# Linear Algebra II (MATH 225) – v 1.23

**Paul Buckingham**

### About these notes

These notes provide the core material for a second course in linear algebra taught at the University of Alberta. Students in this course should use the notes as follows:

- Read the sections according to the schedule posted on eClass.
- Come to class having read the appropriate sections. In class, the instructor will explain parts of the notes further and do some more examples.
- Refer to the notes when reviewing for exams, remembering as well that the examples done in class are crucial to your understanding of the course.

### Proofs

With one exception, I have included proofs of all the results stated, although many of the proofs appear in the Appendix rather than the main text. The exception is the existence of a basis in a vector space that is not finitely generated; I prove the existence of a basis only for finitely generated vector spaces. Occasionally, I prove a result after I have first given some examples of the result's use.

### Correction of typos

There may be some typos in these notes. As the course unfolds, I will replace this document from time to time with updated versions correcting typos. Please bear this in mind if you intend to annotate the notes directly, for there will be no easy way to transfer your annotations from one version of the notes to the next.

While I will correct typos and may make other very small changes, I will not make any substantive changes to the notes during the course, so the content will be essentially stable.

### Notation

When a column vector would take up too much space on the page, I write it as a row with commas. For example,

$$\begin{pmatrix} 3 \\ 5 \\ -1 \\ 2 \end{pmatrix}$$

may be written $(3, 5, -1, 2)$. This row with commas should not be confused with the $1 \times 4$ matrix $\begin{pmatrix} 3 & 5 & -1 & 2 \end{pmatrix}$, which has no commas.

I will use the symbol $\leftrightarrow$ for row equivalence of matrices. That is, for two $m \times n$ matrices $A$ and $B$, $A \leftrightarrow B$ signifies that $A$ and $B$ are row equivalent.

# Contents

# (I) Vector Spaces

## I − 1 What is a vector space?

In MATH 125, we learned about the space $\mathbb{R}^n$. It has two key properties: We can add vectors in $\mathbb{R}^n$, and we can scale vectors in $\mathbb{R}^n$. Roughly speaking, a vector space is a space in which we can perform both addition and scalar multiplication.

Here is the formal definition: A vector space (over $\mathbb{R}$) is a non-empty set $V$ together with operations

$$\begin{array}{rcll} V \times V & \to & V & \\ (\mathbf{u}, \mathbf{v}) & \mapsto & \mathbf{u} + \mathbf{v} & \text{(addition)} \\ \\ \mathbb{R} \times V & \to & V & \\ (a, \mathbf{v}) & \mapsto & a\mathbf{v} & \text{(scalar multiplication)} \end{array}$$

satisfying the following properties:

(i) For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$. (Associativity)

(ii) For all $\mathbf{u}, \mathbf{v} \in V$, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. (Commutativity)

(iii) There is $\mathbf{0} \in V$ such that, for all $\mathbf{u} \in V$, $\mathbf{u} + \mathbf{0} = \mathbf{u}$. (Zero element)

(iv) For all $\mathbf{u} \in V$, there is $\mathbf{v} \in V$ such that $\mathbf{u} + \mathbf{v} = \mathbf{0}$. (Additive inverse)

(v) For all $\mathbf{u}, \mathbf{v} \in V$ and all $c \in \mathbb{R}$, $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$. (Distributivity)

(vi) For all $\mathbf{u} \in V$ and all $c, d \in \mathbb{R}$, $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$. (Distributivity)

(vii) For all $\mathbf{u} \in V$ and all $c, d \in \mathbb{R}$, $(cd)\mathbf{u} = c(d\mathbf{u})$

(viii) For all $\mathbf{u} \in V$, $1\mathbf{u} = \mathbf{u}$.

**Proposition 1.1.** *The vector* $\mathbf{0}$ *satisfying property (iii) is unique. We call it the* zero vector *of* $V$*.*

*Proof.* Suppose $\mathbf{0}$ and $\mathbf{0}'$ both satisfy property (iii). Then

$$\begin{array}{rcll} \mathbf{0}' & = & \mathbf{0}' + \mathbf{0} & \text{by (iii) applied to } \mathbf{0} \text{ with } \mathbf{u} = \mathbf{0}' \\ & = & \mathbf{0} + \mathbf{0}' & \text{by (ii)} \\ & = & \mathbf{0} & \text{by (iii) applied to } \mathbf{0}' \text{ with } \mathbf{u} = \mathbf{0}. \end{array}$$

□

**Proposition 1.2.** *If* $\mathbf{u} \in V$*, then the element* $\mathbf{v} \in V$ *satisfying* $\mathbf{u} + \mathbf{v} = \mathbf{0}$*, as in property (iv), is unique.*

*Proof.* Let $\mathbf{u} \in V$, and suppose $\mathbf{v}, \mathbf{v}' \in V$ satisfy (iv), that is, $\mathbf{u}+\mathbf{v} = \mathbf{0}$ and $\mathbf{u}+\mathbf{v}' = \mathbf{0}$. Then

$$\begin{aligned}
\mathbf{v}' &= \mathbf{v}' + \mathbf{0} \quad \text{by (iii)} \\
&= \mathbf{v}' + (\mathbf{u} + \mathbf{v}) \quad \text{because } \mathbf{0} = \mathbf{u} + \mathbf{v} \\
&= (\mathbf{v}' + \mathbf{u}) + \mathbf{v} \quad \text{by (i)} \\
&= (\mathbf{u} + \mathbf{v}') + \mathbf{v} \quad \text{by (ii)} \\
&= \mathbf{0} + \mathbf{v} \quad \text{because } \mathbf{u} + \mathbf{v}' = \mathbf{0} \\
&= \mathbf{v} + \mathbf{0} \quad \text{by (ii)} \\
&= \mathbf{v} \quad \text{by (iii).}
\end{aligned}$$

□

The unique element $\mathbf{v}$ such that $\mathbf{u}+\mathbf{v} = \mathbf{0}$ is called the *additive inverse* of $\mathbf{u}$ and is written $-\mathbf{u}$. If $\mathbf{u}, \mathbf{w} \in V$, then $\mathbf{w} + (-\mathbf{u})$ is abbreviated to $\mathbf{w} - \mathbf{u}$.

### Linear combinations

If $V$ is a vector space and $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$, then a *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is an expression of the form $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k$ with $a_1, \ldots, a_k \in \mathbb{R}$.

### Vectors

In this course, the word *vector* simply means an element of a vector space. Therefore, if $V$ is a vector space, the statements "$\mathbf{u}$ is a vector in $V$" and "$\mathbf{u}$ is an element of $V$" mean the same thing.

## I – 2 Examples of vector spaces

**Example.** MATH 125 was devoted to the study of the vector space $V = \mathbb{R}^n$, together with the usual addition and scalar multiplication given by

$$(u_1, \ldots, u_n) + (v_1, \ldots, v_n) = (u_1 + v_1, \ldots, u_n + v_n)$$

and

$$c(u_1, \ldots, u_n) = (cu_1, \ldots, cu_n).$$

**Example.** The vector space $\mathcal{P}$ consisting of all real-coefficient polynomials[1]

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

of any degree. For example, all of the following are elements of $\mathcal{P}$:

$$\begin{aligned} &2x^2 - 1 \\ &-x^3 + 4x^2 + x - 10 \\ &x \\ &5 \\ &0 \end{aligned}$$

We add and scale polynomials in the usual way.

**Example.** The vector space $\mathcal{P}_n$ consisting of all real-coefficient polynomials

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

of degree less than or equal to $n$. For example, all of the following are elements of $\mathcal{P}_3$:

$$\begin{aligned} &7x^3 + 1 \\ &x^3 + x^2 + x - 4 \\ &x^2 + 2x \\ &8 \\ &0 \end{aligned}$$

By contrast, the polynomial $x^4 + 7x^2$ is not in $\mathcal{P}_3$, although it is in $\mathcal{P}_4$.

**Example.** The space $M_{m,n}(\mathbb{R})$ of $m \times n$ matrices with real entries. Some examples of elements of $M_{2,3}(\mathbb{R})$ are

$$\begin{pmatrix} \sqrt{2} & 0 & -\pi \\ 1 & 2 & 1 - \sqrt[3]{5} \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We add and scale $m \times n$ matrices in the usual way, as in MATH 125. We abbreviate $M_{n,n}(\mathbb{R})$ to just $M_n(\mathbb{R})$. Thus, $M_4(\mathbb{R})$ is the space of $4 \times 4$ real matrices.

[1] For a discussion of what we mean by a polynomial in this course, see Section 15 in the Appendix. The discussion is technical and may safely be omitted.

**Example.** The space $\mathcal{F}$ consisting of all functions $f : \mathbb{R} \to \mathbb{R}$. Some examples of functions from $\mathbb{R}$ to $\mathbb{R}$ are the following:

$$\begin{aligned} f_1 : x &\mapsto e^x \\ f_2 : x &\mapsto \sin(x) + \cos(x) \\ f_3 : x &\mapsto |x - 2| \end{aligned}$$

Remember, every vector space has an operation of addition and an operation of scalar multiplication. What are these operations for the vector space $\mathcal{F}$? In other words, how do we add two functions together, and how do we scale a function?

One intuitive description of addition, in terms of graphs, is as follows: If $f, g \in \mathcal{F}$, then imagine drawing the graphs of $f$ and $g$. At each point $x$ along the horizontal axis, take the $y$-values of the two graphs and add them together. Then plot this new value above $x$. All these new values, once plotted, form the graph of $f + g$. More formally, the function $f + g$ is defined by

$$(f + g)(x) = f(x) + g(x).$$

Similarly, if $f \in \mathcal{F}$ and $c \in \mathbb{R}$, then $cf$ is the function whose graph is obtained by scaling the entire graph of $f$ by $c$. More formally,

$$(cf)(x) = c \cdot f(x).$$

**Example.** The space $\mathcal{S}$ of sequences $(a_0, a_1, a_2, \ldots)$ with entries $a_i \in \mathbb{R}$. We add and scale sequences as follows:

$$\begin{aligned} (a_0, a_1, a_2, \ldots) + (b_0, b_1, b_2, \ldots) &= (a_0 + b_0, a_1 + b_1, a_2 + b_2, \ldots) \\ c(a_0, a_1, a_2, \ldots) &= (ca_0, ca_1, ca_2, \ldots) \end{aligned}$$

Another way to write a sequence is $(a_n)_{n \geq 0}$, which we may abbreviate further to $(a_n)_n$, with the understanding that a sequence in this course always begins with the $n = 0$ term. In this notation, addition and scalar multiplication are given like this:

$$(a_n)_n + (b_n)_n = (a_n + b_n)_n \tag{2.1}$$

$$c(a_n)_n = (ca_n)_n \tag{2.2}$$

For example, consider the sequences

$$\begin{aligned} s &= (n^2)_n \\ t &= (2n + 1)_n \end{aligned}$$

That is, $s = (0, 1, 4, 9, 16, \ldots)$ and $t = (1, 3, 5, 7, 9, \ldots)$. Then according to the rule in (2.1),

$$s + t = (n^2)_n + (2n + 1)_n = (n^2 + 2n + 1)_n = ((n + 1)^2)_n.$$

That is, $s + t$ is the sequence whose $n$th term (starting at $n = 0$) is $(n + 1)^2$. We can see this by adding together the first few terms of $(0, 1, 4, 9, 16, \ldots)$ and $(1, 3, 5, 7, 9, \ldots)$:

$$(0, 1, 4, 9, 16, \ldots) + (1, 3, 5, 7, 9, \ldots) = (1, 4, 9, 16, 25, \ldots).$$

## I – 3 Subspaces

A *subspace* of a vector space $V$ is a non-empty subset $U$ of $V$ that is a vector space with respect to the same operations of addition and scalar multiplication. Checking that a subset is indeed a subspace would seem, at first, to require checking all of the properties of a vector space, but in fact there is a much quicker way.

**Proposition 3.1** (Subspace criterion)**.** *Let $U$ be a subset of a vector space $V$. Then $U$ is a subspace of $V$ if and only if all of the following hold:*

*(i) $U$ is non-empty.*

*(ii) For all $\mathbf{u}, \mathbf{v} \in U$, $\mathbf{u} + \mathbf{v} \in U$. (Closure under addition)*

*(iii) For all $\mathbf{u} \in U$ and all $c \in \mathbb{R}$, $c\mathbf{u} \in U$. (Closure under scalar multiplication)*

For a proof, see Section 1 in the Appendix.

**Example.** Decide whether the set

$$U = \{p \in \mathcal{P} \mid p(2) = 0\}$$

is a subspace of $\mathcal{P}$. (Here, $p(2)$ means the value of the polynomial when evaluated at 2.)

*Solution:*

(i) The set $U$ is non-empty because the zero polynomial $p_0 = 0$ satisfies $p_0(2) = 0$ and therefore lies in $U$.

(ii) Suppose $p, q \in U$, which is to say $p(2) = 0$ and $q(2) = 0$. Then $(p + q)(2) = p(2) + q(2) = 0 + 0 = 0$. Therefore, $p + q \in U$. This shows that $U$ is closed under addition.

(iii) Suppose $p \in U$ and $c \in \mathbb{R}$. Then $(cp)(2) = c \cdot p(2) = c \cdot 0 = 0$, so $cp \in U$. Thus, $U$ is closed under scalar multiplication.

We have shown that all conditions in the subspace criterion hold, so $U$ is a subspace of $\mathcal{P}$.

**Example.** Decide whether the set

$$U = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R}) \,\middle|\, a + b = c + d \right\}$$

is a subspace of $M_2(\mathbb{R})$.

*Solution:*

(i) The set $U$ is non-empty because the zero matrix $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ satisfies the required condition: $0 + 0 = 0 + 0$.

(ii) Suppose

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

are both in $U$, which is to say $a + b = c + d$ and $a' + b' = c' + d'$. Then

$$A + A' = \begin{pmatrix} a + a' & b + b' \\ c + c' & d + d' \end{pmatrix},$$

and

$$\begin{aligned} (a + a') + (b + b') &= (a + b) + (a' + b') \quad \text{(rearranging)} \\ &= (c + d) + (c' + d') \quad \text{by assumption} \\ &= (c + c') + (d + d') \quad \text{(rearranging)}, \end{aligned}$$

so $A + A'$ satisfies the required condition. Thus, $A + A' \in U$, so $U$ is closed under addition.

(iii) Suppose $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in U$ and $\lambda \in \mathbb{R}$. Then

$$\lambda A = \begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix},$$

and

$$\begin{aligned} \lambda a + \lambda b &= \lambda(a + b) \quad \text{(rearranging)} \\ &= \lambda(c + d) \quad \text{by assumption} \\ &= \lambda c + \lambda d \quad \text{(rearranging)}, \end{aligned}$$

so $\lambda A$ satisfies the required condition. Thus, $\lambda A \in U$, so $U$ is closed under scalar multiplication.

We have shown that all conditions in the subspace criterion hold, so $U$ is a subspace of $M_2(\mathbb{R})$.

**Example.** Let $U$ be the subset of $\mathcal{S}$ consisting of all sequences $(a_n)_n$ such that $a_n \geq 0$ for all $n \geq 0$. Decide whether $U$ is a subspace of $\mathcal{S}$.

*Solution:* We show that $U$ is in fact not a subspace of $\mathcal{S}$. We do so by showing that it is not closed under scalar multiplication. We need just one counterexample. Let $s = (1, 0, 0, 0, \ldots)$. Then $(-1)s = (-1, 0, 0, 0, \ldots)$, which is not in $U$, because the zeroth entry is negative.

This is already enough to show that $U$ is not a subspace. However, just for practice, decide also whether $U$ is closed under addition. If so, justify your answer with a proof. Otherwise, find a counterexample.

## I – 4 Linear independence

Let $V$ be a vector space, and let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$. We say that $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are *linearly independent* if the only solution in scalars $c_1, \ldots, c_k$ to the equation

$$c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$$

is $c_1 = c_2 = \cdots = c_k = 0$. Otherwise, we say that $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are *linearly dependent.*

**Example.** The vectors

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -3 \\ 4 \\ 11 \end{pmatrix}$$

in $\mathbb{R}^3$ are linearly dependent because $3\mathbf{u}_1 - 2\mathbf{u}_2 - \mathbf{u}_3 = \mathbf{0}$, meaning that the equation $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 = \mathbf{0}$ has the non-trivial solution $c_1 = 3$, $c_2 = -2$, $c_3 = -1$.

By contrast, consider the vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

in $\mathbb{R}^3$. Try solving the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}$. You will find that the only solution is $c_1 = c_2 = c_3 = 0$, so the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent.

Examples such as the previous one were considered already in MATH 125. Let us now consider linear independence in a vector space other than $\mathbb{R}^n$.

**Example.** Let $f, g \in \mathcal{F}$ be the functions defined by $f(x) = \sin(x)$ and $g(x) = \cos(x)$. We will determine whether $f$ and $g$ are linearly independent. To do so, we set up the equation

$$c_1 f + c_2 g = 0, \tag{4.1}$$

and find its solutions $c_1, c_2 \in \mathbb{R}$. If the only solution is $c_1 = c_2 = 0$, then $f$ and $g$ are linearly independent. Otherwise, they are linearly dependent. Note that the 0 on the right-hand side of (4.1) is the zero function, which is the zero vector in the vector space $\mathcal{F}$.

Because (4.1) is an equation of functions, the equation also holds if we evaluate both sides at a real number $x$. Thus,

$$(c_1 f + c_2 g)(x) = 0$$

for all $x \in \mathbb{R}$, which is to say $c_1 f(x) + c_2 g(x) = 0$ for all $x \in \mathbb{R}$. Remembering the definitions of the functions $f$ and $g$, we therefore have

$$c_1 \sin(x) + c_2 \cos(x) = 0 \tag{4.2}$$

for all $x \in \mathbb{R}$. The key point is that (4.2) is true for all $x \in \mathbb{R}$ by assumption, so we may choose $x$ to be whatever we want. Taking $x = 0$ gives $c_1 \sin(0) + c_2 \cos(0) = 0$, i.e., $c_2 = 0$. Hence, $c_1 \sin(x) = 0$ for all $x \in \mathbb{R}$. We may now take $x = \pi/2$ to obtain $c_1 \sin(\pi/2) = 0$, i.e., $c_1 = 0$.

Having shown that the only solution to the equation $c_1 f + c_2 g = 0$ is $c_1 = c_2 = 0$, we conclude that $f$ and $g$ are linearly independent.

It is not very often that a set of functions is linearly dependent, but consider the following functions in $\mathcal{F}$:

$$\begin{aligned} f_1(x) &= \sin^2(x) \\ f_2(x) &= \cos^2(x) \\ f_3(x) &= 1 \quad \text{(the function that is identically 1)} \end{aligned}$$

Can you find scalars $c_1, c_2, c_3 \in \mathbb{R}$, not all zero, such that $c_1 f_1 + c_2 f_2 + c_3 f_3$ is the zero function, i.e., such that

$$c_1 \sin^2(x) + c_2 \cos^2(x) + c_3$$

is identically zero? If so, then you will have shown that $f_1, f_2, f_3$ are linearly dependent.

### Linear independence for infinite sets of vectors

An infinite set $S$ of vectors in a vector space $V$ is said to be linearly independent if every finite subset of $S$ is a linearly independent set in the sense defined above for finite sets.

## I – 5 Spanning

Let $V$ be a vector space, and let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$. Recall from Section 1 that a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is an expression of the form $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$ where $c_1, \ldots, c_k \in \mathbb{R}$. We will also call any vector expressible this way a *linear combination* of $\mathbf{u}_1, \ldots, \mathbf{u}_k$. The *span* of $\mathbf{u}_1, \ldots, \mathbf{u}_k$, denoted $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$, is the set of all linear combinations of $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

**Example.** In $\mathcal{P}$, the span of $1, x, x^2, x^3$ is the set of all linear combinations $c_1 \cdot 1 + c_2 x + c_3 x^2 + c_4 x^3$, which is the set of polynomials of degree less than or equal to 3. Thus,

$$\mathrm{Span}(1, x, x^2, x^3) = \mathcal{P}_3.$$

We may even consider spans of infinite sets. If $S$ is a possibly infinite set of vectors in $V$, then the span of $S$ is the set of all (finite) linear combinations $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$ where $\mathbf{u}_1, \ldots, \mathbf{u}_k \in S$.

**Example.** Consider the vector space $\mathcal{P}$ again, and let

$$\begin{aligned} S &= \{x^{2n} \mid n \text{ is a non-negative integer}\} \\ &= \{1, x^2, x^4, x^6, \ldots\}. \end{aligned}$$

Then a polynomial $p$ is in $\mathrm{Span}(S)$ if and only if there are real numbers $a_0, \ldots, a_k$ (for some $k$) such that

$$p = a_k x^{2k} + a_{k-1} x^{2(k-1)} + \cdots + a_1 x^2 + a_0.$$

That is, $\mathrm{Span}(S)$ consists of all those polynomials in which every power of $x$ is an even power.

### Spanning sets

A *spanning set* for a vector space $V$ is a subset $S$ of $V$ such that $\mathrm{Span}(S) = V$. If $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are vectors in a vector space $V$, then to decide whether these vectors span $V$, consider the equation

$$c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{v},$$

where $c_1, \ldots, c_k \in \mathbb{R}$ and $\mathbf{v} \in V$. If, for every $\mathbf{v} \in V$, this equation has a solution $c_1, \ldots, c_k$, then $\mathbf{u}_1, \ldots, \mathbf{u}_k$ span $V$. Otherwise, they do not.

**Example.** Let

$$\begin{aligned} p_1 &= x + 1 \\ p_2 &= x^2 + 1 \\ p_3 &= x^2 + x, \end{aligned}$$

considered as elements of $\mathcal{P}_2$. To decide whether $p_1, p_2, p_3$ span $\mathcal{P}_2$, we consider the equation

$$c_1p_1 + c_2p_2 + c_3p_3 = q, \tag{5.1}$$

where $q$ is a general element of $\mathcal{P}_2$. Let $q = a_2x^2 + a_1x + a_0$. Then (5.1) says

$$\begin{array}{rrcl} & c_1(x+1) + c_2(x^2+1) + c_3(x^2+x) & = & a_2x^2 + a_1x + a_0, \\ \text{i.e.,} & (c_2+c_3)x^2 + (c_1+c_3)x + (c_1+c_2) & = & a_2x^2 + a_1x + a_0, \end{array}$$

which is to say

$$\begin{aligned} c_2 + c_3 &= a_2 \\ c_1 \qquad + c_3 &= a_1 \\ c_1 + c_2 \qquad &= a_0 \end{aligned}$$

This system of linear equations in the three unknowns $c_1, c_2, c_3$ is represented by the augmented matrix

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & a_2 \\ 1 & 0 & 1 & a_1 \\ 1 & 1 & 0 & a_0 \end{array}\right)$$

A few row reductions show that the left-hand side of the matrix (i.e., to the left of the vertical line) has a pivot (leading entry) in every row of a row-echelon form, so the system has a solution no matter what $a_0, a_1, a_2$ are. Therefore, $p_1, p_2, p_3$ span $\mathcal{P}_2$.

For example, let $q = x^2 + 2x - 3$. Solve the above equations for $c_1, c_2, c_3$, and then verify directly that $c_1p_1 + c_2p_2 + c_3p_3 = q$. There is, of course, nothing special about this $q$; via the method above, we could do the same for any $q \in \mathcal{P}_2$.

**Example.** Let

$$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 2 \\ 3 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 2 & 1 \\ -3 & -2 \end{pmatrix}, \quad A_4 = \begin{pmatrix} -1 & 4 \\ 5 & 1 \end{pmatrix}.$$

The matrices $A_1, A_2, A_3, A_4$ do not span $M_2(\mathbb{R})$. To see this, observe that each of the four matrices has trace 0 (i.e., the sum of the diagonal entries is 0), so the same is true of any linear combination of them. That is,

$$\operatorname{Tr}(c_1A_1 + c_2A_2 + c_3A_3 + c_4A_4) = 0$$

for all $c_1, c_2, c_3, c_4 \in \mathbb{R}$. Therefore, it is impossible for the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, which has trace 1, to be a linear combination of $A_1, A_2, A_3, A_4$.

In the coming sections, we will see other tools for deciding whether a set is a spanning set.

## I – 6 Bases and dimension

If $V$ is a vector space, then a *basis* for $V$ is a linearly independent spanning set. Every vector space has a basis; for a proof in the case of a vector space of the form $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, see Proposition 2.2 of the Appendix.

**Example.** In MATH 125, we encountered the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$, where $\mathbf{e}_i$ is the vector that has 1 in the $i$th position and 0 elsewhere. For example,

$$\left\{ \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right\}$$

is the standard basis for $\mathbb{R}^3$. We saw in MATH 125 that this set is indeed both linearly independent and a spanning set for $\mathbb{R}^3$.

**Example.** Let $V = \{A \in M_2(\mathbb{R}) \mid \mathrm{Tr}(A) = 0\}$, the set of real $2 \times 2$ matrices of trace 0, and let

$$A_1 = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}, \quad A_2 = \begin{pmatrix}0 & 1\\0 & 0\end{pmatrix}, \quad A_3 = \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix},$$

which all lie in $V$. We show that $\{A_1, A_2, A_3\}$ is a basis for $V$.

*Linear independence:* Suppose that

$$\begin{aligned} c_1A_1 + c_2A_2 + c_3A_3 &= \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}, \\ \text{i.e.,} \quad c_1\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix} + c_2\begin{pmatrix}0 & 1\\0 & 0\end{pmatrix} + c_3\begin{pmatrix}0 & 0\\1 & 0\end{pmatrix} &= \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}, \\ \text{i.e.,} \quad \begin{pmatrix}c_1 & c_2\\c_3 & -c_1\end{pmatrix} &= \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}. \end{aligned}$$

Then $c_1 = c_2 = c_3 = 0$, so $A_1, A_2, A_3$ are linearly independent.

*Spanning:* Let

$$B = \begin{pmatrix}a & b\\c & d\end{pmatrix}$$

be any matrix in $V$, so $\mathrm{Tr}(B) = 0$ by assumption. This means that $d = -a$, so

$$\begin{aligned} B &= \begin{pmatrix}a & b\\c & -a\end{pmatrix} \\ &= a\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix} + b\begin{pmatrix}0 & 1\\0 & 0\end{pmatrix} + c\begin{pmatrix}0 & 0\\1 & 0\end{pmatrix} \\ &= aA_1 + bA_2 + cA_3, \end{aligned}$$

so $B \in \mathrm{Span}(A_1, A_2, A_3)$. Thus, $\{A_1, A_2, A_3\}$ is a linearly independent spanning set for $V$, i.e., a basis.

**Example.** If $n$ is a natural number, then $\{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathcal{P}_n$. A basis for the space $\mathcal{P}$ of all polynomials (with real coefficients) is the infinite set $\{1, x, x^2, \ldots\}$.

### Dimension

**Theorem 6.1.** *Let $V$ be a vector space that has a finite spanning set. Then $V$ has a finite basis, and any two bases for $V$ contain the same number of elements.*

For a proof, see Section 2 of the Appendix.

**Definition 6.2.** *Let $V$ be a vector space that has a finite spanning set. The* dimension *of $V$, denoted* $\dim(V)$*, is the number of elements in a basis for $V$. By Theorem 6.1, this number is independent of the choice of basis and is therefore well defined.*

**Remark.** A zero space, meaning a vector space with only one element, its zero vector, is taken to have the empty set $\emptyset = \{\}$ as a basis. Because the empty set contains no elements, the dimension of a zero space is 0.

If a vector space $V$ has a finite basis, then we call $V$ *finite dimensional*. Otherwise, we say that $V$ is *infinite dimensional*.

**Example.** The space $\mathbb{R}^n$ has basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and therefore has dimension $n$.

**Example.** The space $\mathcal{P}_n$ has basis $\{1, x, x^2, \ldots, x^n\}$ and therefore has dimension $n+1$.

**Example.** The space $M_{m,n}(\mathbb{R})$ has dimension $mn$. (Can you find a basis?)

**Example.** The space $\mathcal{P}$ has no finite basis and is therefore infinite dimensional.

**Example.** The space $\mathcal{S}$ is also infinite dimensional.

The following proposition is proven in the Appendix; see Corollaries 2.6 and 2.7 there.

**Proposition 6.3.** *If $U$ is a subspace of a finite-dimensional space $V$, then $U$ has finite dimension as well, and* $\dim(U) \leq \dim(V)$.

## I – 7 Coordinate vectors

Let $V$ be a vector space with finite dimension $n$, and let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis.

**Proposition 7.1.** *Every vector in $V$ can be written uniquely as a linear combination of the vectors in the basis $\mathcal{B}$. That is, given any $\mathbf{v} \in V$, there are unique scalars $c_1, \ldots, c_n \in \mathbb{R}$ such that $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$.*

*Proof.* The fact that there exist scalars $c_1, \ldots, c_n$ such that $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ is simply a restatement of the fact that $\mathbf{u}_1, \ldots, \mathbf{u}_n$ span $V$. For uniqueness, suppose we have two representations of $\mathbf{v}$ as a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_n$:

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n \\ \text{and} \quad \mathbf{v} &= d_1\mathbf{u}_1 + \cdots + d_n\mathbf{u}_n. \end{aligned}$$

Rearranging the equation

$$c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n = d_1\mathbf{u}_1 + \cdots + d_n\mathbf{u}_n,$$

we arrive at

$$(c_1 - d_1)\mathbf{u}_1 + \cdots + (c_n - d_n)\mathbf{u}_n = \mathbf{0}.$$

By the linear independence of $\mathbf{u}_1, \ldots, \mathbf{u}_n$, we conclude that $c_i - d_i = 0$ for all $i$, that is, $c_i = d_i$. □

The scalars $c_1, \ldots, c_n$ such that $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ are called the *coordinates* of $\mathbf{v}$ with respect to the basis $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, and the vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{R}^n$$

is called the *coordinate vector* of $\mathbf{v}$ with respect to $\mathcal{B}$.

**Example.** Let

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\},$$

a basis for $M_2(\mathbb{R})$. (If you want some practice in bases, show that $\mathcal{B}$ is indeed a basis for $M_2(\mathbb{R})$.) Find the coordinate vector

$$\left[ \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \right]_{\mathcal{B}}.$$

*Solution:* We solve the equation

$$c_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + c_3 \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + c_4 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

for the scalars $c_1, \ldots, c_4$, i.e.,

$$\begin{pmatrix} c_1 + c_2 + c_3 + c_4 & c_2 + c_3 + c_4 \\ c_3 + c_4 & c_4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

This equation of matrices gives us the four linear equations

$$\begin{aligned} c_1 + c_2 + c_3 + c_4 &= 1 \\ c_2 + c_3 + c_4 &= 2 \\ c_3 + c_4 &= 3 \\ c_4 &= 4, \end{aligned}$$

whose unique solution is $c_1 = c_2 = c_3 = -1$, $c_4 = 4$. Thus,

$$\left[\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\right]_{\mathcal{B}} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 4 \end{pmatrix}.$$

One of the great strengths of coordinate vectors is that addition and scalar multiplication in $V$ correspond to addition and scalar multiplication in $\mathbb{R}^n$. That is, if $\mathbf{u}, \mathbf{v} \in V$ and $c \in \mathbb{R}$, then

$$[\mathbf{u} + \mathbf{v}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}} \tag{7.1}$$

$$\text{and} \quad [c\mathbf{u}]_{\mathcal{B}} = c[\mathbf{u}]_{\mathcal{B}}. \tag{7.2}$$

**Remark.** Note how coordinate vectors make an $n$-dimensional vector space $V$ "look" a lot like $\mathbb{R}^n$, with vectors $\mathbf{v} \in V$ being replaced by their coordinate vectors $[\mathbf{v}]_{\mathcal{B}}$. We will make this connection more precise when we study isomorphisms.

**Caution.** The coordinate vector of a vector in $V$ depends on the choice of basis. For example, if we instead take the basis

$$\mathcal{C} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

of $M_2(\mathbb{R})$, then the coordinate vector of the same matrix

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

as in the above example, but now with respect to $\mathcal{C}$, is

$$\left[\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\right]_{\mathcal{C}} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

## I – 8 Linear independence and spanning via coordinate vectors

Let $V$ be a vector space with finite dimension $n$, and let $\mathcal{B}$ be a basis.

**Proposition 8.1.** *Suppose* $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$.

*(i)* $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *are linearly independent in* $V$ *if and only if* $[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_k]_{\mathcal{B}}$ *are linearly independent in* $\mathbb{R}^n$.

*(ii)* $\mathbf{v}_1, \ldots, \mathbf{v}_k$ *span* $V$ *if and only if* $[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_k]_{\mathcal{B}}$ *span* $\mathbb{R}^n$.

*Proof.* We prove (ii) and leave (i) as an exercise. Assume that $[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_k]_{\mathcal{B}}$ span $\mathbb{R}^n$. We show that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span $V$. Take any $\mathbf{v} \in V$. Because $[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_k]_{\mathcal{B}}$ span $\mathbb{R}^n$, we may choose $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$\begin{aligned}[\mathbf{v}]_{\mathcal{B}} &= c_1[\mathbf{v}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{v}_k]_{\mathcal{B}} \\ &= [c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k]_{\mathcal{B}}.\end{aligned}$$

Since $\mathbf{v}$ and $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ have the same coordinate vector, they are equal, i.e., $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$. This shows that $\mathbf{v}_1, \ldots \mathbf{v}_k$ span $V$.

Conversely, suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span $V$. We show that $[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_k]_{\mathcal{B}}$ span $\mathbb{R}^n$. Take any $\mathbf{w} \in \mathbb{R}^n$, and let $\mathbf{v}$ be the vector in $V$ with coordinate vector $\mathbf{w}$. Because $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span $V$, there are $c_1, \ldots, c_k \in \mathbb{R}$ such that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$, so

$$\begin{aligned}\mathbf{w} &= [\mathbf{v}]_{\mathcal{B}} \\ &= [c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k]_{\mathcal{B}} \\ &= c_1[\mathbf{v}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{v}_k]_{\mathcal{B}},\end{aligned}$$

so $\mathbf{w}$ is in the span of $[\mathbf{v}_1]_{\mathcal{B}}, \ldots, [\mathbf{v}_k]_{\mathcal{B}}$. □

**Example.** Decide whether $x^2+x+3, x^2+2x+4, x^2+x-2 \in \mathcal{P}_2$ are linearly independent.

*Solution:* We work with the basis $\{x^2, x, 1\}$ for $\mathcal{P}_2$. The given polynomials are linearly independent if and only if their coordinate vectors

$$\begin{pmatrix}1\\1\\3\end{pmatrix}, \quad \begin{pmatrix}1\\2\\4\end{pmatrix}, \quad \begin{pmatrix}1\\1\\-2\end{pmatrix}$$

are. We know from MATH 125 how to determine whether vectors in $\mathbb{R}^n$ (in this case, $\mathbb{R}^3$) are linearly independent: put the vectors as columns in a matrix and determine whether there is a pivot in every column of a row-echelon form. Here, we row reduce as follows:

$$\begin{pmatrix}1 & 1 & 1\\1 & 2 & 1\\3 & 4 & -2\end{pmatrix} \leftrightarrow \begin{pmatrix}1 & 1 & 1\\0 & 1 & 0\\0 & 1 & -5\end{pmatrix} \leftrightarrow \begin{pmatrix}1 & 1 & 1\\0 & 1 & 0\\0 & 0 & -5\end{pmatrix}.$$

(The symbol ↔ means "is row equivalent to".) There is a pivot in every column of the above row-echelon form, so the original three polynomials are linearly independent.

**Example.** Decide whether the matrices

$$\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}, \quad \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}, \quad \begin{pmatrix}2 & 1\\ 3 & 0\end{pmatrix}, \quad \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}, \quad \begin{pmatrix}3 & 0\\ 4 & 1\end{pmatrix} \tag{8.1}$$

span $M_2(\mathbb{R})$.

*Solution:* We work with the basis

$$\mathcal{B} = \left\{ \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} \right\}$$

of $M_2(\mathbb{R})$. We put the coordinate vectors of the given matrices, with respect to $\mathcal{B}$, as columns in a matrix and then row reduce:

$$\begin{pmatrix}1 & 1 & 2 & 1 & 3\\ 1 & 0 & 1 & -1 & 0\\ 2 & 1 & 3 & 1 & 4\\ -1 & 1 & 0 & 1 & 1\end{pmatrix} \leftrightarrow \begin{pmatrix}1 & 1 & 2 & 1 & 3\\ 0 & -1 & -1 & -2 & -3\\ 0 & -1 & -1 & -1 & -2\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}.$$

In fact, we do not even need a row-echelon form, because we can see already that there is no pivot in the last row. Therefore, the matrices in (8.1) do not span $M_2(\mathbb{R})$.

## I – 9 Linear independence and spanning in relation to dimension

**Proposition 9.1.** *Let $V$ be a vector space of finite dimension $n$.*

*(i) Any linearly independent set in $V$ contains at most $n$ elements.*

*(ii) Any spanning set for $V$ contains at least $n$ elements.*

For a proof, see Proposition 2.4 of the Appendix.

**Example.** In $\mathcal{P}_3$, which has dimension 4, the 5 polynomials

$$x^3+1, \quad 2x^3-3x^2+x-2, \quad x-4, \quad x^2-2x, \quad -x^3+2x-3$$

are linearly dependent. No calculation is necessary; we simply apply part (i) of Proposition 9.1.

**Example.** Because $M_{3,2}(\mathbb{R})$ has dimension 6, the 4 matrices

$$\begin{pmatrix} 1 & -1 \\ 2 & 5 \\ -1 & 3 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1/2 \\ 6 & -2 \\ 2/3 & 5 \end{pmatrix}, \quad \begin{pmatrix} -10 & 2 \\ 5/3 & 16 \\ 0 & 7 \end{pmatrix}, \quad \begin{pmatrix} 1/2 & 1/3 \\ 2 & 3 \\ 1 & 6 \end{pmatrix}$$

do not span $M_{3,2}(\mathbb{R})$. Again, no calculation is necessary; just apply part (ii) of Proposition 9.1.

**Caution.** If $V$ has dimension $n$, then a set of $n$ or fewer vectors need not be linearly independent. For example, the three vectors

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \tag{9.1}$$

in $\mathbb{R}^3$ are linearly dependent.

Similarly, a set of $n$ or more vectors need not span $V$. For example, the three vectors in (9.1) do not span $\mathbb{R}^3$.

### The case of $n$ vectors in an $n$-dimensional space

**Proposition 9.2.** *Let $V$ be a vector space of finite dimension $n$, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. (Note that the $n$ in $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is the same as the dimension of $V$; this is important.) Then the following are equivalent:*

*(i) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.*

*(ii) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span $V$.*

*(iii) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis for $V$.*

For a proof, see Proposition 2.5 of the Appendix.

**Example.** Let $V$ be a 4-dimensional vector space, and let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \in V$. Suppose that the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + c_4\mathbf{v}_4 = \mathbf{0} \tag{9.2}$$

has no solutions besides the trivial one, $c_1 = c_2 = c_3 = c_4 = 0$. Do the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ span $V$?

*Solution:* Yes. The fact that (9.2) has only the trivial solution means exactly that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ are linearly independent. But they are four vectors in the 4-dimensional space $V$, so by Proposition 9.2, they span $V$ (and form a basis, in fact).

## I – 10 Change of basis

Suppose $\mathcal{B}$ and $\mathcal{C}$ are two bases for an $n$-dimensional vector space $V$. Often, it is useful to be able to change from coordinates with respect to $\mathcal{B}$ to coordinates with respect to $\mathcal{C}$, and vice versa. This procedure is what we turn to now.

If $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, then the *change-of-basis matrix* from $\mathcal{B}$ to $\mathcal{C}$ is the $n \times n$ matrix

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} [\mathbf{u}_1]_{\mathcal{C}} & \cdots & [\mathbf{u}_n]_{\mathcal{C}} \end{pmatrix}.$$

That is, the $j$th column of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is the coordinate vector of $\mathbf{u}_j$ with respect to $\mathcal{C}$.

**Example.** Consider the vector space $\mathcal{P}_2$, and let $\mathcal{B} = \{p_1, p_2, p_3\}$ and $\mathcal{C} = \{q_1, q_2, q_3\}$, where

$$p_1 = 1, \quad p_2 = x - 1 = -1 + x, \quad p_3 = (x-1)^2 = 1 - 2x + x^2$$

and

$$q_1 = 1, \quad q_2 = x, \quad q_3 = x^2$$

We know already that $\mathcal{C}$ is a basis for $\mathcal{P}_2$, and we leave it as an exercise to show that $\mathcal{B}$ is a basis as well. (Use Section 8.) Now,

$$[p_1]_{\mathcal{C}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad [p_2]_{\mathcal{C}} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad [p_3]_{\mathcal{C}} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix},$$

so

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

**Proposition 10.1.** *If $\mathcal{B}$ and $\mathcal{C}$ are bases for a finite-dimensional vector space $V$, then for any $\mathbf{v} \in V$ we have $[\mathbf{v}]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}$.*

*Proof.* Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, and write $\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n$ with $a_1, \ldots, a_n \in \mathbb{R}$. Then

$$\begin{aligned} [\mathbf{v}]_{\mathcal{C}} &= [a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n]_{\mathcal{C}} \\ &= a_1[\mathbf{u}_1]_{\mathcal{C}} + \cdots + a_n[\mathbf{u}_n]_{\mathcal{C}} \quad \text{by (7.1) and (7.2)} \\ &= \begin{pmatrix} [\mathbf{u}_1]_{\mathcal{C}} & \cdots & [\mathbf{u}_n]_{\mathcal{C}} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \text{by a property of matrix multiplication} \\ &= P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}. \end{aligned}$$

□

**Example.** Let $p = 5 - 2(x-1) + 3(x-1)^2$. Use the change-of-basis matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ computed in the previous example to find the coordinates of $p$ with respect to the basis $\mathcal{C}$. Hence, write down $p$ in the form $a_0 + a_1x + a_2x^2$.

*Solution:* By Proposition 10.1,

$$[p]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[p]_{\mathcal{B}} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 10 \\ -8 \\ 3 \end{pmatrix}.$$

Therefore, $p = 10 - 8x + 3x^2$.

The next result tells us how we can stack together a sequence of changes of basis, and also how we reverse a change of basis.

**Proposition 10.2.** *Let $\mathcal{B}, \mathcal{C}, \mathcal{E}$ be bases for a finite-dimensional vector space $V$. Then $P_{\mathcal{E}\leftarrow\mathcal{B}} = P_{\mathcal{E}\leftarrow\mathcal{C}}P_{\mathcal{C}\leftarrow\mathcal{B}}$. In particular, $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible, and $P_{\mathcal{C}\leftarrow\mathcal{B}}^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$.*

For a proof, see Section 3 in the Appendix.

**Example.** We return once more to the bases $\mathcal{B}$ and $\mathcal{C}$ of $\mathcal{P}_2$ in the previous two examples. According to Proposition 10.2, we may compute $P_{\mathcal{B}\leftarrow\mathcal{C}}$ by inverting $P_{\mathcal{C}\leftarrow\mathcal{B}}$:

$$P_{\mathcal{B}\leftarrow\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}^{-1} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Using this new matrix, we may express any polynomial $q = a_0 + a_1x + a_2x^2 \in \mathcal{P}_2$ as a quadratic Taylor polynomial about $x = 1$:

$$[q]_{\mathcal{B}} \overset{\text{Prop. 10.1}}{=} P_{\mathcal{B}\leftarrow\mathcal{C}}[q]_{\mathcal{C}} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_0 + a_1 + a_2 \\ a_1 + 2a_2 \\ a_2 \end{pmatrix},$$

so $a_0 + a_1x + a_2x^2 = (a_0 + a_1 + a_2) + (a_1 + 2a_2)(x - 1) + a_2(x - 1)^2$. For example, if $q = 1 + x + x^2$, then $a_0 = a_1 = a_2 = 1$, so $1 + x + x^2 = 3 + 3(x - 1) + (x - 1)^2$.

Finally, we provide a very computational way to find change-of-basis matrices.

**Proposition 10.3.** *Let $\mathcal{B}$ and $\mathcal{C}$ be bases for an $n$-dimensional vector space $V$, and let $\mathcal{E}$ be another basis for $V$. Then the reduced row-echelon form of*

$$\left( P_{\mathcal{E}\leftarrow\mathcal{C}} \;\middle|\; P_{\mathcal{E}\leftarrow\mathcal{B}} \right)$$

*is*

$$\left( I_n \;\middle|\; P_{\mathcal{C}\leftarrow\mathcal{B}} \right),$$

*where $I_n$ is the $n \times n$ identity matrix.*

For a proof, see Section 4 in the Appendix. We will do an example in class to illustrate this proposition.

# (II) Linear Transformations

## II – 1 Linear transformations: definition and examples

Let $U$ and $V$ be vector spaces. A *linear transformation* from $U$ to $V$ is a map $\varphi : U \to V$ satisfying both of the following properties:

(i) For all $\mathbf{u}_1, \mathbf{u}_2 \in U$, $\varphi(\mathbf{u}_1 + \mathbf{u}_2) = \varphi(\mathbf{u}_1) + \varphi(\mathbf{u}_2)$.

(ii) For all $\mathbf{u} \in U$ and $c \in \mathbb{R}$, $\varphi(c\mathbf{u}) = c\varphi(\mathbf{u})$.

If a map satisfies (i), we say that it *respects addition.* If it satisfies (ii), we say that it *respects scalar multiplication.* Thus, a linear transformation is a map between vector spaces that respects both addition and scalar multiplication.

Linear transformations were an important part of MATH 125, but of course all linear transformations in that course were between $\mathbb{R}^n$ and $\mathbb{R}^m$. Now we consider linear transformations between general vector spaces.

**Example.** Consider

$$\begin{array}{rcl} \varphi : M_n(\mathbb{R}) & \to & \mathbb{R} \\ A & \mapsto & \mathrm{Tr}(A), \end{array}$$

that is, $\varphi$ maps $A$ to its trace. We saw in MATH 125 that the trace satisfies $\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$ and $\mathrm{Tr}(cA) = c\,\mathrm{Tr}(A)$, where $c \in \mathbb{R}$. These properties say exactly that our map $\varphi$ is a linear transformation.

Before giving our next two examples, we note that if $p, q \in \mathcal{P}$ and $a, c \in \mathbb{R}$, then

$$\begin{array}{rrcl} & (p+q)(a) & = & p(a) + q(a) \qquad (1.1) \\ \text{and} & (cp)(a) & = & cp(a) \qquad (1.2) \end{array}$$

In other words, (1.1) says that $p + q$ evaluated at $a$ is the same as $p$ evaluated at $a$ plus $q$ evaluated at $a$, and (1.2) says that $cp$ evaluated at $a$ is the same as $c$ times $p(a)$.

**Example.** Fix $a \in \mathbb{R}$. The map

$$\begin{array}{rcl} \varphi : \mathcal{P} & \to & \mathbb{R} \\ p & \mapsto & p(a) \end{array}$$

is a linear transformation:

*Addition:* If $p, q \in \mathcal{P}$, then

$$\begin{array}{rcll} \varphi(p+q) & = & (p+q)(a) & \text{by definition of } \varphi \\ & = & p(a) + q(a) & \text{by (1.1)} \\ & = & \varphi(p) + \varphi(q) & \text{by definition of } \varphi. \end{array}$$

*Scalar multiplication:* If $p \in \mathcal{P}$ and $c \in \mathbb{R}$, then

$$\begin{array}{rcll} \varphi(cp) & = & (cp)(a) & \text{by definition of } \varphi \\ & = & cp(a) & \text{by (1.2)} \\ & = & c\varphi(p) & \text{by definition of } \varphi. \end{array}$$

**Example.** Recall $\mathcal{S}$, the space of sequences $(a_0, a_1, a_2, \ldots)$ of real numbers $a_i$. The map

$$\begin{aligned} \varphi : \mathcal{P} &\to \mathcal{S} \\ p &\mapsto (p(0), p(1), p(2), \ldots) \end{aligned}$$

is linear:

*Addition:* If $p, q \in \mathcal{P}$, then

$$\begin{aligned} & \varphi(p+q) \\ = \; & ((p+q)(0), (p+q)(1), (p+q)(2), \ldots) \quad \text{by definition of } \varphi \\ = \; & (p(0)+q(0), p(1)+q(1), p(2)+q(2), \ldots) \quad \text{by (1.1)} \\ = \; & (p(0), p(1), p(2), \ldots) + (q(0), q(1), q(2), \ldots) \quad \text{by definition of addition in } \mathcal{S} \\ = \; & \varphi(p) + \varphi(q) \quad \text{by definition of } \varphi. \end{aligned}$$

*Scalar multiplication:* If $p \in \mathcal{P}$ and $c \in \mathbb{R}$, then

$$\begin{aligned} \varphi(cp) &= ((cp)(0), (cp)(1), (cp)(2), \ldots) \quad \text{by definition of } \varphi \\ &= (cp(0), cp(1), cp(2), \ldots) \quad \text{by (1.2)} \\ &= c(p(0), p(1), p(2), \ldots) \quad \text{by definition of scalar multiplication in } \mathcal{S} \\ &= c\varphi(p) \quad \text{by definition of } \varphi. \end{aligned}$$

If we wish to show that a map between vector spaces is not a linear transformation, it is enough to show that one of the two properties (i) or (ii) fails in at least one instance.

**Example.** Consider the map

$$\begin{aligned} \varphi : M_2(\mathbb{R}) &\to \mathbb{R}^2 \\ A &\mapsto \begin{pmatrix} \det(A) \\ \operatorname{Tr}(A) \end{pmatrix}. \end{aligned}$$

This map is not a linear transformation, because it does not respect addition, as we shall now see. Let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Note that $A + B = I$, the $2 \times 2$ identity matrix. Then

$$\varphi(A+B) = \varphi(I) = \begin{pmatrix} \det(I) \\ \operatorname{Tr}(I) \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

while

$$\varphi(A) + \varphi(B) = \begin{pmatrix} \det(A) \\ \operatorname{Tr}(A) \end{pmatrix} + \begin{pmatrix} \det(B) \\ \operatorname{Tr}(B) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Thus, $\varphi(A+B) \neq \varphi(A) + \varphi(B)$ for this pair of matrices $A, B$, so $\varphi$ does not respect addition.

## II – 2 Injectivity and kernel

A map $f : X \to Y$ is said to be *injective* (or *one-to-one*) if the equality $f(x_1) = f(x_2)$ implies that $x_1 = x_2$. For example, the exponential map

$$\begin{array}{rcl} \exp : \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & e^x \end{array}$$

is injective. If $x_1, x_2 \in \mathbb{R}$, then the only way for $\exp(x_1)$ to be equal to $\exp(x_2)$ is for $x_1$ to be equal to $x_2$. By contrast, the map

$$\begin{array}{rcl} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & \sin(x) \end{array}$$

is not injective, because $\sin(0) = \sin(\pi)$, but $0 \neq \pi$.

### The kernel of a linear transformation

If $\varphi : U \to V$ is a linear transformation of vector spaces, then the *kernel* of $\varphi$ is the subset of $U$ defined by

$$\mathrm{Ker}(\varphi) = \{\mathbf{u} \in U \mid \varphi(\mathbf{u}) = \mathbf{0}_V\},$$

that is, $\mathrm{Ker}(\varphi)$ consists of all those vectors in $U$ that $\varphi$ maps to the zero vector in $V$.

**Example.** We will find the kernel of the linear transformation

$$\begin{array}{rcl} \varphi : M_2(\mathbb{R}) & \to & \mathbb{R} \\ A & \mapsto & \mathrm{Tr}(A). \end{array}$$

Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R}).$$

Then $A \in \mathrm{Ker}(\varphi)$ if and only if $\varphi(A) = 0$, if and only if $a + d = 0$, if and only if

$$A = \begin{pmatrix} a & b \\ c & -a \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Thus, $\mathrm{Ker}(\varphi)$ is the span of the three matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

**Proposition 2.1.** *If $\varphi : U \to V$ is a linear transformation, then $\varphi(\mathbf{0}_U) = \mathbf{0}_V$.*

*Proof.*

$$\begin{array}{rcl} \varphi(\mathbf{0}_U) & = & \varphi(\mathbf{0}_U + \mathbf{0}_U) \\ & = & \varphi(\mathbf{0}_U) + \varphi(\mathbf{0}_U), \end{array}$$

so adding $-\varphi(\mathbf{0}_U)$ to both sides leaves $\mathbf{0}_V = \varphi(\mathbf{0}_U)$. □

**Exercise.** Show that the kernel of a linear transformation $\varphi : U \to V$ is a subspace of $U$.

**Proposition 2.2.** *Let $\varphi : U \to V$ be a linear transformation. Then $\varphi$ is injective if and only if* $\mathrm{Ker}(\varphi) = \{\mathbf{0}_U\}$*.*

*Proof.* Suppose first that $\varphi$ is injective. If $\mathbf{u} \in \mathrm{Ker}(\varphi)$, then

$$\begin{aligned} \varphi(\mathbf{u}) &= \mathbf{0}_V \\ &= \varphi(\mathbf{0}_U) \quad \text{by Proposition 2.1,} \end{aligned}$$

so $\mathbf{u} = \mathbf{0}_U$ by the assumption of injectivity.

Conversely, suppose that $\mathrm{Ker}(\varphi) = \{\mathbf{0}_U\}$. If $\mathbf{u}_1, \mathbf{u}_2 \in U$ satisfy $\varphi(\mathbf{u}_1) = \varphi(\mathbf{u}_2)$, then $\varphi(\mathbf{u}_1) - \varphi(\mathbf{u}_2) = \mathbf{0}_V$, i.e., $\varphi(\mathbf{u}_1 - \mathbf{u}_2) = \mathbf{0}_V$ by linearity, so $\mathbf{u}_1 - \mathbf{u}_2 \in \mathrm{Ker}(\varphi)$. But $\mathrm{Ker}(\varphi) = \{\mathbf{0}_U\}$ by assumption, so $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}_U$, i.e., $\mathbf{u}_1 = \mathbf{u}_2$. □

**Example.** Consider the linear transformation

$$\begin{aligned} \varphi : \mathcal{P} &\to \mathcal{P} \\ p &\mapsto (xp)', \end{aligned}$$

where $q'$ denotes the derivative of a polynomial $q$. Decide whether $\varphi$ is injective.

*Solution:* We find $\mathrm{Ker}(\varphi)$. If $p \in \mathcal{P}$, then

$$\begin{aligned} p \in \mathrm{Ker}(\varphi) &\iff \varphi(p) = 0 \\ &\iff (xp)' = 0 \quad \text{by definition of } \varphi \\ &\iff xp \text{ is constant} \quad \text{(standard property of differentiation)} \\ &\iff p = 0, \end{aligned}$$

because if $p$ is a non-zero polynomial, then $xp$ is not constant. Thus, $\mathrm{Ker}(\varphi) = \{0\}$, so $\varphi$ is injective.

## II – 3 Surjectivity and image

A map $f : X \to Y$ is said to be *surjective* (or *onto*) if for every $y \in Y$, there is $x \in X$ such that $f(x) = y$. For example, the map

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^3 \end{aligned}$$

is surjective: every real number has a cube root. However, the map

$$\begin{aligned} g : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2 \end{aligned}$$

is not surjective, because negative numbers do not have (real) square roots.

Another way to characterize whether a map $f$ is surjective is via the *image* of $f$, which is the subset of $Y$ given by

$$\text{Image}(f) = \{f(x) \mid x \in X\}.$$

That is, Image($f$) consists of all those elements $y \in Y$ for which there is $x \in X$ such that $f(x) = y$.

**Remark.** Another common word for the same concept is *range*. We will use the word *image*, but be aware that you are likely also to encounter the word *range* elsewhere.

**Example.** The image of $\sin : \mathbb{R} \to \mathbb{R}$ is $[-1, 1]$: For every $y \in [-1, 1]$, there is $x \in \mathbb{R}$ such that $\sin(x) = y$, and if $y \notin [-1, 1]$, then there is no $x \in \mathbb{R}$ such that $\sin(x) = y$.

Straight from the definitions, we see that a map $f : X \to Y$ is surjective if and only if Image($f$) $= Y$.

Let us see some examples from linear algebra.

**Example.** Find the image of the linear transformation

$$\begin{aligned} \varphi : \mathbb{R}^3 &\to \mathbb{R}^3 \\ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &\mapsto \begin{pmatrix} x_1 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

*Solution:*

$$\text{Image}(\varphi) = \{\varphi(\mathbf{u}) \mid \mathbf{u} \in \mathbb{R}^3\} = \left\{ \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} \,\middle|\, x \in \mathbb{R} \right\} = \left\{ x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \,\middle|\, x \in \mathbb{R} \right\} = \text{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

The image of $\varphi$ is not all of the codomain $\mathbb{R}^3$, so $\varphi$ is not surjective.

**Example.** The image of the linear transformation

$$\begin{aligned} \varphi : \mathbb{R}^3 &\to M_2(\mathbb{R}) \\ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &\mapsto \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \end{aligned}$$

consists of the matrices of the form

$$\begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix}$$

with $x_1, x_2, x_3 \in \mathbb{R}$. These are simply the symmetric matrices, i.e., the matrices $A$ with $A^{\mathrm{T}} = A$. Thus,

$$\mathrm{Image}(\varphi) = \{A \in M_2(\mathbb{R}) \mid A^{\mathrm{T}} = A\}.$$

Again, $\mathrm{Image}(\varphi)$ is not equal to the codomain $M_2(\mathbb{R})$ (not every $2 \times 2$ matrix is symmetric), so $\varphi$ is not surjective.

**Example.** Consider the linear transformation

$$\begin{aligned} \varphi : \mathcal{P}_1 &\to \mathbb{R}^2 \\ p &\mapsto \begin{pmatrix} p(0) \\ p(1) \end{pmatrix}, \end{aligned}$$

that is, $\varphi(a_1x+a_0) = \begin{pmatrix} a_0 \\ a_1 + a_0 \end{pmatrix}$. This map is surjective, because given any $\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in \mathbb{R}^2$, we may find a polynomial in $\mathcal{P}_1$ that maps to it under $\varphi$:

$$\varphi((b_2 - b_1)x + b_1) = \begin{pmatrix} b_1 \\ (b_2 - b_1) + b_1 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

In other words, the image of $\varphi$ is all of $\mathbb{R}^2$.

**Exercise.** Show that the image of a linear transformation $\varphi : U \to V$ is a subspace of $V$.

## II – 4 Rank and nullity

Let $\varphi : U \to V$ be a linear transformation. The *nullity* of $\varphi$, denoted nullity($\varphi$), is the dimension of Ker($\varphi$). The *rank* of $\varphi$, denoted rank($\varphi$), is the dimension of the image of $\varphi$. We have the following important theorem:

**Theorem 4.1** (Rank-nullity)**.** *Let $\varphi : U \to V$ be a linear transformation. If $U$ has finite dimension, then so do* Ker($\varphi$) *and* Image($\varphi$)*, and*

$$\operatorname{rank}(\varphi) + \operatorname{nullity}(\varphi) = \dim(U).$$

See the next page for a proof.

**Example.** Use the rank-nullity theorem to find nullity($\varphi$) where

$$\begin{array}{rcl} \varphi : M_3(\mathbb{R}) & \to & \mathbb{R} \\ A & \mapsto & \operatorname{Tr}(A). \end{array}$$

*Solution:* The map $\varphi$ is surjective, because given any $c \in \mathbb{R}$, we have

$$\varphi \begin{pmatrix} c & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = c.$$

Thus, Image($\varphi$) $= \mathbb{R}$, so rank($\varphi$) $=$ dim(Image($\varphi$)) $=$ dim($\mathbb{R}$) $= 1$. Hence, by the rank-nullity theorem,

$$\operatorname{nullity}(\varphi) = \dim(M_3(\mathbb{R})) - \operatorname{rank}(\varphi) = 9 - 1 = 8.$$

**Exercise.** Find a linearly independent set of 8 matrices in Ker($\varphi$), where $\varphi$ is as in the previous example. Conclude, using the fact that nullity($\varphi$) $= 8$, that your linearly independent set is in fact a basis for Ker($\varphi$). (Use Proposition 9.2 in Section I.)

**Example.** Let $n$ be a positive integer. One finds without much difficulty that the space of symmetric $n \times n$ matrices has dimension $\frac{1}{2}n(n+1)$. Use this fact, together with the rank-nullity theorem, to find the rank of the linear transformation

$$\begin{array}{rcl} \varphi : M_n(\mathbb{R}) & \to & M_n(\mathbb{R}) \\ A & \mapsto & A - A^{\mathrm{T}}. \end{array}$$

*Solution:* The kernel of $\varphi$ consists of those $n \times n$ matrices $A$ such that $A - A^{\mathrm{T}} = 0$, i.e., $A = A^{\mathrm{T}}$. Thus, the kernel is simply the space of symmetric $n \times n$ matrices, which has dimension $\frac{1}{2}n(n+1)$. Hence,

$$\operatorname{rank}(\varphi) = \dim(M_n(\mathbb{R})) - \operatorname{nullity}(\varphi) = n^2 - \frac{1}{2}n(n+1) = \frac{1}{2}n(n-1).$$

### Proof of the rank-nullity theorem

Recall that $\varphi : U \to V$ is a linear transformation. If $U$ has finite dimension, then so does $\operatorname{Ker}(\varphi)$, because it is a subspace of $U$. (See Corollary 2.6 in the Appendix.) Further, $\operatorname{Image}(\varphi)$ is spanned by a finite set, namely, the set of images of a basis for $U$, so $\operatorname{Image}(\varphi)$ is also finite dimensional.

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for $\operatorname{Image}(\varphi)$, choose $\mathbf{u}_i \in U$ such that $\varphi(\mathbf{u}_i) = \mathbf{v}_i$, where $i = 1, \ldots, r$, and let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a basis for $\operatorname{Ker}(\varphi)$. We show that $\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is a basis for $U$.

For linear independence, suppose that

$$a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + b_1\mathbf{u}_1 + \cdots + b_r\mathbf{u}_r = \mathbf{0}_U.$$

Then

$$\begin{aligned} \mathbf{0}_V &= \varphi(a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + b_1\mathbf{u}_1 + \cdots + b_r\mathbf{u}_r) \\ &= b_1\mathbf{v}_1 + \cdots + b_r\mathbf{v}_r, \end{aligned}$$

so $b_i = 0$ for all $i$ by the linear independence of $\mathbf{v}_1, \ldots, \mathbf{v}_r$. Hence,

$$a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n = \mathbf{0}_U,$$

so $a_i = 0$ for all $i$ by the linear independence of $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

For spanning, take any $\mathbf{u} \in U$, and write $\varphi(\mathbf{u}) = b_1\mathbf{v}_1 + \cdots + b_r\mathbf{v}_r$ for some scalars $b_i$, which is possible because the $\mathbf{v}_i$ form a basis for $\operatorname{Image}(\varphi)$. If $\mathbf{x} = \mathbf{u} - (b_1\mathbf{u}_1 + \cdots + b_r\mathbf{u}_r)$, then

$$\begin{aligned} \varphi(\mathbf{x}) &= \varphi(\mathbf{u}) - (b_1\mathbf{v}_1 + \cdots + b_r\mathbf{v}_r) \\ &= \varphi(\mathbf{u}) - \varphi(\mathbf{u}) \\ &= \mathbf{0}_V, \end{aligned}$$

so $\mathbf{x} \in \operatorname{Ker}(\varphi)$. Therefore, we may write $\mathbf{x} = a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n$ for some scalars $a_i$, because the $\mathbf{x}_i$ form a basis for $\operatorname{Ker}(\varphi)$. Thus,

$$\mathbf{u} = a_1\mathbf{x}_1 + \cdots + a_n\mathbf{x}_n + b_1\mathbf{u}_1 + \cdots + b_r\mathbf{u}_r.$$

Having shown that $\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is a basis for $U$, we see that

$$\dim(U) = n + r = \operatorname{nullity}(\varphi) + \operatorname{rank}(\varphi).$$

## II – 5 The matrix of a linear transformation

Let $\varphi : U \to V$ be a linear transformation, and assume that $U$ and $V$ are both finite dimensional. Let $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be a basis for $U$, and let $\mathcal{C}$ be a basis for $V$. Then the matrix of $\varphi$ with respect to $\mathcal{B}$ and $\mathcal{C}$ is the matrix

$$[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} [\varphi(\mathbf{u}_1)]_{\mathcal{C}} & \cdots & [\varphi(\mathbf{u}_n)]_{\mathcal{C}} \end{pmatrix}.$$

That is, the $j$th column of $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is the coordinate vector of $\varphi(\mathbf{u}_j)$ with respect to $\mathcal{C}$.

**Example.** Consider the linear transformation

$$\begin{aligned} \varphi : \mathcal{P}_3 &\to \mathcal{P}_2 \\ p &\mapsto p' + \int_0^1 p(x)\,dx, \end{aligned}$$

where $p'$ denotes the derivative of $p$. (For practice, you may like to show that $\varphi$ is indeed a linear transformation.) Let us find $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ where $\mathcal{B} = \{x^3, x^2, x, 1\}$ and $\mathcal{C} = \{x^2, x, 1\}$, bases for $\mathcal{P}_3$ and $\mathcal{P}_2$ respectively.

$$\varphi(x^3) = 3x^2 + \frac{1}{4}, \quad \varphi(x^2) = 2x + \frac{1}{3}, \quad \varphi(x) = \frac{3}{2}, \quad \varphi(1) = 1,$$

so

$$[\varphi(x^3)]_{\mathcal{C}} = \begin{pmatrix} 3 \\ 0 \\ 1/4 \end{pmatrix}, \quad [\varphi(x^2)]_{\mathcal{C}} = \begin{pmatrix} 0 \\ 2 \\ 1/3 \end{pmatrix}, \quad [\varphi(x)]_{\mathcal{C}} = \begin{pmatrix} 0 \\ 0 \\ 3/2 \end{pmatrix}, \quad [\varphi(1)]_{\mathcal{C}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore,

$$[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 1/4 & 1/3 & 3/2 & 1 \end{pmatrix}.$$

One of the key functions of $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is to translate the computation of a vector $\varphi(\mathbf{u})$ into a matrix multiplication. The following proposition shows how we achieve this.

**Proposition 5.1.** *For any* $\mathbf{u} \in U$, $[\varphi(\mathbf{u})]_{\mathcal{C}} = [\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{u}]_{\mathcal{B}}$.

See the next page for a proof.

**Example.** Let $p = x^3 + 3x^2 - 5x - 2$, and suppose we wish to compute

$$p' + \int_0^1 p(x)\,dx,$$

which is simply $\varphi(p)$, where $\varphi$ is as in the previous example. Rather than compute the derivative and the integral directly, let us use the matrix $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ that we found above:

$$\begin{aligned}
[\varphi(p)]_{\mathcal{C}} &= [\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}[p]_{\mathcal{B}} \quad \text{by Proposition 5.1} \\
&= \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 1/4 & 1/3 & 3/2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -5 \\ -2 \end{pmatrix} \\
&= \begin{pmatrix} 3 \\ 6 \\ -33/4 \end{pmatrix},
\end{aligned}$$

so $\varphi(p) = 3x^2 + 6x - \frac{33}{4}$.

If $\varphi : U \to U$ is a linear transformation from a finite-dimensional vector space $U$ to itself, we may wish to take the same basis $\mathcal{B}$ for the domain and the codomain. Rather than writing $[\varphi]_{\mathcal{B}\leftarrow\mathcal{B}}$, we write simply $[\varphi]_{\mathcal{B}}$. Thus, if $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, then

$$[\varphi]_{\mathcal{B}} = \begin{pmatrix} [\varphi(\mathbf{u}_1)]_{\mathcal{B}} & \cdots & [\varphi(\mathbf{u}_n)]_{\mathcal{B}} \end{pmatrix}.$$

**Example.** If $\varphi$ is the linear transformation

$$\begin{aligned}
\varphi : M_2(\mathbb{R}) &\to M_2(\mathbb{R}) \\
A &\mapsto A^{\mathrm{T}}
\end{aligned}$$

and

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

then

$$[\varphi]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

### Proof of Proposition 5.1

Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, and write $\mathbf{u} = a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n$ with $a_1, \ldots, a_n \in \mathbb{R}$. Then

$$\begin{aligned}
[\varphi(\mathbf{u})]_{\mathcal{C}} &= [\varphi(a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n)]_{\mathcal{C}} \\
&= [a_1\varphi(\mathbf{u}_1) + \cdots + a_n\varphi(\mathbf{u}_n)]_{\mathcal{C}} \quad \text{by linearity of } \varphi \\
&= a_1[\varphi(\mathbf{u}_1)]_{\mathcal{C}} + \cdots + a_n[\varphi(\mathbf{u}_n)]_{\mathcal{C}} \quad \text{by (7.1) and (7.2) in Section I} \\
&= \begin{pmatrix} [\varphi(\mathbf{u}_1)]_{\mathcal{C}} & \cdots & [\varphi(\mathbf{u}_n)]_{\mathcal{C}} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \quad \text{(property of matrix multiplication)} \\
&= [\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{u}]_{\mathcal{B}}.
\end{aligned}$$

## II – 6 The kernel and the image via the matrix of a linear transformation

As above, let $\varphi : U \to V$ be a linear transformation where $U$ and $V$ are finite dimensional, and let $\mathcal{B}$ and $\mathcal{C}$ be bases for $U$ and $V$ respectively.

**Proposition 6.1.**

*(i) If $\mathbf{u} \in U$, then $\mathbf{u} \in \mathrm{Ker}(\varphi)$ if and only if $[\mathbf{u}]_{\mathcal{B}} \in \mathrm{Nul}([\varphi]_{\mathcal{C}\leftarrow\mathcal{B}})$.*

*(ii) If $\mathbf{v} \in V$, then $\mathbf{v} \in \mathrm{Image}(\varphi)$ if and only if $[\mathbf{v}]_{\mathcal{C}} \in \mathrm{Col}([\varphi]_{\mathcal{C}\leftarrow\mathcal{B}})$.*

*Proof.* We prove (i) and leave (ii) as an exercise. Let $m = \dim(V)$. If $\mathbf{u} \in U$, then

$$\begin{aligned}
\mathbf{u} \in \mathrm{Ker}(\varphi) \quad &\Longleftrightarrow \quad \varphi(\mathbf{u}) = \mathbf{0}_V \\
&\Longleftrightarrow \quad [\varphi(\mathbf{u})]_{\mathcal{C}} = \mathbf{0}_{\mathbb{R}^m} \quad \text{(take coordinate vectors of both sides)} \\
&\Longleftrightarrow \quad [\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{u}]_{\mathcal{B}} = \mathbf{0}_{\mathbb{R}^m} \quad \text{by Proposition 5.1} \\
&\Longleftrightarrow \quad [\mathbf{u}]_{\mathcal{B}} \in \mathrm{Nul}([\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}).
\end{aligned}$$

□

In light of this proposition, if we wish to find a basis for $\mathrm{Ker}(\varphi)$, then once we have found $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$, all we have to do is find a basis for $\mathrm{Nul}([\varphi]_{\mathcal{C}\leftarrow\mathcal{B}})$ and translate those basis vectors back to vectors in $U$ via $\mathcal{B}$. Similarly, a basis for $\mathrm{Image}(\varphi)$ may be found by finding a basis for $\mathrm{Col}(\varphi)$ and then translating those basis vectors back to vectors in $V$ via $\mathcal{C}$.

**Example.** Find a basis for $\mathrm{Ker}(\varphi)$ and a basis for $\mathrm{Image}(\varphi)$ where

$$\begin{aligned}
\varphi : M_2(\mathbb{R}) \quad &\to \quad \mathcal{P}_2 \\
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad &\mapsto \quad (a - 2b + c + 3d)x^2 + (2a - 3b + c + 8d)x + (3a - 4b + c + 13d).
\end{aligned}$$

*Solution:* Let

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

and $\mathcal{C} = \{x^2, x, 1\}$. Then

$$\begin{aligned}
\varphi\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} &= x^2 + 2x + 3, \qquad & \varphi\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} &= -2x^2 - 3x - 4, \\
\varphi\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} &= x^2 + x + 1, \qquad & \varphi\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} &= 3x^2 + 8x + 13,
\end{aligned}$$

so

$$[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} 1 & -2 & 1 & 3 \\ 2 & -3 & 1 & 8 \\ 3 & -4 & 1 & 13 \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 & 0 & -1 & 7 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

From the given reduced row-echelon form, we find the basis

$$\left\{\begin{pmatrix}1\\1\\1\\0\end{pmatrix},\begin{pmatrix}-7\\-2\\0\\1\end{pmatrix}\right\}$$

for $\mathrm{Nul}([\varphi]_{\mathcal{C}\leftarrow\mathcal{B}})$. (Please review MATH 125 for the method to find a basis for a null space.) A basis for $\mathrm{Ker}(\varphi)$ is therefore

$$\left\{\begin{pmatrix}1&1\\1&0\end{pmatrix},\begin{pmatrix}-7&-2\\0&1\end{pmatrix}\right\}.$$

From the above reduced row-echelon form (or any row-echelon form, in fact), we see that a basis for $\mathrm{Col}([\varphi]_{\mathcal{C}\leftarrow\mathcal{B}})$ is

$$\left\{\begin{pmatrix}1\\2\\3\end{pmatrix},\begin{pmatrix}-2\\-3\\-4\end{pmatrix}\right\},$$

so a basis for $\mathrm{Image}(\varphi)$ is $\{x^2+2x+3, -2x^2-3x-4\}$.

We return to the more general situation of a linear transformation $\varphi : U \to V$ where $U, V$ are finite-dimensional vector spaces with bases $\mathcal{B}, \mathcal{C}$ respectively. Proposition 6.1 shows us that $\varphi$ is injective if and only if $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ has zero null space (i.e., has a pivot in every column of a row-echelon form), and is surjective if and only if $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ has full rank (i.e., has a pivot in every row).

**Example.** Suppose a linear transformation $\varphi : \mathcal{P}_2 \to \mathcal{P}_1$ has matrix

$$[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix}1&2&-1\\1&-1&3\end{pmatrix}$$

with respect to some bases $\mathcal{B}$ for $\mathcal{P}_2$ and $\mathcal{C}$ for $\mathcal{P}_1$. Is $\varphi$ surjective?

*Solution:* A row-echelon form of $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is

$$\begin{pmatrix}1&2&-1\\0&-3&4\end{pmatrix},$$

which has a pivot in each of its two rows, so $\varphi$ is surjective.

## II – 7 Composing linear transformations

If $\varphi : U \to V$ and $\psi : V \to W$ are linear transformations, then we may compose them to get a map

$$\begin{aligned} \psi \circ \varphi : U &\to W \\ \mathbf{u} &\mapsto \psi(\varphi(\mathbf{u})). \end{aligned}$$

**Exercise.** The composition $\psi \circ \varphi$ is again a linear transformation.

**Example.** Consider the linear transformations

$$\begin{aligned} \varphi : \mathcal{P}_3 &\to M_2(\mathbb{R}) \\ p &\mapsto \begin{pmatrix} p(0) & p(1) \\ p(2) & p(3) \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \psi : M_2(\mathbb{R}) &\to \mathcal{P}_1 \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} &\mapsto (a-b)x + (c-d). \end{aligned}$$

Then the composition $\psi \circ \varphi : \mathcal{P}_3 \to \mathcal{P}_1$ is given by

$$\psi \circ \varphi(p) = \psi \begin{pmatrix} p(0) & p(1) \\ p(2) & p(3) \end{pmatrix} = (p(0) - p(1))x + (p(2) - p(3)).$$

**Proposition 7.1.** *If $\varphi : U \to V$ and $\psi : V \to W$ are linear transformations, where $U, V, W$ are finite-dimensional vector spaces with bases $\mathcal{B}, \mathcal{C}, \mathcal{E}$ respectively, then*

$$[\psi \circ \varphi]_{\mathcal{E} \leftarrow \mathcal{B}} = [\psi]_{\mathcal{E} \leftarrow \mathcal{C}} [\varphi]_{\mathcal{C} \leftarrow \mathcal{B}}.$$

*Proof.* Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$. Then

$$\begin{aligned} & [\psi]_{\mathcal{E} \leftarrow \mathcal{C}} [\varphi]_{\mathcal{C} \leftarrow \mathcal{B}} \\ = \; & [\psi]_{\mathcal{E} \leftarrow \mathcal{C}} \left( [\varphi(\mathbf{u}_1)]_{\mathcal{C}} \quad \cdots \quad [\varphi(\mathbf{u}_n)]_{\mathcal{C}} \right) \\ = \; & \left( [\psi]_{\mathcal{E} \leftarrow \mathcal{C}} [\varphi(\mathbf{u}_1)]_{\mathcal{C}} \quad \cdots \quad [\psi]_{\mathcal{E} \leftarrow \mathcal{C}} [\varphi(\mathbf{u}_n)]_{\mathcal{C}} \right) \quad \text{(property of matrix multiplication)} \\ = \; & \left( [\psi(\varphi(\mathbf{u}_1))]_{\mathcal{E}} \quad \cdots \quad [\psi(\varphi(\mathbf{u}_n))]_{\mathcal{E}} \right) \quad \text{by Proposition 5.1 applied to } \psi \\ = \; & \left( [\psi \circ \varphi(\mathbf{u}_1)]_{\mathcal{E}} \quad \cdots \quad [\psi \circ \varphi(\mathbf{u}_n)]_{\mathcal{E}} \right) \\ = \; & [\psi \circ \varphi]_{\mathcal{E} \leftarrow \mathcal{B}}. \end{aligned}$$

□

This proposition, which may seem to state merely some technical fact, is crucial to our understanding of matrices. Matrix multiplication is defined precisely so as to make Proposition 7.1 hold. That is, matrix multiplication is defined to mimic the composition of linear transformations, so that the latter can be computed via the former, once bases have been chosen.

**Example.** We take the linear transformations $\varphi$ and $\psi$ in the previous example, and choose the following bases for $\mathcal{P}_3$, $M_2(\mathbb{R})$, and $\mathcal{P}_1$ respectively:

$$\begin{aligned}\mathcal{B} &= \{x^3, x^2, x, 1\}\\ \mathcal{C} &= \left\{\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\right\}\\ \mathcal{E} &= \{x, 1\}\end{aligned}$$

Then computing $\varphi(x^3), \varphi(x^2), \varphi(x), \varphi(1)$, we obtain

$$[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix}0 & 0 & 0 & 1\\ 1 & 1 & 1 & 1\\ 8 & 4 & 2 & 1\\ 27 & 9 & 3 & 1\end{pmatrix}.$$

Similarly, computing the effect of $\psi$ on the basis vectors in $\mathcal{C}$, we obtain

$$[\psi]_{\mathcal{E}\leftarrow\mathcal{C}} = \begin{pmatrix}1 & -1 & 0 & 0\\ 0 & 0 & 1 & -1\end{pmatrix}.$$

Therefore, by Proposition 7.1,

$$\begin{aligned}[\psi\circ\varphi]_{\mathcal{E}\leftarrow\mathcal{B}} &= \begin{pmatrix}1 & -1 & 0 & 0\\ 0 & 0 & 1 & -1\end{pmatrix}\begin{pmatrix}0 & 0 & 0 & 1\\ 1 & 1 & 1 & 1\\ 8 & 4 & 2 & 1\\ 27 & 9 & 3 & 1\end{pmatrix}\\ &= \begin{pmatrix}-1 & -1 & -1 & 0\\ -19 & -5 & -1 & 0\end{pmatrix}.\end{aligned}$$

For example, if we wish to find $\psi\circ\varphi(x^3 - x^2 + x - 1)$, then we need only compute

$$[\psi\circ\varphi]_{\mathcal{E}\leftarrow\mathcal{B}}[x^3 - x^2 + x - 1]_{\mathcal{B}} = \begin{pmatrix}-1 & -1 & -1 & 0\\ -19 & -5 & -1 & 0\end{pmatrix}\begin{pmatrix}1\\ -1\\ 1\\ -1\end{pmatrix} = \begin{pmatrix}-1\\ -15\end{pmatrix},$$

from which we read off that $\psi\circ\varphi(x^3 - x^2 + x - 1) = -x - 15$.

## II – 8 Invertible linear transformations

If $U$ is a vector space, the *identity map* on $U$ is the map

$$\begin{aligned} \mathbf{1}_U : U &\to U \\ \mathbf{u} &\mapsto \mathbf{u}. \end{aligned}$$

It is obviously a linear transformation.

A linear transformation $\varphi : U \to V$ is called *invertible* if there is a linear transformation $\psi : V \to U$ such that

$$\psi \circ \varphi = \mathbf{1}_U \quad \text{and} \quad \varphi \circ \psi = \mathbf{1}_V.$$

If the map $\psi$ exists, it is unique and is called the *inverse* of $\varphi$. The inverse of $\varphi$ is denoted $\varphi^{-1}$.

**Remark.** An invertible linear transformation is also called an *isomorphism*.

To find the inverse of a linear transformation $\varphi : U \to V$, if an inverse exists, do the following: For an arbitrary element $\mathbf{v} \in V$, try to solve the equation $\varphi(\mathbf{u}) = \mathbf{v}$ for $\mathbf{u}$ in terms of $\mathbf{v}$. There are two cases:

(a) If, for each $\mathbf{v} \in V$, there is a unique $\mathbf{u} \in U$ such that $\varphi(\mathbf{u}) = \mathbf{v}$, then $\varphi$ is invertible, and the inverse sends $\mathbf{v}$ to that $\mathbf{u}$.

(b) If there is some $\mathbf{v} \in V$ such that the equation $\varphi(\mathbf{u}) = \mathbf{v}$ has either no solutions or more than one solution, then $\varphi$ is not invertible.

**Example.** Decide whether

$$\begin{aligned} \varphi : \mathcal{P}_2 &\to \mathbb{R}^3 \\ p &\mapsto \begin{pmatrix} p(-1) \\ p(0) \\ p(1) \end{pmatrix} \end{aligned}$$

is invertible, and find its inverse if so.

*Solution:* Let $\mathbf{v} = (a_1, a_2, a_3) \in \mathbb{R}^3$. We try to solve $\varphi(p) = \mathbf{v}$ for $p = b_2x^2 + b_1x + b_0 \in \mathcal{P}_2$. Now, $\varphi(p) = \mathbf{v}$ if and only if

$$\begin{pmatrix} p(-1) \\ p(0) \\ p(1) \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

if and only if

$$\begin{aligned} b_2 - b_1 + b_0 &= a_1 \\ b_0 &= a_2 \\ b_2 + b_1 + b_0 &= a_3 \end{aligned}$$

This system has the unique solution

$$\begin{aligned} b_2 &= \frac{1}{2}a_1 - a_2 + \frac{1}{2}a_3 \\ b_1 &= -\frac{1}{2}a_1 + \frac{1}{2}a_3 \\ b_0 &= a_2, \end{aligned}$$

i.e., $p = (\frac{1}{2}a_1 - a_2 + \frac{1}{2}a_3)x^2 + (-\frac{1}{2}a_1 + \frac{1}{2}a_3)x + a_2$. Therefore, $\varphi$ is invertible, and its inverse $\varphi^{-1} : \mathbb{R}^3 \to \mathcal{P}_2$ is given by

$$\varphi^{-1}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = (\tfrac{1}{2}a_1 - a_2 + \tfrac{1}{2}a_3)x^2 + (-\tfrac{1}{2}a_1 + \tfrac{1}{2}a_3)x + a_2.$$

For example, the unique $p \in \mathcal{P}_2$ such that $p(-1) = 0$, $p(0) = 1$, and $p(1) = 0$, i.e., such that $\varphi(p) = (0, 1, 0)$, is $\varphi^{-1}(0, 1, 0) = -x^2 + 1$.

**Proposition 8.1.** *Let $\varphi : U \to V$ be a linear transformation, where $U$ and $V$ are finite dimensional with bases $\mathcal{B}$ and $\mathcal{C}$ respectively. Then $\varphi$ is invertible if and only if $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible, and if this is the case, $[\varphi^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}} = [\varphi]^{-1}_{\mathcal{C}\leftarrow\mathcal{B}}$.*

For a proof, see Section 5 of the Appendix.

**Example.** Consider the linear transformation

$$\begin{aligned} \varphi : \mathcal{P}_3 &\to M_2(\mathbb{R}) \\ p &\mapsto \begin{pmatrix} p(-2) & p'(-1) \\ p''(1) & p'''(2) \end{pmatrix}. \end{aligned}$$

If $\mathcal{B} = \{x^3, x^2, x, 1\}$ and

$$\mathcal{C} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

then

$$[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} -8 & 4 & -2 & 1 \\ 3 & -2 & 1 & 0 \\ 6 & 2 & 0 & 0 \\ 6 & 0 & 0 & 0 \end{pmatrix}.$$

(Recall from Section 5 how to find $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$.) By row reducing the augmented matrix $\left([\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} \;\middle|\; I\right)$, we find that $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible, with inverse

$$[\varphi]^{-1}_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{pmatrix} 0 & 0 & 0 & 1/6 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 1 & 1 & -3/2 \\ 1 & 2 & 0 & 1/3 \end{pmatrix}. \tag{8.1}$$

Therefore, $\varphi$ is invertible, and $[\varphi^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}$ is equal to the matrix in (8.1).

## II – 9 Criterion for invertibility in terms of injectivity and surjectivity

**Proposition 9.1.** *A linear transformation $\varphi : U \to V$ is invertible (i.e., an isomorphism) if and only if it is both injective and surjective.*

We prove this proposition after giving the next two examples.

**Example.** Let $V$ be the subspace of $\mathcal{P}$ consisting of the polynomials $q$ such that $q(0) = 0$, and consider the linear transformation

$$\begin{array}{rcl} \varphi : \mathcal{P} & \to & V \\ p & \mapsto & xp. \end{array}$$

(For example, $\varphi(x^2 - 3x + 5) = x^3 - 3x^2 + 5x$.) We show that $\varphi$ is an isomorphism using Proposition 9.1. First, we show injectivity, which is the same as showing that $\mathrm{Ker}(\varphi) = \{0\}$; see Proposition 2.2. If $p \in \mathcal{P}$ and $xp$ is the zero polynomial, then $p$ must be the zero polynomial as well, since a non-zero polynomial multiplied by $x$ is still non-zero. For surjectivity, take any $q \in V$, i.e., $q$ is a polynomial such that $q(0) = 0$. Writing $q = a_n x^n + \cdots + a_1 x + a_0$, we see by substituting $x = 0$ that $a_0 = 0$, so

$$q = a_n x^n + \cdots + a_1 x = xp$$

where $p = a_n x^{n-1} + \cdots + a_1$. Thus, $q = \varphi(p)$, and $\varphi$ is surjective.

**Example.** Decide whether the linear transformation

$$\begin{array}{rcl} \varphi : \mathcal{P} & \to & \mathcal{S} \\ p & \mapsto & (p(0), p(1), p(2), \ldots) \end{array}$$

is an isomorphism.

*Solution:* It is not an isomorphism, because it is not surjective. To see this, we show that the sequence $s = (1, 0, 0, 0, \ldots)$ is not in $\mathrm{Image}(\varphi)$. Suppose that $p$ is a polynomial such that $\varphi(p) = s$. Then $p(n)$ is zero for all positive integers $n$, so it has infinitely many roots and therefore must be the zero polynomial. (This is something you probably already know about polynomials: The graph of a non-zero polynomial can cross the horizontal axis only finitely many times. For a formal proof, see Lemma 15.1 in the Appendix.) But $p$ cannot be the zero polynomial, because $p(0) = 1$. Therefore, no $p$ satisfying $\varphi(p) = s$ exists, so $s \notin \mathrm{Image}(\varphi)$, and so $\varphi$ is not surjective.

### Proof of Proposition 9.1

Assume first that $\varphi$ is invertible, i.e., there is a linear transformation $\psi : V \to U$ such that $\psi \circ \varphi = \mathbf{1}_U$ and $\varphi \circ \psi = \mathbf{1}_V$. If $\varphi(\mathbf{u}_1) = \varphi(\mathbf{u}_2)$, then

$$\begin{array}{rrcl} & \psi(\varphi(\mathbf{u}_1)) & = & \psi(\varphi(\mathbf{u}_2)), \\ \text{i.e.,} & \psi \circ \varphi(\mathbf{u}_1) & = & \psi \circ \varphi(\mathbf{u}_2), \\ \text{i.e.,} & \mathbf{1}_U(\mathbf{u}_1) & = & \mathbf{1}_U(\mathbf{u}_2), \\ \text{i.e.,} & \mathbf{u}_1 & = & \mathbf{u}_2. \end{array}$$

Thus, $\varphi$ is injective. Further, $\varphi$ is surjective, because if $\mathbf{v} \in V$, then $\mathbf{v} = \mathbf{1}_V(\mathbf{v}) = \varphi \circ \psi(\mathbf{v}) = \varphi(\mathbf{u})$ where $\mathbf{u} = \psi(\mathbf{v})$.

Conversely, assume that $\varphi$ is both injective and surjective. Then if $\mathbf{v} \in V$, there is a unique $\mathbf{u} \in U$ such that $\varphi(\mathbf{u}) = \mathbf{v}$. Existence is just the definition of surjectivity, and for uniqueness, we observe that if $\varphi(\mathbf{u})$ and $\varphi(\mathbf{u}')$ are both equal to $\mathbf{v}$, then $\mathbf{u} = \mathbf{u}'$ by injectivity. Let $\mathbf{u}_\mathbf{v}$ be the unique vector $\mathbf{u}$ such that $\varphi(\mathbf{u}) = \mathbf{v}$. Then we have a map

$$\begin{array}{rcl} \psi : V & \to & U \\ \mathbf{v} & \mapsto & \mathbf{u}_\mathbf{v}. \end{array}$$

By construction, $\psi \circ \varphi = \mathbf{1}_U$ and $\varphi \circ \psi = \mathbf{1}_V$. All that remains is to show that $\psi$ is a linear transformation. To that end, let $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$ and $c \in \mathbb{R}$. Then

$$\begin{aligned} \psi(\mathbf{v}_1 + \mathbf{v}_2) &= \psi\Big(\varphi \circ \psi(\mathbf{v}_1) + \varphi \circ \psi(\mathbf{v}_2)\Big) \quad \text{because } \varphi \circ \psi = \mathbf{1}_V \\ &= \psi\Big(\varphi(\psi(\mathbf{v}_1)) + \varphi(\psi(\mathbf{v}_2))\Big) \\ &= \psi\Big(\varphi(\psi(\mathbf{v}_1) + \psi(\mathbf{v}_2))\Big) \quad \text{because } \varphi \text{ respects addition} \\ &= \psi \circ \varphi(\psi(\mathbf{v}_1) + \psi(\mathbf{v}_2)) \\ &= \psi(\mathbf{v}_1) + \psi(\mathbf{v}_2) \quad \text{because } \psi \circ \varphi = \mathbf{1}_U, \end{aligned}$$

$$\begin{aligned} \text{and} \quad \psi(c\mathbf{v}) &= \psi\Big(c\big(\varphi \circ \psi(\mathbf{v})\big)\Big) \quad \text{because } \varphi \circ \psi = \mathbf{1}_V \\ &= \psi\Big(c\varphi(\psi(\mathbf{v}))\Big) \\ &= \psi\Big(\varphi(c\psi(\mathbf{v}))\Big) \quad \text{because } \varphi \text{ respects scalar multiplication} \\ &= \psi \circ \varphi(c\psi(\mathbf{v})) \\ &= c\psi(\mathbf{v}) \quad \text{because } \psi \circ \varphi = \mathbf{1}_U. \end{aligned}$$

## II – 10 Isomorphic vector spaces

We say that vector spaces $U$ and $V$ are *isomorphic* if there is an isomorphism (i.e., an invertible linear transformation) $\varphi : U \to V$. In this case, we write $U \cong V$.

**Theorem 10.1.** *Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.*

*Proof.* (Sketch) Let $U$ and $V$ be finite-dimensional vector spaces. Suppose first that there is an isomorphism $\varphi : U \to V$. Let $n = \dim(U)$, and let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be a basis for $U$. We leave it as an exercise to show that the vectors $\varphi(\mathbf{u}_1), \dots, \varphi(\mathbf{u}_n) \in V$ are linearly independent and span $V$, so $\dim(V) = n = \dim(U)$.

Conversely, suppose that $U$ and $V$ both have dimension $n$. If $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is a basis for $U$, and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for $V$, then we may define a map

$$\begin{array}{rcl} U & \to & V \\ a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n & \mapsto & a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n, \end{array}$$

and this map is a linear transformation. Further, it is invertible, with inverse given by mapping $b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$ to $b_1\mathbf{u}_1 + \cdots + b_n\mathbf{u}_n$. Thus, $U$ and $V$ are isomorphic. □

**Example.** By Theorem 10.1, $\mathbb{R}^6 \cong M_{2,3}(\mathbb{R})$, because both $\mathbb{R}^6$ and $M_{2,3}(\mathbb{R})$ have dimension 6. By the same proposition, $\mathcal{P}_8$ is not isomorphic to $M_4(\mathbb{R})$, because the former has dimension 9 while the latter has dimension 16.

### Coordinates and isomorphism

We return to an idea hinted at in Section I – 7. Let $U$ be a vector space of finite dimension $n$, and let $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be a basis for $U$. We asserted that, via coordinate vectors, $U$ "looks a lot like" $\mathbb{R}^n$. Now we can make this assertion more precise, namely, by observing that the map

$$\begin{array}{rcl} \varphi : U & \to & \mathbb{R}^n \\ \mathbf{u} & \mapsto & [\mathbf{u}]_{\mathcal{B}} \end{array}$$

is an isomorphism. In fact, it is the isomorphism described in the second half of the proof of Theorem 10.1, where we take for $\mathcal{C}$ the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of $\mathbb{R}^n$. Indeed,

$$\begin{array}{rcll} \varphi(a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n) & = & [a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n]_{\mathcal{B}} & \text{by definition of } \varphi \\ & = & \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} & \\ & = & a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n. & \end{array}$$

# (III) Diagonalization

## III – 1 Review of eigenvalues, eigenvectors, and diagonalization

In MATH 125, we saw an introduction to eigenvalues and eigenvectors, together with a brief study of diagonalization over $\mathbb{R}$. In this course, we will see some further aspects of diagonalization. Our treatment will include orthogonal diagonalization of real symmetric matrices, which we will come to once we have studied inner products in Section IV.

First, we briefly recall some of the key ideas from the topic of eigenvalues and eigenvectors. Let $A$ be an $n \times n$ matrix. An *eigenvector* of $A$ is a non-zero vector $\mathbf{u} \in \mathbb{R}^n$ such that $A\mathbf{u}$ is a scalar multiple of $\mathbf{u}$. That is, $\mathbf{u}$ is an eigenvector of $A$ if it is non-zero and there is a scalar $\lambda \in \mathbb{R}$ such that $A\mathbf{u} = \lambda\mathbf{u}$. We say in this case that $\lambda$ is an *eigenvalue* of $A$, and that $\mathbf{u}$ is an eigenvector of $A$ with eigenvalue $\lambda$.

So, a (real) eigenvalue of $A$ is a scalar $\lambda \in \mathbb{R}$ such that there is a non-zero vector $\mathbf{u} \in \mathbb{R}^n$ satisfying $A\mathbf{u} = \lambda\mathbf{u}$.

The real eigenvalues of $A$ turn out to be the real roots of the *characteristic polynomial* of $A$, which is the monic degree-$n$ polynomial $p_A(x) = \det(xI - A)$.

**Remark.** Some books define the characteristic polynomial to be $\det(A - xI)$, but since $\det(A - xI) = (-1)^n \det(xI - A)$, there is little difference between the two definitions.

If $\lambda$ is an eigenvalue of $A$, then the *eigenspace* of $A$ associated to $\lambda$ is the set $\{\mathbf{u} \in \mathbb{R}^n \mid A\mathbf{u} = \lambda\mathbf{u}\}$, i.e., the set of eigenvectors with eigenvalue $\lambda$ together with the zero vector (which is not an eigenvector). Equivalently, the eigenspace associated to $\lambda$ is $\mathrm{Nul}(A - \lambda I)$.

**Example.** Find the eigenvalues and eigenspaces of the matrix $A = \begin{pmatrix} -8 & 10 \\ -5 & 7 \end{pmatrix}$.

*Solution:* The characteristic polynomial is $\det(xI - A) = x^2 + x - 6 = (x-2)(x+3)$, so the eigenvalues are $-3$ and $2$. The eigenspace associated to $-3$ is the null space of

$$-3I - A = \begin{pmatrix} 5 & -10 \\ 5 & -10 \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 & -2 \\ 0 & 0 \end{pmatrix},$$

so this eigenspace is spanned by $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$. The eigenspace associated to 2 is the null space of

$$2I - A = \begin{pmatrix} 10 & -10 \\ 5 & -5 \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix},$$

i.e., it is the span of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

### Diagonalization

Let us briefly recall diagonalization over $\mathbb{R}$. An $n \times n$ matrix $A$ is *diagonalizable* if it is similar to a diagonal matrix, i.e., if there is an invertible $n \times n$ matrix $P$ such that

$P^{-1}AP$ is diagonal. Before giving the next theorem, which is key, we recall that the *algebraic multiplicity* $m_\lambda$ of an eigenvalue $\lambda \in \mathbb{R}$ is the number of times the factor $x - \lambda$ appears in $p_A(x)$. Also, the *geometric multiplicity* $d_\lambda$ of $\lambda$ is the dimension of the associated eigenspace. We always have $1 \leq d_\lambda \leq m_\lambda$; see Lemma 6.1 in the Appendix.

**Theorem 1.1.** *Let $A \in M_n(\mathbb{R})$. Then the following are equivalent:*

*(i) $A$ is diagonalizable over $\mathbb{R}$.*

*(ii) $\mathbb{R}^n$ has a basis consisting of eigenvectors of $A$.*

*(iii) The sum of the geometric multiplicities of the real eigenvalues of $A$ is equal to $n$.*

*(iv) All roots of $p_A(x)$ are real, and the geometric multiplicity of every eigenvalue of $A$ is equal to its algebraic multiplicity.*

For a proof, see Section 6 of the Appendix.

To diagonalize a diagonalizable matrix $A$, find a basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors of $A$, let $P$ be the invertible matrix $\begin{pmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{pmatrix}$, and let $D$ be the diagonal matrix whose diagonal entries are $\lambda_1, \dots, \lambda_n$, where $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$. Then $P^{-1}AP = D$.

**Example.** Let

$$A = \begin{pmatrix} 1 & 1 & 4 \\ 2 & 2 & -4 \\ -2 & 1 & 7 \end{pmatrix}.$$

We find that $p_A(x) = (x-3)^2(x-4)$, so the eigenvalues are 3 and 4. Now,

$$3I - A = \begin{pmatrix} 2 & -1 & -4 \\ -2 & 1 & 4 \\ 2 & -1 & -4 \end{pmatrix} \leftrightarrow \begin{pmatrix} 2 & -1 & -4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

so the eigenspace associated to 3 has basis

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Similarly, by row reducing $4I - A$, we find that the eigenspace associated to 4 has basis

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\}.$$

Hence, if

$$P = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

then $P$ is invertible and $P^{-1}AP = D$. Thus, $A$ is diagonalizable.

## III – 2 Solving systems of differential equations via diagonalization

We turn to an application of diagonalization to the problem of solving systems of differential equations. A first-order differential equation with constant coefficients is an equation of the form

$$f' = \lambda f, \tag{2.1}$$

where $\lambda$ is some given real number, and $f$ is a differentiable function to be solved for. In fact, the solutions to the equation in (2.1) are straightforward to write down. They are the functions $f(x) = ae^{\lambda x}$ with $a \in \mathbb{R}$. (One direction is easy: Just differentiate the function $f(x) = ae^{\lambda x}$ and observe that you get $\lambda f(x)$. For the other direction, see Section 7 of the Appendix.)

Now consider the following *system* of differential equations:

$$\begin{aligned} f_1' &= 4f_1 - f_2 \\ f_2' &= 2f_1 + f_2 \end{aligned} \tag{2.2}$$

Here, we are solving for two differentiable functions $f_1, f_2$. We may express the system in the form

$$\begin{pmatrix} f_1' \\ f_2' \end{pmatrix} = A \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{2.3}$$

where

$$A = \begin{pmatrix} 4 & -1 \\ 2 & 1 \end{pmatrix}.$$

We will use the fact that $A$ is diagonalizable to solve the system in (2.2). Following the method in Section 1 (or MATH 125), we find that $P^{-1}AP = D$ where

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

Define functions $g_1, g_2$ by

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = P^{-1} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}. \tag{2.4}$$

Differentiating, we obtain

$$\begin{aligned} \begin{pmatrix} g_1' \\ g_2' \end{pmatrix} &= P^{-1} \begin{pmatrix} f_1' \\ f_2' \end{pmatrix} \quad \text{(I will explain this step in class)} \\ &= P^{-1}A \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \quad \text{by (2.3)} \\ &= P^{-1}AP \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \quad \text{by (2.4)} \\ &= D \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 2g_1 \\ 3g_2 \end{pmatrix}. \end{aligned}$$

Thus, $g_1' = 2g_1$ and $g_2' = 3g_2$. These new differential equations are just instances of the differential equation we saw in (2.1) and therefore have the solutions $g_1(x) = a_1e^{2x}$ and $g_2(x) = a_2e^{3x}$ where $a_1, a_2 \in \mathbb{R}$. Now we use (2.4) again to express the functions $f_1$ and $f_2$ in terms of the functions $e^{2x}$ and $e^{3x}$. Specifically,

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \overset{(2.4)}{=} P \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} g_1 + g_2 \\ 2g_1 + g_2 \end{pmatrix}.$$

Thus,

$$f_1(x) \quad = \quad a_1e^{2x} + a_2e^{3x} \tag{2.5}$$

$$f_2(x) \quad = \quad 2a_1e^{2x} + a_2e^{3x} \tag{2.6}$$

The constants $a_1, a_2 \in \mathbb{R}$ are arbitrary. Any choice of these constants gives a solution to the system in (2.2).

### Systems of ODEs with constraints

The solutions $f_1, f_2$ found in the above example were general solutions without any further constraints. Suppose instead that we were asked to solve the same system but now subject to the constraints $f_1(0) = -1$ and $f_2(0) = 4$. Then substituting $x = 0$ in (2.5) and (2.6), we obtain the linear system

$$\begin{aligned} a_1 + a_2 &= -1 \\ 2a_1 + a_2 &= \phantom{-}4, \end{aligned}$$

which has the solution $a_1 = 5$, $a_2 = -6$. Therefore, the solution to (2.2) together with the constraints $f_1(0) = -1$ and $f_2(0) = 4$ is

$$\begin{aligned} f_1(x) \quad &= \quad 5e^{2x} - 6e^{3x} \\ f_2(x) \quad &= \quad 10e^{2x} - 6e^{3x} \end{aligned}$$

In class, we will do an example of a system of three differential equations in three functions $f_1, f_2, f_3$. The method will be identical, but will involve diagonalizing a $3 \times 3$ matrix instead of a $2 \times 2$ one.

## III – 3 Diagonalization over the complex numbers

Complex numbers are introduced in MATH 125, but in case you need an overview of the key points, please refer to the document called *What You Need to Know about Complex Numbers for MATH 225* on eClass.

One point that should be stressed is that **every real number is also a complex number**. Thus, $3, \pi, \sqrt{2}, -\sqrt{5}, 1+2i, \pi - 5i$ are all complex numbers.

### Complex eigenvalues and eigenvectors

A complex eigenvector of $A$ is a non-zero vector $\mathbf{u} \in \mathbb{C}^n$ such that $A\mathbf{u}$ is a complex scalar multiple of $\mathbf{u}$. That is, $\mathbf{u}$ is an eigenvector of $A$ if it is non-zero and there is a scalar $\lambda \in \mathbb{C}$ such that $A\mathbf{u} = \lambda\mathbf{u}$. We say in this case that $\lambda$ is a complex eigenvalue of $A$, and that $\mathbf{u}$ is an eigenvector of $A$ with eigenvalue $\lambda$.

A complex eigenvalue of $A$ is therefore a scalar $\lambda \in \mathbb{C}$ such that there is a non-zero vector $\mathbf{u} \in \mathbb{C}^n$ satisfying $A\mathbf{u} = \lambda\mathbf{u}$.

The complex eigenvalues of $A$ are the complex roots of the characteristic polynomial of $A$.

If $\lambda$ is a complex eigenvalue of $A$, then the eigenspace of $A$ associated to $\lambda$ is the set $\{\mathbf{u} \in \mathbb{C}^n \mid A\mathbf{u} = \lambda\mathbf{u}\}$, i.e., the set of complex eigenvectors with eigenvalue $\lambda$ together with the zero vector (which is not an eigenvector).

**Example.** The matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has characteristic polynomial $p_A(x) = x^2 + 1 = (x-i)(x+i)$, so its complex eigenvalues are $i$ and $-i$. The complex eigenspace associated to the eigenvalue $i$ is found, as in the real situation, by row reducing $\lambda I - A$ where now $\lambda = i$:

$$iI - A = \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix}. \tag{3.1}$$

The associated eigenspace is the complex null space of the matrix in (3.1), which is the complex span of $\begin{pmatrix} i \\ 1 \end{pmatrix}$. The eigenspace associated to $-i$ is the complex span of $\begin{pmatrix} -i \\ 1 \end{pmatrix}$. This uses Proposition 3.1 below.

If $\mathbf{u} \in \mathbb{C}^n$, let $\overline{\mathbf{u}}$ be the vector obtained by taking the complex conjugate of every entry in $\mathbf{u}$.

**Proposition 3.1.** *If $A \in M_n(\mathbb{R})$ and $\lambda \in \mathbb{C}$ is an eigenvalue of $A$, then $\overline{\lambda}$ is also an eigenvalue of $A$. Further, if the complex eigenspace for $\lambda$ is spanned by $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, then the complex eigenspace for $\overline{\lambda}$ is spanned by $\{\overline{\mathbf{u}}_1, \ldots, \overline{\mathbf{u}}_k\}$.*

*Proof.* All we have to show is that if $\mathbf{u}$ is a complex eigenvector with eigenvalue $\lambda$, then $\overline{\mathbf{u}}$ is an eigenvector with eigenvalue $\overline{\lambda}$. This is straightforward, because if $A\mathbf{u} = \lambda\mathbf{u}$, then

$$\overline{\lambda}\,\overline{\mathbf{u}} = \overline{\lambda\mathbf{u}} = \overline{A\mathbf{u}} = A\overline{\mathbf{u}}$$

because $A$ is real. □

**Diagonalization over $\mathbb{C}$**

If $A \in M_n(\mathbb{C})$, i.e., is an $n \times n$ matrix with complex entries, then $A$ is said to be *diagonalizable over* $\mathbb{C}$ if there is an invertible $n \times n$ matrix $P \in M_n(\mathbb{C})$ such that $P^{-1}AP$ is diagonal. The criteria for when a matrix is diagonalizable over $\mathbb{C}$ are similar for the situation over $\mathbb{R}$. If $\lambda \in \mathbb{C}$ is an eigenvalue, the (complex) algebraic multiplicity $m_\lambda$ of $\lambda$ is the number of times the factor $x - \lambda$ appears in the factorization of $p_A(x)$ over $\mathbb{C}$. (For example, because $x^2 + 1 = (x - i)(x + i)$ in the example above, each of the eigenvalues $i, -i$ has algebraic multiplicity 1.)

The (complex) geometric multiplicity $d_\lambda$ of a complex eigenvalue $\lambda$ is the dimension of the associated eigenspace as a complex vector space. We again have $1 \leq d_\lambda \leq m_\lambda$.

**Theorem 3.2.** *Let $A \in M_n(\mathbb{C})$. Then the following are equivalent:*

*(i) $A$ is diagonalizable over $\mathbb{C}$.*

*(ii) $\mathbb{C}^n$ has a basis consisting of eigenvectors of $A$.*

*(iii) The sum of the complex geometric multiplicities of the eigenvalues of $A$ is equal to $n$.*

*(iv) The complex geometric multiplicity of every eigenvalue of $A$ is equal to its complex algebraic multiplicity.*

The proof is identical to that in the real case. See Section 6 of the Appendix.

**Example.** Let

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

which has characteristic polynomial $p_A(x) = (x - 3)(x^2 - 2x + 2)$. The eigenvalues of $A$ are therefore 3, $1 + i$, and $1 - i$. (We have used the usual formula for the roots of a quadratic polynomial to find the two non-real eigenvalues.) The eigenspace associated to the eigenvalue 3 is found just as usual, by row reducing $3I - A$, and we find that it is spanned by $(1, 0, -1)$. For the eigenspace associated to $1+i$, we row reduce $(1+i)I - A$:

$$(1+i)I - A = \begin{pmatrix} i & 3 & 2 \\ -1 & -1+i & -1 \\ 1 & -1 & -1+i \end{pmatrix} \leftrightarrow \begin{pmatrix} 1 & 0 & i \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

(I will do the row-reduction steps in detail in class.) The eigenspace associated to $1 + i$ is therefore spanned by $(i, 1, -1)$. Using Proposition 3.1, we see that the eigenspace associated to $1 - i$ is spanned by $(-i, 1, -1)$. Hence, if

$$P = \begin{pmatrix} 1 & i & -i \\ 0 & 1 & 1 \\ -1 & -1 & -1 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1+i & 0 \\ 0 & 0 & 1-i \end{pmatrix},$$

then $P^{-1}AP = D$. Thus, $A$ is diagonalizable over $\mathbb{C}$.

## III – 4 The $2 \times 2$ case with non-real eigenvalues

Recall from MATH 125 that the matrices describing rotations of the plane are the matrices of the form

$$\begin{pmatrix} c & -s \\ s & c \end{pmatrix},$$

where $c, s \in \mathbb{R}$ satisfy $c^2 + s^2 = 1$. Equivalently, the $2 \times 2$ rotation matrices are those of the form

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$

because if $c^2 + s^2 = 1$, we may always choose $\theta \in \mathbb{R}$ such that $\cos(\theta) = c$ and $\sin(\theta) = s$. The number $\theta \in \mathbb{R}$ is the anticlockwise angle of rotation.

**Proposition 4.1.** *Let $A$ be a real $2 \times 2$ matrix that has a non-real complex eigenvalue $\lambda = a + bi$. If $\mathbf{w}$ is an eigenvector for $\lambda$, then the real matrix $Q = \begin{pmatrix} \mathrm{Re}(\mathbf{w}) & \mathrm{Im}(\mathbf{w}) \end{pmatrix}$ is invertible, and*

$$Q^{-1}AQ = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = s \begin{pmatrix} a/s & b/s \\ -b/s & a/s \end{pmatrix},$$

*where $s = |\lambda| = \sqrt{a^2 + b^2}$.*

For a proof of Proposition 4.1, see Section 8 of the Appendix.

**Remark.** The notation $\mathrm{Re}(\mathbf{w})$ and $\mathrm{Im}(\mathbf{w})$ appearing in the proposition signifies the real part and imaginary part respectively of the complex vector $\mathbf{w}$.

**Remark.** The matrix $\begin{pmatrix} a/s & b/s \\ -b/s & a/s \end{pmatrix}$ appearing in the proposition is a rotation matrix, because $(a/s)^2 + (-b/s)^2 = (a^2 + b^2)/(a^2 + b^2) = 1$. We also point out that the effect of replacing $A$ by $Q^{-1}AQ$ amounts simply to a change of basis. The proposition therefore says that every real $2 \times 2$ matrix with non-real eigenvalues is, after a change of basis, a scalar times a rotation matrix.

**Example.** Let

$$A = \begin{pmatrix} 3 & -17 \\ 1 & -5 \end{pmatrix}.$$

The characteristic polynomial of $A$ is $p_A(x) = \det(xI - A) = x^2 + 2x + 2$, whose complex roots are

$$\frac{1}{2}(-2 \pm \sqrt{-4}) = -1 \pm i.$$

We will apply Proposition 4.1 with $\lambda = -1 + i$. (We may use either eigenvalue.) First, we find an eigenvector with eigenvalue $\lambda = -1+i$:

$$\begin{aligned}
(-1+i)I - A &= \begin{pmatrix} -4+i & 17 \\ -1 & 4+i \end{pmatrix} \\
&\leftrightarrow \begin{pmatrix} 1 & -4-i \\ -4+i & 17 \end{pmatrix} \quad \text{(interchange the rows, then scale row 1 by } -1\text{)} \\
&\leftrightarrow \begin{pmatrix} 1 & -4-i \\ 0 & 0 \end{pmatrix} \quad \text{(add } 4-i \text{ times row 1 to row 2)}
\end{aligned}$$

An eigenvector is therefore $\mathbf{w} = \begin{pmatrix} 4+i \\ 1 \end{pmatrix}$. Hence, we let

$$Q = \begin{pmatrix} \mathrm{Re}(\mathbf{w}) & \mathrm{Im}(\mathbf{w}) \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}.$$

By Proposition 4.1, $Q$ is invertible and

$$\begin{aligned}
Q^{-1}AQ &= \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix} \quad \text{(no calculation necessary; just use Prop. 4.1)} \\
&= |\lambda| \begin{pmatrix} -1/|\lambda| & 1/|\lambda| \\ -1/|\lambda| & -1/|\lambda| \end{pmatrix} \\
&= \sqrt{2} \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \\
&= \sqrt{2} \begin{pmatrix} \cos(5\pi/4) & -\sin(5\pi/4) \\ \sin(5\pi/4) & \cos(5\pi/4) \end{pmatrix} \\
&= \sqrt{2}\, R_{5\pi/4}.
\end{aligned}$$

Thus, after a change of basis, $A$ is $\sqrt{2}$ times rotation anticlockwise by angle $5\pi/4$.

# (IV) Inner Product Spaces

## IV – 1 Inner product spaces: definition and examples

An *inner product* on a vector space $V$ is an operation that assigns to each pair $(\mathbf{u},\mathbf{v}) \in V^2$ a real number $\langle \mathbf{u},\mathbf{v}\rangle$, in such a way that the following hold:

(i) For all $\mathbf{u},\mathbf{v} \in V$, $\langle \mathbf{u},\mathbf{v}\rangle = \langle \mathbf{v},\mathbf{u}\rangle$.

(ii) For all $\mathbf{u},\mathbf{v},\mathbf{w} \in V$, $\langle \mathbf{u},\mathbf{v}+\mathbf{w}\rangle = \langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{u},\mathbf{w}\rangle$.

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

(iv) For all $\mathbf{u} \in V$, $\langle \mathbf{u},\mathbf{u}\rangle \geq 0$, and $\langle \mathbf{u},\mathbf{u}\rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}_V$.

A vector space $V$ together with an inner product $\langle\cdot,\cdot\rangle$ is called an *inner product space*, denoted $(V, \langle\cdot,\cdot\rangle)$.

**Remark.** It follows from the axioms of an inner product that $\langle \mathbf{u}+\mathbf{v},\mathbf{w}\rangle = \langle \mathbf{u},\mathbf{w}\rangle + \langle \mathbf{v},\mathbf{w}\rangle$:

$$\langle \mathbf{u}+\mathbf{v},\mathbf{w}\rangle \overset{\text{(i)}}{=} \langle \mathbf{w},\mathbf{u}+\mathbf{v}\rangle \overset{\text{(ii)}}{=} \langle \mathbf{w},\mathbf{u}\rangle + \langle \mathbf{w},\mathbf{v}\rangle \overset{\text{(i)}}{=} \langle \mathbf{u},\mathbf{w}\rangle + \langle \mathbf{v},\mathbf{w}\rangle.$$

Similarly, we may show that $\langle \mathbf{u},c\mathbf{v}\rangle = c\langle \mathbf{u},\mathbf{v}\rangle$ (short exercise).

**Example.** The dot product studied in MATH 125 is an inner product on $\mathbb{R}^n$. For example, we know that if $\mathbf{u},\mathbf{v} \in \mathbb{R}^n$, then $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$. This is axiom (i) of an inner product. We leave it as an exercise to verify the other three axioms. The dot product is also called the *standard inner product* on $\mathbb{R}^n$.

**Example.** Let us put an inner product on $\mathcal{P}_2$. For $p, q \in \mathcal{P}_2$, define

$$\langle p,q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1).$$

We will verify axiom (iv) and leave the other axioms as an exercise. If $p \in \mathcal{P}_2$, then

$$\langle p,p\rangle = p(-1)^2 + p(0)^2 + p(1)^2 \geq 0.$$

Further, because $p(-1)^2, p(0)^2, p(1)^2$ are all non-negative, $\langle p,p\rangle = 0$ if and only if $p(-1)^2 = p(0)^2 = p(1)^2 = 0$, if and only if $p(-1) = p(0) = p(1) = 0$. But a polynomial of degree at most 2 that vanishes at three different points must be the zero polynomial.

There is nothing special about the numbers $-1, 0, 1$ in the above example. Any three distinct numbers would do for $\mathcal{P}_2$. In fact, as the next example shows, we may similarly put an inner product on $\mathcal{P}_n$ for any $n \geq 1$.

**Example.** If $x_0, \ldots, x_n$ are distinct real numbers, then

$$\langle p,q\rangle = \sum_{i=0}^{n} p(x_i)q(x_i)$$

defines an inner product on $\mathcal{P}_n$.

**Example.** Let $a < b$ be real numbers, and define $C[a,b]$ to be the space of continuous functions $f : [a,b] \to \mathbb{R}$. We may define an inner product on $C[a,b]$ by

$$\langle f, g \rangle = \int_a^b f(x)g(x)\,dx.$$

We will verify axiom (ii). If $f, g, h \in C[a,b]$, then

$$\begin{aligned}
\langle f, g+h \rangle &= \int_a^b f(x)(g+h)(x)\,dx \\
&= \int_a^b f(x)(g(x)+h(x))\,dx \\
&= \int_a^b (f(x)g(x)+f(x)h(x))\,dx \\
&= \int_a^b f(x)g(x)\,dx + \int_a^b f(x)h(x)\,dx \\
&= \langle f, g \rangle + \langle f, h \rangle.
\end{aligned}$$

Axiom (iv) is harder to verify, requiring a little analysis. See Section 9 of the Appendix.

## IV – 2 Length, distance, and orthogonality

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. The *norm* of a vector $\mathbf{u} \in V$ is the non-negative real number $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle}$. A vector of norm 1 is called a *unit vector*. The *distance* $\operatorname{dist}(\mathbf{u}, \mathbf{v})$ between two vectors $\mathbf{u}, \mathbf{v} \in V$ is the norm of their difference, i.e.,

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

**Example.** When $\mathbb{R}^n$ has the standard inner product (the dot product), the norm of a vector is its length in the sense of MATH 125. Thus, in $\mathbb{R}^3$ for example, $\|(x_1, x_2, x_3)\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$.

In a general inner product space, we will use the word *norm* instead of *length*, but remember that they are the same concept in $\mathbb{R}^2$ and $\mathbb{R}^3$ (when these spaces are given the standard inner product).

**Example.** Give $\mathcal{P}_2$ the inner product $\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$. If $p = x^2 + x + 1$ and $q = x^2 - 2$, then

$$\begin{aligned} \operatorname{dist}(p, q) &= \|p - q\| \\ &= \|x + 3\| \\ &= \sqrt{(-1+3)^2 + (0+3)^2 + (1+3)^2} \\ &= \sqrt{29}. \end{aligned}$$

### Orthogonality

Two vectors $\mathbf{u}, \mathbf{v}$ in an inner product space $(V, \langle\cdot,\cdot\rangle)$ are said to be *mutually orthogonal* (or simply *orthogonal*) if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$. This assumption is equivalent to $\langle \mathbf{v}, \mathbf{u}\rangle = 0$ because of symmetry of the inner product.

**Example.** In the inner product space $C[-\pi, \pi]$, the functions cos and sin are orthogonal:

$$\langle \cos, \sin\rangle = \int_{-\pi}^{\pi} \cos(x)\sin(x)\,dx = \frac{1}{2}\int_{-\pi}^{\pi} \sin(2x)\,dx = -\frac{1}{4}[\cos(2x)]_{-\pi}^{\pi} = 0.$$

If $U$ is a subspace of an inner product space $(V, \langle\cdot,\cdot\rangle)$, then the *orthogonal complement* of $U$ in $V$ is the set

$$U^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{u}, \mathbf{v}\rangle = 0 \text{ for all } \mathbf{u} \in U\}.$$

If $U = \operatorname{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ for some vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$, then we may show that

$$U^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{u}_1, \mathbf{v}\rangle = \cdots = \langle \mathbf{u}_k, \mathbf{v}\rangle = 0\}.$$

**Example.** Let $U = \operatorname{Span}(\mathbf{u}_1, \mathbf{u}_2) \subseteq \mathbb{R}^4$ where $\mathbf{u}_1 = (1, 0, 1, 0)$ and $\mathbf{u}_2 = (1, 0, 0, -1)$. Then a vector $\mathbf{v} = (x_1, x_2, x_3, x_4) \in V$ lies in $U^\perp$ if and only if $\mathbf{u}_1 \cdot \mathbf{v} = \mathbf{u}_2 \cdot \mathbf{v} = 0$, i.e.,

$$\begin{aligned} x_1 + x_3 &= 0 \\ x_1 - x_4 &= 0 \end{aligned}$$

The solution to this system is $x_1 = x_4$ and $x_3 = -x_4$, with $x_2, x_4$ being free. Therefore, $U^\perp$ has basis $\{(0,1,0,0),(1,0,-1,1)\}$.

**Proposition 2.1.** *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in an inner product space* $(V, \langle\cdot,\cdot\rangle)$ *such that* $\langle \mathbf{u}, \mathbf{v}\rangle = 0$*, then* $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$*.*

*Proof.*

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

The middle two terms on the right are both zero by assumption, and the outer terms are $\|\mathbf{u}\|^2$ and $\|\mathbf{v}\|^2$. □

### Orthogonal and orthonormal bases

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. A basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $V$ is said to be *orthogonal* if $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = 0$ for all $i \neq j$. If, in addition, all of the $\mathbf{u}_i$ have norm 1, i.e., $\|\mathbf{u}_i\| = 1$, then the basis is said to be *orthonormal*. Thus, a basis is $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is orthonormal if and only if

$$\langle \mathbf{u}_i, \mathbf{u}_j\rangle = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{otherwise.} \end{cases}$$

**Example.** The standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ in $\mathbb{R}^n$ is an orthonormal basis when $\mathbb{R}^n$ has the standard inner product.

**Example.** Give $\mathcal{P}_2$ the inner product $\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$. An orthogonal basis for $\mathcal{P}_2$ with respect to this inner product is $\{p_1, p_2, p_3\}$ where

$$p_1 = 1, \quad p_2 = x, \quad p_3 = x^2 - \frac{2}{3}.$$

We verify that $\langle p_1, p_3\rangle = 0$ and leave the other calculations as an exercise:

$$\langle p_1, p_3\rangle = \langle 1, x^2 - \tfrac{2}{3}\rangle = (1 - \tfrac{2}{3}) - \tfrac{2}{3} + (1 - \tfrac{2}{3}) = 0.$$

Note that the basis $\{p_1, p_2, p_3\}$ is orthogonal but not orthonormal.

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal basis for a given inner product space, then we may produce from it an orthonormal basis by scaling each $\mathbf{u}_i$ by $1/\|\mathbf{u}_i\|$. Thus, $\{\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_n\}$ is an orthonormal basis where

$$\hat{\mathbf{u}}_i = \frac{1}{\|\mathbf{u}_i\|}\mathbf{u}_i.$$

**Example.** Continuing with the previous example, we find that

$$\|1\| = \sqrt{3}, \quad \|x\| = \sqrt{2}, \quad \|x^2 - \tfrac{2}{3}\| = \sqrt{\tfrac{2}{3}},$$

so

$$\left\{ \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{2}}x, \sqrt{\tfrac{3}{2}}(x^2 - \tfrac{2}{3}) \right\}$$

is an orthonormal basis for $\mathcal{P}_2$ (with respect to the inner product given above).

## IV − 3 The Gram–Schmidt process and orthogonal projection

We study a procedure for producing an orthogonal basis for a finite-dimensional inner product space $(V, \langle \cdot, \cdot \rangle)$. Called the Gram–Schmidt process, it begins with any basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for $V$, and successively replaces each $\mathbf{v}_i$ with another vector $\mathbf{u}_i$ such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal basis.

The $\mathbf{u}_j$ are defined in terms of the $\mathbf{v}_i$ as follows:

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1 \\
\mathbf{u}_2 &= \mathbf{v}_2 - \frac{\langle \mathbf{u}_1, \mathbf{v}_2 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 \\
\mathbf{u}_3 &= \mathbf{v}_3 - \frac{\langle \mathbf{u}_1, \mathbf{v}_3 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 - \frac{\langle \mathbf{u}_2, \mathbf{v}_3 \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle} \mathbf{u}_2 \\
&\vdots \\
\mathbf{u}_n &= \mathbf{v}_n - \frac{\langle \mathbf{u}_1, \mathbf{v}_n \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 - \frac{\langle \mathbf{u}_2, \mathbf{v}_n \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle} \mathbf{u}_2 - \cdots - \frac{\langle \mathbf{u}_{n-1}, \mathbf{v}_n \rangle}{\langle \mathbf{u}_{n-1}, \mathbf{u}_{n-1} \rangle} \mathbf{u}_{n-1}
\end{aligned}$$

(Can you see why each $\mathbf{u}_k$ is orthogonal to all the $\mathbf{u}_j$ before it?)

**Example.** Give $\mathcal{P}_2$ the inner product $\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(2)q(2)$. Note that this is a different inner product from the examples in Section 2. Starting with the basis $\{q_1, q_2, q_3\}$ for $\mathcal{P}_2$, where

$$q_1 = 1, \quad q_2 = x, \quad q_3 = x^2,$$

we use the Gram–Schmidt process to produce an orthogonal basis $\{p_1, p_2, p_3\}$ for $\mathcal{P}_2$:

$$\begin{aligned}
&p_1 = q_1 = 1, \\
&p_2 = q_2 - \frac{\langle p_1, q_2 \rangle}{\langle p_1, p_1 \rangle} p_1 = x - \frac{3}{3} \cdot 1 = x - 1, \\
&p_3 = q_3 - \frac{\langle p_1, q_3 \rangle}{\langle p_1, p_1 \rangle} p_1 - \frac{\langle p_2, q_3 \rangle}{\langle p_2, p_2 \rangle} p_2 = x^2 - \frac{5}{3} \cdot 1 - \frac{4}{2}(x-1) = x^2 - 2x + \frac{1}{3}.
\end{aligned}$$

The basis $\{p_1, p_2, p_3\}$ is orthogonal but not orthonormal. The norms of the $p_i$ are

$$\|1\| = \sqrt{3}, \quad \|x - 1\| = \sqrt{2}, \quad \|x^2 - 2x + \tfrac{1}{3}\| = \sqrt{\tfrac{2}{3}},$$

so

$$\left\{ \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{2}}(x-1), \sqrt{\tfrac{3}{2}}(x^2 - 2x + \tfrac{1}{3}) \right\}$$

is an orthonormal basis for $\mathcal{P}_2$ (with respect to the inner product just given).

**Remark.** In the Gram–Schmidt process, having found a given $\mathbf{u}_k$, it is permissible to scale it by a non-zero scalar before moving on to the computation of $\mathbf{u}_{i+1}$. Doing so can make the subsequent computations easier.

### Orthogonal projection

Let $U$ be a finite-dimensional subspace of an inner product space $(V, \langle \cdot, \cdot \rangle)$. The Gram–Schmidt process ensures that $U$ has an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. Then for $\mathbf{v} \in V$, we define

$$\mathrm{proj}_U(\mathbf{v}) = \frac{\langle \mathbf{u}_1, \mathbf{v}\rangle}{\langle \mathbf{u}_1, \mathbf{u}_1\rangle}\,\mathbf{u}_1 + \cdots + \frac{\langle \mathbf{u}_k, \mathbf{v}\rangle}{\langle \mathbf{u}_k, \mathbf{u}_k\rangle}\,\mathbf{u}_k \in U.$$

It appears as though $\mathrm{proj}_U(\mathbf{v})$ depends on the choice of orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$, but in fact it does not; it depends only on the subspace $U$ and the vector $\mathbf{v}$, as we show in Section 10 in the Appendix. We call $\mathrm{proj}_U(\mathbf{v})$ the *orthogonal projection* of $\mathbf{v}$ onto $U$.

The orthogonal projection of a vector has the following important interpretation.

**Proposition 3.1.** *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and $U$ a finite-dimensional subspace of $V$. If $\mathbf{v} \in V$, then $\mathrm{proj}_U(\mathbf{v})$ is the unique vector $\mathbf{u} \in U$ that minimizes $\mathrm{dist}(\mathbf{v}, \mathbf{u})$.*

See Section 10 of the Appendix for a proof.

**Example.** Consider the space $V = C[-\pi, \pi]$ and the subspace $U$ of $V$ spanned by the functions $1, \cos, \sin$. We compute $\mathrm{proj}_U(f)$, where $f(x) = x^2 + x$. For this, we use an orthogonal basis for $U$. In fact, $\{1, \cos, \sin\}$ is already one. (We saw in Section 2 that cos and sin are orthogonal in $C[-\pi, \pi]$, and we leave it as a short exercise to show that $\langle 1, \cos\rangle = \langle 1, \sin\rangle = 0$ as well.) Hence,

$$\mathrm{proj}_U(f) = \frac{\langle 1, f\rangle}{\langle 1, 1\rangle}\cdot 1 + \frac{\langle \cos, f\rangle}{\langle \cos, \cos\rangle}\cos + \frac{\langle \sin, f\rangle}{\langle \sin, \sin\rangle}\sin.$$

Each of the inner products in this expression is an integral. For example,

$$\langle \cos, f\rangle = \int_{-\pi}^{\pi} \cos(x) f(x)\,dx = \int_{-\pi}^{\pi} \cos(x)(x^2 + x)\,dx = -4\pi. \tag{3.1}$$

(If you have not seen techniques for evaluating an integral such as the one in (3.1), then do not worry. I am just illustrating a point.) Once we have evaluated all the relevant integrals, we find that

$$\mathrm{proj}_U(f) = \frac{\pi^2}{3} - 4\cos + 2\sin.$$

According to Proposition 3.1, then, $\frac{\pi^2}{3} - 4\cos + 2\sin$ is the "closest" function in $\mathrm{Span}(1, \cos, \sin)$ to the function $f(x) = x^2 + x$. If you plot the two functions, you will see that they are in fact quite similar on the interval $[-\pi, \pi]$. This is the very beginning of a very big subject called *Fourier analysis*. We could obtain better approximations of $f$ by projecting it not onto $\mathrm{Span}(1, \cos(x), \sin(x))$ but instead onto

$$\mathrm{Span}(1, \cos(x), \sin(x), \cos^2(x), \sin^2(x), \cos^3(x), \sin^3(x), \ldots, \cos^n(x), \sin^n(x))$$

for some large integer $n$. Fourier analysis tells us how to compute these projections. The subject has important applications in, for example, signal processing.

## IV – 4 $QR$-factorization

Suppose $A \in M_{m,n}(\mathbb{R})$ has linearly independent columns (so necessarily $m \geq n$). A *$QR$-factorization* of $A$ is a factorization $A = QR$ where $Q \in M_{m,n}(\mathbb{R})$ has orthonormal columns and $R \in M_n(\mathbb{R})$ is upper triangular. In fact, the diagonal entries of $R$ may be chosen to be positive, and in this case the factorization is unique and is called the *totally positive* $QR$-factorization of $A$. All of this is proven in Section 11 of the Appendix.

### Method to obtain the totally positive $QR$-factorization of $A$

Assume that $A$ has linearly independent columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ (in that order). To find the totally positive $QR$-factorization of $A$, do the following.

(i) Apply the Gram–Schmidt process to $\mathbf{v}_1, \ldots, \mathbf{v}_n$, obtaining an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ for $\mathrm{Col}(A)$. Make sure that, if any of the $\mathbf{u}_i$ are scaled along the way, they are scaled only by positive scalars.

(ii) Let $\mathbf{w}_i = \frac{1}{\|\mathbf{u}_i\|}\mathbf{u}_i$ for $i = 1, \ldots, n$.

(iii) Express each $\mathbf{v}_j$ as $\mathbf{v}_j = r_{1,j}\mathbf{w}_1 + r_{2,j}\mathbf{w}_2 + \cdots + r_{j,j}\mathbf{w}_j$. This is possible because of how the Gram–Schmidt process works. You should find that $r_{j,j} > 0$.

(iv) Let $Q = \begin{pmatrix}\mathbf{w}_1 & \cdots & \mathbf{w}_n\end{pmatrix}$ and $R = (r_{i,j})$, where $r_{i,j} = 0$ if $i > j$. Then $A = QR$ is the totally positive $QR$-factorization of $A$.

**Example.** Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix},$$

which you may verify has linearly independent columns. Let the columns be $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, and define

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \\
\mathbf{u}_2 &= \mathbf{v}_2 - \frac{\mathbf{u}_1 \cdot \mathbf{v}_2}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix} - \frac{3}{3}\begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 1 \\ 1 \end{pmatrix} \\
\mathbf{u}_3 &= \mathbf{v}_3 - \frac{\mathbf{u}_1 \cdot \mathbf{v}_3}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_3}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 \\
&= \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} - \frac{4}{3}\begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} - \frac{-1}{3}\begin{pmatrix} 0 \\ -1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix}
\end{aligned}$$

The basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for $\mathrm{Col}(A)$ is orthogonal but not orthonormal, so we scale these basis vectors by positive scalars to obtain the orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, where

$$\mathbf{w}_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\0\\1\end{pmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{3}}\begin{pmatrix}0\\-1\\1\\1\end{pmatrix}, \quad \mathbf{w}_3 = \frac{1}{\sqrt{3}}\begin{pmatrix}-1\\1\\1\\0\end{pmatrix}.$$

Now, rearranging the equations for $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ on the previous page, we obtain

$$\mathbf{v}_1 = \begin{pmatrix}1\\1\\0\\1\end{pmatrix} = \sqrt{3}\,\mathbf{w}_1$$

$$\mathbf{v}_2 = \begin{pmatrix}1\\1\\0\\1\end{pmatrix} + \begin{pmatrix}0\\-1\\1\\1\end{pmatrix} = \sqrt{3}\,\mathbf{w}_1 + \sqrt{3}\,\mathbf{w}_2$$

$$\mathbf{v}_3 = \frac{4}{3}\begin{pmatrix}1\\1\\0\\1\end{pmatrix} - \frac{1}{3}\begin{pmatrix}0\\-1\\1\\1\end{pmatrix} + \frac{1}{3}\begin{pmatrix}-1\\1\\1\\0\end{pmatrix} = \frac{4\sqrt{3}}{3}\mathbf{w}_1 - \frac{\sqrt{3}}{3}\mathbf{w}_2 + \frac{\sqrt{3}}{3}\mathbf{w}_3$$

(These steps will be explained in more detail in class.) Hence, letting

$$Q = \begin{pmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3\end{pmatrix} = \begin{pmatrix}1/\sqrt{3} & 0 & -1/\sqrt{3}\\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3}\\ 0 & 1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{3} & 1/\sqrt{3} & 0\end{pmatrix}$$

$$R = \begin{pmatrix}\sqrt{3} & \sqrt{3} & 4\sqrt{3}/3\\ 0 & \sqrt{3} & -\sqrt{3}/3\\ 0 & 0 & \sqrt{3}/3\end{pmatrix},$$

we obtain the totally positive $QR$-factorization $A = QR$.

## IV – 5 The Cauchy–Schwarz and triangle inequalities

**Theorem 5.1** (Cauchy–Schwarz)**.** *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. If $\mathbf{u}, \mathbf{v} \in V$, then $|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$, with equality holding if and only if one of the vectors is a scalar multiple of the other.*

For a proof, see Section 12 of the Appendix.

**Example.** Let

$$\mathbf{u} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix},$$

and endow $\mathbb{R}^3$ with the standard inner product (the dot product). Then

$$\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}\cdot\mathbf{v} = -2+6+2 = 6, \quad \|\mathbf{u}\| = \sqrt{1+9+4} = \sqrt{14}, \quad \|\mathbf{v}\| = \sqrt{4+4+1} = 3,$$

so indeed $|\langle \mathbf{u}, \mathbf{v}\rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$.

**Example.** Let $b > 0$. The integral

$$\int_0^b \frac{1}{(1+x+\sin(x))^2}\,dx$$

has no simple expression in terms of $b$, but we can find a lower bound for it using the Cauchy–Schwarz inequality. We work in the inner product space $C[0, b]$. Define functions $f, g \in C[0, b]$ by

$$f(x) = \frac{1}{1+x+\sin(x)} \quad \text{and} \quad g(x) = 1+\cos(x).$$

The Cauchy–Schwarz inequality applied to the functions $f, g$ says $|\langle f, g\rangle| \le \|f\|\,\|g\|$, so $\langle f, g\rangle^2 \le \|f\|^2\|g\|^2$. Hence,

$$\begin{aligned} \int_0^b \frac{1}{(1+x+\sin(x))^2}\,dx &= \|f\|^2 \quad \text{by definition of the norm} \\ &\ge \frac{\langle f, g\rangle^2}{\|g\|^2} \quad \text{(Cauchy–Schwarz inequality rearranged)} \\ &= \frac{(\int_0^b f(x)g(x)\,dx)^2}{\int_0^b g(x)^2\,dx} \\ &= \frac{(\int_0^b \frac{1+\cos(x)}{1+x+\sin(x)}\,dx)^2}{\int_0^b (1+\cos(x))^2\,dx}. \qquad (5.1) \end{aligned}$$

The two integrals appearing in (5.1) may be evaluated easily:

$$\int_0^b \frac{1+\cos(x)}{1+x+\sin(x)}\,dx = [\ln(1+x+\sin(x))]_0^b = \ln(1+b+\sin(b)),$$

$$\int_0^b (1+\cos(x))^2\,dx = \int_0^b (1+2\cos(x)+\cos^2(x))\,dx = \frac{3}{2}b + 2\sin(b) + \frac{1}{4}\sin(2b).$$

Therefore,

$$\int_0^b \frac{1}{(1+x+\sin(x))^2}\,dx \geq \frac{\big(\ln(1+b+\sin(b))\big)^2}{\frac{3}{2}b+2\sin(b)+\frac{1}{4}\sin(2b)}.$$

### Angle

The Cauchy–Schwarz inequality allows us to define the *angle* between two non-zero vectors $\mathbf{u}, \mathbf{v}$ in an inner product space. This angle is the unique $\theta \in [0, \pi]$ such that

$$\cos(\theta) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|}. \tag{5.2}$$

This definition makes sense, because the Cauchy–Schwarz inequality guarantees that the right-hand side of (5.2) lies in the interval $[-1, 1]$, so there is indeed a $\theta$ satisfying the equality.

In the case of $\mathbb{R}^2$ and $\mathbb{R}^3$ endowed with the standard inner product, the angle defined as above is the same as our usual notion of angle in these spaces.

### Interpretation of orthogonality in terms of angle

Suppose that two non-zero vectors $\mathbf{u}, \mathbf{v}$ in an inner product space are orthogonal. Then $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, so the angle $\theta$ between them satisfies $\cos(\theta) = 0$. Thus, $\theta = \pi/2$ (because $\theta \in [0, \pi]$ by assumption). That is to say, the angle between two non-zero orthogonal vectors is $\pi/2$, which fits with our intuitive notion of angle. Conversely, if the angle between two non-zero vectors is $\pi/2$, then their inner product is zero, i.e., they are orthogonal.

### The triangle inequality

The triangle inequality for an inner product space $(V, \langle \cdot, \cdot \rangle)$ states that if $\mathbf{u}, \mathbf{v} \in V$, then $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$. We may prove this using the Cauchy–Schwarz inequality:

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&\leq \langle \mathbf{u}, \mathbf{u} \rangle + 2|\langle \mathbf{u}, \mathbf{v} \rangle| + \langle \mathbf{v}, \mathbf{v} \rangle \\
&\leq \langle \mathbf{u}, \mathbf{u} \rangle + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \langle \mathbf{v}, \mathbf{v} \rangle \quad \text{(Cauchy–Schwarz inequality)} \\
&= \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2.
\end{aligned}$$

Now take square roots.

To see why the triangle inequality is so called, take the case of $\mathbb{R}^2$ (with the standard inner product), and consider the triangle formed by the points $\mathbf{0}$, $\mathbf{u}$, and $\mathbf{u}+\mathbf{v}$. What are the numbers $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{u}+\mathbf{v}\|$ in relation to this triangle?

## IV – 6 An application of Cauchy–Schwarz to constrained optimization

Consider the following problem, a type of *constrained optimization* problem:

> Find the maximum value of $3x_1 - 2x_2 + 4x_3$ subject to $x_1^2 + x_2^2 + x_3^2 = 1$, and find the values of $x_1, x_2, x_3$ where the maximum is attained.

While this looks like a problem of calculus, it has a natural solution in linear algebra via the Cauchy–Schwarz inequality. Let

$$\mathbf{v} = \begin{pmatrix} 3 \\ -2 \\ 4 \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

The given problem is to find the maximum value of $\mathbf{v} \cdot \mathbf{x}$ subject to $\|\mathbf{x}\|^2 = 1$, i.e., $\|\mathbf{x}\| = 1$, and to find where the maximum occurs. Now,

$$\begin{aligned} \mathbf{v} \cdot \mathbf{x} &\leq |\mathbf{v} \cdot \mathbf{x}| \\ &\leq \|\mathbf{v}\| \, \|\mathbf{x}\| \quad \text{(Cauchy–Schwarz inequality)} \\ &= \|\mathbf{v}\| \quad \text{because } \|\mathbf{x}\| = 1 \\ &= \sqrt{3^2 + (-2)^2 + 4^2} \\ &= \sqrt{29}, \end{aligned}$$

with equality holding if and only if $\mathbf{x}$ is a *positive* scalar multiple of $\mathbf{v}$. (If $\mathbf{x}$ were a negative scalar times $\mathbf{v}$, then $\mathbf{v} \cdot \mathbf{x}$ would equal $-\sqrt{29}$, not $\sqrt{29}$.) To find the $\mathbf{x}$ at which the maximum is attained, we use the two facts that $\|\mathbf{x}\| = 1$ and $\mathbf{x} = c\mathbf{v}$ for some $c > 0$. Thus,

$$1 = \|\mathbf{x}\| = \|c\mathbf{v}\| = |c|\|\mathbf{v}\| = c\|\mathbf{v}\| = c\sqrt{29},$$

so $c = 1/\sqrt{29}$. The maximum of $\sqrt{29}$ therefore occurs at

$$\mathbf{x} = \frac{1}{\sqrt{29}} \begin{pmatrix} 3 \\ -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{29} \\ -2/\sqrt{29} \\ 4/\sqrt{29} \end{pmatrix}.$$

Next we consider a related, but slightly different problem:

> Find the minimum value of $x_1^2+x_2^2+x_3^2$ subject to the constraint $3x_1+2x_2+2x_3 = 5$, and find the values of $x_1, x_2, x_3$ where the minimum is attained.

Let

$$\mathbf{v} = \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

We are to find the minimum value of $\|\mathbf{x}\|^2$ subject to $\mathbf{v} \cdot \mathbf{x} = 5$, and we must also find where the minimum occurs. By Cauchy–Schwarz,

$$\|\mathbf{x}\|^2 \geq \frac{|\mathbf{v} \cdot \mathbf{x}|^2}{\|\mathbf{v}\|^2} = \frac{25}{17},$$

with equality holding if and only if $\mathbf{x} \in \mathrm{Span}(\mathbf{v})$. To find this $\mathbf{x}$, we use the facts that $\mathbf{v} \cdot \mathbf{x} = 5$ and $\mathbf{x} = c\mathbf{v}$ for some $c \in \mathbb{R}$:

$$5 = \mathbf{v} \cdot \mathbf{x} = \mathbf{v} \cdot (c\mathbf{v}) = c\|\mathbf{v}\|^2 = 17c,$$

so $c = 5/17$. The minimum of $25/17$ therefore occurs at

$$\mathbf{x} = \frac{5}{17}\begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 15/17 \\ 10/17 \\ 10/17 \end{pmatrix}.$$

## IV – 7 Orthogonal diagonalization of real symmetric matrices

A matrix $P \in M_n(\mathbb{R})$ is called *orthogonal* if $P^{\mathrm{T}}P = I$. In other words, $P$ is orthogonal if it is invertible and $P^{\mathrm{T}} = P^{-1}$. The rotations and reflections in $\mathbb{R}^2$ and $\mathbb{R}^3$ are all described by orthogonal matrices.

Orthogonal matrices are characterized by the property that they have orthonormal columns and orthonormal rows. For example, the matrix

$$\begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix}$$

is orthogonal. Note how its columns are orthonormal, as are its rows.

Recall that a matrix $A \in M_n(\mathbb{R})$ is called *symmetric* if $A^{\mathrm{T}} = A$. Unlike a general square matrix, **real symmetric matrices are always diagonalizable**. This would be interesting by itself, but we can say more: A real symmetric matrix $A$ can be diagonalized by an orthogonal matrix $P$, that is, there is an orthogonal matrix $P$ such that $P^{-1}AP$ is diagonal. Because $P^{-1} = P^{\mathrm{T}}$ for an orthogonal matrix $P$, we usually write $P^{\mathrm{T}}AP$ instead of $P^{-1}AP$. Conversely, $PDP^{\mathrm{T}}$ is symmetric when $D$ is diagonal, and we have:

**Theorem 7.1.** *Let $A \in M_n(\mathbb{R})$. The following are equivalent:*

*(i) $A$ is symmetric.*

*(ii) $A$ is orthogonally diagonalizable, i.e., there is an orthogonal matrix $P$ such that $P^{\mathrm{T}}AP$ is diagonal.*

For a proof, see Section 13 of the Appendix.

### Method to orthogonally diagonalize a real symmetric matrix

Let $A \in M_n(\mathbb{R})$ be symmetric. To orthogonally diagonalize it:

(i) Find the eigenvalues of $A$. (They will all be real; see Section 13 of the Appendix.)

(ii) Find a basis for each eigenspace.

(iii) Apply Gram–Schmidt to each basis found in (ii) to obtain an orthonormal basis for each eigenspace.

(iv) Let $P$ be a matrix whose columns are the $n$ basis vectors found in (iii). Then $P$ is orthogonal. (See Proposition 13.2 in the Appendix.)

(v) Let $D$ be the diagonal matrix whose $i$th diagonal entry is the eigenvalue corresponding to the $i$th column of $P$. Then $P^{\mathrm{T}}AP = D$.

**Example.** Orthogonally diagonalize the real symmetric matrix

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

*Solution:* The characteristic polynomial is $p_A(x) = (x+2)(x-7)^2$. (For a $3 \times 3$ matrix, the characteristic polynomial will usually be given to you in this course.) Row reducing $-2I - A$, we find that the eigenspace associated to $-2$ has basis $\{(-2, -1, 2)\}$. Because this is a one-dimensional eigenspace, finding an orthonormal basis entails simply scaling this basis vector to have norm 1. Thus, an orthonormal basis is $\{(-2/3, -1/3, 2/3)\}$.

Now we turn to the eigenspace associated to 7. Again, row reducing $7I - A$, we find that the eigenspace has basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ where

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

We apply Gram–Schmidt to the vectors $\mathbf{v}_1, \mathbf{v}_2$, letting

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{v}_1 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \\ \mathbf{u}_2 &= \mathbf{v}_2 - \frac{\mathbf{u}_1 \cdot \mathbf{v}_2}{\mathbf{u}_1 \cdot \mathbf{u}_1}\,\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{-1}{5}\begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix}. \end{aligned}$$

The basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ is orthogonal but not orthonormal. Scaling, we obtain the orthonormal basis

$$\left\{ \begin{pmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{pmatrix}, \begin{pmatrix} 4/\sqrt{45} \\ 2/\sqrt{45} \\ 5/\sqrt{45} \end{pmatrix} \right\}.$$

Hence, if

$$P = \begin{pmatrix} -2/3 & -1/\sqrt{5} & 4/\sqrt{45} \\ -1/3 & 2/\sqrt{5} & 2/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{pmatrix},$$

then $P$ is orthogonal and $P^{\mathrm{T}}AP = D$.

## IV – 8 Quadratic forms

A *quadratic form* in $n$ variables is a function

$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R} \\ \mathbf{x} &\mapsto \mathbf{x}^{\mathrm{T}} A \mathbf{x}, \end{aligned}$$

where $A$ is real symmetric matrix. (There is no advantage in allowing a general matrix $A \in M_n(\mathbb{R})$, because the quadratic form associated to any $A \in M_n(\mathbb{R})$ is the same as the quadratic form associated to the symmetric matrix $\frac{1}{2}(A + A^{\mathrm{T}})$.)

**Example.** If $A = \begin{pmatrix} 1 & 5 \\ 5 & -4 \end{pmatrix}$, the associated quadratic form is

$$\begin{aligned} \mathbf{x}^{\mathrm{T}} A \mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 5 \\ 5 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= x_1^2 + 5x_1x_2 + 5x_2x_1 - 4x_2^2 \\ &= x_1^2 + 10x_1x_2 - 4x_2^2. \end{aligned}$$

We may pass back and forth between real symmetric matrices and quadratic forms as follows:

- If $A = (a_{i,j}) \in M_n(\mathbb{R})$ is symmetric, then

$$\mathbf{x}^{\mathrm{T}} A \mathbf{x} = \sum_{i=1}^{n} a_{i,i} x_i^2 + \sum_{i<j} 2a_{i,j} x_i x_j.$$

- The real symmetric matrix associated to the $n$-variable quadratic form

$$\sum_{i=1}^{n} b_{i,i} x_i^2 + \sum_{i<j} b_{i,j} x_i x_j$$

  is $A = (a_{i,j})$ where

$$a_{i,j} = \begin{cases} b_{i,i} & \text{if } i = j \\ \frac{1}{2} b_{i,j} & \text{if } i < j \\ \frac{1}{2} b_{j,i} & \text{if } i > j. \end{cases}$$

**Example.** The real symmetric matrix associated to the quadratic form

$$3x_1^2 - 4x_3^2 + x_1x_2 - 2x_1x_3 - 5x_2x_4$$

is

$$\begin{pmatrix} 3 & 1/2 & -1 & 0 \\ 1/2 & 0 & 0 & -5/2 \\ -1 & 0 & -4 & 0 \\ 0 & -5/2 & 0 & 0 \end{pmatrix}.$$

An important first problem in the study of quadratic forms is to determine whether, for example, a given quadratic form may take negative values. For example, are there real numbers $x_1, x_2$ such that

$$5x_1^2 - 14x_1x_2 + 10x_2^2 < 0? \tag{8.1}$$

At first glance, there appears to be no reason why not. Yet substituting some values for $x_1, x_2$, we find that the left-hand side of (8.1) always appears to be positive (except when $x_1 = x_2 = 0$).

We will investigate this problem further in Section 9. First, we introduce some terminology.

A quadratic form $f : \mathbb{R}^n \to \mathbb{R}$ is called

- *positive definite* if $f(\mathbf{x}) > 0$ for all non-zero $\mathbf{x} \in \mathbb{R}^n$
- *negative definite* if $f(\mathbf{x}) < 0$ for all non-zero $\mathbf{x} \in \mathbb{R}^n$
- *non-negative definite* if $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$
- *non-positive definite* if $f(\mathbf{x}) \leq 0$ for all $\mathbf{x} \in \mathbb{R}^n$
- *indefinite* if $f(\mathbf{x})$ takes both positive and negative values

For example, the quadratic forms $x_1^2 + 4x_2^2$ and $3x_1^2 + 6x_2^2 + 5x_3^2$ are obviously both positive definite, while $x_1^2 - 7x_2^2$ is obviously indefinite. The situation becomes harder to decide when there are cross terms, as in our example $5x_1^2 - 14x_1x_2 + 10x_2^2$ earlier, though we will see that in fact it is positive definite.

Here are a couple more to think about:

$$\begin{aligned} f(x_1, x_2) &= 7x_1^2 + 12x_1x_2 + 5x_2^2 \\ g(x_1, x_2) &= 13x_1^2 - 18x_1x_2 + 10x_2^2 \end{aligned}$$

One of these two quadratic forms is positive definite, and the other is indefinite. Which is which? (I recommend you wait until the next section before spending too long on this!)

## IV – 9 Diagonalizing quadratic forms

As you may have appreciated from the previous section, the presence of cross terms in a quadratic form $f$ can make it difficult to decide whether $f$ is positive definite, negative definite, and so on. A neat way around this problem is to make a change of variables such that, in the new variables, there are no cross terms. This idea leads us to the notion of *diagonalization of a quadratic form.*

**Definition 9.1.** *To* orthogonally diagonalize *a quadratic form $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x}$ is to find an orthogonal matrix $P$ such that the quadratic form $g(\mathbf{y}) = f(P\mathbf{y})$ has no cross terms.*

This amounts to orthogonally diagonalizing the real symmetric matrix $A$, for if $P^{\mathrm{T}}AP = D$ with $D$ diagonal, then the quadratic form $g(\mathbf{y}) = f(P\mathbf{y})$ satisfies

$$\begin{aligned} g(\mathbf{y}) &= (P\mathbf{y})^{\mathrm{T}}A(P\mathbf{y}) \quad \text{because } f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x} \\ &= \mathbf{y}^{\mathrm{T}}P^{\mathrm{T}}AP\mathbf{y} \\ &= \mathbf{y}^{\mathrm{T}}D\mathbf{y} \\ &= \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2, \end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ are the diagonal entries of $D$, and $\mathbf{y} = (y_1, \ldots, y_n)$.

**Example.** Consider the quadratic form $f(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 2x_2^2$. It is given by $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x}$ where

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}.$$

By the method of Section 7, we find that $P^{\mathrm{T}}AP = D$ where

$$P = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}.$$

Hence, the quadratic form $g(\mathbf{y}) = f(P\mathbf{y})$ satisfies

$$g(\mathbf{y}) = \mathbf{y}^{\mathrm{T}}D\mathbf{y} = \begin{pmatrix} y_1 & y_2 \end{pmatrix}\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 6y_1^2 + y_2^2.$$

We see, therefore, that $g(\mathbf{y})$ is positive for all non-zero $\mathbf{y}$, so $f(\mathbf{x})$ is positive for all non-zero $\mathbf{x}$. Thus, $f$ is positive definite.

We may even re-express $f$ in terms of $g$. If we take $\mathbf{y} = P^{\mathrm{T}}\mathbf{x}$ in the equation $g(\mathbf{y}) = f(P\mathbf{y})$ and remember that $PP^{\mathrm{T}} = I$, we obtain

$$f(\mathbf{x}) = g(P^{\mathrm{T}}\mathbf{x}). \tag{9.1}$$

But

$$P^{\mathrm{T}}\mathbf{x} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}}(2x_1 + x_2) \\ \frac{1}{\sqrt{5}}(-x_1 + 2x_2) \end{pmatrix},$$

so (9.1) says

$$\begin{aligned} f(\mathbf{x}) &= 6\Big(\tfrac{1}{\sqrt{5}}(2x_1+x_2)\Big)^2 + \Big(\tfrac{1}{\sqrt{5}}(-x_1+2x_2)\Big)^2 \\ &= \frac{6}{5}(2x_1+x_2)^2 + \frac{1}{5}(-x_1+2x_2)^2. \end{aligned} \tag{9.2}$$

If you would like to appreciate further what we have achieved here, expand out the expression in (9.2). You will find that you get back $5x_1^2 + 4x_1x_2 + 2x_2^2$.

### Quadratic forms and eigenvalues

Let $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x}$ be a quadratic form, where, as usual, $A \in M_n(\mathbb{R})$ is a real symmetric matrix. We know that there exist $P$ orthogonal and $D$ diagonal such that $P^{\mathrm{T}}AP = D$. The diagonal entries of $D$ are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A$ (possibly repeated). So, if all the eigenvalues are positive, then the diagonalized quadratic form $g(\mathbf{y}) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$ will be positive definite. If all the eigenvalues are negative, then of course $g$ will be negative definite. And if there is a mix of positive and negative eigenvalues, then the $g$ will be indefinite. But whatever of these properties $g$ has, the original quadratic form $f$ has as well, because $f$ and $g$ are related simply by a change of variables. We summarize our discussion in the following proposition:

**Proposition 9.2.** *Let $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x}$ be a quadratic form, where $A$ is real symmetric. Then $f$ is*

- *positive definite if and only if all the eigenvalues of $A$ are positive,*
- *negative definite if and only if all the eigenvalues of $A$ are negative,*
- *non-negative definite if and only if all the eigenvalues of $A$ are non-negative,*
- *non-positive definite if and only if all the eigenvalues of $A$ are non-positive,*
- *indefinite if and only if $A$ has a mix of positive and negative eigenvalues.*

**Example.** The quadratic form $f(x_1, x_2) = 5x_1^2 - 14x_1x_2 + 10x_2^2$ from Section 8 has associated real symmetric matrix $A = \begin{pmatrix} 5 & -7 \\ -7 & 10 \end{pmatrix}$. The characteristic polynomial of $A$ is $p_A(x) = x^2 - 15x + 1$, whose roots are

$$\frac{1}{2}(15 \pm \sqrt{15^2 - 4}).$$

Both of these roots are positive, since obviously $\sqrt{15^2-4} < 15$. Therefore, by Proposition 9.2, $f$ is positive definite.

**Exercise.** For the quadratic form $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3$, decide which of the following properties it has: positive definite, negative definite, etc.

## IV – 10 Constrained optimization of quadratic forms

Recall that we saw a couple of constrained-optimization problems in Section 6. Now we consider a new kind of constrained-optimization problem:

> Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form. Find the maximum (or minimum) value of $f(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\| = 1$, and determine where the maximum (or minimum) occurs.

The problem is answered by the following proposition.

**Proposition 10.1.** *Suppose $A \in M_n(\mathbb{R})$ is symmetric, and let $f$ be the associated quadratic form. Let the eigenvalues of $A$ be $\lambda_1 \geq \cdots \geq \lambda_n$ (some may be repeated).*

*(i) The maximum value of $f(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$ is $\lambda_1$ and occurs at any unit eigenvector with eigenvalue $\lambda_1$.*

*(ii) The minimum value of $f(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$ is $\lambda_n$ and occurs at any unit eigenvector with eigenvalue $\lambda_n$.*

For a proof, see Section 14 of the Appendix.

**Example.** Find the maximum value of the quadratic form

$$f(x_1, x_2, x_3) = 3x_1^2 + 6x_2^2 + 3x_3^2 - 4x_1x_2 + 8x_1x_3 + 4x_2x_3$$

subject to the constraint $x_1^2 + x_2^2 + x_3^2 = 1$, and find a triple $(x_1, x_2, x_3)$ where that maximum occurs.

*Solution:* The constraint $x_1^2 + x_2^2 + x_3^2 = 1$ is simply $\|\mathbf{x}\|^2 = 1$, i.e., $\|\mathbf{x}\| = 1$, so we may apply Proposition 10.1. The associated real symmetric matrix is

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

whose eigenvalues we found in Section 7; they are 7 and $-2$. Therefore, the maximum of $f(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$ is 7 and occurs at any unit eigenvector with eigenvalue 7. We found earlier that the eigenspace associated to 7 has orthonormal basis

$$\left\{ \begin{pmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{pmatrix}, \begin{pmatrix} 4/\sqrt{45} \\ 2/\sqrt{45} \\ 5/\sqrt{45} \end{pmatrix} \right\}, \tag{10.1}$$

so a unit eigenvector with eigenvalue 7 is $(x_1, x_2, x_3) = (-1/\sqrt{5}, 2/\sqrt{5}, 0)$, for example.

**Exercise.** In the above example, find *all* of the $\mathbf{x} \in \mathbb{R}^3$ where $f(\mathbf{x})$ takes the maximum value of 7. This amounts to finding all the unit eigenvectors of $A$ with eigenvalue 7. To do this, let $\mathbf{u}_1, \mathbf{u}_2$ be the basis vectors in (10.1), and determine how to choose $a_1, a_2 \in \mathbb{R}$ such that $a_1\mathbf{u}_1 + a_2\mathbf{u}_2$ has norm 1.

**Exercise.** Find the maximum and minimum values of $x_1x_2 + x_1x_3 + x_2x_3$ subject to the constraint $x_1^2 + x_2^2 + x_3^2 = 1$.

# Appendix

## Appendix: 1 Proof of Proposition 3.1 in Section I

We recall the statement to be proven:

> Let $U$ be a subset of a vector space $V$. Then $U$ is a subspace of $V$ if and only if all of the following hold:
>
> (i) $U$ is non-empty.
>
> (ii) For all $\mathbf{u}, \mathbf{v} \in U$, $\mathbf{u} + \mathbf{v} \in U$. (Closure under addition)
>
> (iii) For all $\mathbf{u} \in U$ and all $c \in \mathbb{R}$, $c\mathbf{u} \in U$. (Closure under scalar multiplication)

*Proof.* Assume that $U$ is a non-empty subset of $V$ that is closed under addition and scalar multiplication. The key axioms to prove about $U$ are that it contains the zero vector of $V$ and that it contains additive inverses. (The other axioms are more straightforward.)

So, choose any $\mathbf{u} \in U$, which is possible because $U$ is assumed to be non-empty. Then $(-1)\mathbf{u} \in U$ because $U$ is assumed to be closed under scalar multiplication. But $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$ in $V$, because $(-1)\mathbf{u} + \mathbf{u} = (-1+1)\mathbf{u} = 0\mathbf{u} = \mathbf{0}_V$. (It is a short exercise to show that $0\mathbf{u} = \mathbf{0}_V$.) Thus, $U$ contains additive inverses. But then, because $U$ is closed under addition and contains both $\mathbf{u}$ and $-\mathbf{u}$, it contains $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}_V$.

The converse is immediate: If $U$ is a subspace of $V$, and so is a vector space itself, then it is necessarily non-empty and closed under addition and scalar multiplication. □

## Appendix: 2 Some basics of linear algebra

**Lemma 2.1.** *Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be linearly independent vectors in a vector space $V$, and let $\mathbf{v}$ be any element of $V$. Then the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}$ are linearly independent if and only if $\mathbf{v} \notin \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$.*

*Proof.* Assume that $\mathbf{v} \notin \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, and suppose that

$$\sum_{i=1}^{k+1} a_i \mathbf{v}_i = \mathbf{0},$$

where $\mathbf{v}_{k+1} = \mathbf{v}$. If $a_{k+1} \neq 0$, then $\mathbf{v}$ would be in $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, contrary to our assumption, so $a_{k+1} = 0$. Hence,

$$\sum_{i=1}^{k} a_i \mathbf{v}_i = \mathbf{0},$$

so the remaining $a_i$ are all zero by linear independence of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. The converse is immediate. □

**Proposition 2.2.** *Let $S$ be a finite set of elements in a vector space $V$. Then given any linearly independent subset $S'$ of $S$, there is a basis $\mathcal{B}$ for* $\operatorname{Span}(S)$ *satisfying $S' \subseteq \mathcal{B} \subseteq S$.*

*Proof.* Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in S$ be linearly independent. If $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \neq \operatorname{Span}(S)$, then there is some $\mathbf{v}_{k+1} \in S \smallsetminus \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. By Lemma 2.1, the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}$ are linearly independent. We continue in this way. Because $S$ is finite, this process has to stop with some linearly independent set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ that spans $\operatorname{Span}(S)$, i.e., a basis for $\operatorname{Span}(S)$. □

**Theorem 2.3.** *Let $V$ be a vector space.*

*(i) If $\mathcal{B}, \mathcal{E}$ are bases of $V$, then for any $\mathbf{v} \in \mathcal{B}$, there is $\mathbf{w} \in \mathcal{E}$ such that $(\mathcal{B} \smallsetminus \{\mathbf{v}\}) \cup \{\mathbf{w}\}$ is a basis for $V$.*

*(ii) If $\mathcal{B}, \mathcal{E}$ are finite bases of $V$, then $\#\mathcal{B} = \#\mathcal{E}$.*

*Proof.* We prove the first assertion, from which the second follows. Let $\mathbf{v}_0 \in \mathcal{B}$. We may write

$$\mathbf{v}_0 = \sum_{\mathbf{w} \in T} b_{\mathbf{w}} \mathbf{w}$$

for some finite subset $T$ of $\mathcal{E}$. If every $\mathbf{w} \in T$ were in $\operatorname{Span}(\mathcal{B} \smallsetminus \{\mathbf{v}_0\})$, then so would $\mathbf{v}_0$ be, contradicting the linear independence of $\mathcal{B}$. Therefore, there is some $\mathbf{w}_0 \in T$ that is not in $\operatorname{Span}(\mathcal{B} \smallsetminus \{\mathbf{v}_0\})$. Let $\mathcal{C} = (\mathcal{B} \smallsetminus \{\mathbf{v}_0\}) \cup \{\mathbf{w}_0\}$. We show that $\mathcal{C}$ is a basis of $V$.

For linear independence, we are to show that every finite subset of $\mathcal{C}$ is linearly independent. Such a subset is either a finite subset of $\mathcal{B} \smallsetminus \{\mathbf{v}_0\}$, in which case we are done immediately, or a set of the form $(S \smallsetminus \{\mathbf{v}_0\}) \cup \{\mathbf{w}_0\}$ where $S$ is a finite subset of $\mathcal{B}$. But $\mathbf{w}_0 \notin \operatorname{Span}(S \smallsetminus \{\mathbf{v}_0\})$, so $(S \smallsetminus \{\mathbf{v}_0\}) \cup \{\mathbf{w}_0\}$ is linearly independent by Lemma 2.1.

Now we turn to spanning. Because $\mathcal{C} \cup \{\mathbf{v}_0\} = \mathcal{B} \cup \{\mathbf{w}_0\}$, which spans $V$, it is enough to show that $\mathbf{v}_0 \in \operatorname{Span}(\mathcal{C})$. To that end, write

$$\mathbf{w}_0 = a_0 \mathbf{v}_0 + \sum_{\mathbf{v} \in S} a_{\mathbf{v}} \mathbf{v}$$

for some finite subset $S$ of $\mathcal{B} \smallsetminus \{\mathbf{v}_0\}$, possible because $\mathcal{B}$ spans $V$. The linear independence of $\mathcal{B}$ implies that $a_0 \neq 0$, so we may rearrange the equation above to express $\mathbf{v}_0$ as a linear combination of $\mathbf{w}_0$ and vectors in $\mathcal{B} \smallsetminus \{\mathbf{v}_0\}$, as desired. □

**Proposition 2.4.** *Let $V$ be a vector space, and suppose that $V$ has a finite spanning set (so $V$ has a finite basis by Proposition 2.2).*

*(i) All bases of $V$ are finite and have the same cardinality. Let that cardinality be $n$.*

*(ii) Any linearly independent set in $V$ contains at most $n$ elements.*

*(iii) Any spanning set for $V$ contains at least $n$ elements.*

*Proof.* Without yet knowing that all bases of $V$ are finite, we nonetheless know from Theorem 2.3 that all finite bases have the same cardinality. Let that cardinality be $n$.

Now, let $T$ be a linearly independent set in $V$, let $S'$ be a finite subset of $T$, and let $S = S' \cup X$ where $X$ is some choice of finite spanning set for $V$. Proposition 2.2 says that there is a basis $\mathcal{B}$ for $\operatorname{Span}(S) = V$ such that $S' \subseteq \mathcal{B}$. Then $\#S' \leq \#\mathcal{B} = n$. Therefore, $T$ is finite of cardinality at most $n$.

Having shown that every linearly independent set in $V$ is finite, we now know that all bases of $V$ are indeed finite.

Finally, if $S$ were some spanning set for $V$ that contained fewer than $n$ elements, then by Proposition 2.2, $S$ would contain a basis $\mathcal{B}$ for $\operatorname{Span}(S) = V$. (Just apply the proposition with $S' = \{\mathbf{v}\}$ for some non-zero $\mathbf{v} \in S$.) But then $\mathcal{B}$ too would have cardinality less than $n$, contradicting (i). □

If $V$ is a vector space that has a finite spanning set, the *dimension* of $V$, denoted $\dim(V)$, is the number of elements in a basis for $V$. By Proposition 2.4, this number is independent of the choice of basis and is therefore well defined. If a vector space $V$ has a finite basis, then we call $V$ *finite dimensional*. Otherwise, we say that $V$ is *infinite dimensional*.

We note several useful corollaries of Proposition 2.4.

**Corollary 2.5.** *Let $V$ be a vector space of finite dimension $n$, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. (Note that the $n$ in $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is the same as the dimension of $V$; this is important.) Then the following are equivalent:*

*(i)* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *are linearly independent.*

*(ii)* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *span $V$.*

*(iii)* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *form a basis for $V$.*

*Proof.* (i) ⇒ (ii) If there were some $\mathbf{v} \in V \smallsetminus \operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, then by Lemma 2.1, $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}$ would be linearly independent, contradicting Proposition 2.4. Therefore, $\operatorname{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$.

(ii) ⇒ (iii) If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ were not linearly independent, then some $\mathbf{v}_i$ would be a linear combination of the others, and then $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_n$ would span $V$, contradicting Proposition 2.4. Thus, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent and therefore form a basis.

(iii) ⇒ (i) Immediate. □

**Corollary 2.6.** *A subspace of a finite-dimensional vector space is finite dimensional.*

*Proof.* Let $U$ be a subspace of a finite-dimensional vector space $V$. If $U$ is zero, then we are done. Otherwise, choose a non-zero vector $\mathbf{u}_1 \in U$. If $\mathbf{u}_1$ spans $U$, then stop. Otherwise, we may choose some $\mathbf{u}_2 \in U \smallsetminus \operatorname{Span}(\mathbf{u}_1)$, and then $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly independent by Lemma 2.1. If $\mathbf{u}_1, \mathbf{u}_2$ span $U$, then stop. Otherwise, we may choose some $\mathbf{u}_3 \in U \smallsetminus \operatorname{Span}(\mathbf{u}_1, \mathbf{u}_2)$. We continue in this way until we arrive at a spanning set for $U$, which we must do. Indeed, the vectors in the sequence $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots$ are linearly independent, so the sequence must contain no more than $\dim(V)$ vectors by Proposition 2.4. □

**Corollary 2.7.** *Let $W$ be a subspace of a finite-dimensional vector space $V$.*

*(i)* $\dim(W) \leq \dim(V)$.

*(ii)* $W = V$ *if and only if* $\dim(W) = \dim(V)$.

*Proof.* Note that $W$ has finite dimension by Corollary 2.6.

(i) Let $\mathcal{B}$ be a basis for $W$. Then, being a linearly independent set in $V$, it has cardinality at most $\dim(V)$ by Proposition 2.4. But by definition, the cardinality of $\mathcal{B}$ is $\dim(W)$.

(ii) If $W = V$, then obviously $\dim(W) = \dim(V)$. Suppose conversely that $\dim(W) = \dim(V)$. If $\mathcal{B}$ is a basis for $W$, then extend it to a basis $\mathcal{C}$ of $V$, so that $\mathcal{B} \subseteq \mathcal{C}$. Then

$$\begin{aligned} \#\mathcal{B} &= \dim(W) \\ &= \dim(V) \\ &= \#\mathcal{C}, \end{aligned}$$

so $\mathcal{B} = \mathcal{C}$. Thus, $\mathcal{B}$ is a basis for $V$ and therefore spans $V$, so $W = V$. □

## Appendix: 3 Proof of Proposition 10.2 in Section I

We recall the statement to be proven:

> Let $\mathcal{B}, \mathcal{C}, \mathcal{E}$ be bases for a finite-dimensional vector space $V$. Then $P_{\mathcal{E}\leftarrow\mathcal{B}} = P_{\mathcal{E}\leftarrow\mathcal{C}} P_{\mathcal{C}\leftarrow\mathcal{B}}$. In particular, $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible, and $P_{\mathcal{C}\leftarrow\mathcal{B}}^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$.

*Proof.* Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$. Then

$$\begin{aligned} P_{\mathcal{E}\leftarrow\mathcal{C}} P_{\mathcal{C}\leftarrow\mathcal{B}} &= P_{\mathcal{E}\leftarrow\mathcal{C}} \Big( [\mathbf{u}_1]_{\mathcal{C}} \quad \cdots \quad [\mathbf{u}_n]_{\mathcal{C}} \Big) \\ &= \Big( P_{\mathcal{E}\leftarrow\mathcal{C}}[\mathbf{u}_1]_{\mathcal{C}} \quad \cdots \quad P_{\mathcal{E}\leftarrow\mathcal{C}}[\mathbf{u}_n]_{\mathcal{C}} \Big) \\ &= \Big( [\mathbf{u}_1]_{\mathcal{E}} \quad \cdots \quad [\mathbf{u}_n]_{\mathcal{E}} \Big) \quad \text{by Proposition 10.1 in Section I} \\ &= P_{\mathcal{E}\leftarrow\mathcal{B}}. \end{aligned}$$

The last assertion in the proposition comes about by taking $\mathcal{E} = \mathcal{B}$ and noting that $P_{\mathcal{B}\leftarrow\mathcal{B}}$ is the identity matrix. □

## Appendix: 4 Proof of Proposition 10.3 in Section I

Before proving Proposition 10.3, which we will recall in a moment, let us introduce a useful concept. An *elementary matrix* is a matrix obtained by performing a single elementary row operation on an identity matrix. For example, the matrix

$$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 7 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \tag{4.1}$$

is an elementary matrix, because it is obtained from the $4 \times 4$ identity matrix via the row operation that adds 7 times the first row to the third row. Other examples of elementary matrices are

$$\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1\end{pmatrix}.$$

(Which row operation does each of them correspond to?) Note that every elementary matrix is invertible.

Let $E$ be an $m \times m$ elementary matrix, corresponding to some row operation $\rho$. Then if $A$ is any matrix with $m$ rows, $EA$ is the matrix obtained by performing the row operation $\rho$ on $A$. For example, if $E$ is the elementary matrix in (4.1) and $A$ is a matrix with 4 rows, then $EA$ is the matrix obtained from $A$ by adding 7 times the first row of $A$ to the third. We leave it as an exercise to prove this observation for all elementary matrices.

**Lemma 4.1.**

*(i) Every invertible matrix is a product of elementary matrices.*

*(ii) Two matrices $A, B \in M_{m,n}(\mathbb{R})$ are row equivalent if and only if there is an invertible $m \times m$ matrix $P$ such that $B = PA$.*

*Proof.* (i) Let $P$ be an invertible $m \times m$ matrix. Then the reduced row-echelon form of $P$ is $I_m$, the $m \times m$ identity matrix. Let $(\rho_1, \dots, \rho_k)$ be a sequence of row operations that transforms $I_m$ to $P$, and for each $i$ let $E_i$ be the $m \times m$ elementary matrix corresponding to $\rho_i$. Then

$$P = E_k E_{k-1} \cdots E_2 E_1 I = E_k E_{k-1} \cdots E_2 E_1.$$

(ii) Suppose that $A$ is row equivalent to $B$. Then there are elementary matrices $E_1, \dots, E_k$ such that $B = E_k E_{k-1} \cdots E_2 E_1 A$, i.e., $B = PA$ where $P = E_k \cdots E_1$. Conversely, if $B = PA$ where $P$ is invertible, then by (i), $P = E_k \cdots E_1$ for some elementary matrices $E_i$, so $B = E_k \cdots E_k A$, which is row equivalent to $A$. □

A nice application of elementary matrices is a quick proof of Proposition 10.3 in Section I. Here again is the statement to be proven:

> Let $\mathcal{B}$ and $\mathcal{C}$ be bases for an $n$-dimensional vector space $V$, and let $\mathcal{E}$ be another basis for $V$. Then the reduced row-echelon form of
>
> $$\left(P_{\mathcal{E}\leftarrow\mathcal{C}} \;\middle|\; P_{\mathcal{E}\leftarrow\mathcal{B}}\right)$$
>
> is
>
> $$\left(I_n \;\middle|\; P_{\mathcal{C}\leftarrow\mathcal{B}}\right),$$
>
> where $I_n$ is the $n \times n$ identity matrix.

*Proof.*

$$\begin{aligned}
\left(P_{\mathcal{E}\leftarrow\mathcal{C}} \;\middle|\; P_{\mathcal{E}\leftarrow\mathcal{B}}\right) &\leftrightarrow P_{\mathcal{C}\leftarrow\mathcal{E}}\left(P_{\mathcal{E}\leftarrow\mathcal{C}} \;\middle|\; P_{\mathcal{E}\leftarrow\mathcal{B}}\right) \quad \text{by part (ii) of Lemma 4.1} \\
&= \left(P_{\mathcal{C}\leftarrow\mathcal{E}}P_{\mathcal{E}\leftarrow\mathcal{C}} \;\middle|\; P_{\mathcal{C}\leftarrow\mathcal{E}}P_{\mathcal{E}\leftarrow\mathcal{B}}\right) \\
&= \left(P_{\mathcal{C}\leftarrow\mathcal{C}} \;\middle|\; P_{\mathcal{C}\leftarrow\mathcal{B}}\right) \quad \text{by Proposition 10.2 in Section I} \\
&= \left(I_n \;\middle|\; P_{\mathcal{C}\leftarrow\mathcal{B}}\right).
\end{aligned}$$

□

## Appendix: 5 Proof of Proposition 8.1 in Section II

We recall the statement to be proven:

> Let $\varphi : U \to V$ be a linear transformation, where $U$ and $V$ are finite dimensional with bases $\mathcal{B}$ and $\mathcal{C}$ respectively. Then $\varphi$ is invertible if and only if $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible, and if this is the case, $[\varphi^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}} = [\varphi]^{-1}_{\mathcal{C}\leftarrow\mathcal{B}}$.

*Proof.* Suppose that $\varphi : U \to V$ is invertible, i.e., there is $\psi : V \to U$ such that $\psi \circ \varphi = \mathbf{1}_U$ and $\varphi \circ \psi = \mathbf{1}_V$. Then

$$\begin{aligned}
I_n &= [\mathbf{1}_U]_{\mathcal{B}} \quad \text{where } n = \dim(U) \\
&= [\psi \circ \varphi]_{\mathcal{B}} \\
&= [\psi]_{\mathcal{B}\leftarrow\mathcal{C}}[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} \quad \text{by Proposition 7.1 in Section II,} \\
\text{and} \quad I_m &= [\mathbf{1}_V]_{\mathcal{C}} \quad \text{where } m = \dim(V) \\
&= [\varphi \circ \psi]_{\mathcal{C}} \\
&= [\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}[\psi]_{\mathcal{B}\leftarrow\mathcal{C}} \quad \text{by Proposition 7.1 in Section II again.}
\end{aligned}$$

Thus, $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible, and its inverse is $[\psi]_{\mathcal{B}\leftarrow\mathcal{C}} = [\varphi^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}$.

Conversely, suppose that $[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible. For brevity of notation, let $A = [\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}$. Let $B = A^{-1}$, and let $\psi : V \to U$ be the linear transformation satisfying $[\psi]_{\mathcal{B}\leftarrow\mathcal{C}} = B$. Then

$$\begin{array}{rrcl} & [\psi \circ \varphi]_{\mathcal{B}} & = & [\psi]_{\mathcal{B}\leftarrow\mathcal{C}}[\varphi]_{\mathcal{C}\leftarrow\mathcal{B}} \\ & & = & BA \\ & & = & I_n \quad \text{so } \psi \circ \varphi = \mathbf{1}_U, \\ \\ \text{and} & [\varphi \circ \psi]_{\mathcal{C}} & = & [\varphi]_{\mathcal{C}\leftarrow\mathcal{B}}[\psi]_{\mathcal{B}\leftarrow\mathcal{C}} \\ & & = & AB \\ & & = & I_m \quad \text{so } \varphi \circ \psi = \mathbf{1}_V. \end{array}$$

Thus, $\varphi$ is invertible. □

## Appendix: 6 Some basics of diagonalization

Let $F$ be a field. (We have not formally introduced the concept of a field, but for this section you may take $F$ to stand for either $\mathbb{R}$, the field of real numbers, or $\mathbb{C}$, the field of complex numbers.) If $A \in M_n(F)$, then $p_A(x)$ denotes its characteristic polynomial, i.e., $p_A(x) = \det(xI - A)$.

If $\lambda \in F$ is an eigenvalue of $A$, the geometric multiplicity of $\lambda$ (over $F$) is the dimension of $\mathrm{Nul}(\lambda I - A) \subseteq F^n$, and is denoted $d_\lambda$. The algebraic multiplicity of $\lambda$ is the number of times the factor $x - \lambda$ occurs in $p_A(x)$.

**Lemma 6.1.** *If $\lambda \in F$ is an eigenvalue of $A$, then $1 \le d_\lambda \le m_\lambda$.*

*Proof.* The inequality $1 \le d_\lambda$ follows immediately from the definition of an eigenvalue: there has to be a non-zero vector $\mathbf{v} \in F^n$ such that $A\mathbf{v} = \lambda\mathbf{v}$.

Now let $\{\mathbf{v}_1, \ldots, \mathbf{v}_d\}$ be a basis for the eigenspace associated to $\lambda$, and extend it to a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $F^n$. Let $P = P_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{pmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{pmatrix}$, where $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard basis of $F^n$, and let $B = P^{-1}AP$. Then for $j = 1, \ldots, d$, the $j$th column of $B$ is

$$\begin{array}{rcl} P^{-1}A\mathbf{v}_j & = & P^{-1}\lambda\mathbf{v}_j \\ & = & \lambda P^{-1}\mathbf{v}_j \\ & = & \lambda P_{\mathcal{B}\leftarrow\mathcal{E}}\mathbf{v}_j \\ & = & \lambda\mathbf{e}_j. \end{array}$$

Hence,

$$\begin{array}{rcl} p_A(x) & = & p_B(x) \\ & = & \det(xI - B) \\ & = & (x-\lambda)^d p_C(x), \end{array}$$

where $C$ is the $(n-d)\times(n-d)$ matrix in the bottom right-hand corner of $B$. Thus, $(x-\lambda)^d$ divides $p_A(x)$, as required. □

**Proposition 6.2.** *The sum of the geometric multiplicities of the eigenvalues of $A$ cannot exceed $n$.*

*Proof.* The sum of the geometric multiplicities is at most the sum of the algebraic multiplicities by Lemma 6.1, which is the degree of $p_A(x)$, and this is $n$. □

**Lemma 6.3.** *Suppose that $\lambda_1,\ldots,\lambda_s$ are distinct eigenvalues of $A$, and suppose that for each $i$, $T_i$ is a linearly independent set of eigenvectors of $A$ with eigenvalue $\lambda_i$. Then $T_1\cup\cdots\cup T_s$ is a linearly independent set.*

*Proof.* The proof rests on the observation that if $\lambda,\mu\in F$, and if $\mathbf{u}\in F^n$ satisfies $A\mathbf{u}=\lambda\mathbf{u}$, then $(A-\mu I)\mathbf{u}=(\lambda-\mu)\mathbf{u}$. With this in mind, suppose that

$$\sum_{j=1}^{s}\sum_{\mathbf{u}\in T_j} c_{j,\mathbf{u}}\mathbf{u}=\mathbf{0}, \tag{6.1}$$

where the $c_{j,\mathbf{u}}$ are in $F$. Choose any $i\in\{1,\ldots,s\}$, and let

$$\begin{aligned} B_i &= \prod_{j\neq i}(A-\lambda_j I)\\ \beta_i &= \prod_{j\neq i}(\lambda_i-\lambda_j). \end{aligned}$$

Then by the observation at the start of the proof,

$$B_i\mathbf{u}=\begin{cases}\beta_i\mathbf{u} & \text{if } \mathbf{u}\in T_i\\ \mathbf{0} & \text{if } \mathbf{u}\in T_j \text{ with } j\neq i.\end{cases}$$

Applying $B_i$ to both sides of (6.1), we therefore obtain

$$\mathbf{0}=B_i\sum_{j=1}^{s}\sum_{\mathbf{u}\in T_j} c_{j,\mathbf{u}}\mathbf{u}=\sum_{\mathbf{u}\in T_i} c_{i,\mathbf{u}}\beta_i\mathbf{u}=\beta_i\sum_{\mathbf{u}\in T_i} c_{i,\mathbf{u}}\mathbf{u},$$

and hence $\sum_{\mathbf{u}\in T_i} c_{i,\mathbf{u}}\mathbf{u}=\mathbf{0}$ because $\beta_i\neq 0$. The linear independence of the set $T_i$ then gives $c_{i,\mathbf{u}}=0$ for all $\mathbf{u}\in T_i$. □

**Lemma 6.4.** *Let $A,B\in M_n(F)$ with $B=(b_{i,j})$, let $\mathbf{u}_1,\ldots,\mathbf{u}_n\in F^n$, and let $P=\begin{pmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{pmatrix}\in M_n(F)$. Then the following are equivalent:*

*(i)* $A\mathbf{u}_j=\sum_{i=1}^n b_{i,j}\mathbf{u}_i$ *for* $j=1,\ldots,n$.

*(ii)* $AP=PB$.

*Proof.* By definition of matrix multiplication, the $j$th column of $AP$ is $A\mathbf{u}_j$, and the $j$th column of $PB$ is $\begin{pmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_n\end{pmatrix}\mathbf{b}_j=\sum_{i=1}^n b_{i,j}\mathbf{u}_i$, where $\mathbf{b}_j$ is the $j$th column of $B$. Therefore, the $j$th column of $AP$ is equal to that of $PB$ if and only if $A\mathbf{u}_j=\sum_{i=1}^n b_{i,j}\mathbf{u}_i$. □

**Theorem 6.5.** *Let $A \in M_n(F)$. Then the following are equivalent:*

*(i) $A$ is diagonalizable over $F$.*

*(ii) $F^n$ has a basis consisting of eigenvectors of $A$.*

*(iii) The sum of the geometric multiplicities (over $F$) of the eigenvalues of $A$ is equal to $n$.*

*(iv) All roots of $p_A(x)$ are in $F$, and the geometric multiplicity of every eigenvalue of $A$ is equal to its algebraic multiplicity.*

*Proof.* We first prove the equivalence of (i) and (ii). Suppose that $F^n$ has a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, where $A\mathbf{u}_j = \lambda_j \mathbf{u}_j$. Then Lemma 6.4 says that $AP = PD$, where $P = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{pmatrix}$ and $D$ is the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Because $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a basis for $F^n$, $P$ is invertible, so the equality $AP = PD$ implies that $P^{-1}AP = D$.

Conversely, suppose that there is an invertible matrix $P \in M_n(F)$ such that $P^{-1}AP = D$, diagonal. Let the diagonal entries of $D$ be $\lambda_1, \ldots, \lambda_n$, and let the columns of $P$ be $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Then because $AP = PD$, Lemma 6.4 tells us that $A\mathbf{u}_j = \lambda_j \mathbf{u}_j$ for all $j$. The $\mathbf{u}_j$ are therefore eigenvectors of $A$, and they form a basis for $F^n$ because $P$ is invertible.

To complete the proof, we establish the implications (ii) ⇒ (iii) ⇒ (iv) ⇒ (ii). Assume (ii), i.e., that $F^n$ has a basis $\mathcal{B}$ consisting of eigenvectors of $A$. Without loss of generality, we may assume that

$$\mathcal{B} = \{\mathbf{u}_{1,1}, \ldots, \mathbf{u}_{1,r_1}, \ldots, \mathbf{u}_{s,1}, \ldots, \mathbf{u}_{s,r_s}\},$$

where $A\mathbf{u}_{i,j} = \lambda_i \mathbf{u}_{i,j}$, the $\lambda_i \in F$ being distinct eigenvalues. Then $d_{\lambda_i} \geq r_i$ for each $i$, so

$$\sum_{i=1}^{s} d_{\lambda_i} \geq \sum_{i=1}^{s} r_i = n.$$

The sum of the geometric multiplicities of all the eigenvalues in $F$ is therefore at least $n$, but then it has to be exactly $n$ by Proposition 6.2.

Next, assume (iii), i.e., that the sum of the geometric multiplicities over $F$ is $n$. Let the eigenvalues of $A$ in $F$ be $\lambda_1, \ldots, \lambda_t$. Then

$$\begin{aligned} \sum_{i=1}^{t} d_{\lambda_i} &= n \quad \text{by assumption} \\ &\geq \sum_{i=1}^{t} m_{\lambda_i}. \end{aligned}$$

Hence, because $d_{\lambda_i} \leq m_{\lambda_i}$ for all $i$ by Lemma 6.1, we must have $d_{\lambda_i} = m_{\lambda_i}$ for all $i$, and we must also have $\sum_{i=1}^{t} m_{\lambda_i} = n$, so all the roots of $p_A(x)$ are in $F$.

Finally, assume (iv), i.e., that all the roots of $p_A(x)$ are in $F$ and that $d_{\lambda_i} = m_{\lambda_i}$ for all $i$, where $\lambda_1, \ldots, \lambda_t$ are the eigenvalues of $A$. Then

$$\sum_{i=1}^{t} d_{\lambda_i} = \sum_{i=1}^{t} m_{\lambda_i} = n.$$

Hence, if $\mathcal{B}_i$ is a basis for the eigenspace associated to $\lambda_i$, then the set $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_t$ is linearly independent by Lemma 6.3 and has $n$ elements. It is therefore a basis for $F^n$ consisting of eigenvectors of $A$. □

## Appendix: 7 The differential equation $f' = \lambda f$

We solve the differential equation $f' = \lambda f$, where $\lambda$ is some given constant. If $f'(x) = \lambda f(x)$, then $0 = f'(x) - \lambda f(x)$, so multiplying both sides by $e^{-\lambda x}$ gives

$$\begin{aligned} 0 &= e^{-\lambda x} f'(x) - e^{-\lambda x} f(x) \\ &= \frac{d}{dx}(e^{-\lambda x} f(x)). \end{aligned}$$

Hence, $e^{-\lambda x} f(x) = a$ for some constant $a$, so $f(x) = ae^{\lambda x}$. Conversely, the function $f(x) = ae^{\lambda x}$ obviously satisfies $f' = \lambda f$.

## Appendix: 8 Proof of Proposition 4.1 in Section III

We recall the statement to be proven:

> Let $A$ be a real $2 \times 2$ matrix that has a non-real complex eigenvalue $\lambda = a + bi$. If $\mathbf{w}$ is an eigenvector for $\lambda$, then the real matrix $Q = \begin{pmatrix} \mathrm{Re}(\mathbf{w}) & \mathrm{Im}(\mathbf{w}) \end{pmatrix}$ is invertible, and
>
> $$Q^{-1}AQ = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = s \begin{pmatrix} a/s & b/s \\ -b/s & a/s \end{pmatrix},$$
>
> where $s = |\lambda| = \sqrt{a^2 + b^2}$.

*Proof.* We start from the equality $A\mathbf{w} = \lambda \mathbf{w}$. If $\mathbf{u} = \mathrm{Re}(\mathbf{w})$ and $\mathbf{v} = \mathrm{Im}(\mathbf{w})$, then the equality says

$$\begin{aligned} A(\mathbf{u} + i\mathbf{v}) &= (a + bi)(\mathbf{u} + i\mathbf{v}), \\ \text{i.e.,} \quad A\mathbf{u} + iA\mathbf{v} &= a\mathbf{u} - b\mathbf{v} + i(b\mathbf{u} + a\mathbf{v}). \end{aligned}$$

Equating real and imaginary parts gives

$$\begin{aligned} A\mathbf{u} &= a\mathbf{u} - b\mathbf{v} \\ A\mathbf{v} &= b\mathbf{u} + a\mathbf{v}. \end{aligned}$$

These two equations can be expressed by the single equation

$$\begin{aligned} A\begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix} &= \begin{pmatrix} \mathbf{u} & \mathbf{v} \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \\ \text{i.e.,} \quad AQ &= Q \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \end{aligned}$$

It remains to show that $Q$ is invertible, for then we obtain

$$Q^{-1}AQ = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

To that end, note that $\mathbf{w}$ and $\overline{\mathbf{w}}$ are linearly independent over $\mathbb{C}$, since they are eigenvectors with different eigenvalues ($\lambda$ and $\overline{\lambda}$). Because $\mathbf{u} = \frac{1}{2}(\mathbf{w}+\overline{\mathbf{w}})$ and $\mathbf{v} = \frac{1}{2i}(\mathbf{w}-\overline{\mathbf{w}})$, the coordinate vectors of $\mathbf{u}$ and $\mathbf{v}$ with respect to the basis $\{\mathbf{w}, \overline{\mathbf{w}}\}$ of $\mathbb{C}^2$ are

$$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1/(2i) \\ -1/(2i) \end{pmatrix},$$

which are linearly independent over $\mathbb{C}$. Hence, $\mathbf{u}$ and $\mathbf{v}$ are linearly independent over $\mathbb{C}$ (not only over $\mathbb{R}$). It follows that $Q$ is invertible, as desired. □

## Appendix: 9 The inner product space $C[a, b]$

We establish axiom (iv) of an inner product for the inner product we defined on $C[a, b]$, namely,

$$\langle f, g \rangle = \int_a^b f(x)g(x)\,dx.$$

We are to show that, if $f \in C[a, b]$, then $\langle f, f \rangle \geq 0$, with equality holding if and only if $f$ is the zero function. The first assertion is clear, because

$$\langle f, f \rangle = \int_a^b f(x)^2\,dx \geq 0,$$

and it is also clear that $\langle z, z \rangle = 0$ where $z$ here denotes the zero function. Now suppose that $f \in C[a, b]$ is not the zero function, meaning that there is $x_0 \in [a, b]$ such that $f(x_0) \neq 0$. Then $g(x_0) > 0$ where $g : [a, b] \to \mathbb{R}$ is the function defined by $g(x) = f(x)^2$. The function $g$ is again continuous, so because $g(x_0) > 0$, we may find a positive real number $y_0$ and an interval $I$ of width $c > 0$ in $[a, b]$ such that $g(x) \geq y_0$ for $x \in I$. Hence,

$$\langle f, f \rangle = \int_a^b g(x)\,dx \geq cy_0 > 0.$$

## Appendix: 10 Orthogonal projection

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and $U$ a subspace of $V$.

**Lemma 10.1.** *$U \cap U^{\perp} = \{\mathbf{0}\}$.*

*Proof.* If $\mathbf{u}$ is in both $U$ and $U^{\perp}$, then it must be orthogonal to itself, i.e., $\langle \mathbf{u}, \mathbf{u} \rangle = 0$. But then $\mathbf{u} = \mathbf{0}$. □

Now assume that $U$ has finite dimension $n$. Because of the Gram–Schmidt process, we know that $U$ has an orthogonal basis. If $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is such a basis for $U$, and if $\mathbf{v} \in V$, then we define

$$\operatorname{proj}_{\mathcal{B}}(\mathbf{v}) = \sum_{i=1}^{n} \frac{\langle \mathbf{u}_i, \mathbf{v}\rangle}{\langle \mathbf{u}_i, \mathbf{u}_i\rangle}\, \mathbf{u}_i \in U.$$

**Lemma 10.2.** *With notation as above,* $\mathbf{v} - \operatorname{proj}_{\mathcal{B}}(\mathbf{v}) \in U^{\perp}$*.*

*Proof.* Observe that, for all $i \in \{1, \ldots, n\}$,

$$\langle \mathbf{u}_i, \operatorname{proj}_{\mathcal{B}}(\mathbf{v})\rangle = \frac{\langle \mathbf{u}_i, \mathbf{v}\rangle}{\langle \mathbf{u}_i, \mathbf{u}_i\rangle}\, \langle \mathbf{u}_i, \mathbf{u}_i\rangle = \langle \mathbf{u}_i, \mathbf{v}\rangle,$$

so

$$\langle \mathbf{u}_i, \mathbf{v} - \operatorname{proj}_{\mathcal{B}}(\mathbf{v})\rangle = \langle \mathbf{u}_i, \mathbf{v}\rangle - \langle \mathbf{u}_i, \operatorname{proj}_{\mathcal{B}}(\mathbf{v})\rangle = \langle \mathbf{u}_i, \mathbf{v}\rangle - \langle \mathbf{u}_i, \mathbf{v}\rangle = 0.$$

Being orthogonal to each $\mathbf{u}_i$, $\mathbf{v} - \operatorname{proj}_{\mathcal{B}}(\mathbf{v})$ is therefore orthogonal to every element of $U$. □

**Proposition 10.3.** *Suppose that $\mathcal{B}$ and $\mathcal{C}$ are two orthogonal bases for $U$. If $\mathbf{v} \in V$, then* $\operatorname{proj}_{\mathcal{B}}(\mathbf{v}) = \operatorname{proj}_{\mathcal{C}}(\mathbf{v})$*.*

*Proof.* We start with the equality

$$\operatorname{proj}_{\mathcal{B}}(\mathbf{v}) - \operatorname{proj}_{\mathcal{C}}(\mathbf{v}) = (\mathbf{v} - \operatorname{proj}_{\mathcal{C}}(\mathbf{v})) - (\mathbf{v} - \operatorname{proj}_{\mathcal{B}}(\mathbf{v})).$$

The right-hand side is in $U^{\perp}$ by Lemma 10.2, so $\operatorname{proj}_{\mathcal{B}}(\mathbf{v}) - \operatorname{proj}_{\mathcal{C}}(\mathbf{v})$ is in $U^{\perp}$. But $\operatorname{proj}_{\mathcal{B}}(\mathbf{v}) - \operatorname{proj}_{\mathcal{C}}(\mathbf{v})$ is in $U$ as well, so it is zero by Lemma 10.1. Thus, $\operatorname{proj}_{\mathcal{B}}(\mathbf{v}) = \operatorname{proj}_{\mathcal{C}}(\mathbf{v})$. □

In light of Proposition 10.3, given $\mathbf{v} \in V$, we may define the orthogonal projection of $\mathbf{v}$ onto $U$ to be the vector $\operatorname{proj}_U(\mathbf{v}) = \operatorname{proj}_{\mathcal{B}}(\mathbf{v})$, where $\mathcal{B}$ is any choice of orthogonal basis for $U$.

We now prove Proposition 3.1 in Section IV. We recall the statement to be proven:

> Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and $U$ a finite-dimensional subspace of $V$. If $\mathbf{v} \in V$, then $\operatorname{proj}_U(\mathbf{v})$ is the unique vector $\mathbf{u} \in U$ that minimizes $\operatorname{dist}(\mathbf{v}, \mathbf{u})$.

*Proof.* If $\mathbf{u}$ is any vector in $U$, then

$$\operatorname{dist}(\mathbf{v}, \mathbf{u})^2 = \|\mathbf{v} - \mathbf{u}\|^2 = \|(\mathbf{v} - \operatorname{proj}_U(\mathbf{v})) + (\operatorname{proj}_U(\mathbf{v}) - \mathbf{u})\|^2. \tag{10.1}$$

Now, $\mathbf{v} - \operatorname{proj}_U(\mathbf{v}) \in U^{\perp}$ by Lemma 10.2, and of course $\operatorname{proj}_U(\mathbf{v}) - \mathbf{u} \in U$, so Proposition 2.1 in Section IV shows that the expression on the right of (10.1) is equal to

$$\|\mathbf{v} - \operatorname{proj}_U(\mathbf{v})\|^2 + \|\operatorname{proj}_U(\mathbf{v}) - \mathbf{u}\|^2.$$

The left-hand term here is independent of $\mathbf{u}$, and the right-hand term is zero if and only if $\mathbf{u} = \operatorname{proj}_U(\mathbf{v})$. □

## Appendix: 11 Existence and uniqueness of the totally positive $QR$-factorization

We outlined the existence of the totally positive $QR$-factorization of a matrix $A$ (with linearly independent columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$) in Section IV − 4. All that was missing was the justification that, with $Q$ and $R$ defined there, one has $A = QR$. But the $j$th column of $QR$ is $Q$ times the $j$th column of $R$, which is equal to

$$\begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_n \end{pmatrix} \begin{pmatrix} r_{1,j} \\ r_{2,j} \\ \vdots \\ r_{j,j} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = r_{1,j}\mathbf{w}_1 + r_{2,j}\mathbf{w}_2 + \cdots + r_{j,j}\mathbf{w}_j,$$

and this is simply $\mathbf{v}_j$ (the $j$th column of $A$) by the choice of the $r_{i,j}$.

For uniqueness, suppose that $Q, Q' \in M_{m,n}(\mathbb{R})$ have orthonormal columns, that $R, R' \in M_n(\mathbb{R})$ are upper triangular with positive diagonal entries, and that $QR = Q'R'$. We show that $Q = Q'$ and $R = R'$. Write

$$Q = \begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_n \end{pmatrix}, \quad Q' = \begin{pmatrix} \mathbf{w}'_1 & \cdots & \mathbf{w}'_n \end{pmatrix}, \quad R = (r_{i,j}), \quad R' = (r'_{i,j}).$$

We prove by induction on $j \geq 1$ that $\mathbf{w}_j = \mathbf{w}'_j$ and that $r_{i,j} = r'_{i,j}$ for $i \leq j$. For the case $j = 1$, we see by considering the first column of $QR$ and of $Q'R'$ that $r_{1,1}\mathbf{w}_1 = r'_{1,1}\mathbf{w}'_1$. Taking norms and remembering that $\mathbf{w}_1$ and $\mathbf{w}'_1$ are unit vectors, we obtain $|r_{1,1}| = |r'_{1,1}|$, and hence $r_{1,1} = r'_{1,1}$ because both are positive by assumption. It follows that $\mathbf{w}_1 = \mathbf{w}'_1$ as well.

Now take $j \geq 2$, and assume that the statement to be proven is true up to, but not yet including, that $j$. By looking at the $j$th column of each of $QR$ and $Q'R'$, we see that

$$r_{1,j}\mathbf{w}_1 + \cdots + r_{j,j}\mathbf{w}_j = r'_{1,j}\mathbf{w}'_1 + \cdots + r'_{j,j}\mathbf{w}'_j.$$

But $\mathbf{w}_i = \mathbf{w}'_i$ for $i < j$ by the inductive hypothesis, so we in fact have

$$r_{1,j}\mathbf{w}_1 + \cdots + r_{j-1,j}\mathbf{w}_{j-1} + r_{j,j}\mathbf{w}_j = r'_{1,j}\mathbf{w}_1 + \cdots + r'_{j-1,j}\mathbf{w}_{j-1} + r'_{j,j}\mathbf{w}'_j. \tag{11.1}$$

Dotting both sides with $\mathbf{w}_i$ for $i < j$ gives $r_{i,j} = r'_{i,j}$, and (11.1) then becomes simply $r_{j,j}\mathbf{w}_j = r'_{j,j}\mathbf{w}'_j$. Taking norms again and remembering that $r_{j,j}$ and $r'_{j,j}$ are positive, we arrive at $r_{j,j} = r'_{j,j}$ and $\mathbf{w}_j = \mathbf{w}'_j$. The induction is complete.

## Appendix: 12 Proof of the Cauchy–Schwarz inequality

We recall the statement to be proven:

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. If $\mathbf{u}, \mathbf{v} \in V$, then $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$, with equality holding if and only if one of the vectors is a scalar multiple of the other.

*Proof.* If both $\mathbf{u}$ and $\mathbf{v}$ are zero, then the inequality holds trivially. Therefore, we may assume that one of them, $\mathbf{u}$ say, is non-zero. Let $\mathbf{w} = \mathbf{v} - \lambda \mathbf{u}$, where

$$\lambda = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Note that $\mathbf{u}$ and $\mathbf{w}$ are orthogonal, because

$$\langle \mathbf{u}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} - \lambda \mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \lambda \langle \mathbf{u}, \mathbf{u} \rangle = 0.$$

Now,

$$\begin{aligned}
\|\mathbf{v}\|^2 &= \langle \mathbf{v}, \mathbf{v} \rangle \\
&= \langle \lambda \mathbf{u} + \mathbf{w}, \lambda \mathbf{u} + \mathbf{w} \rangle \\
&= \lambda^2 \langle \mathbf{u}, \mathbf{u} \rangle + \langle \lambda \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{w}, \lambda \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\
&= \lambda^2 \langle \mathbf{u}, \mathbf{u} \rangle + 2\lambda \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\
&= \lambda^2 \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \quad \text{because } \langle \mathbf{u}, \mathbf{w} \rangle = 0 \\
&\geq \lambda^2 \langle \mathbf{u}, \mathbf{u} \rangle \quad \text{because } \langle \mathbf{w}, \mathbf{w} \rangle \geq 0 \\
&= \frac{\langle \mathbf{u}, \mathbf{v} \rangle^2}{\langle \mathbf{u}, \mathbf{u} \rangle} \\
&= \frac{\langle \mathbf{u}, \mathbf{v} \rangle^2}{\|\mathbf{u}\|^2}.
\end{aligned}$$

Rearranging and taking square roots, we obtain $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$. Further, equality holds if and only if $\langle \mathbf{w}, \mathbf{w} \rangle = 0$, if and only if $\mathbf{w} = \mathbf{0}$, if and only if $\mathbf{v} \in \operatorname{Span}(\mathbf{u})$. □

## Appendix: 13 Real symmetric matrices

We prove that a real square matrix is symmetric if and only if it is orthogonally diagonalizable. This is Theorem 7.1 in Section IV.

First, we need a lemma.

**Lemma 13.1.** *Let $A \in M_n(\mathbb{R})$ be symmetric. If $\lambda \in \mathbb{C}$ is an eigenvalue of $A$, then in fact $\lambda \in \mathbb{R}$.*

*Proof.* Let $\mathbf{u} \in \mathbb{C}^n$ be an eigenvector with eigenvalue $\lambda$. Then

$$\begin{array}{rcl}
\overline{\lambda}(\overline{\mathbf{u}}^{\mathrm{T}}\mathbf{u}) & = & (\overline{\lambda\mathbf{u}})^{\mathrm{T}}\mathbf{u} \\
& = & (\overline{A\mathbf{u}})^{\mathrm{T}}\mathbf{u} \\
& = & (\overline{A}\,\overline{\mathbf{u}})^{\mathrm{T}}\mathbf{u} \\
& = & \overline{\mathbf{u}}^{\mathrm{T}}\,\overline{A}^{\mathrm{T}}\mathbf{u} \\
& = & \overline{\mathbf{u}}^{\mathrm{T}}A^{\mathrm{T}}\mathbf{u} \quad \text{because } A \text{ is real} \\
& = & \overline{\mathbf{u}}^{\mathrm{T}}A\mathbf{u} \quad \text{because } A \text{ is symmetric} \\
& = & \overline{\mathbf{u}}^{\mathrm{T}}(\lambda\mathbf{u}) \\
& = & \lambda(\overline{\mathbf{u}}^{\mathrm{T}}\mathbf{u}).
\end{array}$$

But $\overline{\mathbf{u}}^{\mathrm{T}}\mathbf{u}$ is non-zero, because $\mathbf{u}$ is non-zero by assumption, so $\overline{\lambda} = \lambda$, as desired. □

Now we come to the proof of the theorem. One direction is straightforward: If $A \in M_n(\mathbb{R})$ is orthogonally diagonalizable, then it is symmetric: Choose $P$ orthogonal and $D$ diagonal such that $P^{\mathrm{T}}AP = D$, i.e., $A = PDP^{\mathrm{T}}$. Then $A^{\mathrm{T}} = (PDP^{\mathrm{T}})^{\mathrm{T}} = PD^{\mathrm{T}}P^{\mathrm{T}} = PDP^{\mathrm{T}} = A$. (We have used the fact that a diagonal matrix is symmetric.)

We prove the converse by induction. Specifically, we show by induction on $n \geq 1$ the statement that every symmetric $A \in M_n(\mathbb{R})$ is orthogonally diagonalizable. The case $n = 1$ is immediate, since every $1 \times 1$ matrix is already diagonal, and so is orthogonally diagonalizable in a trivial way.

Now let $n \geq 2$, and assume that all real symmetric $(n-1) \times (n-1)$ matrices are orthogonally diagonalizable. Let $A \in M_n(\mathbb{R})$ be symmetric. Choose an eigenvalue $\lambda \in \mathbb{C}$ of $A$, which is possible by the Fundamental Theorem of Algebra. By Lemma 13.1, $\lambda \in \mathbb{R}$, so we may choose $\mathbf{u}_1 \in \mathbb{R}^n \smallsetminus \{\mathbf{0}\}$ such that $A\mathbf{u}_1 = \lambda\mathbf{u}_1$. Scaling if necessary, we may assume that $\|\mathbf{u}_1\| = 1$, and then we may extend $\{\mathbf{u}_1\}$ to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of $\mathbb{R}^n$. (Use Proposition 2.2 in the Appendix, together with the Gram–Schmidt process.)

Let

$$P = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{pmatrix},$$

an orthogonal matrix. Because $A\mathbf{u}_1 = \lambda\mathbf{u}_1$, Lemma 6.4 in the Appendix shows that

$$AP = P \begin{pmatrix} \lambda & a_2 & \cdots & a_n \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix}$$

for some $a_2, \ldots, a_n \in \mathbb{R}$ and some $B \in M_{n-1}(\mathbb{R})$. Hence,

$$P^{\mathrm{T}}AP = \begin{pmatrix} \lambda & a_2 & \cdots & a_n \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix}. \tag{13.1}$$

But $(P^{\mathrm{T}}AP)^{\mathrm{T}} = P^{\mathrm{T}}A^{\mathrm{T}}P = P^{\mathrm{T}}AP$ because $A$ is symmetric, so the matrix on the right-hand side of (13.1) is also symmetric. Thus, $a_2 = \cdots = a_n = 0$, and $B$ is symmetric.

Knowing that $B$ is a real symmetric $(n-1)\times(n-1)$ matrix, we may apply the inductive hypothesis to it to deduce the existence of an orthogonal matrix $Q \in M_{n-1}(\mathbb{R})$ and a diagonal matrix $D \in M_{n-1}(\mathbb{R})$ such that $Q^{\mathrm{T}}BQ = D$. The matrix

$$\tilde{Q} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q & \\ 0 & & & \end{pmatrix}$$

is orthogonal, so $P\tilde{Q}$ is also orthogonal. Further,

$$\begin{aligned}(P\tilde{Q})^{\mathrm{T}}A(P\tilde{Q}) &= \tilde{Q}^{\mathrm{T}}P^{\mathrm{T}}AP\tilde{Q} \\ &= \tilde{Q}^{\mathrm{T}}\begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix}\tilde{Q} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q^{\mathrm{T}}BQ & \\ 0 & & & \end{pmatrix} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & D & \\ 0 & & & \end{pmatrix},\end{aligned}$$

a diagonal matrix. The induction is complete.

### Orthogonality of eigenvectors from different eigenspaces

Recall from Section IV−7 that, in the process of orthogonally diagonalizing a real symmetric matrix, we construct a matrix $P$ from vectors in the orthonormal bases found for the various eigenspaces. We claimed in that section that $P$ will always be an orthogonal matrix, i.e., its columns will be orthonormal. In particular, the columns are mutually orthogonal. This is clear for columns in the same eigenspace, because they are the output of the Gram–Schmidt process, which generates orthogonal bases. But what about columns of $P$ in different eigenspaces? Why should they necessarily be orthogonal to one another? This is answered in the following proposition (and its proof).

**Proposition 13.2.** *Let $A$ be a real symmetric matrix. If $\mathbf{u}, \mathbf{v}$ are eigenvectors of $A$ in different eigenspaces, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal to each other.*

*Proof.* We could deduce the statement from the theorem we have just proven. Specifically, if $A$ is a real symmetric matrix, then we know that it is orthogonally diagonalizable, and from this the proposition follows via a short argument. However, there is also a direct proof, which we now present.

Suppose that $A\mathbf{u} = \lambda\mathbf{u}$ and $A\mathbf{v} = \mu\mathbf{v}$, where $\lambda \neq \mu$. Then

$$\begin{aligned}(\lambda - \mu)(\mathbf{u}\cdot\mathbf{v}) &= \lambda(\mathbf{u}\cdot\mathbf{v}) - \mu(\mathbf{u}\cdot\mathbf{v}) \\ &= (\lambda\mathbf{u})\cdot\mathbf{v} - \mathbf{u}\cdot(\mu\mathbf{v}) \\ &= (A\mathbf{u})\cdot\mathbf{v} - \mathbf{u}\cdot(A\mathbf{v}) \\ &= (A\mathbf{u})^{\mathrm{T}}\mathbf{v} - \mathbf{u}^{\mathrm{T}}(A\mathbf{v}) \\ &= \mathbf{u}^{\mathrm{T}}A^{\mathrm{T}}\mathbf{v} - \mathbf{u}^{\mathrm{T}}A\mathbf{v} \\ &= \mathbf{u}^{\mathrm{T}}A\mathbf{v} - \mathbf{u}^{\mathrm{T}}A\mathbf{v} \quad \text{because } A \text{ is symmetric} \\ &= 0.\end{aligned}$$

Therefore, because $\lambda \neq \mu$, it follows that $\mathbf{u}\cdot\mathbf{v} = 0$. □

## Appendix: 14 Proof of Proposition 10.1 in Section IV

We recall the statement to be proven:

> Suppose $A \in M_n(\mathbb{R})$ is symmetric, and let $f$ be the associated quadratic form. Let the eigenvalues of $A$ be $\lambda_1 \geq \cdots \geq \lambda_n$ (some may be repeated).
>
> (i) The maximum value of $f(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$ is $\lambda_1$. It occurs at any unit eigenvector with eigenvalue $\lambda_1$, and at no other unit vector.
>
> (ii) The minimum value of $f(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$ is $\lambda_n$. It occurs at any unit eigenvector with eigenvalue $\lambda_n$, and at no other unit vector.

*Proof.* We prove the part concerning the maximum. The part concerning the minimum is proven similarly.

Let $g(\mathbf{y}) = f(P\mathbf{y})$ where $P^{\mathrm{T}}AP = D$ is an orthogonal diagonalization in which the diagonal entries of $D$ are $\lambda_1, \ldots, \lambda_n$ in that order. Note that, if $\mathbf{y} = (y_1, \ldots, y_n)$, then $g(\mathbf{y}) = \mathbf{y}^{\mathrm{T}}D\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$. Note also that the change of variables $\mathbf{y} = P^{\mathrm{T}}\mathbf{x}$ preserves norms since $P$ is orthogonal, so $\mathbf{y}$ has norm 1 if $\mathbf{x}$ does. Hence, if $\|\mathbf{x}\| = 1$, then letting $\mathbf{y} = P^{\mathrm{T}}\mathbf{x}$, we have

$$f(\mathbf{x}) = g(\mathbf{y}) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \leq \lambda_1(y_1^2 + \cdots + y_n^2) = \lambda_1\|\mathbf{y}\|^2 = \lambda_1.$$

Now, if $\mathbf{x}$ is an eigenvector with eigenvalue $\lambda_1$ (and has norm 1), then

$$f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x} = \mathbf{x}^{\mathrm{T}}(\lambda_1\mathbf{x}) = \lambda_1\|\mathbf{x}\|^2 = \lambda_1.$$

Conversely, suppose that $\mathbf{x}$ is a unit vector such that $f(\mathbf{x}) = \lambda_1$. Let $\mathbf{y} = P^{\mathrm{T}}\mathbf{x}$, and write $\mathbf{y} = (y_1, \ldots, y_n)$. We claim that

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 = \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2. \tag{14.1}$$

Indeed,

$$\begin{aligned}
\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 &= g(\mathbf{y}) \\
&= f(P\mathbf{y}) \\
&= f(\mathbf{x}) \\
&= \lambda_1 \\
&= \lambda_1 \|\mathbf{y}\|^2 \\
&= \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2.
\end{aligned}$$

But because $\lambda_i y_i^2 \leq \lambda_1 y_i^2$ for all $i$, the equality in (14.1) implies that $\lambda_i y_i^2 = \lambda_1 y_i^2$ for all $i$, i.e., $(\lambda_1 - \lambda_i) y_i^2 = 0$. This in turn implies that $y_i = 0$ if $\lambda_i < \lambda_1$, so the vector $\mathbf{x} = P\mathbf{y}$ is a linear combination of the columns of $P$ corresponding to the eigenvalue $\lambda_1$, in other words, columns $1, \ldots, k$ where $k$ is greatest such that $\lambda_k = \lambda_1$. These columns are eigenvectors with eigenvalue $\lambda_1$, so $\mathbf{x}$ is an eigenvector with eigenvalue $\lambda_1$. □

## Appendix: 15 Polynomials and polynomial functions

There can be some confusion over what is meant by a polynomial. Here, we discuss two definitions and show that they amount to essentially the same objects. We restrict ourselves to polynomials with real coefficients, although the discussion would be identical for polynomials over any infinite field.

Algebra books usually define a polynomial in a formal way, in which the "powers" $x^n$ are initially not powers of a variable $x$ but simply symbols indexed by non-negative integers $n$. Once this formal definition is made, it is subsequently possible to define products of polynomials such that $x^m x^n = x^{m+n}$, so that the symbols do behave just like powers of a variable. And it is possible to evaluate a polynomial $p = \sum_{n=0}^{N} a_n x^n$ at a given $c \in \mathbb{R}$ by defining $p(c) = \sum_{n=0}^{N} a_n c^n$. A polynomial also has a degree, namely, the largest $n$ such that the coefficient $a_n$ of $x^n$ is non-zero. (This works for non-zero polynomials. The zero polynomial is often assigned the degree $-\infty$.)

In these notes, it is this formal definition of polynomials that I implicitly use, and I use the notation $\mathcal{P}$ for the space of polynomials defined in this formal way.

However, another common interpretation of a polynomial is as a *function* $p : \mathbb{R} \to \mathbb{R}$ of the form

$$p : x \mapsto \sum_{n=0}^{N} a_n x^n.$$

Thus, in this interpretation, a polynomial is a function first, albeit a function of a rather specific type. Let us use the notation $\tilde{\mathcal{P}}$ for the space of polynomials defined in this way, i.e., as functions.

Both $\mathcal{P}$ and $\tilde{\mathcal{P}}$ are vector spaces over $\mathbb{R}$, and there is a linear transformation

$$\begin{aligned}
\varphi : \mathcal{P} &\to \tilde{\mathcal{P}} \\
p &\mapsto (x \mapsto p(x)).
\end{aligned}$$

By definition of $\tilde{\mathcal{P}}$, the map $\varphi$ is surjective. Further, as we shall see below, it is injective as well and therefore an isomorphism.

Recall that if $p \in \mathcal{P}$, then a (real) *root* of $p$ is a real number $c$ such that $p(c) = 0$.

**Lemma 15.1.** *A non-zero polynomial $p \in \mathcal{P}$ of degree $n$ has at most $n$ roots.*

*Proof.* We prove this by induction on $n \geq 0$. If $p$ has degree 0, then $p$ is a non-zero constant and therefore has no roots.

Now choose $n \geq 1$, and assume that every polynomial of degree $n-1$ has at most $n-1$ roots. Let $p \in \mathcal{P}$ have degree $n$. If $p$ has no roots, then we are done. Otherwise, let $c$ be a root. Via long division of polynomials, we may write $p = (x-c)q + r$ where $q, r \in \mathcal{P}$ and $r$ has degree less than the degree of $x - c$, i.e., $r$ is constant. Therefore, evaluating $p = (x-c)q + r$ at $c$, we obtain $0 = (c-c)q(c) + r = r$, i.e., $r$ is zero. Hence, $p = (x-c)q$. Because $q$ has degree $n-1$, the inductive hypothesis says that $q$ has at most $n-1$ roots, so $p$, being the product of $x - c$ and $q$, has at most $1 + (n-1) = n$ roots. The induction is complete. □

**Proposition 15.2.** *The map $\varphi : \mathcal{P} \to \tilde{\mathcal{P}}$ defined above is an isomorphism.*

*Proof.* We have already remarked that $\varphi$ is surjective. It remains to show that it is injective. Suppose $p \in \mathrm{Ker}(\varphi)$. Then $p(x) = 0$ for all $x \in \mathbb{R}$, so $p$ has infinitely many roots because $\mathbb{R}$ is infinite. But a non-zero polynomial has at most finitely many roots by Lemma 15.1, so $p$ must be the zero polynomial. □