2. Vector spaces

2.1 Notation

As far as possible, the notation follows that of the textbook [1]. Only a small subset is required for this chapter. Lower case letters e.g., \( x, y, z \) denote vectors, upper case letters e.g., \( A, B, C \) denote matrices. The superscript \( T \) denotes the transpose of a vector or matrix. Let \( \mathbb{R} \) and \( \mathbb{C} \) denote the sets of real numbers and complex numbers respectively. The space \( \mathbb{R}^n \) consists of all column vectors containing \( n \) real components. For example

\[
\begin{bmatrix}
2 \\
-\pi \\
6
\end{bmatrix}
\text{ and } \begin{bmatrix}
3 & e & 0
\end{bmatrix}^T
\]

are vectors in \( \mathbb{R}^3 \). The spaces \( \mathbb{R}, \mathbb{R}^2, \mathbb{R}^3 \) can be easily visualized as points on a line, the \( x,y \)-plane and the \( x,y,z \)-plane respectively. To reduce wasted space, we will also use parentheses to denote column vectors, e.g., \( (2, -\pi, 6) \) is used to denote the first vector above. We will use the symbol, 0, to represent the scalar quantity zero, vectors of all zeros, and matrices of all zeros. It will be clear from context which is intended (and what the dimension is) The identity matrix (a square matrix with ones on the diagonal and zeros elsewhere) is denoted by \( I \).

Clearly, as we progress beyond three dimensions visualization becomes problematic. Provided we know the rules for working with vectors in these easy to view spaces, linear algebra seamlessly allows us to work with vectors of arbitrary dimension – even if they can’t be visualized!

2.2 Fields and vector spaces

Throughout these notes and the course (and beyond) we will refer to specific sets of numbers. From these sets we will build up the concept of “spaces”. The most useful sets of numbers satisfy a few simple properties relating to addition and multiplication. We refer to these sets as “fields”.

Chapter 2. Vector spaces

Definition 2.2.1 — Field. To every pair \( \alpha \) and \( \beta \) of scalars belonging to the set \( F \), if addition and multiplication operations as defined below are satisfied, then the set \( F \) is called a field.

1. For every \( \alpha \) and \( \beta \) in \( F \), there corresponds a scalar \( \alpha + \beta \) called the sum such that
   \[ \alpha + \beta = \beta + \alpha, \]
   \[ \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma, \]
   \[ \text{there exists a scalar 0 such that } \alpha + 0 = \alpha, \]
   \[ \text{for every } \alpha \text{ there corresponds an unique scalar } -\alpha \text{ such that } \alpha + (-\alpha) = 0. \]

2. To every pair \( \alpha \) and \( \beta \), there corresponds a scalar \( \alpha\beta \) called the product such that
   \[ \alpha \beta = \beta \alpha, \]
   \[ \alpha(\beta \gamma) = (\alpha \beta) \gamma, \]
   \[ \text{there exists a unique scalar 1 (one) such that } \alpha 1 = \alpha \text{ for every } \alpha, \]
   \[ \text{to every non-zero scalar } \alpha \text{ there is a corresponding scalar } \alpha^{-1} \text{ such that } \alpha\alpha^{-1} = 1. \]

3. Multiplication is distributive with respect to addition, i.e., \( \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma. \)

Some standard examples of fields are the reals \( \mathbb{R} \), the set of complex numbers \( \mathbb{C} \), and rational numbers. These sets can easily be verified to be fields by direct application of definition 2.2.1.

The concept of a field essentially defines a set of scalars for which addition and multiplication are well defined. We will now define a “vector space”, one of the most fundamental objects in mathematics.

For the rest of these notes we will simply use the word scalar to denote an element belonging to a field. Here we assume that the field is either the set of real or complex numbers (of course more abstract cases are possible, but we are not concerned with these). It will be clear from context which field we are referring to.

Definition 2.2.2 — Vector space. Given a field \( F \), a vector space is a set \( V \) of elements called vectors that satisfy the following axioms.

1. Let \( x, y, z \) be vectors belonging to the \( V \). There exists a vector \( x + y \) that belongs to \( V \) called the sum such that
   \[ x + y = y + x, \]
   \[ x + (y + z) = (x + y) + z, \]
   \[ \text{there exists in } V \text{ a unique vector 0 (the origin) such that } x + 0 = x \text{ for every } x, \]
   \[ \text{to every vector } x \text{ in } V \text{ there corresponds a unique vector } -x \text{ such that } x + (-x) = 0. \]

2. Let \( \alpha \) be a scalar belonging to the field \( F \) and \( x \) be vector in \( V \). To every pair \( \alpha, x \) there is an element \( \alpha x \) in \( V \) called the product such that
   \[ \alpha(\beta x) = (\alpha \beta)x \text{ for any } \beta \text{ in } F, \]
   \[ 1x = x \text{ for every vector } x. \]

3. Multiplication by scalars is distributive with respect to vector addition, \( \alpha(x + y) = \alpha x + \alpha y \), and

4. multiplication by vectors is distributive with respect to scalar addition, \((\alpha + \beta)x = \alpha x + \beta x. \)

Formally we refer to a “vector space \( V \) defined over \( F \)”. This is a bit of a mouthful so we will typically just use “vector space” and the field will be obvious from context. When \( F = \mathbb{R}, V \)
2.3 Dimension

is called a *real vector space*. In this course we will almost exclusively work with real vector spaces. However, the generalization to complex vector spaces is seamless.

A vector space describes how vectors interact with other vectors (through addition) as well as how vectors interact with scalars (through multiplication). There is no vector-vector notion of *multiplication*. While this may be possible for some vector spaces it is not necessary.

Some examples of vector spaces that you have likely encountered include $\mathbb{R}^n$, the set of $n$-dimensional column vectors of real numbers. The set of matrices with $m$ columns and $n$ rows comprised of complex numbers $\mathbb{C}^{m \times n}$. Also note that $\mathbb{R}^1 = \mathbb{R}$ is both a vector space and a field. The same is true for $\mathbb{C}$.

A trivial example of a vector field is the set $F = \{0\}$, i.e, a set which only contains the origin.

### 2.2.1 Subspaces

“A subspace is like Las Vegas. What happens in Vegas stays in Vegas.” – source unknown

It is often the case that we are interested in just a specific region of a vector space. Geometrically, we may be interested in a just a line or surface contained in a higher dimensional vector space. When the subset of interest also satisfies the definition of a vector space (definition 2.2.2) that subset is referred to as a *subspace*.

**Definition 2.2.3 — Subspace.** A subset of a non-empty vector space that is itself a vector space.

**Example 2.1** The set

$$V = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \mid x_1 = 2x_2 \right\}$$

is a subspace of $\mathbb{R}^2$. The subspace $V$ is a line through the origin.

Subspaces are closed under under scalar multiplication and addition. Here, “closed” means that all vectors in $V$ remain in $V$ after scalar multiplication and addition is performed. More precisely, if $a$ and $b$ belong to $V$, then so does $a + b$. If $\alpha$ belongs to the field in which $V$ is defined over, then $\alpha x$ is in $V$.

**Example 2.2 — A non-example.** The vector space $\mathbb{R}^n$ is not a subspace of $\mathbb{R}^{n+i}$ for any integer $i > 1$. [explanation here...]

### 2.3 Dimension

We will now introduce a few definitions and concepts that allow us to precisely quantify the size (dimension) of a vector space. Qualitatively we have seen that a subspace is smaller than (or equal to) the size of its parent space. Intuitively you probably realize that the dimension of $\mathbb{R}^3$ is three! But what about more complicated vector spaces such as the set of symmetric $n \times n$ matrices, or the set of polynomials with complex coefficients (yes that is a vector space!)?

### 2.3.1 Linear independence

Suppose we have a set of $m$ vectors $x_1, x_2, x_3, \ldots, x_m$ for which use the shorthand notation $\{x_i\}_{i=1}^m$. Frequently we will omit the index set and just use $\{x_i\}$ to represent this set. A *linear combination* of vectors is given by

$$\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_m x_m$$ (2.1)
where each $\lambda$ is a real scalar.

**Definition 2.3.1 — Linear independence.** A finite set $\{x_i\}_{i=1}^m$ of vectors is *linearly independent* if all nontrivial combinations of the vectors are nonzero, i.e.,

$$\lambda_1x_1 + \lambda_2x_2 + \ldots + \lambda_mx_m \neq 0 \quad \text{unless } \lambda_i = 0 \text{ for all } i \text{ (the trivial combination)}.$$

Else the vectors are said to be *linearly dependent*.

A set of $d$ vectors in $\mathbb{R}^n$ must be linearly dependent if $d > n$.

Intuitively, if a set of vectors is linearly independent, then it isn’t possible to write one element of the set as a linear combination of the others.

Although we encounter sets with an infinite number of elements less frequently in this course, the concept of independence extends to this setting. Let $S$ be a set containing an infinite number of vectors. We say that $S$ is linearly independent if every finite subset of $S$ is linearly independent, else it is linearly independent. For the remainder of these notes, whenever we use the term vector space we are referring to a finite-dimensional vector space.

**Definition 2.3.2 — Span.** If a vector space $V$ consists of all linear combinations of the vectors $x_1, x_2, \ldots, x_l$, then these vectors are said to span $V$. Equivalently, every vector $v$ in $V$ can be expressed as a linear combination of the set $\{x_i\}_{i=1}^l$, i.e.,

$$v = \lambda_1x_1 + \ldots + \lambda_lx_l \quad \text{for some coefficients } \lambda_i.$$

Equivalently, define

$$\text{span}(x_1, \ldots, x_m) = \left\{ \sum_{i=1}^m \lambda_i x_i \mid \lambda \in \mathbb{R}^m \right\},$$

then $\{x_i\}$ spans $V$ if $V = \text{span}(x_1, \ldots, x_m)$.

The definition above does not specify the value of $l$, i.e., the number of vectors in the spanning set. We will see later that a value related to this number corresponds to the dimension of the vector space.

It is possible that more than one set of coefficients (the set of $\lambda_i$’s can produce the same vector $v$ – there is no assumption of uniqueness.)

■ **Example 2.3** The vectors $x_1 = (3, 0, 0)$, $x_2 = (0, -1, 0)$ and $x_3 = (4, 0, 0)$ span the x-y plane in a 3-dimensional space. The first two vectors, $x_1$ and $x_2$ also span the plane (as do the pair $x_2, x_3$). In contrast, $x_1$ alone only spans a line. No combination of the three vectors spans $\mathbb{R}^3$. This is because no combination of $\lambda$’s can produce the vector $(0, 0, z)$ for any value of $z$.

■ **Example 2.4** The coordinate vectors in $\mathbb{R}^n$,

$$e_1 = (1, 0, \ldots, 0), \quad e_2 = (0, 1, 0, \ldots, 0), \quad \ldots, \quad e_n = (0, \ldots, 0, 1)$$

span $\mathbb{R}^n$.
2.3 Dimension

2.3.2 Basis

Looking back at example 2.4, the vectors \( \{ e_i \} \) are linearly independent. One interpretation of this is that there are no wasted vectors in the spanning set. If we appended the vector \( e_{12} = (1, 1, 0, \ldots, 0) \) to the set, it would still span \( \mathbb{R}^n \). This new vector however doesn’t add anything new to the spanning set. If we were to remove a vector (and not include the vector \( e_{12} \)) the set would no longer span \( \mathbb{R}^n \).

Spanning sets of minimal dimension have a special name:

**Definition 2.3.3 — Basis.** A basis for a vector space is a set of vectors with the properties:

1. The set is linearly independent, and
2. it spans the vector space.

Together these properties provide the minimal number of elements required to span a vector space. A basis also ensures that an element of \( v \) in \( V \) can only be expressed by a single linear combination of the basis vectors. This is in contrast to a spanning set where numerous coefficient choices are possible.

The following result (which we won’t prove) allows us to expand or shrink a set of vectors in order to form a basis.

**Theorem 2.3.1** Any linearly independent set in \( V \) can be extended to a basis by adding new vectors. Conversely, any spanning set in \( V \) can be shrunk to a basis by discarding vectors if necessary.

2.3.3 Dimension of a vector space

The notion of a basis is fundamental to formally defining the dimension of a vector space. Indeed, all the heavy lifting has already been done.

**Definition 2.3.4** The dimension of a vector space \( V \) is the number of elements in the basis of \( V \).

In these notes we will only be concerned with vector spaces such as \( \mathbb{R}^n \) and \( \mathbb{C}^n \) that have a finite number of basis vectors. Not surprisingly, we refer to such spaces as finite dimensional vector spaces.

Don’t let the word “dimension” confuse you. It appears in two contexts. When talking about vectors, dimension refers to the number of elements in the vector, for example, \( y = (2, -5, 0, 4) \) is 4-dimensional vector. Here \( y \) belongs to the vector space \( \mathbb{R}^4 \). Definition 2.3.4 defines the dimension of a subspace. For example the set of vectors in \( \mathbb{R}^5 \) whose first two components are zero corresponds to a subspace of dimension 4, but whose elements are 6-dimensional vectors, e.g., \( (0, 0, \pi, -1, 1, 7) \).

The notion of the dimension of the vector space expresses the number of degrees of freedom of the space.

It should be clear that a basis for a vector space is not unique. Consider again example 2.4. The vectors \( \{ e_i \}_{i=1}^n \) define a basis for \( \mathbb{R}^n \) but so to does the set \( \{ 4e_i \}_{i=1}^n \). In fact, an infinite number of choices of basis exist.

While the choice of basis is not unique, the number of elements in the basis is.

**Theorem 2.3.2** The number of elements in any basis of a finite-dimensional vector space \( V \) is the same as in any other choice of basis.
Chapter 2. Vector spaces

2.4 Linear transformations (aka linear maps)

A transformation is a function that maps vectors from one vector space to a new vector space. We are specifically interested in linear transformations. We will see later that all linear transformations can be represented by a matrix (the converse is also true). After you have read this subsection it will be helpful to think of matrices not as a grid of numbers, but as transformations or maps from one vector space to another.

Recall that a function $f$ is said to be linear if

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$  \hspace{1cm} (2.2)

where $\alpha, \beta$ are scalars and $x, y$ are vectors.

**Definition 2.4.1 — Transformation.** A transformation is a rule (function) $f$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ that assigns to each vector in $\mathbb{R}^n$ a vector $f(x)$ in $\mathbb{R}^m$. The transformation is denoted $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

**Definition 2.4.2 — Linear transformation.** A linear transformation is a transformation that additionally satisfies the linearity property (2.2).

Consider now a matrix $A \in \mathbb{R}^{m \times n}$. This matrix defines a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ via matrix-vector multiplication. Using the terminology of definition 2.4.1, the transformation is

$$f(x) = Ax.$$ 

The $n$-dimensional vector can be thought of as the input and the $m$-dimensional vector $Ax$ is the output. Moreover, matrix transformations are linear transformations. Matrix-vector multiplication $y = Ax$ thus defines a linear transformation from the vector space $x$ belongs to, to the vector space $y$ belongs to.

**Corollary 2.4.1** Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear transform, then

1. $f(0) = 0$, and
2. for any vectors $v_1, v_2, \ldots, v_p$ and scalars $c_1, c_2, \ldots, c_p$, we have

$$f(c_1 v_1 + c_2 v_2 + \ldots + c_p v_p) = c_1 f(v_1) + c_2 f(v_2) + \ldots + c_p f(v_p).$$

**Corollary 2.4.2** The set of linear transformations on a vector space is itself a vector space.

---

You should be aware that the term map and transformation are often used interchangeably. A third term, operator, is also commonly used. From now on, all maps/transformations/operators are assumed to be linear unless otherwise stated.

2.5 The Big 4

While it’s easy (and fun?) to just make up vector and subspaces, the most important spaces are derived from matrices. There are four fundamental subspaces that completely characterize the behavior of a matrix (linear transformation).

The following definitions are made with respect to a matrix $A \in \mathbb{R}^{m \times n}$.
Definition 2.5.1 — Range. The range of the matrix $A$ is defined as the following subset of $\mathbb{R}^m$:

$$\text{range}(A) := \{Ax \mid x \in \mathbb{R}^n\}.$$ 

The range of a matrix is also referred to as the column space.

In words, the range of a matrix is the set of vectors that $A$ can map all vectors in $\mathbb{R}^n$ to. Put another way, it defines the set of vectors $b \in \mathbb{R}^m$ for which $Ax = b$ can be solved. To see this last point, consider the linear system

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

which can be written as

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} x_1 + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} x_2. \quad (2.3)$$

From the viewpoint of (2.3), the achievable vectors $b$ that $A$ can reach are linear combinations (see equation (2.1)) of the columns of $A$.

Definition 2.5.2 — Nullspace. The nullspace of $A$ is a subset of $\mathbb{R}^n$ given by

$$\text{null}(A) := \{x \in \mathbb{R}^n \mid Ax = 0\}.$$ 

The nullspace is also referred to as the kernel.

Simply put, the nullspace is the set of vectors $x \in \mathbb{R}^n$ that $A$ maps to the origin.

Proposition 2.5.1 The range and nullspace of a linear transformation (matrix) are subspaces. Specifically, if $A \in \mathbb{R}^{m \times n}$ then range($A$) is a subspace of $\mathbb{R}^m$ and null($A$) is a subspace of $\mathbb{R}^n$.

The dimension of the subspace range($A$) is by definition (see section 2.3.3), given by the number of elements in its basis. An equivalent definition is that the dimension of the range is given by the number of linearly independent columns in $A$ (refer to equation (2.3) and imagine the two column vectors are not independent. What does this tell you about the range of $A$). There is a technical term for the number of linearly independent columns in a matrix: rank. Although stated as a proposition below, this is also a(nother) definition for the rank of a matrix.

Proposition 2.5.2 Given $A \in \mathbb{R}^{m \times n}$, with rank($A$) = $r$ then

- $\dim(\text{range}(A)) = r$, and
- $\dim(\text{null}(A)) = n - r$.

The above proposition leads to the following corollary (just rearrange the second statement) relating the number of columns to the dimension of the two subspaces.

Corollary 2.5.3 The number of columns in $A$ is equal to the dimension of the range (rank of $A$) plus the dimension of the nullspace.

$$\dim(\text{range}(A)) + \dim(\text{null}(A)) = n.$$ 

The remaining two subspaces are the row space and the unimaginatively named left nullspace of $A$. 
Definition 2.5.3 — Row space. The row space of \( A \) contains all linear combinations of the rows of \( A \), which are columns of \( A^T \). The row space is thus defined as the range(\( A^T \)), it is a subset of \( \mathbb{R}^n \).

The following result is one of the most fundamental statements in linear algebra. We will prove it later, but for now it will just be stated.

Theorem 2.5.4 The dimension of the column space is equal to the dimension of the row space:

\[
\dim(\text{range}(A)) = \dim(\text{range}(A^T)).
\] (2.4)

The following statements are equivalent to (2.4):
- \( \text{rank}(A) = \text{rank}(A^T) \).
- The number of linearly independent columns of a matrix is equal to the number of linearly independent rows.

Definition 2.5.4 — Left nullspace. The left nullspace of \( A \) is a subset of \( \mathbb{R}^m \) given by \( \text{null}(A^T) \).

Proposition 2.5.5 The left nullspace \( \text{null}(A^T) \) has dimension \( m - r \), where \( \text{rank}(A) = r \).

Proof. Apply the second statement of proposition 2.5.2 to \( A^T \) and use theorem 2.5.4. □

In summary we have defined four subspaces generated from the \( m \times n \) matrix \( A \):
- Column space: range(\( A \)), with dimension \( r \),
- Null space: null(\( A \)), with dimension \( n - r \),
- Row space: range(\( A^T \)), with dimension \( r \),
- Left nullspace: null(\( A^T \)), with dimension \( m - r \).

Everything in this section carries over exactly to the setting where \( A \) is complex. These four subspaces together with their associated dimensions forms part of the fundamental theorem of linear algebra. The full theorem includes some geometric insight into how the spaces fit together. We cover this next.

2.6 Geometry

Now that we have defined a vector space, we can now study some geometric properties (and relationships) of objects (i.e., vectors) in \( V \). Specifically care about size and distance of an between vectors, and angles between vectors.

2.6.1 Vector Norms

The concept of length and angles are generalized to vectors using norms and inner products respectively. Inner products are discussed later in section 2.6.

Definition 2.6.1 — Norm. A norm on \( V \) is a function \( \| \cdot \| : V \to \mathbb{R} \) satisfying

1. \( \| x \| \geq 0 \) for all \( x \) in \( V \), with \( \| x \| = 0 \) if and only if \( x = 0 \),
2. Given a scalar \( \alpha \), \( \| \alpha x \| = |\alpha|\| x \| \) for all \( x \) in \( V \), and
3. \( \| x + y \| \leq \| x \| + \| y \| \) for all \( x, y \) in \( V \).

The most common set of norms is the \( \ell_p \)-norm family. For \( 1 \leq p < \infty \), the \( \ell_p \) norm is given by

\[
\| x \|_p := \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.
\]
We will mostly deal with the case where \( V = \mathbb{R}^n \) and \( p = 1, 2, \infty \).

**Example 2.5** Let \( V = \mathbb{R}^n \) with \( x \) belonging to \( V \).

- For \( p = 2 \), the \( \ell_2 \) norm defines the familiar *Euclidean norm*:
  \[
  \|x\|_2 := \sqrt{\sum_{i=1}^{n} |x_i|^2}.
  \]
  Typically for the Euclidean norm we just write \( \|x\| \).
- The case of \( p = 1 \) is used a lot in optimization:
  \[
  \|x\|_1 := \sum_{i=1}^{n} |x_i|.
  \]
- The \( p = \infty \) case isn’t strictly covered by definition 2.6.1, it is however the (convergent) limit of \( p \to \infty \) given by
  \[
  \|x\|_\infty := \max_{1 \leq i \leq n} |x_i|.
  \]
  The \( \ell_\infty \)-norm is simply the element of \( x \) with maximum absolute value.

When the vector space is \( \mathbb{R}^n \) or \( \mathbb{C}^n \) (or any finite-dimensional vector space) the choice of \( p \) can’t dramatically change the output.

*Theorem 2.6.1 — Equivalence of norms.* For all \( \ell_p \) norms on \( \mathbb{R}^n \) and \( \mathbb{C}^n \) there exist constants \( 0 < c_1 \leq c_2 < \infty \) such that
\[
c_1 \|x\|_a \leq \|x\|_b \leq c_2 \|x\|_a
\]
for all \( x \).

We won’t prove this result as it requires some analysis. Note that the optimal values of the constants will depend on which norms are considered. It isn’t difficult to show that for any \( x \in \mathbb{R}^n \)
\[
\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2 \leq n \|x\|_\infty.
\]

Later in the notes we will encounter norms defined with respect to matrices.

*Theorem 2.6.1 holds for all \( \ell_p \)-norms, defined on \( \mathbb{R}^n \) and \( \mathbb{C}^n \), however, as we shall see later, it does not hold on infinite dimensional spaces.*

**2.6.2 Orthogonality**

We will now look at some geometric properties of vectors that belong to these spaces. In particular we care about the angle between vectors and the length of a vector in a vector space \( V \). There are many different measures that can be used to define length. For now we will consider the standard Euclidean norm on \( \mathbb{R}^n \): Given an \( n \)-dimensional vector \( x \) we define
\[
\|x\| = (x_1^2 + x_2^2 + \ldots + x_n^2)^{\frac{1}{2}}.
\]

The Euclidean norm is natural extension of Pythagoras’ theorem to \( n \)-dimensions. In the next section we can see that it is related to an inner product. For now we are more interested in studying the angle between two two vectors. The full discussion on angles is deferred to the next section.
For now we consider the problem of determining when two vectors in $V$ are orthogonal to each other, i.e., the angle between the is $90^\circ$.

Suppose we are given two vectors $x$ and $y$. If $x$ is orthogonal to $y$ then Pythagoras’ theorem tells us that
\[
\|x\|^2 + \|y\|^2 = \|x - y\|^2,
\] (2.5)
where $x - y$ is the vector from $y$ to $x$. Expanding out both sides of (2.5) gives
\[
(x_1^2 + x_2^2 + \ldots + x_n^2) + (y_1^2 + y_2^2 + \ldots + y_n^2) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2,
\]
the right hand side of which can be written as
\[
(x_1^2 + x_2^2 + \ldots + x_n^2) + (y_1^2 + y_2^2 + \ldots + y_n^2) - 2(x_1y_1 + x_2y_2 + \ldots x_ny_n).
\]
It is now clear that (2.5) can only be satisfied when the cross-product terms sum to zero, i.e.
\[
x_1y_1 + x_2y_2 + \ldots x_ny_n = 0.
\]
The left hand side can be written in vector form as $x^T y$ is an example of an inner product on $\mathbb{R}^n$. 

**Definition 2.6.2** Two vectors $x$ and $y$ belonging to a vector space $V$ are orthogonal if their inner product is zero. When $V = \mathbb{R}^n$ this corresponds to $x^T y = 0$.

The function $x^T y$ is one example of an inner product on $\mathbb{R}^n$. In the next section we will define a set of conditions for a function that define an inner product, and we do this for arbitrary vector spaces. In addition to orthogonality, inner products allow us to calculate angles between vectors (or more specifically, the cosine of an angle between two vectors).

### 2.6.3 Inner products

Given a complex number $x = \alpha + \beta j$, its conjugate is $\bar{x} = \alpha - \beta j$. Now, suppose we have a vector of complex numbers, $z$. The conjugate transpose of $z$, denoted $z^*$, transposes the vector and then takes the conjugate of each element.

**Example 2.6** The conjugate transpose of a column vector:

\[
w = \begin{bmatrix} 2 - 4 j \\ -6 + 2 j \end{bmatrix}, \quad w^* = \begin{bmatrix} 2 + 4 j \\ -6 - 2 j \end{bmatrix}
\]

The conjugate transpose of a 2x2 matrix:

\[
W = \begin{bmatrix} 3 - 2 j & -7 j \\ 2 & -4 + j \end{bmatrix}, \quad W^* = \begin{bmatrix} 3 + 2 j & 2 \\ 7 j & -4 - j \end{bmatrix}.
\]

**Definition 2.6.3 — Inner product.** An inner product in a vector space $V$ is a function denoted by $\langle \cdot, \cdot \rangle$ on vectors $x, y$ belonging to $V$ such that

1. $\langle x, y \rangle = \langle y, x \rangle$,
2. Let $\alpha$ and $\beta$ be scalars, and $z$ a vector in $V$, then
   \[
   \langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle,
   \]
3. $\langle x, x \rangle \geq 0$ with $\langle x, x \rangle = 0$ if and only if $x = 0$. 

Example 2.7 When \( V = \mathbb{R}^n \) or \( \mathbb{C}^n \) the standard inner products are given by
\[
x^T y = \sum_{i=1}^{n} x_i y_i \quad \text{and} \quad x^* y = \sum_{i=1}^{n} \bar{x}_i y_i
\]
respectively. Sometimes it maybe important to specify the vector space, in which case we use subscripts. To explicitly specify that we are using the inner product on a complex vector space we can write \( \langle a, b \rangle_{\mathbb{C}^n} \). Most of the time it will be clear from context so we omit this.

The inner product naturally defines the length of a vector. Consider the vector \( x \in \mathbb{R}^n \). Apply the inner product where both arguments are \( x \) to get
\[
\sum_{i=1}^{n} x_i^2
\]
This quantity is the Euclidean norm of a vector, and we denote it by \( ||x||^2 \). Later on we will encounter other functions that define a norm.

![Figure 2.1: Calculating the angle \( \theta \) between vectors \( a \) and \( b \).](image)

On the vector space \( \mathbb{R}^n \) the inner product captions the notion of the cosine of the angle between two vectors as follows. Consider the two vectors depicted in figure 2.6.3. Reading from the figure we have that
\[
\cos \alpha = \frac{a_1}{||a||}, \quad \sin \alpha = \frac{a_2}{||a||}, \quad \cos \beta = \frac{b_1}{||b||}, \quad \sin \beta = \frac{b_2}{||b||}.
\]
Noting that \( \theta = \beta - \alpha \), and recalling the identity \( \cos(\beta - \alpha) = \cos \alpha \cos \beta + \sin \alpha \sin \beta \) we arrive at the following result.

**Proposition 2.6.2** Let \( a \) and \( b \) be any two vectors in \( \mathbb{R}^n \). The cosine of the angle \( \theta \) between \( a \) and \( b \) is related to the inner product by
\[
\cos \theta = \frac{a_1 b_1 + a_2 b_2}{||a|| ||b||} = \frac{\langle a, b \rangle}{||a|| ||b||}.
\]
(2.6)

From the above proposition it can be seen that the inner product is positive when the angle between vectors is less than 90°, negative when it is greater than 90°, and 0 when equal to 90°.

The intuition behind the definition of the inner product is summarized below.

- The formula (2.6) confirms our notion of orthogonality: When \( a^T b = 0 \) the values for \( \theta \) that make \( \cos \theta = 0 \) are 90 and 180 degrees.
- If the length of \( a \) doubles, both the numerator and denominator double, but the angle remains the same.
• If the sign of $b$ is reversed, the angle changes by $180^\circ$.

The following lemma relates norms (section 2.6.1) to inner products and is used heavily throughout the course.

**Theorem 2.6.3 — Cauchy-Schwartz inequality.** Let $x$ and $y$ be two vectors belonging to an inner product space (vector space with an inner product) then,

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

holds for all $x$. Equality is attained if and only if $x$ and $y$ are linearly dependent.

As both sides of the inequality in theorem 2.6.3 are non-negative we can take square roots to get:

$$|\langle x, y \rangle| \leq \|x\|\|y\|,$$

where $\|x\| = \sqrt{\langle x, x \rangle}$.

Theorem 2.6.3 is sometimes presented in the above form.

### 2.6.4 The fundamental theorem of linear algebra

We are now ready to tie together the concepts described in this chapter. Before we do so, we need to extend the concept of orthogonality from vectors to subspaces.

**Definition 2.6.4 — Orthogonal subspaces.** Let $V$ and $W$ be subspaces of $X$. We say that $V$ and $W$ are orthogonal if for every vector $v$ in $V$ and $w$ in $W$, $\langle v, w \rangle = 0$ for all $v$ and $w$.

**Definition 2.6.5 — Orthogonal complement.** Given a subspace $V$ of $X$, the space of all vectors orthogonal to $V$ is called the orthogonal complement of $V$ and is denoted by $V^\perp$.

Formally,

$$V^\perp := \{x \in X \mid \langle x, y \rangle = 0 \text{ for all } y \in V\}.$$

An important point needs to be made here. It is possible for two subspaces to be orthogonal but not be orthogonal complements of each other. For example, consider $X = \mathbb{R}^3$. The line spanned by $(0, 1, 0)$ is a subspace of $X$, call it $V$. Similarly, let the subspace $W$ be defined by the span of $(0, 0, 1)$. Clearly $V$ and $W$ are orthogonal, however, $V \neq W^\perp$. The subspace $V$ is only a part of $W^\perp$, the dimensions don’t match. In general, for two subspaces $V$ and $W$ of $X$ to be orthogonal complements, they must be orthogonal, and

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

We now present the fundamental theorem of linear algebra. The theorem is stated with respect to complex matrices, it follows through exactly for real matrices by swapping conjugate transpose with transpose.

**Theorem 2.6.4 — The fundamental theorem of linear algebra.** Let $A \in \mathbb{C}^{m \times n}$ be a matrix of rank $r$. Then the following are true:

1. Dimensions:
   a. $\dim(\text{range}(A)) = \dim(\text{range}(A^*)) = r$,
   b. $\dim(\text{range}(A)) + \dim(\text{null}(A)) = n$, 

(c) \( \dim(\text{range}(A^*)) + \dim(\text{null}(A^*)) = m. \)

2. Orthogonality:

(a) \( \text{null}(A) = (\text{range}(A^*))^\perp, \)
(b) \( \text{null}(A^*) = (\text{range}(A))^\perp, \)
(c) \( \text{range}(A) = \text{null}(A^*)^\perp, \)
(d) \( \text{range}(A^*) = \text{null}(A)^\perp. \)

**Proof.** We will only prove the orthogonality claims of the theorem. Note that because \((A^*)^* = A\) we only need to prove part 2a and 2c as parts b and d are obtained through transposing the arguments.

We first prove 2a: Suppose \( w \) is in \( \text{null}(A) \) and \( v \) is in \( \text{range}(A^*) \). By definition \( Aw = 0 \) and \( v = A^*x \) for some vector \( x \). We then have that

\[
w^*v = w^*(A^*x) \\
= (w^*A^*)x \\
= (Aw)^*x \\
= 0^*x = 0.
\]

We have shown that \( \text{null}(A) \subseteq (\text{range}(A^*))^\perp \). To finish the proof we show that \( (\text{range}(A^*))^\perp \subseteq \text{null}(A) \): Let \( u \in (\text{range}(A^*))^\perp \), it follows that \( \langle u, v \rangle = 0 \) if \( v \in \text{range}(A^*) \), which is equivalent to \( \langle u, A^*x \rangle = 0 \) for any \( x \in \mathbb{C}^m \). Thus,

\[
0 = \langle u, A^*x \rangle = \langle Au, x \rangle \quad \text{for all } x \in \mathbb{C}^m.
\]

Let \( x \) be a vector in the basis for \( \mathbb{C}^m \) then it must be the case that \( Au = 0 \), i.e. \( u \in \text{null}(A) \) and so \( (\text{range}(A^*))^\perp \subseteq \text{null}(A) \).

To show part c holds, we use the fact that for closed sets \( X, (\mathbb{R}^n \text{ and } \mathbb{C}^n \text{ are closed}), (X^\perp)^\perp = X \).

Applying this identity to the left and right sides of 2b produces 2c.

The FTLA can be extended to include constructions of orthonormal bases for each of the four subspaces. We defer this result to the section on the *singular value decomposition.*
Chapter 2. Vector spaces

2.7 Exercises

Each question or sub-question is worth 3 marks. You need to justify your answers: Writing “yes” or “no” only scores one mark even if correct. Similarly, if asked to provide a basis, just writing $(x, y, z), (0, 1, 1)$ will only score one mark. A well written answer will be a statement that can be read and makes sense without the question.

Marking scheme:
• 0 marks: Incorrect answer, logic is incorrect or inconsistent.
• 1 mark: Attempted and made some progress towards a correct solution. Or, correct answer, no working provided.
• 2 marks: Mostly correct, clearly understands what needs to be done.
• 3 marks: Correct answer or so close to correct that full marks are deserved.

Problem 2.1 — Fields. Determine which (if any) of the following sets define a field:

1. The set of non-negative integers: \{0, 1, 2, 3, \ldots\}.
2. The set of all integers: \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\}.
3. The set of all rational numbers: numbers that can be written as $\frac{a}{b}$ where $a$ and $b$ are integers.

Problem 2.2 — Vector spaces. Is the set of $n \times n$ real symmetric matrices a vector space?

Problem 2.3 — Vector spaces. Show that the set of all $x$’s that satisfy the system of linear equation $Ax = b$ where $A \in \mathbb{R}^{m \times n}$ and $m > n$ is a subspace of $\mathbb{R}^n$.

Problem 2.4 — Subspaces. Determine which (if any) of the sets described below forms a subspace of the x,y-plane $\mathbb{R}^2$:

1. The set of all vectors $x = (x_1, x_2)$, with $x_1 \geq 0$ and $x_2 \geq 0$ a subspace of $\mathbb{R}^2$.
2. The set that includes all the vectors $x_i \geq 0$ and all vectors such that $x_1 \leq 0$ and $x_2 \leq 0$.

Problem 2.5 — Null space. Describe the null spaces and state the rank of the following matrices:

$$X = \begin{pmatrix} 1 & 0 \\ 4 & 6 \\ 2 & 6 \end{pmatrix} \quad \text{and} \quad Y = \begin{pmatrix} 1 & 0 & 1 \\ 4 & 6 & 10 \\ 2 & 6 & 8 \end{pmatrix}.$$ 

Problem 2.6 — Subspaces, span, basis. Describe geometrically the subspace of $\mathbb{R}^6$ spanned by

1. $(0, 0, 0), (0, 1, 0), (0, -4, 0)$.
2. $(0, 0, 1), (0, 1, 1), (0, 2, 1)$.
3. All six vectors above. Of the 6, which form a basis?
4. The set of all non-negative vectors.
Problem 2.7 — Bases. Find a basis (and thus determine the dimension) of the subspace of $\mathbb{R}^4$ given by

$$\left\{ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \bigg| x - w + z = 0 \right\}.$$ 

Problem 2.8 — Dimension. What is the dimension of the space of $\mathcal{S}^n$, i.e., the set of $n \times n$ symmetric matrices. Provide a basis for this space.

Problem 2.9 — Dimension. Find a basis and hence the dimension of the space of matrices such that the sum of the elements in the first column and the trace of the matrix are both equal to zero.

Problem 2.10 — Norms. Let $u$ be a scalar function such that its first derivative with respect to time ($t$) exists. Denote the derivative by $\dot{u}(t)$.

1. Look up the definition of the supremum of a set. Describe it in your own words.

Determine (if any) of the functions described below define a norm:\footnote{The solution doesn’t depend on any subtleties of using $\sup$ instead of $\max$. If you prefer, just treat $\sup$ as $\max$. It will be useful later in the course to understand the difference.}

\begin{align*}
\text{start}=2 & \quad \sup |\dot{u}(t)| \\
\text{stbrt}=2 & \quad |u(0)| + \sup |\dot{u}(t)|
\end{align*}

Problem 2.11 — Inner products. As defined, inner products are linear in the first argument. Use properties 1) and 2) from the definition to show that

$$\langle x, \alpha y + \beta z \rangle = \bar{\alpha} \langle x, y \rangle + \bar{\beta} \langle x, z \rangle,$$

where $\alpha$ and $\beta$ are scalars and $x,y,z$ are vectors.

Problem 2.12 — Cauchy-Schwartz. Prove the Cauchy-Schwartz inequality. Hint: to begin, start by defining the scalar

$$\alpha = \frac{\langle x, y \rangle}{\langle y, y \rangle}$$

and expand $\|x - \alpha y\|^2$.

Problem 2.13 — Fundamental theorem of linear algebra. Consider the matrix

$$X = \begin{pmatrix} 0 & 1 & 5 & 0 \\ 0 & 2 & 10 & 0 \end{pmatrix}.$$ 

Find the dimension and a basis for each of the four subspaces associated with $X$. 

\begin{align*}
\text{start}=2 & \quad \sup |\dot{u}(t)| \\
\text{stbrt}=2 & \quad |u(0)| + \sup |\dot{u}(t)|
\end{align*}
3. Matrices

In this section we shift our focus to matrices. We define norms and inner products for matrices, describe some special types of matrices, as well as introduce some useful matrix factorizations.

3.1 Basics

The sets of matrices with $m$ rows and $n$ columns is denoted by $M^{m \times n}$. It is often necessary to define the field from which entries are drawn. The vector space of matrices with complex entries is denoted by $C^{m \times n}$ and the subset of $C^{m \times n}$ with real entries is $R^{m \times n}$. The element in the $i^{th}$ row and $j^{th}$ column of the matrix $A$ is denoted by $A_{ij}$.

The transpose of a matrix is a function which takes a matrix as its input and returns a new matrix which reverses the column and row index: For $A \in R^{m \times n}$ the transpose, denoted by $A^T$ is given by $A^T_{ij} = A_{ji}$. Note that this produces a matrix $A^T \in R^{n \times m}$. For matrices in $C^{m \times n}$ we use the conjugate transpose denoted by a superscript $\ast$: for $B \in C^{m \times n}$, $B^\ast_{ij} = \overline{B}_{ji}$, where the bar denotes conjugation.

Associated to all square matrices is the trace function. The trace will be used to define inner products on matrices.

**Definition 3.1.1 — Trace function.** Given an matrix $X \in M^{m \times n}$, the trace of $X$ is a scalar function that returns the sum of the diagonal elements of $X$:

$$\text{tr} \ X := \sum_{i=1}^{n} X_{ii}.$$ 

**Proposition 3.1.1 — Trace linearity.** The trace function is linear: i.e., it satisfies

- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$,
- $\text{tr}(\alpha A) = \alpha \cdot \text{tr}(A)$, for all scalars $\alpha$.

Furthermore the function is symmetric, i.e., $\text{tr}(A^\ast) = \text{tr}(A)$.

**Proposition 3.1.2 — Inner products on matrix vector spaces.** Given two matrices $X, Y \in R^{m \times n}$
the function
\[ \text{tr } X^T Y = \sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij} Y_{ij} =: \langle X, Y \rangle. \]

defines an inner product. For the more general case of complex \( X \) and \( Y \) (of appropriate dimension), this generalizes to
\[ \text{tr } X^* Y =: \langle X, Y \rangle. \]

This definition of the inner product can be viewed in terms of vectors. Let \( \text{vec}(X) \) be function such that \( \text{vec} : \mathbb{C}^{m \times n} \to \mathbb{C}^{mn} \), that takes the columns of \( X \) and stacks them into a vector.

**Example 3.1** The \( \text{vec} \) operator applied to a \( 2 \times 2 \) matrix produces a column vector with 4 elements:
\[
W = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \text{then} \quad \text{vec}(W) = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}.
\]

**Corollary 3.1.3** Using the definition of the \( \text{vec} \) function, the inner products defined in definition 3.1.2 acting on \( m \times n \) matrices can be treated inner products on vectors of length \( mn \). Specifically, for the real case we have
\[
\langle X, Y \rangle_{\mathbb{R}^{m \times n}} = \langle \text{vec}(X), \text{vec}(Y) \rangle_{\mathbb{R}^{mn}} = \text{vec}(X)^T \text{vec}(Y).
\]

Recall that we use the subscript on \( \langle \cdot \rangle \) to make explicit which vector space the inner product acts on. Here it’s not really needed as our notation (capital letters denote matrices and we’ve just defined the \( \text{vec} \)) makes clear the two distinct cases. The relationship between these two inner products is known as an isomorphism.

**Proposition 3.1.4** — Cyclic trace function property. The trace function is cyclic meaning that it satisfies
\[
\text{tr}(ABCD) = \text{tr}(BCDA) = \text{tr}(CDAB) = \text{tr}(DABC).
\]

This property can be extended to an arbitrary number of input matrices. Note that arbitrary combinations are not permitted e.g. \( CABD \).

### 3.2 Special Matrices

A very important subset of matrices are symmetric matrices.

**Definition 3.2.1** — Symmetric and Hermitian matrices. Let \( X \) and \( Y \) be matrices belonging to \( \mathbb{R}^{n \times n} \) and \( \mathbb{C}^{n \times n} \) respectively.

- \( X \) is symmetric if and only if \( X = X^T \).
- \( Y \) is Hermitian if and only if \( Y = Y^* \).

The set of \( n \times n \) symmetric matrices is denoted by \( \mathbb{S}^n \). Similarly \( \mathbb{H}^n \) is used to denote Hermitian matrices.

Unlike scalars, the inverse of a matrix only exists for some matrices. A necessary condition is that the matrix must be square.
3.2 Special Matrices

Definition 3.2.2 — Matrix inverse. The matrix $X \in M^{n \times n}$ is invertible if there exists a matrix $Y \in M^{n \times n}$ such that $XY = I$ and $YX = I$. There is at most one such $Y$ and we call $Y$ the inverse of $X$ and denote it by $Y = X^{-1}$. Thus

$$XX^{-1} = I \quad \text{and} \quad X^{-1}X = I.$$ 

If $A$ and $B$ are invertible matrices, then their product is also invertible. The inverse is given by reversing the order of multiplication and inverting each matrix, i.e., $(AB)^{-1} = B^{-1}A^{-1}$. The rule extends to multiple matrices.

For computation it is useful to deal with sets of vectors that are mutually orthogonal and have unit length. Such sets are said to be orthonormal. We will give the definition for bases, but it equally applies to any other set.

Definition 3.2.3 — Orthonormal basis. A basis $v_1, \ldots, v_k$ for the vector space $V$ is orthonormal if

$$\langle v_i, v_j \rangle = \begin{cases} 0 & \text{for all } i \neq j, \\ 1 & \text{when } i = j. \end{cases} \quad \text{(orthonormality)}$$

Recall that to normalize any vector (make it’s length equal to one) replace $x$ with $x/\|x\|$.

Definition 3.2.4 — Unitary matrices. A matrix $U \in M^{n \times n}$ is unitary if and only if $U^*U = I$. In addition, if $U$ is real, then it is said to be orthogonal.

Recall that for a real matrix $X$, $X^* = X^T$. Thus if $X$ is real and unitary $X^TX = I$ and $X$ is orthogonal.

Theorem 3.2.1 If $U \in M^{n \times n}$, the following statements are equivalent:

1. $U$ is unitary.
2. $U^{-1}$ exists and $U^* = U^{-1}$.
3. $UU^* = I$.
4. $U^*$ is unitary.
5. The columns of $U$ form an orthonormal set.
6. The rows of $U$ form an orthonormal set.

Proof. Statement (1) implies (2) by applying the definition $U^*U = I$ satisfies the definition of a matrix inverse. Since $BA = I$ if and only if $AB = I$, (2) implies (3). As $(U^*)^* = U$ we have that (3) implies (4). Each of the implications can be reversed, thus (1)-(4) are equivalent. To show (1) implies (5), we need to consider how matrix multiplication works. Separating $U$ into columns:

$$U = \begin{bmatrix} \vert & \vert & \vdots & \vert \\ u_1 & u_2 & \ldots & u_n \end{bmatrix}, \quad \text{with } u_i \in \mathbb{C}^n \text{ for } i = 1, \ldots, n.$$ 

By definition we have that $U^*U = I$, i.e.,

$$U^*U = \begin{bmatrix} u_1^*u_1 & u_1^*u_2 & \ldots & u_1^*u_n \\ u_2^*u_1 & u_2^*u_2 & \ldots & u_2^*u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n^*u_1 & \ldots & u_n^*u_{n-1} & u_n^*u_n \end{bmatrix} = I.$$
In order for this to hold $u_i^* u_i = 1$ thus columns have length one, and $u_i^* u_j = 0$ for $i \neq j$, so they are orthogonal. (4) implies (6) follows the same argument.

An important property of unitary (and orthogonal) matrices is that when treated as linear transforms, they do not change the magnitude or angle of the vector they act on.

**Proposition 3.2.2** The linear map $Q$ acting on vector $x$ preserves the length of $x$, when $Q$ is unitary, i.e.,

$$\|Qx\| = \|x\|$$

for any vector $x$.

In addition, the inner-product is preserved:

$$\langle Qx, Qy \rangle = \langle x, y \rangle,$$

for all vectors $x$ and $y$.

**Proof.** Simply apply the definitions the inner product on $\mathbb{R}^n$, and the properties of an orthogonal matrix:

$$\langle Qx, Qy \rangle = x^T Q^T Qy = x^T y = x^T y = \langle x, y \rangle.$$

To see that the norm is preserved we follow the same argument. Square the left hand side of the first equation to get

$$\|Qx\|^2 = x^T Q^T Qx = x^T x = \|x\|^2.$$

Swapping transpose for conjugate transpose in the above arguments covers the complex case.

The concept of non-negativity (and by extension, positivity) is well defined for real numbers. It is also true that for all real numbers that either $x \geq 0$ or $x < 0$. In the matrix setting things are a little more complicated. We first have to define what it means to be non-negative.

**Definition 3.2.5 — Positive semidefinite matrices.** A matrix $Y \in \mathbb{S}^n$ is said to be **positive semidefinite** if

$$x^T Y x \geq 0 \quad \text{for all } x \in \mathbb{R}^n.$$

Furthermore, if

$$x^T Y x > 0 \quad \text{for all } x \in \mathbb{R}^n \text{ and } x \neq 0$$

then $Y$ is said to be **positive definite.** We write $Y \succeq 0$ ($Y \succ 0$) to denote that it is positive semidefinite (positive definite).

Unfortunately definition 3.2.5 doesn’t provide us with a practical method for testing if $Y \succeq 0$ as to do so would mean evaluating $x^T Y x$ for an infinite number of choices of $x$. In the next section we will see how eigenvalues and orthogonality will provide a simple test.

**Example 3.2** Every positive definite matrix $P \succ 0$ defines a norm on $\mathbb{R}^n$ via $\|x\|_P := x^T P x$.

### 3.3 Matrix norms

We now consider extending the concept of a norm to the setting where elements of our vector space $V$ are matrices. The function that defines a norm must still satisfy the properties of definition 2.6.1 and we still use $\| \cdot \|$ to denote a matrix norm.
3.3 Matrix norms

One of the most frequently used matrix norms is the Frobenius norm denoted \( \| \cdot \|_F \): Given a matrix \( A \in \mathbb{M}^{m \times n} \),

\[
\| A \|_F := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |X_{ij}|^2} = \sqrt{\text{trace}(A^*A)}.
\]

The Frobenius norm can be viewed as measuring the size of the matrix through the magnitude of its \( mn \) components. An alternative and often more useful approach is to define the norm in terms of the behavior of the linear map between the domain of the matrix and its range space. Such norms are referred to as “induced norms”.

**Definition 3.3.1 — Induced norm.** For a matrix \( A \in \mathbb{M}^{m \times n} \) if \( \| \cdot \|_p \) and \( \| \cdot \|_q \) are norms defined on \( \mathbb{C}^n \) and \( \mathbb{C}^m \) respectively, then the induced norm is defined as

\[
\| A \|_{p \rightarrow q} := \sup_{x \neq 0, x \in \mathbb{C}^n} \frac{\| Ax \|_q}{\| x \|_p} = \sup_{\| x \|_p \leq 1} \| Ax \|_q = \sup_{\| x \|_p = 1} \| Ax \|_q.
\]

When \( p \) and \( q \) are chosen such that \( p = q \) we simply write \( \| A \|_p \).

Note that the definitions on the left and right follow from property 2 (homogeneity) of the norm function in definition 2.6.1. To show equivalence to the middle definition requires a little extra work (which we omit for now).

In words, the \( \| \cdot \|_{p \rightarrow q} \) norm is the maximum ratio of “output” (\( \| Ax \|_q \)) to “input” (\( \| x \|_p \)) that any vector \( x \) can achieve. Another interpretation is that \( \| A \|_p \) bounds the “gain” of the matrix in the sense that

\[
\| Ax \|_p \leq \| A \|_p \| x \|_p \quad \text{for all vectors } x,
\]

and equality holds for at least one non-zero \( x \). For induced norms where \( p \neq q \), inequality (3.1) generalizes to

\[
\| Ax \|_q \leq \| A \|_{p \rightarrow q} \| x \|_p \quad \text{for all vectors } x.
\]

The “induced” name is now clear as the vector \( x \) is used to induce the norm on the matrix \( A \).

Clearly there is great flexibility in choice of matrix norms. However, computing \( \| \cdot \|_{p \rightarrow q} \) (and defining closed from expressions for) many choices of \( p \) and \( q \) is a challenging task. There are however three choices that are straightforward to characterize and compute\(^1\); namely the cases for \( p = 1, 2, \infty \).

**Example 3.3** Let \( A \) be an \( m \times n \) matrix. Then we have:

- **Maximum column sum:** \( p = 1 \)

\[
\| A \|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |A_{ij}| = \max_{1 \leq j \leq n} \| a_j \|_1, \quad \text{where } a_j \text{ is the } j^{\text{th}} \text{ column of } A.
\]

- **Maximum row sum:** \( p = \infty \):

\[
\| A \|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |A_{ij}| = \max_{1 \leq i \leq m} \| a'_i \|_1, \quad \text{where } a'_i \text{ is the } i^{\text{th}} \text{ row of } A.
\]

- **Spectral norm:** \( p = 2 \)

\[
\| A \|_2 = \sigma_1(A), \quad \text{where } \sigma_1(\cdot) \text{ denotes the largest singular value.}
\]

The spectral norm is often also referred to as the **operator norm**.

\(^1\)For the \( p = 2 \) case, computation is straightforward for medium sized matrices, for very large matrices we run into problems.
Chapter 3. Matrices

The examples given above need to be proven. This will be deferred to the exercises as the end of the chapter. The singular value decomposition required in the definition of the spectral norm will be introduced in the next chapter.

Certain matrix norms satisfy the sub-multiplicative property:

$$
\|AB\| \leq \|A\| \|B\|. 
$$

(3.3)

The Frobenius norm is sub-multiplicative, as are induced norms where $p = q$. Induced norms with $p \neq q$ are sub-multiplicative, but care needs to be taken with regards to the the norms on the right hand side.

**Theorem 3.3.1** Given matrices $A \in \mathbb{R}^{l \times m}$ and $B \in \mathbb{R}^{m \times n}$ and let $\| \cdot \|_l$, $\| \cdot \|_m$, and $\| \cdot \|_n$ be norms on $\mathbb{R}^l$, $\mathbb{R}^m$, and $\mathbb{R}^n$ respectively. Then

$$
\|AB\|_n \leq \|A\|_m \|B\|_n.
$$

**Proof.** For any $x \in \mathbb{R}^n$ we have

$$
\|ABx\|_l \leq \|A\|_m \|Bx\|_m \leq \|A\|_m \|B\|_n \|x\|_n
$$

which follows from applying the inequality (3.2) twice. Dividing through by $\|x\|_n$ gives

$$
\frac{\|ABx\|_l}{\|x\|_n} \leq \|A\|_m \|B\|_n.
$$

As the choice of $x$ was arbitrary, selecting the vector that maximizes the right hand side gives the desired result. 

**3.3.1 Dual norms**

This could go in previous chapter, but a simple interpretation comes from treating the vector $y^T$ as a matrix and then taking norms...