a vector space over a field F and let $T \in L(V, V)$. We say $\lambda \in F$ is an* eigenvalue *of T if there is a nonzero vector $\mathbf{x} \in V$ such that $T(\mathbf{x}) = \lambda\mathbf{x}$. Given an eigenvalue λ of T, any nonzero vector $\mathbf{x} \in V$ satisfying $T(\mathbf{x}) = \lambda\mathbf{x}$ is called an* eigenvector *of T corresponding to λ. Given an eigenvalue λ of T, the set of eigenvectors corresponding to λ (together with the zero vector) is called the $\lambda-$*eigenspace *of T and is denoted V_λ.*

Definition 10.7 *Let $A \in M_n(F)$. We say $\lambda \in F$ is an* eigenvalue *of A if there is a nonzero vector $\mathbf{x} \in F^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Given an eigenvalue λ of A, any nonzero vector $\mathbf{x} \in F^n$ satisfying $A\mathbf{x} = \lambda\mathbf{x}$ is called an* eigenvector *of A corresponding to λ. Given an eigenvalue λ of A, the set of eigenvectors corresponding to λ (together with the zero vector) is called the $\lambda-$*eigenspace *of A and is denoted V_λ.*

Theorem 10.8 (Elementary properties of eigenvalues and eigenvectors) *Let $T \in L(V, V)$. Then:*

1. *The eigenspace V_λ of any eigenvalue λ of T is a subspace of V.*
2. *If $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ are eigenvectors of T corresponding to* ***different*** *eigenvalues $\lambda_1, ..., \lambda_n$ of T, then $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is a linearly independent set.*
3. *If $\dim V = n$, then T can have at most n different eigenvalues.*

4. *If W is a one-dimensional subspace of V such that $T(W) \subseteq W$, then W is spanned by an eigenvector of T.*

Theorem 10.9 (Elementary properties of eigenvalues and eigenvectors) *Let $A \in M_n(F)$. Then:*

1. *The eigenspace V_λ of any eigenvalue λ of A is a subspace of F^n.*
2. *If $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ are eigenvectors of T corresponding to* ***different*** *eigenvalues $\lambda_1, ..., \lambda_n$ of A, then $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is a linearly independent set.*
3. *A can have at most n different eigenvalues.*
4. *λ is an eigenvalue of A if and only if $N(A - \lambda I) \neq \{\mathbf{0}\}$, in which case any nonzero element in $N(A - \lambda I)$ is an eigenvector corresponding to λ.*
5. *The eigenvalues of a triangular (or diagonal) matrix are its diagonal entries.*

Based on the last statement of the above theorem, eigenvalues of a matrix are computed by finding values λ for which $A - \lambda I$ is not invertible:

Definition 10.10 *Given a matrix $A \in M_n(F)$, the* characteristic polynomial *of A is the degree n polynomial* $\det(A - xI)$. *(Here, we are considering this as a polynomial in the variable x.)*

Theorem 10.11 *Let $A \in M_n(F)$. The following are equivalent:*

1. *λ is an eigenvalue of A;*
2. *λ is a root of the characteristic polynomial of A;*
3. *$(x - \lambda)^m$ is a factor of the characteristic polynomial of A.*

Definition 10.12 *Let λ be an eigenvalue of matrix $A \in M_n(F)$. The largest integer m such that $(x - \lambda)^m$ is a factor of the characteristic polynomial of A is called the* multiplicity *of λ.*

For example, if the characteristic polynomial of a matrix is $\det(A - xI) = (x-3)(x+2)^3$, then 3 is an eigenvalue of multiplicity 1, and -2 is an eigenvalue of multiplicity 3.

Theorem 10.13 *Let $A \in M_n(F)$.*

1. *A and A^T have the same eigenvalues.*
2. *If $F = \mathbb{C}$, then A has at least one eigenvalue.*
3. *If A is Hermitian, then every eigenvalue of A is real.*
4. *If $F = \mathbb{R}$ and n is odd, then A has at least one real eigenvalue.*
5. *The trace of A is the sum of its eigenvalues (where each eigenvalue is counted the number of times as its multiplicity).*
6. *The determinant of A is the product of its eigenvalues (again, counting multiplicities).*
7. *If A is a real matrix and $\lambda \in \mathbb{C}$ is an eigenvalue of A with eigenvector $\mathbf{x}$, then $\overline{\lambda}$ is also an eigenvalue of A, with eigenvector $\overline{\mathbf{x}}$.*
8. *A is invertible if and only if 0 is not an eigenvalue of A.*

Proposition 10.14 *Let $A \in M_n(F)$. Then if the eigenvalues of A are $\lambda_1, ..., \lambda_n$, then the eigenvalues of A^k are $\lambda_1^k, ..., \lambda_n^k$. The eigenvectors of A are also eigenvectors of A^k for any k.*

Although we gave two different definitions of eigenvalue/eigenvector/eigenspace above (one for linear transformations and one for matrices), these two definitions are really the same concept. To find eigenvalues/eigenvectors of a linear transformation $T : V \to V$, if V is finite-dimensional one can find a matrix A of T relative to some basis of V; the eigenvalues of A (computed by finding roots of the characteristic polynomial of A) are the eigenvalues of T. More precisely:

Theorem 10.15 *Let T and S be linear transformations, and let $A, B \in M_n(F)$. Then:*

1. *If T and S are conjugate, then T and S have the same eigenvalues. More precisely, if ϕ is an isomorphism such that $S = \phi^{-1} \circ T \circ \phi$, then $\mathbf{x}$ is an eigenvector of T corresponding to λ if and only if $\phi(\mathbf{x})$ is an eigenvector of S corresponding to λ.*
2. *Suppose $T \cong S$. Then for any eigenvalue λ of T and S, $\dim(V_\lambda)$ is the same for both T and S.*
3. *The eigenvalues of the linear transformation $F^n \to F^n$ defined by $\mathbf{x} \mapsto A\mathbf{x}$ are the eigenvalues of A.*
4. *If A is any matrix of a linear transformation T, then A and T have the same eigenvalues.*
5. *If $A \sim B$, then A and B have the same eigenvalues. More precisely, if $P \in M_n(F)$ is an invertible matrix such that $B = P^{-1}AP$, then $\mathbf{x}$ is an eigenvector $\mathbf{x}$ of A corresponding to λ if and only if $P\mathbf{x}$ is an eigenvector of B corresponding to λ.*
6. *Suppose $A \sim B$. Then for any eigenvalue λ of A and B, $\dim(V_\lambda)$ is the same for both A and B.*

10.3 Diagonalization

Theorem 10.16 *Every linear transformation is represented by a triangular matrix, and every square matrix is conjugate to a triangular matrix. More precisely:*

1. *Let V be a finite-dimensional vector space over F and let $T : V \to V$ be a linear transformation. Then there is a basis $\mathcal{B}$ of V such that the matrix of T relative to $\mathcal{B}$ is upper triangular. Furthermore, if V is an inner product space, $\mathcal{B}$ can be taken to be an orthonormal basis.*
2. *Let $A \in M_n(F)$. Then there is an invertible matrix P such that PAP^{-1} is upper triangular. Furthermore, the matrix P can be assumed to be orthogonal (if $F = \mathbb{R}$) or unitary (if $F = \mathbb{C}$).*

Definition 10.17 *A matrix $A \in M_n(F)$ is called* diagonalizable *if it is conjugate to a diagonal matrix.*

Theorem 10.18 *Let $A \in M_n(F)$. The following are equivalent:*

1. *A is diagonalizable;*
2. *there is a basis of F^n consisting of eigenvectors of A;*
3. *$A = P\Lambda P^{-1}$ where P is an invertible matrix (writing $A = P\Lambda P^{-1}$ is called* diagonalizing *A).*

In particular, if A is diagonalizable, the diagonal matrix Λ in the above decomposition $A = P\Lambda P^{-1}$ must have as its diagonal entries the eigenvalues of A, and the columns of P are the corresponding eigenvectors of A.

Theorem 10.19 *Let V be a finite-dimensional vector space over F and let $T : V \to V$ be linear. The following are equivalent:*

1. *there is a basis of V consisting of eigenvectors of T;*
2. *there is a matrix A of T which is diagonalizable;*
3. *every matrix representing T is diagonalizable.*
4. *there is a diagonal matrix representing T.*

Theorem 10.20 *If an $n \times n$ matrix A has n* ***distinct*** *eigenvalues, then A is diagonalizable. If $\dim V = n$ and $T : V \to V$ has n* ***distinct*** *eigenvalues, then some matrix of T is diagonal.*

10.4 Matrix powers and exponentials

Proposition 10.21 *If D is an $n \times n$ diagonal matrix with diagonal entries $d_1, ..., d_n$, then for any nonnegative integer k, D^k is a diagonal matrix with entries $d_1^k, ..., d_n^k$.*

Proposition 10.22 *Suppose $B = PAP^{-1}$. Then for every nonnegative integer n, $B^n = PA^nP^{-n}$.*

The preceding two results tell us how to compute powers of a diagonalizable matrix. Given a diagonalizable matrix A, write $A = P\Lambda P^{-1}$ where Λ is diagonal; then $A^k = P\Lambda^k P^{-1}$.

Definition 10.23 *Let $A \in M_n(F)$. The* matrix exponential *of A, denoted* $\exp(A)$ *or* e^A*, is the $n \times n$ matrix*

$$\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \frac{1}{4!}A^4 + ...$$

(In fact, this infinite sum converges for all (real and complex) matrices A, although exactly what is meant by "convergence" for infinite sums of matrices is beyond the scope of this course.)

Theorem 10.24 (Properties of matrix exponentials) *Let $A, B \in M_n(F)$ where $F = \mathbb{R}$ or $\mathbb{C}$. Then:*

1. *if A is diagonal with entries $\lambda_1, ..., \lambda_n$, then $\exp(A)$ is diagonal with entries $e^{\lambda_1}, ..., e^{\lambda_n}$;*
2. $\exp(PAP^{-1}) = P\exp(A)P^{-1}$ *for any invertible $P \in M_n(F)$;*
3. *if $AB = BA$, then $\exp(A + B) = \exp(A)\exp(B)$ (this equation does not hold in general);*
4. $\det(e^A) = e^{tr(A)}$.

To compute the matrix exponential of a diagonal matrix A, write $A = P\Lambda P^{-1}$; then $\exp(A) = Pe^{\Lambda}P^{-1}$.

Theorem 10.25 *Let $x_1(t), x_2(t), ..., x_n(t)$ be a collection of n differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. Write $\mathbf{x}(t) = (x_1(t), ..., x_n(t))$ and set $\mathbf{x}'(t) = (x_1'(t), ..., x_n'(t))$. Suppose $A \in M_n(F)$ is such that the following differential equation holds:*

$$\mathbf{x}'(t) = A\mathbf{x}(t).$$

Then every solution of this differential equation is of the form $\mathbf{x}(t) = e^{At}\mathbf{x}_0$, where $\mathbf{x}(0) = \mathbf{x}_0$.

10.5 A word about non-diagonalizable matrices

Suppose $A \in M_n(F)$ is a matrix which is not diagonalizable. Then A must have an eigenvalue λ of multiplicity $m > 1$ where the dimension of the corresponding eigenspace V_λ is less than m. For example, the matrix

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

is the "classic" example of a non-diagonalizable matrix, because it has only one eigenvalue $\lambda = 1$ and the only eigenvectors corresponding to $\lambda = 1$ are those in the span of $(1, 0)$. Thus V_1 is one-dimensional (but $\lambda = 1$ has multiplicity two since the characteristic polynomial is $(1 - x)^2$). Therefore F^2 has no basis of eigenvectors of A and A is therefore not diagonalizable. This means we cannot write $A = P\Lambda P^{-1}$ where Λ is diagonal and therefore computing exponentials and powers of a matrix becomes somewhat harder (fortunately this is a relatively rare occurrence).

In this situation, one uses what is called the *Jordan canonical form* of a matrix. Roughly speaking, the Jordan canonical form of a matrix A is a matrix J (conjugate to A) whose powers and exponentials are relatively easy to compute. More precisely:

Definition 10.26 *A matrix J is said to be in* Jordan canonical form *if there are square matrices $B_1, B_2, ...$ such that J has the following block form:*

$$J = \begin{pmatrix} B_1 & & & \\ & B_2 & & \\ & & \ddots & \\ & & & B_k \end{pmatrix}$$

where each B_j is a Jordan block matrix*, which means it is of the form*

$$B_j = \begin{pmatrix} \lambda_j & 1 & & & \\ & \lambda_j & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ & & & & \lambda_j \end{pmatrix}$$

where λ is some constant.

Notice that each block B_j has one λ_j associated to it; the numbers $\lambda_1, ..., \lambda_k$ are the eigenvalues of A. For a diagonalizable matrix A, each block B_j is the 1×1 block (λ_j), so J is the diagonal matrix Λ.

Related to the ideas of the Jordan canonical form is the following fact, which is used (via the machinery of Jordan canoncial forms) to compute powers and exponents of non-diagonalizable matrices:

Theorem 10.27 *Given $A \in M_n(F)$, there exist $D, N \in M_n(F)$ such that $A = D + N$, $DN = ND$, D is diagonalizable and N is nilpotent (i.e. $N^k = 0$ for some $k \geq 1$).*

11 Spectral theory

Definition 11.1 *A matrix $A \in M_n(\mathbb{C})$ is called* normal *if $AA^H = A^H A$. (If A is real, this means $A^T A = AA^T$.)*

Proposition 11.2 *Here are some facts about normal matrices:*

1. *Every Hermitian matrix is normal.*
2. *Every real symmetric matrix is normal.*
3. *Every unitary matrix is normal.*
4. *For any $A \in M_n(\mathbb{C})$, both AA^H and $A^H A$ are normal.*
5. *A is normal if and only if $||A\mathbf{x}|| = ||A^H\mathbf{x}||$ for every $\mathbf{x} \in \mathbb{C}^n$ (where $||\cdot||$ be the norm associated to the Hermitian inner product on $\mathbb{C}^n$).*
6. *If A is normal, so is $A - tI$ for any scalar t.*
7. *If A is normal, then if $\mathbf{x}$ is an eigenvector of A corresponding to eigenvalue λ, then $\mathbf{x}$ is also an eigenvector of A^H corresponding to eigenvalue $\overline{\lambda}$.*
8. *If A is normal, then eigenvectors of A corresponding to different eigenvalues of A are orthogonal (w.r.t. Hermitian inner product).*

11.1 Complex spectral theory

Proposition 11.3 *Let $A \in M_n(\mathbb{C})$ and let $<,>$ be the Hermitian inner product on $\mathbb{C}^n$. A is Hermitian if and only if for every $\mathbf{x} \in \mathbb{C}^n$, $< A\mathbf{x}, \mathbf{x} > \in \mathbb{R}$.*

Theorem 11.4 (Spectral theorem, complex version) *Let $A \in M_n(\mathbb{C})$. The following are equivalent:*

1. *A is normal.*
2. *There is an orthonormal (w.r.t. to Hermitian inner product) basis of $\mathbb{C}^n$ of eigenvectors of A.*

3. *A is "unitariliy diagonalizable", i.e. there exists a unitary matrix* $U \in U_n$ *and a diagonal matrix* $\Lambda \in M_n(\mathbb{C})$ *such that* $A = U\Lambda U^{-1} = U\Lambda U^H$.
4. *There is a unitary matrix* U *(the inverse of* U *in part (3)) such that* $UAU^H = UAU^{-1}$ *is diagonal.*

11.2 Real spectral theory

Proposition 11.5 *Let* $A \in M_n(\mathbb{C})$ *and let* $<,>$ *be dot product on* $\mathbb{R}^n$. *If* A *is symmetric and* $< A\mathbf{x}, \mathbf{x} >= 0$ *for every* $\mathbf{x} \in \mathbb{R}^n$, *then* $\mathbf{x} = \mathbf{0}$.

Remark: The hypothesis that A is symmetric is necessary here (otherwise, consider a matrix which rotates vectors by $\pi/2$). No hypothesis of symmetry/Hermiticity was necessary in the complex case.

Theorem 11.6 (Spectral theorem, real version) *Let* $A \in M_n(\mathbb{R})$. *The following are equivalent:*

1. *A is symmetric.*
2. *There is an orthonormal (w.r.t. to dot product) basis of* $\mathbb{R}^n$ *of eigenvectors of* A.
3. *A is "orthogonally diagonalizable", i.e. there exists an orthogonal matrix* $Q \in O_n$ *and a diagonal matrix* $\Lambda \in M_n(\mathbb{R})$ *such that* $A = Q\Lambda Q^{-1} = Q\Lambda Q^T$.
4. *There is an orthogonal matrix* Q *(the inverse of the* Q *in part (3)) such that* $QAQ^T = QAQ^{-1}$ *is diagonal.*

11.3 Positive and positive definite matrices

Definition 11.7 *A Hermitian matrix* $A \in M_n(\mathbb{C})$ *is called* positive *if* $\mathbf{x}^H A\mathbf{x} \geq 0$ *for every* $\mathbf{x} \in \mathbb{C}^n$. *A Hermitian matrix* $A \in M_n(\mathbb{C})$ *is called* positive definite *if if is positive, and if the only vector* $\mathbf{x} \in \mathbb{C}^n$ *satisfying* $\mathbf{x}^H A\mathbf{x} = 0$ *is* $\mathbf{x} = \mathbf{0}$.

Theorem 11.8 *Let* $A \in M_n(\mathbb{C})$. *The following are equivalent:*

1. *A is positive.*
2. *A is Hermitian and the eigenvalues of A are nonnegative.*
3. *There is a positive matrix* B *such that* $B^2 = A$ *(B is called the* square root *of A and is denoted* $B = \sqrt{A}$*).*
4. *There is a Hermitian matrix* B *such that* $B^2 = A$.
5. *There is a matrix* $B \in M_n(F)$ *such that* $A = B^H B$.
6. *There is a matrix* $B \in M_n(F)$ *such that* $A = BB^H$.

Corollary 11.9 *Every positive matrix has a unique positive square root.*

Theorem 11.10 *Let* $A \in M_n(\mathbb{C})$. *The following are equivalent:*

1. *A is positive definite.*

2. *The expression* $< \mathbf{x}, \mathbf{y} >= \mathbf{y}^H A\mathbf{x}$ *defines an inner product on* $\mathbb{C}^n$*;*
3. A *is Hermitian and all its eigenvalues are positive.*
4. *If for each* $k = 1, ..., n$*, one takes the upper-leftmost* $k \times k$ *entries of* A *and calls that matrix* A_k *the* upper left submatrix *of* A*, then* $\det A_k > 0$ *for all* k*.*

11.4 Singular value decomposition

The set $M_n(\mathbb{C})$ of $n \times n$ complex matrices can be thought of as a generalization of the complex numbers. Many ideas about complex numbers generalize naturally to complex square matrices. In particular:

COMPLEX NUMBER CONCEPT	COMPLEX MATRIX CONCEPT
complex number $z \in \mathbb{C}$	complex matrix $M \in M_n(\mathbb{C})$
conjugation $z \mapsto \overline{z}$	taking the Hermitian of a matrix $M \mapsto M^H$
real number	Hermitian matrix
nonnegative real number	positive matrix
positive real number	positive definite matrix
nonnegative real numbers have unique nonnegative square roots	positive matrices have unique positive square roots
unit circle (complex numbers of modulus 1) ($\overline{z} = z^{-1}$)	unitary matrices ($U^H = U^{-1}$)
every $z \in \mathbb{C}$ can be written $z = a + ib$ where $a, b \in \mathbb{R}$	every $M \in M_n(\mathbb{C})$ can be written $M = A + iB$ where A, B are Hermitian
every $z \in \mathbb{C}$ is $z = \lvert z\rvert e^{i\theta} = e^{i\theta}\sqrt{z\overline{z}}$	see Theorem 11.11 below
modulus	singular values (see below)

Theorem 11.11 (Polar decomposition) *Let* $A \in M_n(\mathbb{C})$*. Then* $A = U\sqrt{A^H A}$ *for some unitary matrix* U*.*

Definition 11.12 *Let* $A \in M_{mn}(\mathbb{C})$*. The* singular values *of* A *are the eigenvalues of* $\sqrt{A^H A}$ *(equivalently, the square roots of the eigenvalues of the positive matrix* $A^H A$*).*

Theorem 11.13 (Singular value decomposition (complex version)) *Let* $A \in M_{mn}(\mathbb{C})$*. Then there exist unitary matrices* $U \in U_m$ *and* $V \in U_n$ *and a diagonal matrix* $\Sigma \in M_{mn}(\mathbb{C})$ *such that* $A = U\Sigma V^H$*.*

Theorem 11.14 (Singular value decomposition (complex version)) *Let* $A \in M_{mn}(\mathbb{R})$*. Then there exist orthogonal matrices* $Q \in O_m$ *and* $R \in O_n$ *and a diagonal matrix* $\Sigma \in M_{mn}(\mathbb{R})$ *such that* $A = Q\Sigma R^T$*.*

The entries of Σ are the singular values of A.

Proposition 11.15 *Let A have SVD $A = U\Sigma V^H$. Then, if A has rank r:*

1. *The first r columns of U form an orthonormal basis of $C(A)$.*
2. *The last $m-r$ columns of U form an orthonormal basis of $N(A^H)$.*
3. *The first r columns of V form an orthonormal basis of $R(A)$.*
4. *The last $n-r$ columns of V form an orthonormal basis of $N(A)$.*

11.5 Pseudoinverses and least-squares solutions

Definition 11.16 *Let $A \in M_{mn}(\mathbb{C})$ have SVD $A = U\Sigma V^H$ and rank r. Without loss of generality we can write $\Sigma = \left(\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right)$ where D is a diagonal $r \times r$ matrix with diagonal entries equal to the nonzero singular values of A. Similarly, partition U and V into their first r columns and their last $m-r$ (respectively $n-r$) columns as*

$$U = \left(\begin{array}{c|c} U_r & U_{m-r} \end{array} \right) \text{ and } V = \left(\begin{array}{c|c} V_r & V_{n-r} \end{array} \right)$$

The (Moore-Penrose) pseudoinverse *of A, denoted A^+, is the matrix $A^+ \in M_{nm}(\mathbb{C})$ defined by*

$$A^+ = V_r D^{-1}(U_r)^H.$$

Theorem 11.17 (Properties of pseudoinverses) *Let $A \in M_{mn}(\mathbb{C})$ and let $A^+ \in M_{nm}(\mathbb{C})$ be its pseudoinverse. Then:*

1. *If A is real, so is A^+.*
2. *If A has full column rank (i.e. if the columns of A are linearly independent), then $A^+ = (A^H A)^{-1} A^H$.*
3. *If A has full row rank (i.e. if the rows of A are linearly independent), then $A^+ = A^H(AA^H)^{-1}$.*
4. *If A is invertible, then $A^+ = A^{-1}$.*
5. *For any $\mathbf{x} \in \mathbb{C}^m$, $AA^+\mathbf{x}$ is the projection of $\mathbf{x}$ onto $C(A)$.*
6. *For any $\mathbf{x} \in \mathbb{C}^n$, $A^+A\mathbf{x}$ is the projection of $\mathbf{x}$ onto $R(A)$.*
7. *$AA^+A = A^+$ and $A^+AA^+ = A$.*
8. *AA^+ and A^+A are Hermitian.*
9. *$(A^+)^+ = A$.*
10. *$(A^+)^T = (A^T)^+$; $(A^+)^H = (A^H)^+$; $\overline{A^+} = (\overline{A})^+$.*
11. *If $r \neq 0$, then $(rA)^+ = \frac{1}{r}A^+$.*
12. *$N(A^+) = N(A^H)$ and $C(A^+) = C(A^H)$.*

Definition 11.18 *Let $A \in M_{mn}(\mathbb{C})$ and let $\mathbf{b} \in \mathbb{C}^m$. The vector $\mathbf{x}^+ = A^+\mathbf{b} \in \mathbb{C}^n$ is called the* least-squares solution *to the system $A\mathbf{x} = \mathbf{b}$.*

Theorem 11.19 *Let $A \in M_{mn}(\mathbb{C})$ and let $\mathbf{b} \in \mathbb{C}^m$. Let $\mathbf{x}^+$ be the least-squares solution of $A\mathbf{x} = \mathbf{b}$ and let $\mathbf{b}^+ = A\mathbf{x}^+$. Then*

$$||\mathbf{b}^+ - \mathbf{b}|| \leq ||A\mathbf{x} - \mathbf{b}||$$

for any $\mathbf{x} \in \mathbb{C}^n$ (where $||\cdot||$ is the norm coming from the Hermitian inner product).

Restated, this means that the closest (with respect to the usual Euclidean distance) one can come to solving $A\mathbf{x} = \mathbf{b}$ when $\mathbf{b} \notin C(A)$ is to set $\mathbf{x} = \mathbf{x}^+ = A^+\mathbf{b}$. Furthermore, if $\mathbf{u} \in \mathbb{C}^n$ is any other solution of $A\mathbf{x} = \mathbf{b}^+$, then $||\mathbf{u}|| \geq ||\mathbf{x}^+||$.