Lecture notes for Math 61CM, Linear algebra, Version 2018

Lenya Ryzhik*

October 4, 2018

Nothing found here is original except for a few mistakes and misprints here and there. These notes are simply a record of what I cover in class, to spare the students some of the necessity of taking the lecture notes and compensate for my bad handwriting. The material comes mostly from the book Leon Simon “An Introduction to Multivariable Mathematics” and lecture notes by Andras Vasy.

1 Notes on algebra

This section goes a little bit more generally about the notion of a field we have introduced in the analysis class when we defined the axioms for the real numbers. We start with the definition of a group.

Definition 1.1 A group \((G, \ast)\) is a set \(G\) together with a map \(\ast : G \times G \to G\) with the properties

1. (Associativity) For all \(x, y, z \in G\), \(x \ast (y \ast z) = (x \ast y) \ast z\).

2. (Units) There exists \(e \in G\) such that for all \(x \in G\), \(x \ast e = x = e \ast x\).

3. (Inverses) For all \(x \in G\) there exists \(y \in G\) such that \(x \ast y = e = y \ast x\).

A basic property is that one can talk about the unit, that is, given (1) and (2), \(e\) is unique:

Lemma 1.2 In any group \((G, \ast)\), the unit \(e\) is unique.

Proof: Suppose \(e, f \in G\) are units. Then \(e = e \ast f\) since \(f\) is a unit, and \(e \ast f = f\) since \(e\) is a unit. Combining these, \(e = f\). □

Note that this proof used only (1) and (2), so it is useful to define a more general notion than that of a group.

Definition 1.3 A semigroup \((G, \ast)\) is a set \(G\) together with a map \(\ast : G \times G \to G\) with the properties

1. (Associativity) For all \(x, y, z \in G\), \(x \ast (y \ast z) = (x \ast y) \ast z\).

2. (Units) There exists \(e \in G\) such that for all \(x \in G\), \(x \ast e = x = e \ast x\).

Thus, a semigroup would be a group if each element had an inverse. Notice also that the proof of the above lemma shows that even in a semigroup, the unit is unique.

Inverses are also unique in a group, or, when they exist, in a semigroup.

*Department of Mathematics, Stanford University, Stanford, CA 94305, USA; ryzhik@math.stanford.edu
Lemma 1.4 Suppose that \((G,\ast)\) is a semigroup with unit \(e\), \(x \in G\), and suppose that there exist \(y, z \in G\) such that \(y \ast x = e = x \ast z\). Then \(y = z\).

Notice that if \(G\) is a group, the existence of such a \(y, z\) is guaranteed, even with \(y = z\), by (3). Thus, this lemma says in particular that in a group, inverses are unique.

However, it says more: in a semigroup, any left inverse (if it exists) equals any right inverse (if it exists). In particular, if both left and right inverses exist, they are both unique: e.g. if \(y, y'\) are left inverses, they are both equal to any left inverse \(z\), and thus to each other.

**Proof:** We have

\[
y = y \ast e = y \ast (x \ast z),
\]

where we used that \(e\) is the unit and \(x \ast z = e\). Similarly, we have

\[
z = e \ast z = (y \ast x) \ast z.
\]

But by the associativity, \(y \ast (x \ast z) = (y \ast x) \ast z\), so combining these three equations shows that \(z = y\), as desired. \(\Box\)

There are many interesting groups, such as \((\mathbb{R}, +)\), \((\mathbb{Z}, +)\), \((\mathbb{Q}, +)\), \((\mathbb{R}^n, +)\), \((\mathbb{R}^+, \cdot)\), where \(\mathbb{R}^+\) consists of the positive reals, as well as semigroups, such as \((\mathbb{R}, \cdot)\) (all non-zero elements have inverses), \((\mathbb{Z}, \cdot)\) (only \(\pm 1\) have inverses). Another group with a different flavor is \((\mathbb{Z}, 1)\), the integers modulo \(n \geq 2\): as a set, this can be identified with \([0, 1, \ldots, n - 1]\) (the remainders when dividing by \(n\)), and addition gives the usual sum in \(\mathbb{Z}\), reduced modulo \(n\). For instance, in \((\mathbb{Z}_5, +)\), we have \(2 + 4 = 1\). It may be less confusing though to write \([0], [1], \ldots, [n - 1]\) for the set, and represent the last identity in \(\mathbb{Z}_5\) as \([2] + [4] = [1]\). In general, when the operation is understood, one might just write the set for a group or semigroup, without indicating the operation.

An important and slightly different class of groups are groups of transformations. Fix a set \(X\) and consider maps \(G : X \to X\) with composition serving as group multiplication: \(G_1 \circ G_2\) is a map from \(X\) to \(X\) such that \(G_1 \circ G_2(x) = G_1(G_2(x))\). A collection \(\mathcal{G}\) of such maps forms a group if (1) it is closed under composition: if \(G_1, G_2 \in \mathcal{G}\) then \(G_1 \circ G_2 \in \mathcal{G}\), (2) the identity map \(\text{Id}\) is in \(\mathcal{G}\)—recall that \(\text{Id}(x) = x\) for all \(x \in X\), and (3) for each \(G \in \mathcal{G}\) there is a map \(G^{-1}\) in \(\mathcal{G}\) such that \(G \circ G^{-1} = G^{-1} \circ G = \text{Id}\).

**Exercise 1.5** Show that a map \(G : X \to X\) has a left-and-right inverse \(G^{-1}\) such that \(G \circ G^{-1} = G^{-1} \circ G = \text{Id}\) if and only if \(G\) is one-to-one and onto.

**Exercise 1.6** Show that the maps \(\mathbb{R}^2 \to \mathbb{R}^2\) of the form \((x_1, x_2) \to (ax_1 + bx_2, cx_2 + dx_2)\) with \(ad - bc = 1\) form group.

**Exercise 1.7** Let \(\ell_\infty\) be the set of all infinite sequences \(a = (a_1, a_2, \ldots, a_n, \ldots)\) such that there exists a constant \(M\) that depends on the sequence but not on \(n\) so that

\[
|a_n| \leq M \text{ for all } n \in \mathbb{N}.
\]

Define a map \(S_k\) acting on such sequences as \(S_k : \ell_\infty \to \ell_\infty\) as \(S_k \cdot a = b\), with the sequence \(b_n\) having the entries

\[
b_n = e^{-k}a_n.
\]

Show that each \(S_k\) maps \(\ell_\infty\) to \(\ell_\infty\). In other words, that if the sequence \(a_n\) is in \(\ell_\infty\), then the sequence \(b_n\) is also in \(\ell_\infty\), and also show that \(S_k, k \in \mathbb{Z}\) form a group, with \(\text{Id} = S_0\), and \(S_k \circ S_n = S_{k+n}\).
An example of a semi-group that may be less familiar is as follows. Again, let \( \ell_\infty \) be the set of all infinite sequences \( a = (a_1, a_2, \ldots, a_n, \ldots) \) such that there exists a constant \( M \) that depends on the sequence but not on \( n \) so that 
\[
|a_n| \leq M \text{ for all } n \in \mathbb{N}.
\]
Given any \( k \geq 0 \), we may define a map \( D_k : \ell_0 \to \ell_0 \) as \( D_k a = b \), with the sequence \( b_n \) having the entries 
\[
b_n = e^{-kn}a_n.
\]
If \( a \in \ell_\infty \), there exists \( M \) so that \( |a_n| \leq M \) for all \( n \). Then we have 
\[
|b_n| \leq e^{-kn}|a_n| \leq e^{-kn}M \leq M,
\]
hence the sequence \( b_n \) is also in \( l_\infty \). Thus, \( S_k \) is, indeed, a map from \( \ell_\infty \) to \( \ell_\infty \). It is also easy to check that the set \( \mathcal{G} \) of all \( S_k \), with \( k \geq 0 \), is a commutative semi-group, with the product given by the composition of maps:
\[
S_k \circ S_n = S_{k+n}, \tag{1.1}
\]
and \( \text{Id} = S_0 \) being the unit.

**Exercise 1.8** Check that (1.1), indeed, holds, where the left side is the composition of \( S_n \) and \( S_k \).

However, \( S_k \) can not have an inverse in \( \mathcal{G} \). Indeed, if there is a map \( D_k \) that maps \( \ell_\infty \) to \( \ell_\infty \) such that \( D_k \circ S_k = \text{Id} \), then it would act as \( D_k a = c \), with the entries 
\[
c_n = e^{kn}a_n.
\]
In other words, we would have to set \( D_k = S_{-k} \) for \( k \geq 0 \). However, such \( D_k \) is not a map in \( \mathcal{G} \). Moreover, we can not add to \( \mathcal{G} \) maps "\( S_k \) with \( k < 0 \)" if we want to keep them as maps from \( \ell_\infty \) to \( \ell_\infty \). This is because of the following exercise.

**Exercise 1.9** Fix \( k \in \mathbb{N} \) and find a sequence \( a = (a_1, a_2, \ldots, a_n, \ldots) \) that belongs to \( \ell_\infty \) such that the sequence \( b = (b_1, b_2, \ldots, b_n, \ldots) \) with the entries 
\[
b_n = e^{kn}a_n
\]
does not belong to \( \ell_\infty \).

Many (semi)groups are commutative; in fact, all of the above examples are:

**Definition 1.10** A commutative, or abelian, semigroup \( (G, \ast) \) is one in which \( x \ast y = y \ast x \) for all \( x, y \in G \).

Noncommutative semigroups will play a role in this class, including the set \( M_n \) of \( n \times n \) matrices with matrix multiplication as the operation, which is non-commutative if \( n \geq 2 \), and permutations of a finite set \( S \) which is non-commutative if the set has at least 3 elements (this will be discussed when we talk about determinants).

We then can make the following definition:

**Definition 1.11** A field \( (F, +, \cdot) \) is a set \( F \) with two maps \( + : F \times F \to F \) and \( \cdot : F \times F \to F \) such that

1. \( (F, +) \) is a commutative group, with unit \( 0 \).
2. \((F, \cdot)\) is a commutative semigroup with unit 1 such that \(1 \neq 0\) and such that \(x \neq 0\) implies that \(x\) has a multiplicative inverse (i.e. \(y\) such that \(x \cdot y = 1 = y \cdot x\)).

3. The distributive law holds:
\[
x \cdot (y + z) = x \cdot y + x \cdot z.
\]

One usually writes \(-x\) for the additive inverse (inverse with respect to \(+\)), \(x^{-1}\) for the multiplicative inverse. The common examples of a field include \((\mathbb{R}, +, \cdot)\), \((\mathbb{Q}, +, \cdot)\), and complex numbers \((\mathbb{C}, +, \cdot)\). A more interesting field is the subset of \(\mathbb{R}\) of numbers of the form
\[
\{a + b \sqrt{2} : a, b \in \mathbb{Q}\}.
\]
The most interesting part in showing that this is a field is that multiplicative inverses exist; that these exist (within this set!) when \(a + b \sqrt{2} \neq 0\) follows from the following computation in \(\mathbb{R}\):
\[
(a + b \sqrt{2})^{-1} = \frac{a - b \sqrt{2}}{a^2 - 2b^2} = (a^2 - 2b^2)^{-1}a - (a^2 - 2b^2)^{-1}b \sqrt{2}.
\]
Note that both \((a^2 - 2b^2)^{-1}a, -(a^2 - 2b^2)^{-1}b\) are, indeed, rational, and the denominator \(a^2 - 2b^2 \neq 0\), since \(\sqrt{2}\) is an irrational number.

Finally, \(\mathbb{Z}_n\) is not a field for a general \(n\). For instance, if \(n = 6\), \([2] \cdot [3] = [0]\). However, if \(n\) is a prime \(p\), then it is – it is the finite field of \(p = n\) elements.

**Exercise 1.12** Check that \(\mathbb{Z}_n\) is a field if and only if \(n\) is a prime number.

Another very important example is the set \(\mathbb{C}\) of complex numbers \(z = x + iy\), with \(x, y \in \mathbb{R}\) and the rules of addition
\[
(x + iy) + (u + iw) = (x + u) + i(y + w), \quad (1.2)
\]
and multiplication
\[
(x + iy)(u + iw) = (xu - yw) + i(yu + xw). \quad (1.3)
\]

**Exercise 1.13** Show that \(\mathbb{C}\) is a field, and the inverse of a complex number \(x + iy\) is given by
\[
(x + iy)^{-1} = \frac{x}{x^2 + y^2} - i \frac{y}{x^2 + y^2}, \quad (1.4)
\]
provided that not both \(x = 0\) and \(y = 0\).

As an example of a general result in a field, let us show the following.

**Lemma 1.14** If \((F, +, \cdot)\) is a field, then \(0 \cdot x = 0\) for all \(x \in F\).

**Proof.** Since \(0 = 0 + 0\), we have
\[
0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x, \quad (1.5)
\]
so
\[
0 = -(0 \cdot x) + (0 \cdot x) = -(0 \cdot x) + (0 \cdot x + 0 \cdot x) = (-0 \cdot x) + 0 \cdot x + 0 \cdot x = 0 + 0 \cdot x = 0 \cdot x,
\]
as desired. On the last line, the first equality uses that \(-0 \cdot x\) is the additive inverse of \(0 \cdot x\), the second uses (1.5), the third uses associativity of addition, the fourth uses again that \(-0 \cdot x\) is the additive inverse of \(0 \cdot x\), while the fifth uses that \(0\) is the additive unit. \(\square\)

Notice that this proof uses the distributive law crucially: this is what links addition (0 is the additive unit!) to multiplication.

For more examples, see Appendix A, Problem 1.1 in the Simon book, and the analysis lecture notes.
2 Vector spaces

2.1 The definition of a vector space

One usually encounters first vectors as "arrows" on the plane that can be added and multiplied by a real number. The general definition of a vector space generalizes this concept in a dramatic fashion.

**Definition 2.1** A vector space $V$ over a field $F$ is a set $V$ such that the sum $v + w \in V$ is defined for any two elements $v, w \in V$, and for $\lambda \in F$ and $v \in V$ the product $\lambda v$ is defined as an element of $V$. These operations satisfy the following properties.

1. The set $(V, +)$ is a commutative group, with unit 0 and with the inverse of $x$ denoted by $-x$,

2. For all $x \in V$, $\lambda, \mu \in F$, we have

$$ (\lambda \star \mu)x = \lambda(\mu x), \ 1x = x. $$

Here, 1 is the multiplicative unit of the field $F$ and $\star$ denotes the product in $F$. Note that in the expression $\lambda \star \mu$ in the left side, we are using the field multiplication in $F$; everywhere else the product is by an element in $F$ and in $V$.

3. For all $x, y \in V$, $\lambda, \mu \in F$,

$$ (\lambda + \mu)x = \lambda x + \mu x $$

and

$$ \lambda(x + y) = \lambda x + \lambda y, $$

so that the two distributive laws hold.

Note that the two distributive laws are different: in the first, in the left side, $+$ is in $F$, in the second, it is in $V$. In the right side, it is in $V$ in both distributive laws.

The ‘same’ argument as for fields shows that $0x = 0$, and $\lambda 0 = 0$, where, in the first case, in the left 0 is the additive unit of the field, and all other 0’s are the zero vector, the additive unit of the vector space. Another example of what one can prove in a vector space is

**Lemma 2.2** Suppose $V$ is a vector space over a field $F$. Then $(-1)x = -x$.

**Proof.** We have, using the distributive law, and that $0x = 0$, observed above,

$$ 0 = 0x = (1 + (-1))x = 1x + (-1)x = x + (-1)x, $$

which says exactly that $(-1)x$ is the additive inverse of $x$, which we denote by $-x$. □

**Remark 2.3** In these lectures, we will usually take $F = \mathbb{R}$, sometimes $F = \mathbb{C}$ and rarely any other field. Whenever we do not specify the field, that means that either the result holds for all $F$ (most often), or $F = \mathbb{R}$ (less often and should be clear in the context), or we have forgotten to specify $F$ completely. If in doubt, ask!
2.2 The space $\mathbb{R}^n$

A standard example of a vector space is the space $\mathbb{R}^n$ of $n$-tuples $x = (x_1, x_2, \ldots, x_n)$ with each $x_k \in \mathbb{R}$. Addition of vectors in $\mathbb{R}^n$ is defined component-wise as

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n).$$

Multiplication by a number $\lambda \in \mathbb{R}$ is also defined component-wise:

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

A simple observation is that if we define the vectors

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

then every vector $x = (x_1, x_2, \ldots, x_n)$ can be written as

$$x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n.$$ 

A natural notion coming from the elementary geometry is the norm of a vector $x = (x_1, \ldots, x_n)$:

$$\|x\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}.$$ 

More generally, if $F$ is a field, $F^n$ is the set of ordered $n$-tuples of elements of $F$. Then $F^n$ is a vector space with the definition of component-wise addition and component-wise multiplication by scalars $\lambda \in F$, exactly as for $\mathbb{R}^n$, as is easy to check. An important special case is the space $\mathbb{C}^n$.

2.3 Some vector spaces of functions

A less familiar vector space, over $\mathbb{R}$, is the set $C([0, 1])$ of continuous real-valued functions on the interval $[0, 1]$ (that officially do not know yet in 61CM). Here, addition and multiplication by elements of $\mathbb{R}$ is defined by:

$$(f + g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x), \quad \lambda \in \mathbb{R}, \quad f, g \in C([0, 1]), \quad x \in [0, 1].$$

That is, $f + g$ is the continuous function on $[0, 1]$ whose value at any $x \in [0, 1]$ is the sum of the values of $f$ and $g$ at that point, and $\lambda f$ is the continuous function on $[0, 1]$ whose value at any $x \in [0, 1]$ is the product of $\lambda$ and the value of $f$ at $x$ (this is a product in $\mathbb{R}$). In a sense, this is a generalization of what we have seen for $\mathbb{R}^n$ or $\mathbb{F}^n$. However, we are not taking $n$-tuples but collections parametrized by the coordinates $x \in [0, 1]$. In other words, instead of having $n$ coordinates for a vector, "we have $[0, 1]$ of the coordinates".

Another common example of a set of functions that form a vector space is the set $P$ of all polynomials $p(x)$ on $\mathbb{R}$, with real valued coefficients, with the usual addition and multiplication by a number. This is simply because the sum of two polynomials is also a polynomial, and if $\lambda \in \mathbb{R}$ and $p(x)$ is a polynomial with real valued coefficients, then $\lambda p(x)$ is a polynomial with real valued coefficients.

2.4 Subspaces and linear dependence of vectors

Let $V$ be a vector space. We say that a subset $W$ of $V$ is a subspace of $V$ if $W$ forms a vector space itself. In other words, $W$ is a subspace of $V$ if $0 \in W$, and $W$ is closed under addition: if $v, w \in W$ then $v + w \in W$, and under multiplication by an element of $F$: if $v \in W$ and $\lambda \in F$ then $\lambda v \in W$.

The most trivial and also most boring example of a subspace is $W = \{0\}$. Here are some less trivial examples.
Exercise 2.4 (1) Let $V$ be the set of all vectors $x$ in $\mathbb{R}^n$ such that $x_1 = 0$. Show that $V$ is a subspace of $\mathbb{R}^n$.

(2) Let $V$ be the set of all vectors $x$ in $\mathbb{R}^n$, $n \geq 3$, such that $x_1 + 2x_2 + 17x_3 = 0$. Show that $V$ is a subspace of $\mathbb{R}^n$.

(3) Let $P$ be the vector space of all polynomials $p(x)$, $x \in \mathbb{R}$, and $V$ be the collection of all polynomials $p(x)$ such that $p(2) = 0$. Show that $V$ is a subspace of $P$.

(4) Let $P$ be the vector space of all polynomials $p(x)$, $x \in \mathbb{R}$, and $W$ be the collection of all polynomials $p(x)$ such that $p(2) = 1$. Show that $W$ is not a subspace of $P$.

To give more general examples of subspaces, we introduce some terminology.

Definition 2.5 The span of a collection of vectors $v_1, \ldots, v_N \in V$ is the collection of all vectors $y \in V$ that can be represented in the form

$$y = c_1 v_1 + c_2 v_2 + \cdots + c_N v_N,$$

with $c_k \in \mathbb{F}$, $k = 1, \ldots, N$. We denote the span of $v_1, \ldots, v_N$ as $\text{span}[v_1, \ldots, v_N]$.

A standard example of a span is a two-dimensional plane in $\mathbb{R}^3$: you fixed two vectors $v_1, v_2 \in \mathbb{R}^3$ and consider all vectors of the form $\lambda v_1 + \mu v_2$, with $\lambda, \mu \in \mathbb{R}$ — this is the plane spanned by $v_1$ and $v_2$.

More generally, we may start with any collection $W$, finite or infinite, of vectors in $V$ and define the span of $W$ as the collection of all vectors in $V$ that can be written as a finite sum

$$y = c_1 w_1 + c_2 w_2 + \cdots + c_N w_N,$$

with $c_k \in \mathbb{F}$, and $w_k \in W$, for all $k = 1, \ldots, N$.

Lemma 2.6 Let $W$ be a non-empty collection of vectors in a vector space $V$. Then $\text{span}[W]$ is a sub-space of $V$.

Proof. First, note that if we take any $v \in W$, then $0 = 0v$ is in $\text{span}[W]$. Next, if $v, w \in \text{span}[W]$, then there exist $c_1, \ldots, c_N \in \mathbb{F}$ and $u_1, \ldots, u_N \in W$, and also $d_1, \ldots, d_M \in \mathbb{F}$ and $\tilde{w}_1, \ldots, \tilde{w}_M \in W$ such that

$$v = c_1 u_1 + \cdots + c_N u_N, \quad w = d_1 \tilde{w}_1 + \cdots + d_M \tilde{w}_M.$$

Then we simply write

$$v + w = c_1 u_1 + \cdots + c_N u_N + d_1 \tilde{w}_1 + \cdots + d_M \tilde{w}_M,$$

and see that $v + w$ is in $\text{span}[W]$, and we also have, for any $\lambda \in \mathbb{F}$:

$$\lambda v = (\lambda c_1) u_1 + \cdots + (\lambda c_N) u_N,$$

showing that $\lambda v$ is in $\text{span}[W]$. □

Later, we will show that any sub-space of $\mathbb{R}^n$ is a span of a finite collection of vectors $v_1, \ldots, v_N \in \mathbb{R}^n$, with $N \leq n$. This is not true for all subspaces of all vector spaces.

Exercise 2.7 Let $P$ be the vector space of all polynomials $p(x)$, $x \in \mathbb{R}$, and $V$ be the collection of all polynomials $p(x)$ such that $p(2) = 0$. We have seen in Exercise 2.4 that $V$ is a sub-space of $P$. Show that $V$ is not a span of a finite collection of polynomials in $P$. 


Definition 2.8  Vectors \( v_1, \ldots, v_N \in V \) are linearly dependent if there exist \( c_k \in \mathbb{F} \), \( k = 1, \ldots, N \) such that not all \( c_k = 0 \) and

\[
c_1v_1 + \cdots + c_Nv_N = 0.
\]  

We say that the vectors \( v_1, \ldots, v_N \in V \) are linearly independent if they are not linearly dependent. In other words, \( v_1, \ldots, v_N \in V \) are linearly independent if (2.1) implies that all \( c_k = 0 \).

Exercise 2.9  (1) Show that the collection of vectors

\[ e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (1, 0, \ldots, 0), \]

is a collection of linearly independent vectors in \( \mathbb{R}^n \).

(2) Show that the vectors \( (1, 0, 0), (1, 2, 1), (0, 0, 1), (1, 1, 1) \) form a linearly dependent collection in \( \mathbb{R}^3 \).

The next lemma shows that if \( v_1, \ldots v_N \) are linearly dependent then one of these vectors can be expressed as a linear combination of the rest.

Lemma 2.10  Vectors \( v_1, \ldots v_N \) are linearly dependent if and only if there exists \( k \in \{1, \ldots, N\} \) and \( c_1, \ldots, c_{k-1}, c_{k+1}, \ldots, c_N \in \mathbb{F}, \) such that

\[
v_k = c_1v_1 + \cdots + c_{k-1}v_{k-1} + c_{k+1}v_{k+1} + \cdots + c_Nv_N. \tag{2.2}
\]

Proof. Let us first assume that \( v_1, \ldots, v_N \) are linearly dependent, so that there exist \( c_k, k = 1, \ldots, N \) such that not all \( c_k = 0 \) and

\[
c_1v_1 + \cdots + c_Nv_N = 0. \tag{2.3}
\]

Then, choose some \( k \) such that \( k \neq 0 \) and write

\[
v_k = \frac{c_1}{c_k}v_1 - \frac{c_2}{c_k}v_2 - \cdots - \frac{c_{k-1}}{c_k}v_{k-1} - \frac{c_{k+1}}{c_k}v_{k+1} - \cdots - \frac{c_N}{c_k}v_N,
\]

which shows that \( v_k \) is a linear combination of \( v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_N \).

Next, assume that (2.2) holds. Then we can write

\[
c_1v_1 + \cdots + c_{k-1}v_{k-1} + (-1)v_k + c_{k+1}v_{k+1} + \cdots + c_Nv_N,
\]

showing that the vectors \( v_1, \ldots, v_N \in V \) are linearly dependent. \( \square \)

3 Inner products

3.1 The inner product on \( \mathbb{R}^n \)

The dot product of two vectors \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) and \( y = (y_1, \ldots, y_n) \in \mathbb{R}^n \) is defined as

\[
(x \cdot y) = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \sum_{k=1}^{n} x_ky_k.
\]

The dot product has some simple basic properties: first,

\[
(x \cdot y) = (y \cdot x), \quad \text{for all } x, y \in \mathbb{R}^n, \tag{3.1}
\]

and, second, we have

\[
(\lambda x + \mu y) \cdot z = \lambda (x \cdot z) + \mu (y \cdot z), \quad \text{for all } x, y, z \in \mathbb{R}^n \text{ and } \lambda, \mu \in \mathbb{R}. \tag{3.2}
\]
We may also define the norm of a vector \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) as
\[
\|x\| = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2} = \left( \sum_{k=1}^{n} x_k^2 \right)^{1/2}.
\] (3.3)
Alternatively, we may write
\[
\|x\| = (x \cdot x)^{1/2},
\] (3.4)
which shows that
\[
(x \cdot x) \geq 0 \text{ for all } x \in \mathbb{R}^n.
\] (3.5)
A simple but extremely important consequence of these properties is the Cauchy-Schwartz inequality.

**Theorem 3.1** *(The Cauchy-Schwartz inequality)* For any vectors \( x, y \in \mathbb{R}^n \) we have
\[
|\langle x, y \rangle| \leq \|x\| \|y\|.
\] (3.6)

**Proof.** While this claim seems a tedious exercise in algebra, the proof is surprisingly short and elegant. Note that if \( x = 0 \) or \( y = 0 \) then the claim is trivial. We fix \( x \neq 0 \) and \( y \neq 0 \) in \( \mathbb{R}^n \) and consider the function
\[
p(t) = \|x + ty\|^2,
\]
where \( t \in \mathbb{R} \). Writing
\[
p(t) = \|x + ty\|^2 = (x + ty) \cdot (x + ty) = (x \cdot x) + 2t(x \cdot y) + t^2(y \cdot y),
\] (3.7)
shows that \( p(t) \) is a quadratic polynomial in \( t \) – recall that \( x \) and \( y \) are fixed. However, \( p(t) \geq 0 \) because of the property (3.5) of the dot product – it is the product of the vector \( x + ty \) with itself. A quadratic polynomial of the form
\[
a + bt + ct^2
\]
is non-negative for all \( t \in \mathbb{R} \) if and only if \( b^2 \leq 4ac \). Translating this into the coefficients of \( p(t) \) in (3.7), with
\[
a = (x \cdot x) = \|x\|^2, \quad b = 2(x \cdot y), \quad c = (y \cdot y) = \|y\|^2,
\]
gives
\[
|\langle x, y \rangle|^2 \leq \|x\|^2 \|y\|^2,
\]
which is (3.6). □

An important consequence of the Cauchy-Schwartz inequality is the triangle inequality.

**Theorem 3.2** *(The triangle inequality)* For any \( x, y \in \mathbb{R}^n \) we have
\[
\|x + y\| \leq \|x\| + \|y\|.
\] (3.8)

**Proof.** Since both the left and the right side in (3.8) are non-negative, it suffices to establish this inequality for the squares of the left and right sides. However, we have, using the Cauchy-Schwartz inequality
\[
\|x + y\|^2 = \|x\|^2 + 2(x \cdot y) + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,
\]
which proves (3.8). □
3.2 General inner products

The standard dot product on $\mathbb{R}^n$ gives an incredibly useful structure when it can be defined on other vector spaces. Here is the general definition that takes the basic properties of the standard dot product as the starting point.

**Definition 3.3** An inner product space $V$ is a vector space over $\mathbb{R}$ with a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that

1. (Positive definiteness) $\langle x, x \rangle \geq 0$ for all $x \in V$, with $\langle x, x \rangle = 0$ if and only if $x = 0$.
2. (Linearity in the first slot) $\langle (\lambda x + \mu y), z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$ for all $x, y, z \in V$, $\lambda, \mu \in \mathbb{R}$.
3. (Symmetry) $\langle x, y \rangle = \langle y, x \rangle$.

One often writes $x \cdot y = \langle x, y \rangle$ for an inner product. The standard dot product on $\mathbb{R}^n$ is an example of an inner product:

$$x \cdot y = \sum_{j=1}^{n} x_j y_j, \ x = (x_1, \ldots, x_n), \ y = (y_1, \ldots, y_n).$$

It is easy to verify that conditions (1)-(3) hold for the standard dot product.

A more interesting inner product is defined on the space $P$ of polynomials with real valued coefficients, as

$$\langle f, g \rangle = \int_{0}^{1} f(x)g(x) \, dx. \quad (3.9)$$

**Exercise 3.4** Check that (3.9) defines an inner product on $P$. In particular, explain why $\langle f, f \rangle = 0$ implies that $f(x) \equiv 0$ is a zero polynomial.

There is an extension of the definition when the underlying field is not $\mathbb{R}$ but $\mathbb{C}$. The only change is that the symmetry condition in Definition 3.3 is replaced by the Hermitian symmetry:

$$\langle x, y \rangle = \overline{\langle y, x \rangle},$$

where the bar denotes complex conjugate. Recall that the complex conjugate of a number $z = a + ib$, with $a, b \in \mathbb{R}$, is $\overline{z} = a - ib$. The main property of the complex conjugate we will use is that if $z = x + iy$, so that $\overline{z} = x - iy$, then

$$z\overline{z} = (x + iy)(x - iy) = x^2 + y^2 \geq 0. \quad (3.10)$$

If $z = x + iy$, we define $|z| = \sqrt{x^2 + y^2}$, and (3.10) says that $|z|^2 = z\overline{z}$.

The basic example of a space with a complex inner product is $\mathbb{C}^n$, with the inner product

$$x \cdot y = \sum_{j=1}^{n} x_j \overline{y_j}, \ x = (x_1, \ldots, x_n) \in \mathbb{C}^n, \ y = (y_1, \ldots, y_n) \in \mathbb{C}^n.$$ 

The complex conjugate is needed to ensure that

$$x \cdot x = \sum_{j=1}^{n} x_j \overline{x_j} = \sum_{j=1}^{n} |x_j|^2 \geq 0,$$

and $(x \cdot x) = 0$ if and only if $x = 0$ (i.e. $x_j = 0$ for all $j$).
Exercise 3.5 Show that if we define the product \( x \star y \) on \( \mathbb{C}^n \) as
\[
x \star y = x_1 y_1 + \cdots + x_n y_n, \quad x = (x_1, \ldots, x_n) \in \mathbb{C}^n, \quad y = (y_1, \ldots, y_n) \in \mathbb{C}^n,
\]
then there exist vectors \( x \in \mathbb{C}^n \) such that \( x \star x = 0 \).

We may also define the inner product on the space \( C([0,1]; \mathbb{R}) \) of real-valued continuous functions on \([0,1]\) as
\[
\langle f, g \rangle = \int_0^1 f(x)g(x) \, dx, \quad (3.11)
\]
and on the space \( C([0,1]; \mathbb{C}) \) of continuous complex valued functions on \([0,1]\) as
\[
\langle f, g \rangle = \int_0^1 f(x)\overline{g(x)} \, dx. \quad (3.12)
\]

Exercise 3.6 Check that \((3.11)\) and \((3.12)\) define the inner product on the corresponding spaces.

Note that when the field is \( \mathbb{R} \), symmetry plus linearity in the first slot give linearity in the second slot as well:
\[
\langle x, \lambda y + \mu z \rangle = \lambda \langle x, y \rangle + \mu \langle x, z \rangle.
\]
If the field is \( \mathbb{C} \), they give conjugate linearity in the second slot:
\[
\langle x, (\lambda y + \mu z) \rangle = \overline{\lambda} \langle x, z \rangle + \overline{\mu} \langle x, z \rangle \quad \text{for all } x, y, z \in V, \ \lambda, \mu \in \mathbb{C}.
\]
This linearity also gives \( \langle 0_V, x \rangle = 0 \) for all \( x \in V \), as follows by writing \( 0_V = 0 \star 0_V \) (with the first 0 in the right side being the real number 0, and the two other zeros are the zero vector in \( V \), which we temporarily denote by \( 0_V \) for clarity, and \( \star \) denoting multiplication of the vector \( 0_V \) by the real number 0):
\[
\langle 0_V, v \rangle = \langle 0 \star 0_V, v \rangle = 0 \langle 0_V, v \rangle = 0, \ \forall v \in V,
\]
and by symmetry then
\[
\langle v, 0_V \rangle = \langle 0_V, v \rangle = 0,
\]
with Hermitian symmetry working similarly in the complex case.

In inner product spaces one defines the norm (which is just a notation rather than a notion at this moment) by
\[
\|x\| = \sqrt{\langle x, x \rangle},
\]
with the square root being the non-negative square root of a non-negative number (the latter being the case by positive definiteness). Note that \( \|x\| = 0 \) if and only if \( x = 0 \). We also note a useful property of the norm
\[
\|cv\|^2 = \langle cv, cv \rangle = c \langle v, cv \rangle = c^2 \langle v, v \rangle = |c|^2 \|v\|^2, \ c \in \mathbb{R}, \ v \in V,
\]
so
\[
\|cv\| = |c| \|v\|. \quad (3.13)
\]
This property of the norm is called absolute homogeneity (of degree 1). The same statement, \((3.13)\), is valid if the field is \( \mathbb{C} \), but in that case the proof is
\[
\|cv\|^2 = \langle cv, cv \rangle = c \langle v, cv \rangle = c \overline{c} \langle v, v \rangle = |c|^2 \|v\|^2, \ c \in \mathbb{C}, \ v \in V,
\]
because \( c\overline{c} = |c|^2 \) for a complex number \( v \in \mathbb{C} \).

One concept that is tremendously useful in inner product spaces is orthogonality:
Definition 3.7 Suppose $V$ is an inner product space. For $v, w \in V$ we say that $v$ is orthogonal to $w$ if $\langle v, w \rangle = 0$.

Exercise 3.8 Check that in $\mathbb{R}^2$ the fact that $v \cdot w = 0$ means exactly that the angle between the vectors $v, w \in \mathbb{R}^2$ is 90 degrees, without using the cosine theorem.

Note that $\langle v, w \rangle = 0$ if and only if $\langle w, v \rangle = 0$, both in real and complex inner product spaces, so $v$ is orthogonal to $w$ if and only if $w$ is orthogonal to $v$ – so we often say simply that $v$ and $w$ are orthogonal.

As an illustration of its use, let’s prove the Pythagorean theorem.

Lemma 3.9 Suppose $V$ is an inner product space, $v, w \in V$ and $v$ and $w$ are orthogonal. Then

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

Proof. Since $v - w = v + (-w)$, the statement about $v - w$ follows from the statement for $v + w$ and $\|v - w\| = \|w\|$. Now, we write

$$\langle v + w, v + w \rangle = \langle v, v + w \rangle + \langle w, v + w \rangle = \langle v, v \rangle + \langle v, w \rangle + \langle w, v \rangle + \langle w, w \rangle = \langle v, v \rangle + \langle w, w \rangle$$

by the orthogonality of $v$ and $w$, proving the result. □

An important use of orthogonality is the following decomposition.

Lemma 3.10 Suppose $v, w \in V$, $w \neq 0$. Then there exist unique $v_\parallel, v_\perp \in V$ such that $v = v_\parallel + v_\perp$, with $v_\parallel = cw$ for some $c \in \mathbb{R}$ (or $\mathbb{C}$ if the field is $\mathbb{C}$) and $\langle v_\perp, w \rangle = 0$; see Figure 3.1.

![Figure 3.1: The decomposition $v = v_\parallel + v_\perp$.](image)

Proof: In order to find a a candidate for $c$, let us assume that we have found a decomposition $v = v_\parallel + v_\perp$, with $v_\parallel = cw$ and $\langle v_\perp, w \rangle = 0$. Then taking the inner product with $w$ of both sides, we deduce

$$\langle v, w \rangle = \langle v_\parallel, w \rangle + \langle v_\perp, w \rangle = \langle cw, w \rangle + \langle v_\perp, w \rangle = c\|w\|^2.$$

This, as $w \neq 0$, gives

$$c = \frac{\langle v, w \rangle}{\|w\|^2}.$$

Thus, $c$ (if it exists) is uniquely defined by $v$ and $w$, and $v_\parallel = cw$ and $v_\perp = v - cw$, giving uniqueness of such orthogonal decomposition (if it exists).
To show that the orthogonal decomposition does exist, we take
\[ c = \frac{\langle v, w \rangle}{\|w\|^2}, \quad v_{\parallel} = cw, \quad v_{\perp} = v - cw, \]
then both \( v_{\perp} + v_{\parallel} = v \) and \( v_{\parallel} = cw \) are satisfied automatically, so we merely need to check that \( \langle v_{\perp}, w \rangle = 0 \). But we have
\[ \langle v_{\perp}, w \rangle = \langle v, w \rangle - c \langle w, w \rangle = \langle v, w \rangle - \frac{\langle v, w \rangle}{\|w\|^2} \|w\|^2 = 0, \]
so the desired vectors \( v_{\perp} \) and \( v_{\parallel} \) indeed exist. \( \square \)

One calls \( v_{\parallel} \) the orthogonal projection of \( v \) onto the span of \( w \). We will soon see the following general version of this statement.

**Lemma 3.11** Let \( V \) be an inner product vector space, and \( W \) a subspace of \( V \) spanned by a finite collection of vectors \( \{w_1, \ldots, w_N\} \). Then for any \( v \in V \) there exists a unique pair of vectors \( v_{\parallel} \) and \( v_{\perp} \) so that, \( v = v_{\parallel} + v_{\perp} \), \( v_{\parallel} \in W \) and the vector \( v_{\perp} \) has the property that \( \langle w, v_{\perp} \rangle = 0 \) for any \( w \in W \).

We will not prove this lemma now but later when it will be actually useful.

It will be often very useful to be able to estimate the inner product using the norm. This is achieved by the Cauchy-Schwarz inequality, a generalization of the Cauchy-Schwarz inequality we have seen for the dot product in \( \mathbb{R}^n \).

**Lemma 3.12 (Cauchy-Schwarz)** In an inner product space \( V \),
\[ |\langle v, w \rangle| \leq \|v\| \|w\|, \quad v, w \in V. \] (3.14)

For real-valued function spaces, the Cauchy-Schwarz inequality says explicitly that
\[ \left| \int_0^1 f(x)g(x) \, dx \right| \leq \left( \int_0^1 f(x)^2 \, dx \right)^{1/2} \left( \int_0^1 g(x)^2 \, dx \right)^{1/2}. \] (3.15)
A direct proof of the functional inequality (3.15), without using the Cauchy-Schwartz inequality, is actually not so trivial!

**Exercise 3.13** Verify that the proof of the Cauchy-Schwartz inequality we presented in \( \mathbb{R}^n \) works verbatim in any vector space. We present below an alternative, more geometric proof that uses the orthogonal decomposition.

**Proof.** If \( w = 0 \), then both sides of (3.14) vanish, so we may assume \( w \neq 0 \). Write \( v = v_{\parallel} + v_{\perp} \) as in Lemma 3.10, so that
\[ v_{\parallel} = cw, \quad c = \frac{\langle v, w \rangle}{\|w\|^2}, \quad \langle v_{\parallel}, v_{\perp} \rangle = 0. \]
The last condition above allows us to use the Pythagorean theorem (Lemma 3.9), which gives
\[ \|v\|^2 = \|v_{\parallel}\|^2 + \|v_{\perp}\|^2 \geq \|v_{\parallel}\|^2 = |c|^2 \|w\|^2 = \frac{\|v, w\|^2}{\|w\|^4} \|w\|^2 = \frac{\|v, w\|^2}{\|w\|^2}. \]
Multiplying through by \( \|w\|^2 \) and taking the non-negative square root completes the proof of the lemma. \( \square \)

A useful consequence of the Cauchy-Schwarz inequality is the triangle inequality for the norm:
Lemma 3.14 In an inner product space $V$, we have
\[ \|v + w\| \leq \|v\| + \|w\|, \text{ for all } v, w \in V. \]

Exercise 3.15 Verify that the proof of the triangle inequality in $\mathbb{R}^n$ applies verbatim.

In general, regardless of the inner product structure, one defines the notion of a norm on a vector space $V$ as follows.

Definition 3.16 Suppose $V$ is a vector space. A norm on $V$ is a map $\| \cdot \| : V \to \mathbb{R}$ such that

1. (Positive definiteness) $\|v\| \geq 0$ for all $v \in V$, and $v = 0$ if and only if $\|v\| = 0$.

2. (Absolute homogeneity) $\|cv\| = |c| \|v\|$, $v \in V$, and $c$ a scalar (so $c \in \mathbb{R}$ or $c \in \mathbb{C}$, depending on whether $V$ is real or complex).

3. (Triangle inequality) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

Thus, Lemma 3.14 shows that the map $\| \cdot \| : V \to \mathbb{R}$ we defined on an inner product space is indeed a norm in this sense, so our use of the word norm was justified. Note that (2) implies immediately that $\|0 \in V\| = 0$, where $0 \in V$ in the left is the zero in $V$, and 0 in the right is the number zero. Unless otherwise specified, when we talk of a norm on an inner product space $V$, we always mean $\|v\| = \langle v, v \rangle^{1/2}$.

Here are some other examples of a norm on $\mathbb{R}^n$, other than the norm coming from the standard dot product. First, for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, set
\[ \|(x_1, \ldots, x_n)\|_1 = \sum_{j=1}^{n} |x_j|. \]

The triangle inequality for this norm follows from that on $\mathbb{R}$, the other two properties (positivity and absolute homogeneity) are immediate. This norm is known as the $\ell_1$-norm on $\mathbb{R}^n$.

The second example is the norm, also on $\mathbb{R}^n$:
\[ \|(x_1, \ldots, x_n)\|_\infty = \max\{|x_j| : j = 1, \ldots, n\}. \]

This norm is known as the $\ell_\infty$-norm on $\mathbb{R}^n$. The triangle inequality comes from the fact that for each $j$, we have
\[ |x_j + y_j| \leq |x_j| + |y_j|, \]
so this is also true for any $j$ maximizing $|x_j + y_j|$ (there may be several such $j$). Using that $j$, we can write
\[ \|x + y\|_\infty = |x_j + y_j| \leq |x_j| + |y_j| \leq \|x\|_\infty + \|y\|_\infty. \]

As a side remark, more generally, given any $p \geq 1$, we can define the following norm on $\mathbb{R}^n$:
\[ \|(x_1, \ldots, x_n)\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}. \tag{3.16} \]

Positive and positive homogeneity are easy to check for this norm, known as the $\ell_p$-norm on $\mathbb{R}^n$. The triangle inequality is slightly trickier to prove, and we will do it later. The norm coming from the standard dot product in $\mathbb{R}^n$ is known as the $\ell_2$-norm.

Note also that if $V$ is an inner product space then the inner product of a pair of vectors can be expressed in terms of the corresponding norm as
\[ \langle x, y \rangle = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2). \tag{3.17} \]
However, not all norms on vector spaces come from an inner product. One way to see if the norm comes from the inner product or not, is to verify if (3.17) defines an inner product, that is, if the operation defined by (3.17) satisfies properties (1)-(3) in Definition 3.3. Note that positivity and symmetry are automatic: if $\| \cdot \|$ is a norm satisfying Definition 3.16 then properties (1) and (2) in this definition immediately imply positivity and symmetry of the "inner product" defined by (3.17) (the quotation marks refer to the fact that we do not know for a given norm if (3.17) defines an inner product or not). However, the linearity in the first slot, property (2) in Definition 3.3 does not hold for all norms. For example, for the $\ell_\infty$ norm, we can take $v = (1, 1), w = (0, 1)$

and observe that, with the definition

$$
\langle x, y \rangle_\infty = \frac{1}{4}(\|x + y\|_\infty^2 - \|x - y\|_\infty^2),
$$

we would have

$$
\langle v, w \rangle_\infty = \frac{1}{4}(\|(1, 2)\|_\infty^2 - \|(1, 0)\|_\infty^2) = \frac{1}{4}(4 - 1) = \frac{3}{4}
$$

but note that

$$
-\frac{3}{4}v + w = \left(-\frac{3}{4}, \frac{1}{4}\right), \quad -\frac{3}{4}v - w = \left(-\frac{3}{4}, -\frac{7}{4}\right)
$$

hence

$$
\langle \left(-\frac{3}{4}\right)v, w \rangle_\infty = \frac{1}{4}(\left\|\left(-\frac{3}{4}, \frac{1}{4}\right)\right\|_\infty^2 - \left\|\left(-\frac{3}{4}, -\frac{7}{4}\right)\right\|_\infty^2) = \frac{1}{4}\left(\frac{9}{16} - \frac{49}{16}\right) = -\frac{5}{8} \neq -\frac{3}{4} \cdot \frac{3}{4} = -\frac{3}{4} \langle v, w \rangle_\infty.
$$

Thus, the $\ell_\infty$-norm does not come from an inner product on $\mathbb{R}^n$.

**Exercise 3.17** Show that the $\ell_1$-norm also does not come from an inner product on $\mathbb{R}^n$.

## 4 Gaussian elimination and the linear dependence lemma

Our goal in this section is to prove the following lemma.

**Lemma 4.1** (The linear dependence lemma) Let $v_1, \ldots, v_k$ be a collection of $k$ vectors in a vector space $V$. Then any collection $w_1, w_2, \ldots, w_{k+1}$ of $k + 1$ vectors in the span of $v_1, \ldots, v_k$ is linearly dependent.

We will assume that $V$ is a vector space over $\mathbb{R}$ but the proof is verbatim the same for a vector space over any field $F$. It may be helpful for the reader to focus on the situation when $V = \mathbb{R}^n$ as this does not change anything in the arguments or statements in this section.

In order to show that the vectors $w_1, w_2, \ldots, w_{k+1}$ are linearly dependent, we need to find real numbers $x_1, x_2, \ldots, x_{k+1}$ such that not all of $x_j$ equal to zero, and

$$
x_1w_1 + x_2w_2 + \cdots + x_{k+1}w_{k+1} = 0.
$$

Note that this is a vector identity: the left side is a vector in $V$, and the $0$ in the right side is the zero vector in $V$. Let us translate (4.1) into a system of linear equations for the unknowns $x_1, \ldots, x_{k+1}$. Since each vector $w_j$ is in the span of $v_1, \ldots, v_k$, it is a linear combination of the vectors $v_1, \ldots, v_k$. Thus, there exist real numbers $a_{jm}, j = 1, \ldots, k, m = 1, \ldots, k + 1$, so that

$$
w_m = a_{1m}v_1 + a_{2m}v_2 + \cdots + a_{km}v_k.
$$
Substituting these expressions into (4.1) gives

\[ x_1(a_{11}v_1 + a_{21}v_2 + \cdots + a_{k1}v_k) + x_2(a_{12}v_1 + a_{22}v_2 + \cdots + a_{k2}v_k) + \cdots + x_{k+1}(a_{1,k+1}v_1 + a_{2,k+1,2}v_2 + \cdots + a_{k,k+1}v_k) = 0, \]

which can be re-written as

\[ (a_{11}x_1 + a_{21}x_2 + \cdots + a_{1,k+1}x_{k+1})v_1 + (a_{21}x_1 + a_{22}x_2 + \cdots + a_{2,k+1}x_{k+1})v_2 + \cdots + (a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{k,k+1}x_{k+1})v_k = 0. \]

Equality (4.4) holds, clearly, if we find real numbers \( x_1, \ldots, x_{k+1} \) such that

\[ a_{11}x_1 + a_{12}x_2 + \cdots + a_{1,k+1}x_{k+1} = 0 \]
\[ a_{22}x_1 + a_{22}x_2 + \cdots + a_{2,k+1}x_{k+1} = 0 \]
\[ \vdots \]
\[ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{k,k+1}x_{k+1} = 0. \]

Note that while (4.4) is an equality in \( V \) – the left side is a vector in \( V \) and the right side is the zero vector in \( V \), the system (4.5) involves only real numbers. This is a homogeneous system of \( k \) equations for \( (k+1) \) unknowns – recall that the coefficients \( a_{mj} \) are given. Our task is to show that it always has a solution, such that not all of \( x_j \) equal to zero, no matter what the numbers \( a_{mj} \) are. More generally, we will consider systems of \( m \) equations for \( n \) unknown real numbers \( x_1, \ldots, x_n \):

\[ a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \]
\[ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \]
\[ \vdots \]
\[ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0, \]

with \( m < n \) and show that it always has a non-zero solution. Such systems are known as homogeneous because the right side is zero. This means that if \( (x_1, \ldots, x_n) \) is a solution of (4.6) then for any \( \lambda \in \mathbb{R} \), the numbers \( x_1' = \lambda x_1, x_2' = \lambda x_2, \ldots, x_n' = \lambda x_n \) also provide a solution, hence the name "homogeneous".

**Lemma 4.2** A homogeneous system (4.6) of \( m \) linear equations for \( n \) unknown real numbers \( x_1, \ldots, x_n \), with \( m < n \), always has a solution such that not all of \( x_j \) equal to zero.

As we have explained, Lemma 4.1 follows from Lemma 4.2, so our goal is to prove the latter.

### 4.1 Gaussian elimination

We will prove Lemma 4.2 using what is known as the Gaussian elimination procedure. The first observation is that the set of solutions of the system (4.6) does not change under the following operations:

1. Interchange any pair of equations.
2. Multiply one of the equations by a non-zero constant.
3. Add a multiple of one equation to another. For example, take equation \( j \) and add it to equation \( i \) with \( i \neq j \).
Gaussian elimination is the following process of finding a solution to (4.6). If all coefficients in the first column vanish: \( a_{j1} = 0 \) for all \( 1 \leq j \leq m \), so that the system has the form

\[
\begin{align*}
0x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\
0x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\
\vdots & & \vdots \\
0x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0,
\end{align*}
\]

(4.7)

then a solution to (4.7) is simply \( x_1 = 1 \) and \( x_2 = \cdots = x_n = 0 \) – hence, we found a solution. If not all \( a_{j1} = 0 \), then by interchanging the rows we may assume without loss of generality that \( a_{11} \neq 0 \).

We divide the first equation by \( a_{11} \) – this does not affect all other equations, and re-write (4.6) as

\[
\begin{align*}
x_1 + \tilde{a}_{12}x_2 + \cdots + \tilde{a}_{1n}x_n &= 0 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\
\vdots & & \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0,
\end{align*}
\]

(4.8)

where \( \tilde{a}_{1j} = a_{1j}/a_{11} \). Next, we subtract from each equation \( j \geq 2 \) the first equation multiplied by \( a_{j1} \). Then the system will have the form

\[
\begin{align*}
x_1 + \tilde{a}_{12}x_2 + \cdots + \tilde{a}_{1n}x_n &= 0 \\
0x_1 + \tilde{a}_{22}x_2 + \cdots + \tilde{a}_{2n}x_n &= 0 \\
\vdots & & \vdots \\
0x_1 + \tilde{a}_{m2}x_2 + \cdots + \tilde{a}_{mn}x_n &= 0,
\end{align*}
\]

(4.9)

with the coefficients in the first row unchanged from the previous step: \( \tilde{a}_{1j} = \tilde{a}_{1j} \), and the coefficients in the other rows changed, so that, for instance, \( \tilde{a}_{22} = a_{22} - a_{21}\tilde{a}_{12} \), and generally,

\[
\tilde{a}_{jk} = a_{jk} - a_{j1}\tilde{a}_{1k}.
\]

Again, if all \( \tilde{a}_{j2} = 0 \), we can take \( x_2 = 1 \) and all other \( x_j = 0 \), otherwise, we can find \( j \geq 2 \) so that \( \tilde{a}_{j2} \neq 0 \), and move that equation into the second row. In other words, we may assume without loss of generality that \( \tilde{a}_{22} \neq 0 \). We divide now the second row by \( \tilde{a}_{22} \):

\[
\begin{align*}
x_1 + \tilde{a}_{12}x_2 + \cdots + \tilde{a}_{1n}x_n &= 0 \\
0x_1 + x_2 + \cdots + \tilde{a}_{2n}x_n &= 0 \\
\vdots & & \vdots \\
0x_1 + \tilde{a}_{m2}x_2 + \cdots + \tilde{a}_{mn}x_n &= 0.
\end{align*}
\]

(4.10)

We can now eliminate \( x_2 \) from all but the first two equations, in the same way as we eliminated \( x_1 \) from all but the first equation, and then continue this process with \( x_3 \), and so on, known as the Gaussian elimination.

4.2 The proof of Lemma 4.2

The proof of Lemma 4.2 is by induction, and does not use the full Gaussian elimination. Instead, we observe that when \( m = 1 \), the system (4.6) consists of just one equation:

\[
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0,
\]

(4.11)
for $n > 1$ unknowns. If all $a_{1j} = 0$ then we simply set $x_1 = 1$ and all other $x_j = 0$. Otherwise, if, say, $a_{11} \neq 0$, then we take any $x_2, \ldots, x_n$ and set
\[ x_1 = \frac{a_{12}}{a_{11}}x_2 - \cdots - \frac{a_{1n}}{a_{11}}x_n. \]
Thus, the conclusion of the lemma holds for $m = 1$ and all $n > 1$. Next, assume that it holds for some $m$ and all $n > m$. Then, given a system of $m+1$ equations with $n > m+1$ unknowns, we apply the Gaussian elimination to obtain a system of $m$ equations for $n-1 > m$ unknowns $x_2, \ldots, x_n$. It has a solution by the induction assumption, and $x_1$ can then be found from the first equation in the system obtained after one step of the Gaussian elimination, and we are done. □

As we have discussed, Lemma 4.1 is a consequence of Lemma 4.2, thus Lemma 4.1 is proved as well.

5 The basis theorem

We now turn to the notion of a basis of a vector space. Let us first give a general definition of a basis of a vector space.

**Definition 5.1** A collection of vectors $\{v_1, v_2, \ldots, v_N\}$ is a basis for a vector space $V$ if $\{v_1, v_2, \ldots, v_N\}$ are linearly independent and $V = \text{span}\{v_1, v_2, \ldots, v_N\}$.

A canonical example of a basis is the collection of vectors
\[ e_1 = (1,0,\ldots,0), \quad e_2 = (0,1,0,\ldots,0), \ldots, e_n = (0,0,\ldots,0,1) \]
(5.1) in $\mathbb{R}^n$. Note that for real numbers $x_1, x_2, \ldots, x_n \in \mathbb{R}$, we have
\[ x_1 e_1 + x_2 e_2 + \cdots + x_n e_n = (x_1, x_2, \ldots, x_n). \]
(5.2)
This identity implies that $e_k$, $k = 1, \ldots, n$ are linearly independent, for if we have
\[ x_1 e_1 + \cdots + x_n e_n = 0, \]
for some real numbers $x_1, x_2, \ldots, x_n \in \mathbb{R}$, then (5.2) shows that all $x_k = 0$. Moreover, $e_1, \ldots, e_n$, form a basis of $\mathbb{R}^n$ because any vector $(c_1, \ldots, c_n) \in \mathbb{R}^n$ can be written as
\[ c = c_1 e_1 + c_2 e_2 + \cdots + c_n e_n, \]
hence $\mathbb{R}^n$ is spanned by $e_1, \ldots, e_n$. In particular, $\mathbb{R}^n$ is the span of the vectors $e_1, \ldots, e_n$ and thus no collection of more than $n$ vectors in $\mathbb{R}^n$ can be linearly independent, according to Lemma 4.1.

**Exercise 5.2** Another example of a basis in a vector space is the following. Let $P_m$ be the vector space of all polynomials of degree less or equal to $m$. Show that the polynomials
\[ p_0(x) = 1, \quad p_1(x) = x, \ldots, p_m(x) = x^m \]
form a basis of $P_m$. 

18
5.1 The basis theorem in $\mathbb{R}^n$

The first main result of this section is the following.

**Theorem 5.3** Any non-trivial subspace of $\mathbb{R}^n$ has a basis.

**Proof.** Let $V$ be a subspace of $\mathbb{R}^n$ and let $q$ be the largest integer such that there exist $q$ linearly independent vectors $v_1, \ldots, v_q$ in $V$. Note that $q \leq n$ – as we have noted above, Lemma 4.1 implies that no collection of more than $n$ vectors can be linearly independent in $\mathbb{R}^n$. We claim that $V$ is the span of $\{v_1, \ldots, v_q\}$. Indeed, as all $v_j$ belong to $V$, we know that $\text{span}\{v_1, \ldots, v_q\}$ is a sub-space of $V$. Let us assume that there is a vector $w \in V$ such that $w$ is not in $\text{span}\{v_1, \ldots, v_q\}$. It follows from the way $q$ was chosen, that the collection $\{v_1, \ldots, v_q, w\}$ is linearly dependent. Hence, there exist real numbers $a_1, \ldots, a_q$ and $b$, such that they are not all equal to zero, and

$$a_1 v_1 + \cdots + a_q v_q + bw = 0.$$ 

If $b = 0$ then not all of $a_j$ are zero, and

$$a_1 v_1 + \cdots + a_q v_q = 0,$$

which is a contradiction to the linear independence of $\{v_1, \ldots, v_q\}$. Hence, $b \neq 0$, but then we can write

$$w = -\frac{a_1}{b} v_1 - \cdots - \frac{a_q}{b} v_q,$$

showing that $w$ is in $\text{span}\{v_1, \ldots, v_q\}$, which is a contradiction. Thus, we have $V = \text{span}\{v_1, \ldots, v_q\}$, and, as these vectors are linearly independent, they form a basis for $V$. □

The next result shows that any linearly independent set of vectors in a sub-space can be completed to form a basis of the sub-space.

**Theorem 5.4** Let $V$ be a sub-space of $\mathbb{R}^n$ and $u_1, \ldots, u_k$ be a collection of linearly independent vectors in $V$. Then it can be extended to form a basis, that is, there exists a basis $v_1, \ldots, v_q$ of $V$ with $q \geq k$ such that $v_1 = u_1, v_2 = u_2, \ldots, v_k = u_k$.

**Proof.** Similar to the proof of Theorem 5.3, let $q$ be the largest integer so that there exists a collection of $q$ linearly independent vectors in $V$ such that $v_1 = u_1, v_2 = u_2, \ldots, v_k = u_k$. Arguing exactly as in the previous proof, we can show that $V$ is the span of $\{v_1, \ldots, v_q\}$. As these vectors are linearly independent they form a basis for $V$, which, by construction, satisfies the requirement of the present theorem. □

**Theorem 5.5** Let $V$ be a sub-space of $\mathbb{R}^n$. Then every basis of $V$ has the same number of vectors.

**Proof.** Assume that $V$ has two bases, $\{v_1, \ldots, v_q\}$ and $\{w_1, \ldots, w_k\}$ with $q > k$. As $\{w_1, \ldots, w_k\}$ is a basis for $V$, we know that $V = \text{span}\{w_1, \ldots, w_k\}$. As $q > k$, Lemma 4.1 implies that the vectors $\{v_1, \ldots, v_q\}$ are linearly dependent, which is a contradiction. □

5.2 Finite-dimensional vector spaces

**Definition 5.6** A vector space $V$ is called finite dimensional if there is a finite subset $\{v_1, \ldots, v_n\}$ of elements of $V$ such that $\text{Span}\{v_1, \ldots, v_n\} = V$. We say that $V$ is $n$-dimensional if it has a basis that contains exactly $n$ vectors.

Not every vector space is finite-dimensional.
Exercise 5.7 Show that the vector space $P$ of all polynomials is not finite-dimensional.

Exactly as in $\mathbb{R}^n$, we have the following versions of Theorems 5.3 and 5.4, with essentially identical proofs to those of these two theorems.

**Theorem 5.8** Any non-trivial subspace of a finite-dimensional vector space $V$ has a basis.

**Theorem 5.9** Let $V$ be a subspace of a finite-dimensional space $W$ and $u_1, \ldots, u_k$ be a collection of linearly independent vectors in $V$. Then there exists a basis $v_1, \ldots, v_q$ of $V$ with $q \geq k$ such that $v_1 = u_1, v_2 = u_2, \ldots, v_k = u_k$.

**Exercise 5.10** Explain how the proofs of Theorems 5.3 and 5.4 rely on the fact that $\mathbb{R}^n$ is a finite-dimensional vector space.

**Theorem 5.11** Let $V$ be an $n$-dimensional vector space. Then (i) any collection of $n$ linearly independent vectors $\{v_1, \ldots, v_n\}$ is a basis for $V$, and (ii) any collection of $n$ vectors that span $V$ must be linearly independent.

**Proof.** For (i), note that since $V$ is $n$-dimensional, it is a span of $n$ basis vectors. Lemma 4.1 implies that given any $w \in V$, the collection $\{v_1, \ldots, v_n, w\}$ is linearly dependent. As $\{v_1, \ldots, v_n\}$ are linearly independent, as in the proof of Theorem 5.3, it follows that $w$ can be written as

$$w = a_1 v_1 + \cdots + a_n v_n.$$  

As this is true for any vector $w \in V$, we conclude that $\{v_1, \ldots, v_n\}$ span $V$, and, as they are linearly independent, they form a basis for $V$. As for part (ii), assume that $\{v_1, \ldots, v_n\}$ span $V$ but are linearly dependent. It follows then from Lemma 2.10 that there exists some $j \neq n$ and real numbers $c_1, \ldots, c_{j-1}, c_{j+1}, \ldots, c_n$, such that

$$v_j = c_1 v_1 + \cdots + c_{j-1} v_{k-1} + c_{k+1} v_{k+1} + \cdots + c_n v_n. \quad (5.3)$$

It follows that $V$ is the span of $(n - 1)$ vectors $\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n\}$ and thus $V$ can not be $n$-dimensional by Lemma 4.1. $\square$