LINEAR ALGEBRA. PART 2.

LECTURERS: DR K. KIRKINA AND DR E. ZORIN

CONTENTS

6. ABSTRACT LINEAR ALGEBRA

6.1. **Vector Spaces.** All of the ideas and the reasoning underlying the study of subspaces of $\mathbb{F}^n$ in Part 1 can be applied to any set of objects for which the idea of linear combination makes sense. In other words, if we have a set V for which, given two elements $u, v \in V$ and scalars α, β, we can make sense of $\alpha u + \beta v$ as an element of V then we could try to do linear algebra in V. To do this, we need to be able to *add* and *scale* the objects. Here are some examples.

Example 6.1.

(i) Let $\mathbb{R}_n[x]$ denote the set of all polynomials in x of degree no more than n and with real coefficients, i.e.,

$$\mathbb{R}_n[x] = \{\alpha_0 + \alpha_1 x + \ldots + \alpha_n x^n \mid \alpha_j \in \mathbb{R},\ 0 \le j \le n\}.$$

Version: April 23, 2024.

It is easy to define linear combinations here: for $\alpha, \beta \in \mathbb{R}$ we have

$$\alpha(\alpha_0 + \alpha_1 x + \ldots + \alpha_n x^n) + \beta(\beta_0 + \beta_1 x + \ldots + \beta_n x^n) = \sum_{j=0}^{n} (\alpha\alpha_j + \beta\beta_j)x^j.$$

Notice that there is a "zero polynomial" given by $\alpha_j = 0$ for all j. It is the result of multiplying any polynomial by the zero scalar. Notice also that we can easily replace $\mathbb{R}$ by $\mathbb{C}$ to get $\mathbb{C}_n[x]$ with linear combinations over $\mathbb{C}$.
Note that there are other operations we can perform on polynomials, e.g. multiplication, but this is not relevant here - at the moment we are only considering adding and scaling, and ignoring any additional structure.

(ii) For any interval $[a, b] \subseteq \mathbb{R}$ let $\mathcal{F}([a, b], \mathbb{R})$ denote the set of all real-valued functions $f : [a, b] \to \mathbb{R}$.* We can make linear combinations of these using pointwise addition. For $f, g \in \mathcal{F}([a, b], \mathbb{R})$ and $\alpha, \beta \in \mathbb{R}$ define the function $\alpha f + \beta g : [a, b] \to \mathbb{R}$ by

$$(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) \text{ for all } x \in [a, b].$$

Notice that there is a "zero function" defined by $\mathbf{0}(x) = 0$ for all $x \in [a, b]$. It is obtained by multiplying any function by the zero scalar. Notice also that $\mathbb{R}_n[x] \subseteq \mathcal{F}([a, b], \mathbb{R})$ for any n † and that linear combinations in $\mathbb{R}_n[x]$ agree with linear combinations in $\mathcal{F}([a, b], \mathbb{R})$. We could replace the codomain $\mathbb{R}$ in this example by $\mathbb{C}$ to allow linear combinations over $\mathbb{C}$.

(iii) The set $M_{p\times n}(\mathbb{F})$ of all $p \times n$ matrices for either $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ has an obvious notion of linear combination using the natural matrix sum and scalar multiplication: for two such matrices $A = (a_{jk})$ and $B = (b_{jk})$ and $\alpha, \beta \in \mathbb{R}$, the jk-entry of the matrix $\alpha A + \beta B$ is given by $\alpha a_{jk} + \beta b_{jk}$. Note that we also have a zero matrix here, which again can be obtained by scaling any matrix by zero.

The properties we need to make linear combinations work are encapsulated in the following definition of an abstract vector space. By convention elements of a vector space are referred to as vectors, even when the vector space is nothing like $\mathbb{R}^n$ or $\mathbb{C}^n$. For these notes we will mainly consider vector spaces over the fields of real or complex numbers, but other interesting examples of fields come up in number theory and abstract algebra and the definition below works for any field.

*Note that we are using a slightly different notation convention to Part 1: we will use the symbol $\subseteq$ to denote an arbitrary subset and $\subsetneq$ if we want to emphasise that it is a proper subset (i.e. one that is not equal to the whole set). This is analogous to how the symbols $\leq$ and $<$ are used.

†Technically this is not quite true, since the elements of $\mathbb{R}_n[x]$ are not yet functions since we have not specified a domain for them, but polynomials are defined on all of $\mathbb{R}$, hence we can consider them as functions on the domain $[a, b]$, making the statement true. In general functions with a domain containing $[a, b]$ can be considered as functions with the restricted domain $[a, b]$.

Definition 6.2. *Let $\mathbb{F}$ be a field (for our purposes, usually $\mathbb{R}$ or $\mathbb{C}$). A* ***vector space over*** $\mathbb{F}$ *is a set V together with two binary operations:*

vector addition	***scalar multiplication***
$V \times V \to V$	$\mathbb{F} \times V \to V$
$(u, v) \mapsto u + v$	$(\alpha, u) \mapsto \alpha u$

satisfying the following axioms:

- *(A1)* $u + v = v + u$ *for all* $u, v \in V$ *(commutativity)*
- *(A2)* $u + (v + w) = (u + v) + w$ *for all* $u, v, w \in V$ *(associativity)*
- *(A3) there exists* $\mathbf{0} \in V$ *such that* $\mathbf{0} + v = v$ *for all* $v \in V$ *(additive identity)*
- *(A4) given any* $v \in V$ *there exists* $-v \in V$ *with* $(-v) + v = \mathbf{0}$ *(additive inverse)*
- *(M1)* $\alpha(u + v) = \alpha u + \alpha v$ *for all* $\alpha \in \mathbb{F}$ *and* $u, v \in V$
- *(M2)* $\alpha(\beta v) = (\alpha\beta)v$ *for all* $\alpha, \beta \in \mathbb{F}$ *and* $v \in V$
- *(M3)* $(\alpha + \beta)v = \alpha v + \beta v$ *for all* $\alpha, \beta \in \mathbb{F}$ *and* $v \in V$
- *(M4)* $1v = v$ *for all* $v \in V$ *(where* $1 \in \mathbb{F}$ *is the usual 1).*

The additive identity $\mathbf{0} \in V$ *is called the* ***zero vector****.*
A ***real vector space*** *is a vector space over* $\mathbb{R}$*. A* ***complex vector space*** *is a vector space over* $\mathbb{C}$*.*
A ***vector*** *is an element of a vector space.*
Given a vector space V over a field $\mathbb{F}$, a ***scalar*** *is an element of* $\mathbb{F}$*.*

Note, the operations being *binary operations* implies that V is closed under them: for any $u, v \in V$ and any $\alpha \in \mathbb{F}$ we have $u + v \in V$ and $\alpha u \in V$.

Axioms A2-A4, together with the fact that vector addition is a binary operation, say that a vector space is a **group** under vector addition, and together with axiom A1 (commutativity) we in fact have that a vector space is an **abelian group** under vector addition.

Example 6.3.
- (i) Clearly $\mathbb{F}^n$ is a vector space over $\mathbb{F}$ with the natural definitions of vector addition and scalar multiplication. Any subspace $S \subseteq \mathbb{F}^n$ is a vector space over $\mathbb{F}$ in its own right.
- (ii) It is an exercise (somewhat tedious but worthwhile) to check that all the examples in Example 6.1 are vector spaces using the operations of vector addition and scalar multiplication given.
- (iii) There is one particular example we need to define: the **trivial vector space**. This consists of one element, called $\mathbf{0}$, for which $\mathbf{0} + \mathbf{0} = \mathbf{0}$, and scalar multiplication $\alpha\mathbf{0} = \mathbf{0}$ for all $\alpha \in \mathbb{F}$. We will write the trivial vector space as $\{\mathbf{0}\}$. We do not think of $\mathbf{0}$ as necessarily being a coordinate vector of zeroes (because $(0, 0) \neq (0, 0, 0)$, for example), but the vector space $\{\mathbf{0}\}$ can be identified with the trivial subspace of every vector space (see below).

Notice that pretty much all these axioms are "obvious". The purpose here is to make sure that even for very unusual choices of V and operations of vector addition and scalar multiplication, the algebra still behaves very nicely.

Example 6.4. In this module we will mainly consider vector spaces over the infinite fields $\mathbb{R}$ and $\mathbb{C}$, but the definition of a vector space works just as well for other infinite fields (like $\mathbb{Q}$), and over finite fields. Recall that the *order* of a field is the number of elements in it. A fundamental fact about finite fields is that they must have prime power order, i.e. they must have order p^k, for p a prime and $k \in \mathbb{N}$. Also, for each natural number q of the form p^k, there is (up to isomorphism) a unique field with that number of elements. We denote this field by $\mathbb{F}_q$.
The easiest examples of finite fields occur when q is itself a prime. Then the algebraic operations in the field are given by modular arithmetic.
For a basic example, let us consider the field of two elements, $\mathbb{F}_2 = \{0, 1\}$. The operations in this field are given by multiplication and addition modulo 2:

$$\begin{array}{c|c|c} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline 1 & 1 & 0 \end{array} \qquad \begin{array}{c|c|c} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 0 & 1 \end{array}$$

Next, we can consider a vector space over this field, for example we can consider $\mathbb{F}_2^3$, which consists of 3-dimensional coordinate vectors with entries in $\mathbb{F}_2$. There are 3 entries in each such vector, and each entry can be either 0 or 1, so we see that there are $2^3 = 8$ elements in this vector space:

$$\mathbb{F}_2^3 = \left\{ \begin{pmatrix} 0\\0\\0 \end{pmatrix}, \begin{pmatrix} 1\\0\\0 \end{pmatrix}, \begin{pmatrix} 0\\1\\0 \end{pmatrix}, \begin{pmatrix} 0\\0\\1 \end{pmatrix}, \begin{pmatrix} 1\\1\\0 \end{pmatrix}, \begin{pmatrix} 1\\0\\1 \end{pmatrix}, \begin{pmatrix} 0\\1\\1 \end{pmatrix}, \begin{pmatrix} 1\\1\\1 \end{pmatrix} \right\}.$$

These vectors can be added and scaled (by elements of the field, 0 and 1) in the obvious way, for example:

$$\begin{pmatrix} 0\\1\\1 \end{pmatrix} + \begin{pmatrix} 1\\0\\1 \end{pmatrix} = \begin{pmatrix} 1\\1\\0 \end{pmatrix} \qquad \text{and} \qquad 0 \cdot \begin{pmatrix} 1\\0\\1 \end{pmatrix} = \begin{pmatrix} 0\cdot 1\\0\cdot 0\\0\cdot 1 \end{pmatrix} = \begin{pmatrix} 0\\0\\0 \end{pmatrix}.$$

Check that this set with these operations satisfies all the axioms of a vector space (this follows from the properties of modular arithmetic).
This very easy example nonetheless leads to much rich mathematics - the field of order 2 is especially useful since computers work in binary and vector spaces over this field can be used to encode strings of data in cryptography, and $\mathbb{F}_2^3$ in particular is related to an interesting object in projective geometry called the *Fano plane*.

The above example still had the familiar form of $V = \mathbb{F}^n$, albeit for a less familiar field $\mathbb{F}$. The following example is more counter-intuitive:

Example 6.5. Let V_1 be the set of $n \times n$ real diagonal matrices. Any two elements of V_1 can be written as

$$u = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n \end{pmatrix}, \quad v = \begin{pmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \beta_n \end{pmatrix}.$$

Take $\mathbb{F} = \mathbb{R}$. There is an obvious way to define addition and scalar multiplication so that V_1 becomes a vector space over $\mathbb{R}$. The sum of two such "vectors" u and v in V_1, and the scalar product with an element $\gamma \in \mathbb{R}$, are given by:

$$u + v = \begin{pmatrix} \alpha_1 + \beta_1 & 0 & \dots & 0 \\ 0 & \alpha_2 + \beta_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n + \beta_n \end{pmatrix}, \quad \gamma u = \begin{pmatrix} \gamma\alpha_1 & 0 & \dots & 0 \\ 0 & \gamma\alpha_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \gamma\alpha_n \end{pmatrix}.$$

Note that these are the same matrix operations as we saw in Example 6.1(iii), just applied to the special case of diagonal matrices.
(Note that V_1 with these operations is "basically" just the vector space $\mathbb{R}^n$ again since all we have really changed is the notation: when we define isomorphisms of vector spaces later it will be clear that the map φ given by

$$\varphi : V_1 \to \mathbb{R}^n, \qquad \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n \end{pmatrix} \mapsto \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

is a vector space isomorphism, i.e. a structure-preserving bijective map.)

Now let us consider a similar example, but with very different operations. Let V_2 be the subset of V_1 consisting only of matrices with (strictly) positive diagonal entries. We again take our field $\mathbb{F}$ to be $\mathbb{R}$, and we can now define "vector addition" and "scalar multiplication" on V_2 by

$$u + v := \begin{pmatrix} \alpha_1\beta_1 & 0 & \dots & 0 \\ 0 & \alpha_2\beta_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n\beta_n \end{pmatrix}, \quad \gamma u := \begin{pmatrix} \alpha_1^\gamma & 0 & \dots & 0 \\ 0 & \alpha_2^\gamma & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n^\gamma \end{pmatrix},$$

for u, v as above but now with $\alpha_j > 0$ and $\beta_j > 0$, and $\gamma \in \mathbb{R}$. (Recall that arbitrary real powers of a positive real number are defined by $\alpha_j^\gamma = \exp(\gamma \log(\alpha_j))$. This is the reason we need $\alpha_j > 0$ now.)

So here we are using matrix multiplication to define "addition" in V. The "zero vector" for this addition is the identity matrix I_n.
(Check that these operations satisfy the axioms, even though "vector addition" and "scalar multiplication" do not match their usual meanings. For example,

$$u + v = \begin{pmatrix} \alpha_1\beta_1 & 0 & \dots & 0 \\ 0 & \alpha_2\beta_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n\beta_n \end{pmatrix} = \begin{pmatrix} \beta_1\alpha_1 & 0 & \dots & 0 \\ 0 & \beta_2\alpha_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \beta_n\alpha_n \end{pmatrix} = v + u,$$

so A1 is satisfied. And for $\gamma, \lambda \in \mathbb{R}$ we get

$$(\gamma + \lambda)u = \begin{pmatrix} \alpha_1^{\gamma+\lambda} & 0 & \dots & 0 \\ 0 & \alpha_2^{\gamma+\lambda} & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n^{\gamma+\lambda} \end{pmatrix} = \begin{pmatrix} \alpha_1^{\gamma} + \alpha_1^{\lambda} & 0 & \dots & 0 \\ 0 & \alpha_2^{\gamma} + \alpha_2^{\lambda} & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \alpha_n^{\gamma} + \alpha_n^{\lambda} \end{pmatrix} = \gamma u + \lambda u,$$

so M3 is satisfied.)
Moreover, defining "addition" in terms of matrix multiplication wouldn't work for non-diagonal matrices, as matrix multiplication is typically not commutative, so we wouldn't satisfy axiom A1.
Also, note that the operations we used on V_1 would not make V_2 into a vector space: the zero vector for that addition would need to be the zero matrix, which is not in V_2, and the usual additive inverse of a diagonal matrix with positive diagonal entries would need to have negative diagonal entries, so it would not be in V_2, so axioms A3 and A4 would fail. In the language of the definition below, this shows that V_2 is not a subspace of V_1.

Once we have a vector space V over a field $\mathbb{F}$, all the definitions we made for subspaces of $\mathbb{F}^n$ earlier can be imported to V with no change. We summarise them here.

Definition 6.6. *Let V be a vector space over a field $\mathbb{F}$.*

(i) *Given vectors $v_1, \dots, v_q \in V$ and scalars $\alpha_1, \dots, \alpha_q \in \mathbb{F}$, the sum*

$$\alpha_1 v_1 + \dots + \alpha_q v_q = \sum_{j=1}^{q} \alpha_j v_j,$$

*is called a **linear combination** of $v_1, \dots, v_q$.*
Note that, by definition, a linear combination is always a finite sum.

(ii) *A subset $S \subseteq V$ is called a **subspace** (or **linear subspace**) of V if it contains the zero vector $\mathbf{0} \in V$ and it is closed under taking all possible linear combinations.*

(iii) *A collection of vectors $C \subseteq V$ is* ***linearly dependent*** *if there are $v_1, \ldots, v_q \in C$ and scalars $\alpha_1, \ldots, \alpha_q \in \mathbb{F}$, not all zero, for which*

$$\alpha_1 v_1 + \ldots + \alpha_q v_q = \mathbf{0}.$$

Otherwise we say C is a ***linearly independent*** *set.*

(iv) *The* ***span*** *of a collection of vectors $C \subseteq V$ is the set of all linear combinations of vectors from C, denoted $\mathrm{Sp}(C)$. When C is a finite collection $v_1, \ldots, v_q$ we may also write $\mathrm{Sp}(v_1, \ldots, v_q)$. We say C spans V when $V = \mathrm{Sp}(C)$.*

(v) *For a non-trivial vector space $V \neq \{\mathbf{0}\}$ a collection $\mathcal{B} \subseteq V$ of vectors, not necessarily finite, in V is called a* ***basis of*** *V if it is both linearly independent and spans V.*

As in Part 1, we define the empty set $\emptyset$ to be the basis for the trivial vector space $\{\mathbf{0}\}$. For any collection $C \subseteq V$, $\mathrm{Sp}(C)$ is a subspace of V, and it is the smallest subspace that contains C (i.e., if $S \subseteq V$ is any subspace with $C \subseteq S$ then $\mathrm{Sp}(C) \subseteq S$).
As before, an *ordered* set $\mathcal{B}$ is a basis for V if and only if every vector in V has a *unique* expression as a linear combination of elements of $\mathcal{B}$.

Example 6.7*.*

(i) Every vector space V has a zero vector $\mathbf{0}$ and $\{\mathbf{0}\} \subseteq V$ is a subspace, called the trivial subspace.

(ii) Let $\mathbb{R}[x]$ denote the set of all polynomials with real coefficients and of any degree, with addition and scalar multiplication defined as earlier. Then $\mathbb{R}_n[x] \subsetneq \mathbb{R}[x]$ is a subspace for every choice of n. (Recall that $\mathbb{R}_n[x]$ is the set of such polynomials of degree at most n. We need the degree to be "at most" n here, rather than equal to n, to ensure that the zero polynomial $\mathbf{0}$ is in the space, and since the sum of two polynomials of degree n could have degree less than n if the relevant coefficients cancel.)
It is clear that $\{1, x, \ldots, x^n\}$ is a basis for $\mathbb{R}_n[x]$ and that the infinite set $\{1, x, x^2, \ldots\}$ is a basis for $\mathbb{R}[x]$. It is an exercise to show that $\mathbb{R}[x]$ cannot have a finite basis.

(iii) The vector space of all $p \times n$ matrices $M_{p\times n}(\mathbb{F})$ has a standard basis given by $\{E_{jk} \mid 1 \leq j \leq p,\ 1 \leq k \leq n\}$ where the matrix E_{jk} has zeroes everywhere except for a 1 in the jk-entry (j-th row, k-th column). For example, $M_{2\times 2}(\mathbb{F})$ has standard basis

$$E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

(iv) Let $C([a,b],\mathbb{R}) \subseteq \mathcal{F}([a,b],\mathbb{R})$ denote the set of all continuous functions $f : [a,b] \to \mathbb{R}$ on the interval $[a,b]$. The zero function $\mathbf{0} : [a,b] \to \mathbb{R}$, $\mathbf{0}(x) = 0$ for all x, is constant, hence continuous. Elementary real analysis shows that if $f, g \in$

$C([a,b],\mathbb{R})$ then for any $\alpha,\beta \in \mathbb{R}$ the linear combination $\alpha f + \beta g$ is also continuous, hence in $C([a,b],\mathbb{R})$. Hence $C([a,b],\mathbb{R})$ is a subspace of $\mathcal{F}([a,b],\mathbb{R})$.

(v) Continuing on with the idea of the previous example, let $\mathcal{I}([a,b],\mathbb{R}) \subseteq \mathcal{F}([a,b],\mathbb{R})$ denote the set of all functions whose integral $\int_a^b f(x)dx$ exists. Then clearly $\mathbf{0}$ (the zero function above) is integrable, and elementary real analysis shows that a linear combination of integrable functions is integrable, so $\mathcal{I}([a,b],\mathbb{R})$ is also a subspace of $\mathcal{F}([a,b],\mathbb{R})$. Since every continuous function is integrable, we could also think of $C([a,b],\mathbb{R})$ as a subspace of $\mathcal{I}([a,b],\mathbb{R})$. Indeed, all the inclusions

$$\mathbb{R}_n[x] \subseteq \mathbb{R}[x] \subseteq C([a,b],\mathbb{R}) \subseteq \mathcal{I}([a,b],\mathbb{R}) \subseteq \mathcal{F}([a,b],\mathbb{R})$$

are subspace inclusions. Because $\mathbb{R}[x]$ has no finite basis, it follows (using the same reasoning as Theorem 1.11) that every vector space containing it cannot have a finite basis.*

We have written these as arbitrary inclusions, but in fact (as long as we are in the usual situation where $a < b$) we could have written $\subsetneq$ for all of them since each is a proper subspace of the next.

(vi) What does independence mean in a space of functions? Consider $\mathcal{F}([0,2\pi],\mathbb{R})$, and the three functions x, $\sin(x)$ and e^x (for example). They are linearly independent if

$$\alpha x + \beta \sin(x) + \gamma e^x = \mathbf{0}(x) \quad \text{implies} \quad \alpha = \beta = \gamma = 0.$$

Note that this is an equation of functions and so it has to hold for all $x \in [0,2\pi]$. We can evaluate at $x = 0$ to get $0 + 0 + \gamma e^0 = 0$, hence $\gamma = 0$.

Now we have $\alpha x + \beta \sin(x) = 0$. Differentiating once gives us $\alpha + \beta \cos(x) = \mathbf{0}(x)$, and setting $x = \frac{\pi}{2}$ gives $\alpha = 0$. Hence we are left with $\beta \sin(x) = 0$ for all $x \in [0,2\pi]$, which can only happen if $\beta = 0$.

So all the coefficients are 0, hence these three functions are linearly independent.

With these definitions, all of the results we proved in Chapter 1 for subspaces of $\mathbb{F}^n$ still hold, with exactly the same proofs. In particular the Steinitz Exchange Lemma (Lemma

---

*As an aside, does $C([a,b],\mathbb{R})$ have a basis? It can be shown that *every vector space has a basis*, so yes. But the proof of this fact in the general case is non-constructive and relies on Zorn's Lemma, a variant of the Axiom of Choice. So while a basis exists, it is uncountable and impossible to describe explicitly. One might be tempted to try using series (e.g. Taylor series or Fourier series) to decompose some continuous functions, but this would not work - an arbitrary function would need to be a linear combination of the basis functions, and in our definition a linear combination needs to be a *finite* sum. So bases, as we have defined them, are a less useful concept on infinite dimensional spaces. But in future subjects like Functional Analysis you can meet other concepts of "basis" that are more suitable for infinite dimensions.

1.10) holds for any vector space with a finite basis, which means we can define the dimension of a vector space.

Theorem 6.8. *Let V be a vector space over $\mathbb{F}$ with a finite basis. Then every basis of V has the same number of elements. This number is called the **dimension of** V, denoted $\dim(V)$.*

Now, using the same arguments as in chapter 1 (for Theorem 1.11 and Lemma 1.12) we have the same facts about subspaces of a finite dimensional vector space as we do for subspaces of $\mathbb{F}^n$.

Lemma 6.9. *Let V be an n-dimensional vector space and $S \subseteq V$ be a subspace. Then S has a finite basis. Let $q = \dim(S)$.*

(i) *Any linearly independent subset of S has no more than q elements. Hence, any subset of S containing more than q vectors is linearly dependent.*
(ii) *Any linearly independent subset $C \subseteq V$ can be extended to a basis of V (i.e., there is a basis $\mathcal{B}$ which contains C). In particular, any basis of S can be extended to a basis of V.*
(iii) *Any finite spanning set for S contains a basis. Hence no subset containing fewer than q vectors can span S.*
(iv) *Any linearly independent set containing q vectors must span S (hence is a basis for S). Similarly, any spanning set of size q must be linearly independent (hence is a basis for S).*

Example 6.10.

(i) Since $\mathbb{R}_n[x]$ has basis $\{1, x, \dots, x^n\}$ it has dimension $n+1$.
(ii) Using the standard basis for $M_{p\times n}(\mathbb{F})$ above we see this vector space has dimension pn.
(iii) Let $V = M_{2\times 2}(\mathbb{R})$ and let $S \subseteq V$ denote the subset of symmetric 2×2 matrices. This is a linear subspace because for any $A, B \in S$ and $\alpha, \beta \in \mathbb{R}$ we have

$$(\alpha A + \beta B)^\top = \alpha A^\top + \beta B^\top = \alpha A + \beta B,$$

so S is closed under linear combinations, and clearly the zero matrix O is symmetric. Since $M_{2\times 2}(\mathbb{R})$ is 4-dimensional S must have a finite basis. Indeed, every matrix in S can be written in the form

$$\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} = \alpha \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \beta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

so the three matrices on the right hand side span S and are clearly linearly independent. So they provide a basis for S. Hence $\dim(S) = 3$. We can extend this basis of S to a basis of V by adding one more linearly independent matrix. To find a linearly independent matrix, we just need to take one that is not in the

span of the first three, i.e. not in S, so not a symmetric matrix. So we could for example use $E_{12} - E_{21} = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, or just $E_{12} = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$.
By the lemma above, any three independent symmetric 2×2 matrices will span S and therefore be a basis, while four or more symmetric 2×2 matrices must be linearly dependent.

Remark 6.11. The definitions of sum and direct sum given earlier for subspaces of $\mathbb{F}^n$ translate immediately to subspaces $S_1, \ldots, S_q \subseteq V$ of an abstract vector space V. Lemmas 1.16 and 1.17 still hold, using exactly the same proofs.

6.2. **Linear Maps.** It is easy to see how to extend the definition of a linear map given earlier to the case of maps between vector spaces over the same field $\mathbb{F}$.

Definition 6.12. *Let V, W be vector spaces over the same field $\mathbb{F}$. A map $L : V \to W$ is called a* ***linear map (or linear transformation)*** *if it maps linear combinations to linear combinations in the following way:*

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v), \quad \textit{for all } u, v \in V,\ \alpha, \beta \in \mathbb{F}.$$

(Note that this is equivalent to $L(u + v) = L(u) + L(v)$ and $L(\alpha v) = \alpha L(v)$, so a linear map is a map that respects vector addition and scalar multiplication.)

In abstract algebra linear maps are also referred to as **vector space homomorphisms**, since they, like other homomorphisms (e.g. of groups, rings, fields) are structure-preserving maps. For this reason we will denote the set of all linear maps from V to W by $\mathrm{Hom}(V, W)$. The following lemma shows that we can make $\mathrm{Hom}(V, W)$ into a vector space over $\mathbb{F}$ too. The proof is left as an exercise.

Lemma 6.13. *Let U, V, W be vector spaces over the same field $\mathbb{F}$. If $L, M \in \mathrm{Hom}(V, W)$ and $\alpha, \beta \in \mathbb{F}$ then $\alpha L + \beta M$, defined by*

$$(\alpha L + \beta M)(v) = \alpha L(v) + \beta M(v), \quad v \in V,$$

is also a linear map. Also, if $L \in Hom(V, W)$ and $K \in \mathrm{Hom}(U, V)$ then the composite $L \circ K \in \mathrm{Hom}(U, W)$.

Remark 6.14. Since V and W are not necessarily finite dimensional, we don't expect all linear maps to be represented by matrices (unlike in Chapter 2).
If V and W are finite dimensional, and we have picked an ordered basis $(v_1, v_2, \ldots, v_n)$ for V and an ordered basis $(w_1, w_2, \ldots, w_m)$ for W, then the linear map $L : V \to W$ can be represented (with respect to these bases) by an $m \times n$ matrix, whose j-th column is given by $L(v_j)$, written as a column vector:

$$L(v_j) = \alpha_1 w_1 + \alpha_2 w_2 + \ldots + \alpha_m w_m \ \mapsto \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}.$$

The reason this works is because there is an isomorphism between any vector space over $\mathbb{F}$ of dimension m and $\mathbb{F}^m$, as we will see below.

Here are some examples of where linear maps can and cannot be represented by matrices.

Example 6.15.

(i) Let $V = \mathbb{R}_2[x]$, and define $L : V \to V$ by $p(x) \mapsto p'(x)$ (i.e., differentiate the polynomial once). The basic rules of differentiation show that it is clearly linear (since $(f+g)' = f' + g'$ and $(\alpha f)' = \alpha f'$). We can write this map explicitly as

$$L(\alpha_0 + \alpha_1 x + \alpha_2 x^2) = \alpha_1 + 2\alpha_2 x.$$

We pick the ordered basis $(v_1, v_2, v_3) = (1, x, x^2)$ for V, calculate the effect of L on this basis, and write the answer out with all the coefficients for clarity:

$$\begin{aligned} L(v_1) &= L(1) = 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ L(v_2) &= L(x) = 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ L(v_3) &= L(x^2) = 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2. \end{aligned}$$

So using the method described in the remark above, we use these linear combinations as our column vectors, resulting in the matrix of L with respect to this basis being

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that if we had picked a different ordered basis, we would have ended up with a different matrix.

(ii) Let $V = M_{2\times 2}(\mathbb{R})$, and let $L : V \to V$ given by $L(A) = A^\top$. We saw above that the transpose respects linear combinations, so L is a linear map.
We pick an ordered basis for V (here we are just using the standard basis, but we are additionally choosing an ordering on it):

$$v_1 = E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},\ v_2 = E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},\ v_3 = E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\ v_4 = E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then we calculate the effect L has on this basis:

$$\begin{aligned} L(v_1) &= v_1 = 1 \cdot v_1 + 0 \cdot v_2 + 0 \cdot v_3 + 0 \cdot v_4 \\ L(v_2) &= v_3 = 0 \cdot v_1 + 0 \cdot v_2 + 1 \cdot v_3 + 0 \cdot v_4 \\ L(v_3) &= v_2 = 0 \cdot v_1 + 1 \cdot v_2 + 0 \cdot v_3 + 0 \cdot v_4 \\ L(v_4) &= v_4 = 0 \cdot v_1 + 0 \cdot v_2 + 0 \cdot v_3 + 1 \cdot v_4. \end{aligned}$$

So the matrix representing L with respect to this basis is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(iii) Let $V = \mathbb{R}[x]$ and define $L : V \to V$ by $p(x) \mapsto p'(x)$ again. This map is still linear, but since $\mathbb{R}[x]$ does not have a finite basis we cannot represent the map by a matrix.

(iv) Let $V = \mathcal{I}([a,b],\mathbb{R})$, and define

$$L : V \to V, \quad L(f) = \int_a^x f(t)dt.$$

$\mathcal{I}([a,b],\mathbb{R})$ is the space of real integrable functions on the closed interval $[a,b]$, and the Fundamental Theorem of Calculus tells us that $L(f)$ itself is a differentiable, hence continuous, hence integrable function, so this map is well-defined. Integration is a linear operation, so L is a linear map. But $\mathcal{I}([a,b],\mathbb{R})$ is not finite dimensional, so we cannot represent L by a matrix.

The ideas of **image** and **kernel** (or null-space) apply to any linear map $L : V \to W$. We define

$$\mathrm{Ker}(L) = \{v \in V \mid L(v) = \mathbf{0}\}, \quad \mathrm{Im}(L) = \{w \in W \mid w = L(v), \text{ for } v \in V\}.$$

The proof of Lemma 2.4 still works, with minor alterations:

Lemma 6.16. *For a linear map $L : V \to W$ between vector spaces over the same field $\mathbb{F}$, the image $\mathrm{Im}(L)$ is a subspace of the codomain W, while the kernel $\mathrm{Ker}(L)$ is a subspace of domain V (note that these might not be finite dimensional).*
If V has a finite basis $\{v_1, \dots, v_n\}$ then $\mathrm{Im}(L)$ is spanned by $L(v_1), \dots, L(v_n)$.

Proof. We only need to prove the claim about the image being spanned by the $L(v_j)$. Let w be in the image of L, so $w = L(v)$ for some $v \in V$. We can write $v = \alpha_1 v_1 + \dots + \alpha_n v_n$ for some $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ and then by linearity of L we get

$$w = L(v) = L(\alpha_1 v_1 + \dots + \alpha_n v_n) = \alpha_1 L(v_1) + \dots + \alpha L(v_n).$$

So w is in the span of $L(v_1), \dots, L(v_n)$, so $\mathrm{Im}(L) \subseteq \mathrm{Sp}(L(v_1), \dots, L(v_n))$. The inclusion in the other direction is clearly true since by definition each $L(v_j) \in \mathrm{Im}(L)$, so we get $\mathrm{Im}(L) = \mathrm{Sp}(L(v_1), \dots, L(v_n))$. □

Recall that a map $L : V \to W$ is **injective** (or one-to-one) when $L(u) = L(v)$ implies $u = v$, and **surjective** (or onto) when $\mathrm{Im}(L) = W$. We say it is **bijective** when it is both injective and surjective.
The following result (adapted from Lemma 2.5) still holds.

Lemma 6.17. *For any linear map $L : V \to W$ between vector spaces over the same field $\mathbb{F}$, L is injective precisely when $\mathrm{Ker}(L) = \{\mathbf{0}\}$. Hence L is bijective when $\mathrm{Ker}(L) = \{\mathbf{0}\}$ and $W = \mathrm{Im}(L)$. When L is bijective it has an inverse $L^{-1} : W \to V$ which is also a linear map.*

Definition 6.18. *An invertible linear map $L : V \to W$ is called a **vector space isomorphism**, or simply an **isomorphism**, and we say V **is isomorphic to** W, denoted $V \simeq W$, when such a map exists.*

Remark 6.19. Because composition of two linear maps gives a linear map (check this!), and composition of two bijections gives a bijection (check this!), if $U \simeq V$ and $V \simeq W$ then $U \simeq W$ (recall that this means the relation of isomorphism is *transitive*). Every vector space is isomorphic to itself, $V \simeq V$, with one obvious example of an isomorphism being the **identity map** $I_V : V \to V$, $I_V(v) = v$.
If $V \simeq W$ then clearly $W \simeq V$, so together with the previous facts we get that isomorphism is an *equivalence relation* (since it is reflexive, symmetric and transitive) on the set of vector spaces.

Proof. Most of Lemma 6.17 was proved earlier, but it remains to show that if $L : V \to W$ is a vector space isomorphism of vector spaces over the same field $\mathbb{F}$, then $L^{-1} : W \to V$ is also a linear map. This was clear in the $\mathbb{F}^n$ case, since we could represent a linear map by a matrix, and if the map was bijective then the matrix was invertible, so its inverse matrix would represent the inverse map (which would automatically be linear since it was represented by a matrix). The following proof is more general since doesn't rely on matrices, so it also applies to infinite dimensional spaces.
Let $\alpha, \beta \in \mathbb{F}$ and $w_1, w_2 \in W$. To show that L^{-1} is linear we need to show

$$L^{-1}(\alpha w_1 + \beta w_2) = \alpha L^{-1}(w_1) + \beta L^{-1}(w_2). \tag{$*$}$$

We apply L to the left side of $(*)$:

$$L(L^{-1}(\alpha w_1 + \beta w_2)) = \alpha w_1 + \beta w_2,$$

since L and L^{-1} are inverses of each other, so $L \circ L^{-1} = I_W$, the identity map on W. Then we apply L to the right side of $(*)$:

$$L(\alpha L^{-1}(w_1) + \beta L^{-1}(w_2)) = \alpha L(L^{-1}(w_1)) + \beta L(L^{-1}(w_2)) = \alpha w_1 + \beta w_2,$$

by linearity of L and since $L \circ L^{-1} = I_W$. So we have shown that

$$L(L^{-1}(\alpha w_1 + \beta w_2)) = L(\alpha L^{-1}(w_1) + \beta L^{-1}(w_2)).$$

To proceed we can either observe that L is bijective, hence injective, or we can apply L^{-1} to both sides and use the fact that $L^{-1} \circ L = I_V$. Either way we get

$$L^{-1}(\alpha w_1 + \beta w_2) = \alpha L^{-1}(w_1) + \beta L^{-1}(w_2),$$

as required. So L^{-1} is linear.
(Note that we could also have done the vector addition and scalar multiplication parts of this proof separately, in which case we would need to show $L^{-1}(\alpha w) = \alpha L^{-1}(w)$ and $L^{-1}(w_1 + w_2) = L^{-1}(w_1) + L^{-1}(w_2)$, for all $\alpha \in \mathbb{F}$, $w, w_1, w_2 \in W$.) $\square$

If V and W are both finite dimensional then $\mathrm{Ker}(L)$ and $\mathrm{Im}(L)$ must also be finite dimensional by Lemma 6.9. As before we define the **rank of** L to be $\mathrm{rank}(L) = \dim(\mathrm{Im}(L))$ and the **nullity of** L to be $\mathrm{null}(L) = \dim(\mathrm{Ker}(L))$, when this makes sense. In that case the Rank-Nullity Theorem applies, with the same proof as before.

Rank-Nullity Theorem for Vector Spaces. *Let V, W be finite dimensional vector spaces over the same field $\mathbb{F}$ and $L : V \to W$ be a linear map. Then* $\mathrm{rank}(L) + \mathrm{null}(L) = \dim(V)$.

Example 6.20. Let $V = W = \mathbb{R}_n[x]$, and define $L : V \to V$ by $p(x) \mapsto p'(x)$ again.
We can write an arbitrary polynomial in this space as $p(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_n x^n$, for some $\alpha_j \in \mathbb{R}$. Then we have

$$L(p(x)) = p'(x) = \alpha_1 + 2\alpha_2 x + 3\alpha_3 x^2 + \ldots + n\alpha_n x^{n-1} \in \mathbb{R}_{n-1}[x].$$

So $\mathrm{Im}(L) \subseteq \mathbb{R}_{n-1}[x]$. But note that every polynomial in $\mathbb{R}_{n-1}[x]$ can be obtained like this. So in fact we have $\mathrm{Im}(L) = \mathbb{R}_{n-1}[x]$.
For the kernel, we observe that $L(p(x)) = p'(x) = 0$ if and only if $p(x)$ is a constant polynomial, i.e. $p(x) = \alpha_0$ for all x, for some $\alpha_0 \in \mathbb{F}$. So we can describe the kernel as the span of some constant polynomial, e.g. $\mathrm{Ker}(L) = \mathrm{Sp}(1) = \{\alpha_0 \mid \alpha_0 \in \mathbb{R}\}$.
Note that we have found that $\mathrm{rank}(L) = \dim(\mathrm{Im}(L)) = n$, $\mathrm{null}(L) = \dim(\mathrm{Ker}(L)) = 1$, and $\dim(V) = n + 1$, which is what we would expect from the Rank-Nullity Theorem.
We can also find the matrix of this linear map, as we did earlier with the $n = 2$ case. Using the ordered basis $(v_1, v_2, \ldots, v_{n+1}) = (1, x, \ldots, x^n)$ for V, we can observe that

$$L(v_j) = L(x^{j-1}) = (x^{j-1})' = (j-1)x^{j-2} = (j-1)v_{j-1}.$$

So, as before, we can use this to represent L with respect to this basis as the $(n+1) \times (n+1)$ matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & n \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Note that the rank of this matrix is n since there are clearly n linearly independent columns, and the nullity can be found by the Rank-Nullity Theorem to be $(n+1) - \mathrm{rank}(A) = (n+1) - n = 1$. So we have $\mathrm{rank}(L) = \mathrm{rank}(A)$ and $\mathrm{null}(L) = \mathrm{null}(A)$. We will see below that this happens in general when we can represent a linear map using a matrix.

Corollary 6.21. *If V and W are vector spaces over the same field and $\dim(V) = \dim(W)$ then $V \simeq W$ (they are isomorphic). In particular, every n-dimensional vector space over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.*

The proof is constructive and of practical use.

Proof. It suffices to show that whenever $n = \dim(V)$ then there is an isomorphism $\psi : V \to \mathbb{F}^n$. Then both $V \simeq \mathbb{F}^n$ and $W \simeq \mathbb{F}^n$ hence $V \simeq W$ by the Remark above. Take any ordered basis $\mathcal{B} = (v_1, \ldots, v_n)$ of V and define

$$\psi_{\mathcal{B}}\Big(\sum_{j=1}^{n} \alpha_j v_j\Big) = \sum_{j=1}^{n} \alpha_j \mathbf{e}_j = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

It is easy the check that this map is linear. We can check addition and scalar multiplication separately:

$$\begin{aligned}\psi_{\mathcal{B}}\Big(\sum_{j=1}^{n} \alpha_j v_j + \sum_{j=1}^{n} \beta_j v_j\Big) = \psi_{\mathcal{B}}\Big(\sum_{j=1}^{n} (\alpha_j + \beta_j) v_j\Big) = \sum_{j=1}^{n} (\alpha_j + \beta_j)\mathbf{e}_j = \begin{pmatrix} \alpha_1 + \beta_1 \\ \vdots \\ \alpha_n + \beta_n \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \\ = \psi_{\mathcal{B}}\Big(\sum_{j=1}^{n} \alpha_j v_j\Big) + \psi_{\mathcal{B}}\Big(\sum_{j=1}^{n} \beta_j v_j\Big),\end{aligned}$$

$$\psi_{\mathcal{B}}\Big(\gamma\Big(\sum_{j=1}^{n} \alpha_j v_j\Big)\Big) = \psi_{\mathcal{B}}\Big(\sum_{j=1}^{n} (\gamma\alpha_j) v_j\Big) = \sum_{j=1}^{n} (\gamma\alpha_j)\mathbf{e}_j = \begin{pmatrix} \gamma\alpha_1 \\ \vdots \\ \gamma\alpha_n \end{pmatrix} = \gamma \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \gamma\Big(\psi_{\mathcal{B}}\Big(\sum_{j=1}^{n} \alpha_j v_j\Big)\Big).$$

(We could also have observed that $\psi_{\mathcal{B}}(v_j) = \mathbf{e}_j$ for $j = 1, \ldots n$ and that the map has been defined on the rest of V in such a way as to be linear. This is called **extending by linearity**: if we have specified what a map does to a basis for V, then this automatically defines a linear map on all of V, since, as we saw in Lemma 6.16, the images of the rest of the elements of V can be obtained as linear combinations of the images of the basis vectors.)

We call $\psi_{\mathcal{B}}$ the **coordinate map with respect to the ordered basis** $\mathcal{B}$. Clearly the kernel of this is $\{\mathbf{0}\} \subseteq V$ (since $\psi_{\mathcal{B}}(v) = \mathbf{0}$ implies all of v's coordinates α_j are zero), hence $\mathrm{null}(\psi_{\mathcal{B}}) = 0$. By the Rank-Nullity Theorem $\mathrm{rank}(\psi_{\mathcal{B}}) = n$ and therefore $\mathrm{Im}(\psi_{\mathcal{B}}) = \mathbb{F}^n$ (the only n-dimensional subspace of $\mathbb{F}^n$ is $\mathbb{F}^n$ itself). Hence by Lemma 6.17 $\psi_{\mathcal{B}}$ is an isomorphism. □

A consequence of this result is that when $L : V \to W$ is represented by a $p \times n$ matrix A the rank and nullity of L are equal to the rank and nullity of the matrix A.

Note that the proof above gives an isomorphism $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$ *for each* ordered basis $\mathcal{B}$ of V. So while any vector space V of dimension n over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$, to go between the two we need to *choose* a basis to work with. This makes it easier to do concrete matrix calculations, but the more abstract vector space approach (see e.g. the proof of Lemma 6.17) has its advantages - for instance proofs are often simpler and more "elegant" if we avoid having to choose a basis, and we can be certain that the results we prove do not depend on the arbitrary choice we made.

6.3. **Change of basis.** We have seen (Corollary 6.21) that any finite dimensional vector space V of dimension n over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$. In the proof of that result we used the isomorphism $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$ (the coordinate map with respect to the ordered basis $\mathcal{B}$).

Given an ordered basis $\mathcal{B} = (v_1, \dots, v_n)$, we defined $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$ by

$$\psi_{\mathcal{B}}(\alpha_1 v_1 + \ldots + \alpha_n v_n) = \alpha_1 \mathbf{e}_1 + \ldots + \alpha_n \mathbf{e}_n = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

Recall that this is the unique linear map for which $\psi(v_j) = \mathbf{e}_j$. Conversely, by a generalisation of Lemma 2.9, vector space isomorphisms map bases to bases, so given a vector space isomorphism $\psi : V \to \mathbb{F}^n$, its inverse $\psi^{-1} : \mathbb{F}^n \to V$ is also an isomorphism, and ψ^{-1} maps the standard basis $(\mathbf{e}_1, \dots \mathbf{e}_n)$ of $\mathbb{F}^n$ to an ordered basis $(v_1, \dots, v_n)$ for V, with $v_j = \psi^{-1}(\mathbf{e}_j)$.

Thus, we have the following result.

Lemma 6.22. *For a finite dimensional vector space V, there is a bijective correspondence between coordinate maps (i.e. isomorphisms) $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$ and ordered bases $\mathcal{B} = (v_1, \dots, v_n)$ of V.*

Note that the above reasoning shows that *all* vector space isomorphisms from V to $\mathbb{F}^n$ have the form of a coordinate map for some ordered basis.

Now suppose we choose two different ordered bases $\mathcal{A} = (w_1, \dots, w_n)$ and $\mathcal{B} = (v_1, \dots, v_n)$ for a vector space V. Each gives a coordinate map, $\psi_{\mathcal{A}} : V \to \mathbb{F}^n$ and $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$. How are these related? Since each coordinate map is an isomorphism, it is invertible, so we have $\psi_{\mathcal{A}}^{-1} \circ \psi_{\mathcal{A}} = I_V$, the identity map on V. Applying $\psi_{\mathcal{B}}$ to both sides gives us

$$\psi_{\mathcal{B}} = \psi_{\mathcal{B}} \circ \psi_{\mathcal{A}}^{-1} \circ \psi_{\mathcal{A}}.$$

But now observe that $\psi_{\mathcal{A}}^{-1}$ is a linear map from $\mathbb{F}^n$ to V, and $\psi_{\mathcal{B}}$ is a linear map from V to $\mathbb{F}^n$, so their composite $\psi_{\mathcal{B}} \circ \psi_{\mathcal{A}}^{-1}$ is a linear map from $\mathbb{F}^n$ to $\mathbb{F}^n$, and is therefore represented by a matrix which we will call the **change of basis matrix** (or the **transition matrix**) and denote by $C_{\mathcal{A}}^{\mathcal{B}}$. The previous equation tells us that

$$\psi_{\mathcal{B}}(v) = C_{\mathcal{A}}^{\mathcal{B}} \psi_{\mathcal{A}}(v), \quad \text{for all } v \in V.$$

To find the matrix explicitly, we apply the linear map $\psi_{\mathcal{B}} \circ \psi_{\mathcal{A}}^{-1}$ to the standard basis $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ of $\mathbb{F}^n$:

$$(\psi_{\mathcal{B}} \circ \psi_{\mathcal{A}}^{-1})(\mathbf{e}_j) = \psi_{\mathcal{B}}(\psi_{\mathcal{A}}^{-1}(\mathbf{e}_j)) = \psi_{\mathcal{B}}(w_j),$$

since by definition $\psi_{\mathcal{A}}(w_j) = \mathbf{e}_j$, so $\psi_{\mathcal{A}}^{-1}(\mathbf{e}_j) = w_j$. This calculation gives us the j-th column of the matrix $C_{\mathcal{A}}^{\mathcal{B}}$ representing $\psi_{\mathcal{B}} \circ \psi_{\mathcal{A}}^{-1}$. Note that $\psi_{\mathcal{B}}(w_j)$ is by definition the column vector of the coordinates of the vector w_j written in terms of the basis $\mathcal{B}$. So the j-th column of the matrix $C_{\mathcal{A}}^{\mathcal{B}}$ is given by coordinates of the vector w_j (from the basis $\mathcal{A}$) written in terms of the basis $\mathcal{B}$.

In other words, if we write the basis vector $w_j \in \mathcal{A}$ in terms of the other basis $\mathcal{B}$,

$$w_j = C_{1j}v_1 + \ldots + C_{nj}v_n,$$

for some $C_{1j}, \cdots, C_{nj} \in \mathbb{F}$, then we see that the j-th column of the matrix $C_{\mathcal{A}}^{\mathcal{B}}$ is given by

$$\psi_{\mathcal{B}}(w_j) = C_{1j}\mathbf{e}_1 + \ldots + C_{nj}\mathbf{e}_n = \begin{pmatrix} C_{1j} \\ \vdots \\ C_{nj} \end{pmatrix}.$$

Then the change of basis matrix is given by

$$C_{\mathcal{A}}^{\mathcal{B}} = \begin{pmatrix} C_{11} & \ldots & C_{1n} \\ \vdots & & \vdots \\ C_{n1} & \ldots & C_{nn} \end{pmatrix}.$$

We can summarise this as follows.

Lemma 6.23. *Let $\psi_{\mathcal{A}}$ and $\psi_{\mathcal{B}}$ be two coordinate maps on a finite dimensional vector space V, corresponding to (ordered) bases $\mathcal{A} = (w_1, \dots, w_n)$ and $\mathcal{B} = (v_1, \dots, v_n)$. Then $\psi_{\mathcal{B}}(v) = C_{\mathcal{A}}^{\mathcal{B}}\, \psi_{\mathcal{A}}(v)$ for all $v \in V$, where $C_{\mathcal{A}}^{\mathcal{B}}$ is the change of basis matrix, whose columns are the coordinates of basis $\mathcal{A}$ vectors in terms of basis $\mathcal{B}$ vectors.*
Interpreting $C_{\mathcal{A}}^{\mathcal{B}}$ as a linear map $\mathbb{F}^n \to \mathbb{F}^n$, we have $\psi_{\mathcal{B}} = C_{\mathcal{A}}^{\mathcal{B}} \circ \psi_{\mathcal{A}}$.*

Change of basis matrices possess some natural properties, which are easily proved from the defining equation $C_{\mathcal{A}}^{\mathcal{B}}\mathbf{e}_j = \psi_{\mathcal{B}}(w_j)$.

Lemma 6.24. *For three bases $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ we have $C_{\mathcal{A}}^{\mathcal{C}} = C_{\mathcal{B}}^{\mathcal{C}} C_{\mathcal{A}}^{\mathcal{B}}$. Since $C_{\mathcal{A}}^{\mathcal{A}} = I_n$, the identity matrix, it follows that $C_{\mathcal{B}}^{\mathcal{A}} = (C_{\mathcal{A}}^{\mathcal{B}})^{-1}$.*

---

*This is known as “abuse of notation”, since strictly speaking we should use a different symbol for a matrix and the linear map $\mathbb{F}^n \to \mathbb{F}^n$ represented by it. Abuse of notation is common in advanced mathematics, and physics, since it often makes expressions easier to read. It is used when there is an obvious one-to-one correspondence between the two objects being given the same name, and the intended meaning should be clear from context.

Example 6.25. Let $V = \mathbb{R}_2[x]$, the space of polynomials with real coefficients of degree at most 2. We have seen that the standard basis $\mathcal{B} = (1, x, x^2)$ is a basis for V.
Consider $\mathcal{A} = (1+x, x, 1+x^2)$. We claim that this is also a basis for V. To verify this, we check that the three vectors are linearly independent. Assume $\alpha(1+x)+\beta x+\gamma(1+x^2) = 0$ for some $\alpha, \beta, \gamma \in \mathbb{R}$. Then $(\alpha+\gamma)+(\alpha+\beta)x+\gamma x^2 = 0$, which is only true when $\alpha+\gamma = 0$, $\alpha+\beta = 0$ and $\gamma = 0$. Combining the first and last equations gives us $\alpha = 0$, and together with the second equation we have $\beta = 0$. So the vectors are linearly independent, and we know that any three linearly independent vectors in V must be a basis for V since $\dim(V) = 3$.
The coordinate map for $\mathcal{B}$ is easy to write down:

$$\psi_{\mathcal{B}} : V \to \mathbb{R}^3, \quad \psi_{\mathcal{B}}(\alpha_0 + \alpha_1 x + \alpha_2 x^2) = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

From this we can work out $\psi_{\mathcal{A}}$ by first working out $C^{\mathcal{A}}_{\mathcal{B}}$. To do this, we write the elements of $\mathcal{B}$ in terms of $\mathcal{A}$:

$$\begin{aligned} 1 &= 1 \cdot (1+x) + (-1) \cdot x + 0 \cdot (1+x^2) \\ x &= 0 \cdot (1+x) + 1 \cdot x + 0 \cdot (1+x^2) \\ x^2 &= (-1) \cdot (1+x) + 1 \cdot x + 1 \cdot (1+x^2). \end{aligned}$$

Each column of $C^{\mathcal{A}}_{\mathcal{B}}$ then comes from the coordinates in these equations:

$$C^{\mathcal{A}}_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

To find $C^{\mathcal{B}}_{\mathcal{A}}$, we can either compute the inverse of $C^{\mathcal{A}}_{\mathcal{B}}$, or proceed as above since it is easy to express the vectors in $\mathcal{A}$ in terms of the basis $\mathcal{B}$:

$$\begin{aligned} 1+x &= 1 \cdot 1 + 1 \cdot x + 0 \cdot x^2, \\ x &= 0 \cdot 1 + 1 \cdot x + 0 \cdot x^2, \\ 1+x^2 &= 1 \cdot 1 + 0 \cdot x + 1 \cdot x^2. \end{aligned}$$

Thus

$$C^{\mathcal{B}}_{\mathcal{A}} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is straightforward to check that $C^{\mathcal{A}}_{\mathcal{B}} C^{\mathcal{B}}_{\mathcal{A}} = I_3$.

Since $\psi_{\mathcal{A}}(v) = C_{\mathcal{B}}^{\mathcal{A}}\psi_{\mathcal{B}}(v)$ for all $v \in V$, it follows that the coordinate map $\psi_{\mathcal{A}} : V \to \mathbb{R}^3$ is given by

$$\psi_{\mathcal{A}}(\alpha_0 + \alpha_1 x + \alpha_2 x^2) = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_0 - \alpha_2 \\ -\alpha_0 + \alpha_1 + \alpha_2 \\ \alpha_2 \end{pmatrix}.$$

Multiplication by $C_{\mathcal{B}}^{\mathcal{A}}$ takes coordinate vectors written in terms of $\mathcal{B}$ to coordinate vectors written in terms of $\mathcal{A}$. For a concrete example, let $p(x) = 1 + 2x + 3x^2$. We can easily write $p(x)$ as a coordinate vector in terms of $\mathcal{B}$:

$$\psi_{\mathcal{B}}(p(x)) = \psi_{\mathcal{B}}(1 \cdot 1 + 2 \cdot x + 3 \cdot x^2) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Multiplying by $C_{\mathcal{B}}^{\mathcal{A}}$ gives us

$$C_{\mathcal{B}}^{\mathcal{A}}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \\ 3 \end{pmatrix},$$

which is the coordinate vector of $p(x)$ in terms of $\mathcal{A}$, since

$$(-2) \cdot (1 + x) + 4 \cdot x + 3 \cdot (1 + x^2) = -2 - 2x + 4x + 3 + 3x^2 = 1 + 2x + 3x^2 = p(x).$$

So we have verified that

$$C_{\mathcal{B}}^{\mathcal{A}}\psi_{\mathcal{B}}(p(x)) = C_{\mathcal{B}}^{\mathcal{A}}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \\ 3 \end{pmatrix} = \psi_{\mathcal{A}}(p(x)).$$

Recall from Example 6.15 the differentiation map $L : V \to V$, $p(x) \mapsto p'(x)$. We saw that this linear map is represented with respect to the standard basis $\mathcal{B}$ by the matrix

$$M_{\mathcal{B}}(L) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Let us now find the matrix of this map with respect to the basis $\mathcal{A}$. To do this, we apply L to the basis vectors in $\mathcal{A}$, and write the answer in terms of $\mathcal{A}$:

$$\begin{aligned} L(1 + x) &= (1 + x)' = 1 = 1 \cdot (1 + x) + (-1) \cdot x + 0 \cdot (1 + x^2) \\ L(x) &= x' = 1 = 1 \cdot (1 + x) + (-1) \cdot x + 0 \cdot (1 + x^2) \\ L(1 + x^2) &= (1 + x^2)' = 2x = 0 \cdot (1 + x) + 2 \cdot x + 0 \cdot (1 + x^2). \end{aligned}$$

So the matrix of L with respect to the basis $\mathcal{A}$ is:

$$M_{\mathcal{A}}(L) = \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

How are the matrices $M_{\mathcal{B}}(L)$ and $M_{\mathcal{A}}(L)$ related? To describe this we need to study some more theory.

Suppose $L : V \to V$ is a linear map on a finite dimensional vector space. We have seen that by choosing an ordered basis $\mathcal{B} = (v_1, \ldots, v_n)$ for V we can represent L as a matrix, the **matrix representing of L with respect to the basis** $\mathcal{B}$, which we denote by $M_{\mathcal{B}}(L)$. There are two equivalent ways of describing this matrix.

First, L is uniquely determined by its action on the basis vectors of $\mathcal{B}$, so, as we have seen in earlier examples, the j-th column of $M_{\mathcal{B}}(L)$ can be computed by applying L to the basis vector v_j, and writing the coefficients (with respect to $\mathcal{B}$) as a column vector.

A more abstract (but equivalent) description of this matrix is given by noting that the basis $\mathcal{B}$ of V defines the coordinate map $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$, and thus we obtain a linear map

$$M_{\mathcal{B}}(L) : \mathbb{F}^n \to \mathbb{F}^n, \quad M_{\mathcal{B}}(L) : \mathbb{F}^n \xrightarrow{\psi_{\mathcal{B}}^{-1}} V \xrightarrow{L} V \xrightarrow{\psi_{\mathcal{B}}} \mathbb{F}^n.$$

So we can define the map $M_{\mathcal{B}}(L)$ by $M_{\mathcal{B}}(L) = \psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1}$, and also use the notation $M_{\mathcal{B}}(L)$ to denote the matrix representing this map with respect to the standard basis of $\mathbb{F}^n$.*

Why are these two descriptions of $M_{\mathcal{B}}(L)$ equivalent? We have seen that the matrix $M_{\mathcal{B}}(L)$, for a given basis $\mathcal{B} = (v_1, \ldots, v_n)$, is obtained by applying L to v_j, writing the result in terms of $\mathcal{B}$ and then writing the coordinates obtained as the j-th column vector. We can describe this in terms of the coordinate map $\psi_{\mathcal{B}}$: we start with $\mathbf{e}_j$ in the standard basis of $\mathbb{F}^n$ and apply $\psi_{\mathcal{B}}^{-1}$ to get v_j; then we apply L, and then apply $\psi_{\mathcal{B}}$ to get the coordinates in terms of $\mathcal{B}$.

More concretely, since $\mathcal{B}$ is a basis, we can write

$$L(v_j) = A_{1j}v_1 + \ldots + A_{nj}v_n,$$

for some $A_{1j} \cdots, A_{nj} \in \mathbb{F}$.

*This is, as mentioned in the footnote above, abuse of notation, since strictly speaking we should be using different notation for a linear map and the matrix representing it.

Then $M_{\mathcal{B}}(L)$ is the $n \times n$ matrix (A_{ij}). Let us verify this. Recall that $M_{\mathcal{B}}(L)\mathbf{e}_j$ gives the j-th column of the matrix $M_{\mathcal{B}}(L)$. Now

$$\begin{aligned} M_{\mathcal{B}}(L)\mathbf{e}_j &= (\psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1})(\mathbf{e}_j) \\ &= \psi_{\mathcal{B}}\left(L\left(\psi_{\mathcal{B}}^{-1}(\mathbf{e}_j)\right)\right) \\ &= \psi_{\mathcal{B}}\left(L\left(v_j\right)\right) \\ &= \psi_{\mathcal{B}}(A_{1j}v_1 + \ldots + A_{nj}v_n) \\ &= A_{1j}\psi_{\mathcal{B}}(v_1) + \ldots A_{nj}\psi_{\mathcal{B}}(v_n) \\ &= A_{1j}\mathbf{e}_1 + \ldots A_{nj}\mathbf{e}_n. \end{aligned}$$

Therefore

$$M_{\mathcal{B}}(L) = \begin{pmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \ldots & A_{nn} \end{pmatrix}.$$

Proposition 6.26. * *Let V be a vector space over a field $\mathbb{F}$, $L_1, L_2 : V \to V$ be two linear transformations, and $\mathcal{B}$ be a basis for V. Then:*

(i) *for any scalars $\alpha, \beta \in \mathbb{F}$*

$$M_{\mathcal{B}}(\alpha L_1 + \beta L_2) = \alpha M_{\mathcal{B}}(L_1) + \beta M_{\mathcal{B}}(L_2),$$

(ii) $M_{\mathcal{B}}(L_1 \circ L_2) = M_{\mathcal{B}}(L_1)M_{\mathcal{B}}(L_2)$.
In particular, when $L : V \to V$ is invertible, $M_{\mathcal{B}}(L^{-1}) = M_{\mathcal{B}}(L)^{-1}$.

Proof. (i) By definition, viewed as a linear map from $\mathbb{F}^n$ to $\mathbb{F}^n$, we have $M_{\mathcal{B}}(L) = \psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1}$. Let $\mathbf{v}$ be an arbitrary column vector in $\mathbb{F}^n$, and let v be the corresponding vector in

*For those on the Pure Maths stream, this proposition is saying that the map $L \mapsto M_{\mathcal{B}}(L)$ is both linear and respects function composition, so it is an algebra homomorphism (where an *algebra* is, roughly speaking, an object with three operations: addition, multiplication/composition and scalar multiplication). You will see more maps like this if you go on to study Representation Theory, which involves studying abstract objects (like groups) by representing them with more concrete objects (like matrices), in such a way that the concrete objects behave like the abstract ones do.

V with respect to the coordinate map $\psi_{\mathcal{B}}$, i.e. we have $v = \psi_{\mathcal{B}}^{-1}(\mathbf{v})$, so $\mathbf{v} = \psi_{\mathcal{B}}(v)$. Then

$$\begin{aligned}
M_{\mathcal{B}}(\alpha L_1 + \beta L_2)(\mathbf{v}) &= (\psi_{\mathcal{B}} \circ (\alpha L_1 + \beta L_2) \circ \psi_{\mathcal{B}}^{-1})(\mathbf{v}) \\
&= (\psi_{\mathcal{B}} \circ (\alpha L_1 + \beta L_2) \circ \psi_{\mathcal{B}}^{-1})(\psi_{\mathcal{B}}(v)) \\
&= (\psi_{\mathcal{B}} \circ (\alpha L_1 + \beta L_2) \circ \psi_{\mathcal{B}}^{-1} \circ \psi_{\mathcal{B}})(v) \\
&= (\psi_{\mathcal{B}} \circ (\alpha L_1 + \beta L_2))(v) \\
&= \psi_{\mathcal{B}}(\alpha L_1(v) + \beta L_2(v)) \\
&= \alpha\psi_{\mathcal{B}}(L_1(v)) + \beta\psi_{\mathcal{B}}(L_2(v)) \\
&= \alpha\psi_{\mathcal{B}}(L_1(\psi_{\mathcal{B}}^{-1}(\mathbf{v})) + \beta\psi_{\mathcal{B}}(L_2(\psi_{\mathcal{B}}^{-1}(\mathbf{v})) \\
&= \alpha(\psi_{\mathcal{B}} \circ L_1 \circ \psi_{\mathcal{B}}^{-1})(\mathbf{v}) + \beta(\psi_{\mathcal{B}} \circ L_2 \circ \psi_{\mathcal{B}}^{-1})(\mathbf{v}) \\
&= \alpha M_{\mathcal{B}}(L_1)(\mathbf{v}) + \beta M_{\mathcal{B}}(L_2)(\mathbf{v}) \\
&= (\alpha M_{\mathcal{B}}(L_1) + \beta M_{\mathcal{B}}(L_2))(\mathbf{v}).
\end{aligned}$$

Note that we have used the fact that composition of functions is associative here, to allow us to rearrange the brackets. This equality true for all $\mathbf{v} \in \mathbb{F}^n$, so we have an equality of linear maps $\mathbb{F}^n \to \mathbb{F}^n$:

$$M_{\mathcal{B}}(\alpha L_1 + \beta L_2) = \alpha M_{\mathcal{B}}(L_1) + \beta M_{\mathcal{B}}(L_2),$$

and hence also an equality of the corresponding matrices of these maps (with respect to the standard basis of $\mathbb{F}^n$).

(ii) Again, we first interpret $M_{\mathcal{B}}(L)$ as a linear map $\mathbb{F}^n \to \mathbb{F}^n$, and use the definition $M_{\mathcal{B}}(L) = \psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1}$.

We have

$$\begin{aligned}
M_{\mathcal{B}}(L_1) \circ M_{\mathcal{B}}(L_2) &= (\psi_{\mathcal{B}} \circ L_1 \circ \psi_{\mathcal{B}}^{-1}) \circ (\psi_{\mathcal{B}} \circ L_2 \circ \psi_{\mathcal{B}}^{-1}) \\
&= \psi_{\mathcal{B}} \circ L_1 \circ (\psi_{\mathcal{B}}^{-1} \circ \psi_{\mathcal{B}}) \circ L_2 \circ \psi_{\mathcal{B}}^{-1} \\
&= \psi_{\mathcal{B}} \circ L_1 \circ L_2 \circ \psi_{\mathcal{B}}^{-1} \\
&= M_{\mathcal{B}}(L_1 \circ L_2),
\end{aligned}$$

where we again have used associativity of function composition. Since composition of linear maps corresponds to matrix multiplication, viewed instead as matrices we get the desired expression

$$M_{\mathcal{B}}(L_1)M_{\mathcal{B}}(L_2) = M_{\mathcal{B}}(L_1 \circ L_2).$$

Now when L is invertible we can apply this to the identity

$$L^{-1} \circ L = L \circ L^{-1} = I_V,$$

where I_V is the identity map on V, to get

$$M_{\mathcal{B}}(L^{-1})M_{\mathcal{B}}(L) = M_{\mathcal{B}}(L^{-1} \circ L) = M_{\mathcal{B}}(I_V) = I_n,$$
$$M_{\mathcal{B}}(L)M_{\mathcal{B}}(L^{-1}) = M_{\mathcal{B}}(L \circ L^{-1}) = M_{\mathcal{B}}(I_V) = I_n.$$

Hence the inverse of the matrix $M_{\mathcal{B}}(L)$ is the matrix $M_{\mathcal{B}}(L^{-1})$. □

Finally, we want to know how the matrix representation changes if we change bases.

Theorem 6.27. *Let $L : V \to V$ be a linear transformation, and $\mathcal{A}, \mathcal{B}$ be two bases for V. Then*

$$M_{\mathcal{B}}(L) = C^{\mathcal{B}}_{\mathcal{A}} M_{\mathcal{A}}(L)\, (C^{\mathcal{B}}_{\mathcal{A}})^{-1} = (C^{\mathcal{A}}_{\mathcal{B}})^{-1} M_{\mathcal{A}}(L)\, C^{\mathcal{A}}_{\mathcal{B}}.$$

In particular, $M_{\mathcal{A}}(L)$ and $M_{\mathcal{B}}(L)$ are similar matrices.

Note: the above expression should remind you of expressions like $PAP^{-1} = D$, which you saw in Section 2.5 when discussing diagonalising matrices. Diagonalising an $n \times n$ matrix A is a special case of Theorem 6.27, since it involves finding a basis for $\mathbb{F}^n$ consisting of eigenvectors for A (if it exists) and performing a change of basis. The matrix P you encountered there was exactly the change of basis matrix $C^{\mathcal{B}}_{\mathcal{A}}$ going from the standard basis of $\mathbb{F}^n$ to one consisting of eigenvectors of A.

Proof. Again, we start by considering $M_{\mathcal{A}}(L)$ as a linear map $\mathbb{F}^n \to \mathbb{F}^n$ defined by $M_{\mathcal{B}}(L) = \psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1}$.

Recall that the change of basis matrix $C^{\mathcal{B}}_{\mathcal{A}}$, viewed as a linear map $\mathbb{F}^n \to \mathbb{F}^n$, satisfies $\psi_{\mathcal{B}} = C^{\mathcal{B}}_{\mathcal{A}} \circ \psi_{\mathcal{A}}$. Then we have

$$\begin{aligned} M_{\mathcal{B}}(L) &= \psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1} \\ &= (C^{\mathcal{B}}_{\mathcal{A}} \circ \psi_{\mathcal{A}}) \circ L \circ (C^{\mathcal{B}}_{\mathcal{A}} \circ \psi_{\mathcal{A}})^{-1} \\ &= C^{\mathcal{B}}_{\mathcal{A}} \circ \psi_{\mathcal{A}} \circ L \circ (\psi_{\mathcal{A}})^{-1} \circ (C^{\mathcal{B}}_{\mathcal{A}})^{-1} \\ &= C^{\mathcal{B}}_{\mathcal{A}} \circ (\psi_{\mathcal{A}} \circ L \circ \psi_{\mathcal{A}}^{-1}) \circ (C^{\mathcal{B}}_{\mathcal{A}})^{-1} \\ &= C^{\mathcal{B}}_{\mathcal{A}} \circ M_{\mathcal{A}}(L) \circ (C^{\mathcal{B}}_{\mathcal{A}})^{-1}. \end{aligned}$$

Note here we have used the basic fact that if f and g are any invertible maps, then $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.

We have now found an equality of linear maps $\mathbb{F}^n \to \mathbb{F}^n$, so we get a corresponding equality of matrices (with respect to the standard basis of $\mathbb{F}^n$), where again function composition becomes matrix multiplication:

$$M_{\mathcal{B}}(L) = C^{\mathcal{B}}_{\mathcal{A}} M_{\mathcal{A}}(L)\, (C^{\mathcal{B}}_{\mathcal{A}})^{-1}.$$

The second equality is obtained by using the fact that $C^{\mathcal{B}}_{\mathcal{A}} = (C^{\mathcal{A}}_{\mathcal{B}})^{-1}$. □

Example 6.28. We return now to our previous example, Example 6.25. Recall that we had $V = \mathbb{R}_2[x]$, and two ordered bases for V, the standard basis and $\mathcal{B} = (1, x, x^2)$ and $\mathcal{A} = (1 + x, x, 1 + x^2)$. Previously we found the two change of basis matrices with respect to these bases:

$$C_{\mathcal{A}}^{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad C_{\mathcal{B}}^{\mathcal{A}} = (C_{\mathcal{A}}^{\mathcal{B}})^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

For the differentiation map $L : V \to V$ given by $p(x) \mapsto p'(x)$ we have computed the matrices of L with respect to these two bases:

$$M_{\mathcal{B}}(L) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \qquad M_{\mathcal{A}}(L) = \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

We can now verify that the matrix equation in Theorem 6.27 holds:

$$\begin{aligned} C_{\mathcal{A}}^{\mathcal{B}} M_{\mathcal{A}}(L) (C_{\mathcal{A}}^{\mathcal{B}})^{-1} &= \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \\ &= M_{\mathcal{B}}(L). \end{aligned}$$

Remark 6.29. In this section we have considered linear maps from a vector space to itself. More generally, any linear map $L : W \to V$ between finite dimensional vector spaces can be represented by a matrix, defined by

$$M_{\mathcal{A}}^{\mathcal{B}}(L) : \mathbb{F}^n \to \mathbb{F}^m, \quad M_{\mathcal{A}}^{\mathcal{B}}(L) = \psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{A}}^{-1},$$

by choosing a basis $\mathcal{A} = (w_1, \dots, w_n)$ for W and a basis $\mathcal{B} = (v_1, \dots, v_m)$ for V. When $W = V$ and $\mathcal{A} = \mathcal{B}$ this recovers our definition above, i.e., $M_{\mathcal{A}}^{\mathcal{A}}(L) = M_{\mathcal{A}}(L)$. It is an exercise to check that a different choice of bases, say $\mathcal{A}'$ for W and $\mathcal{B}'$ for V, results in a change of bases formula

$$M_{\mathcal{A}'}^{\mathcal{B}'}(L) = C_{\mathcal{B}}^{\mathcal{B}'} M_{\mathcal{A}}^{\mathcal{B}}(L) C_{\mathcal{A}'}^{\mathcal{A}}.$$

6.4. **Eigenvectors and Eigenvalues.** In this subsection we consider linear maps from a vector space V to itself. The definitions of eigenvectors and eigenvalues are almost exactly as in Chapter 2.

Definition 6.30. *For a linear map $L : V \to V$ on a vector space V over a field $\mathbb{F}$, an **eigenvector of** L is a non-zero vector $v \in V$ for which $L(v) = \lambda v$ for some scalar $\lambda \in \mathbb{F}$. In that case λ is called an **eigenvalue of** L.*

*The set of all eigenvalues of L is called the **spectrum** of L. It can be thought of as the set of all $\lambda \in \mathbb{F}$ for which $L - \lambda I_V$ is not invertible (i.e. has non-trivial kernel).*

When V is n-dimensional we can represent every linear map on V by an $n \times n$ matrix, and then all the theory of eigenvectors and eigenvalues from earlier carries over.

Example 6.31. You have seen that similar matrices have the same eigenvalues, so all matrices representing the same linear map have the same eigenvalues. We will also see that a linear map $L : V \to V$ has the same eigenvalues as the matrix $M_{\mathcal{B}}(L)$ for any basis $\mathcal{B}$ of V.
Consider again $V = \mathbb{R}_2[x]$ together with the differentiation map $L : V \to V$, $p(x) \mapsto p'(x)$. With respect to the standard basis $\mathcal{B} = (1, x, x^2)$ we have seen that L is represented by the matrix

$$M_{\mathcal{B}}(L) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

This matrix is upper triangular, so we know its eigenvalues are given by its diagonal entries. So this matrix (and hence also L) has only one eigenvalue, $\lambda = 0$ (with algebraic multiplicity 3 since the characteristic polynomial is λ^3).
Using methods seen earlier we find that the corresponding eigenspace is spanned by one vector, $\mathbf{e}_1 = \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)$, so it has dimension 1, hence the geometric multiplicity of the eigenvalue $\lambda = 0$ is 1.
We can then use the coordinate map to get a corresponding eigenvector for L: $\psi_{\mathcal{B}}^{-1}(\mathbf{e}_1) = v_1 = 1$.
This is indeed an eigenvector for the linear map L since we have $1 \neq 0$ and

$$L(1) = 1' = 0 = 0 \cdot 1 = \lambda \cdot 1.$$

Note that this is to be expected, since we know that eigenvectors v with eigenvalue 0 satisfy $L(v) = 0 \cdot v = 0$, so they are precisely elements of the kernel of L, and we have previously seen that the kernel of L consists of constant polynomials, i.e. scalar multiples of 1.

The method shown above relied on the fact that a linear map and any matrix representing it have the same eigenvalues. We state this as a theorem.

Theorem 6.32. *Let $L : V \to V$ be an n-dimensional vector space over a field $\mathbb{F}$. Then for each matrix representation $A = M_{\mathcal{B}}(L)$ of L, $v \in V$ is an eigenvector of L with eigenvalue λ if and only if the coordinate vector $\mathbf{v} = \psi_{\mathcal{B}}(v)$ is an eigenvector of A with eigenvalue λ.*
Moreover, the characteristic polynomial $\det(\lambda I_n - A)$ depends only on L, not $\mathcal{B}$. Hence we can define this to be the characteristic polynomial $c_L(\lambda)$ of L.

So all the previous results about eigenvalues and eigenvectors of $n \times n$ matrices apply to the more general setting of linear maps from V to V.

Proof. For $\mathcal{B} = (v_1, \dots, v_n)$ an ordered basis of V, recall that $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$ is the coordinate map, which is a vector space isomorphism and satisfies $\psi_{\mathcal{B}}(v_j) = \mathbf{e}_j$. For $v \in V$, let $\mathbf{v} = \psi_{\mathcal{B}}(v)$, so $\mathbf{v}$ is the column vector in $\mathbb{F}^n$ corresponding to $v \in V$.
Recall that, viewed as linear maps $\mathbb{F}^n \to \mathbb{F}^n$, we have $M_{\mathcal{B}}(L) = \psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1}$.
Let $\lambda \in \mathbb{F}$. Note that, applying $\psi_{\mathcal{B}}$ to both sides of the eigenvector equation $L(v) = \lambda v$, we have

$$\begin{aligned}\psi_{\mathcal{B}}(L(v)) &= (\psi_{\mathcal{B}} \circ L)(v)\\ &= (\psi_{\mathcal{B}} \circ L \circ \psi_{\mathcal{B}}^{-1} \circ \psi_{\mathcal{B}})(v)\\ &= (M_{\mathcal{B}}(L) \circ \psi_{\mathcal{B}})(v)\\ &= M_{\mathcal{B}}(L)(\psi_{\mathcal{B}}(v))\\ &= M_{\mathcal{B}}(L)(\mathbf{v})\end{aligned}$$

and

$$\psi_{\mathcal{B}}(\lambda v) = \lambda \psi_{\mathcal{B}}(v) = \lambda \mathbf{v},$$

where we have used the fact that $\psi_{\mathcal{B}}$ is linear. Since $\psi_{\mathcal{B}}$ is injective (since it is bijective) we have

$$Lv = \lambda v \;\Leftrightarrow\; \psi_{\mathcal{B}}(L(v)) = \psi_{\mathcal{B}}(\lambda v) \Leftrightarrow\; M_{\mathcal{B}}(L)(\mathbf{v}) = \lambda \mathbf{v} \Leftrightarrow\; A\mathbf{v} = \lambda \mathbf{v}.$$

So v is an eigenvector of L with eigenvalue λ if and only if $\mathbf{v} = \psi_{\mathcal{B}}(v)$ is an eigenvector of the matrix $A = M_{\mathcal{B}}(L)$ with eigenvalue λ.
Finally we have to show that the characteristic polynomial $\det(\lambda I_n - A)$ depends only on L and not the choice of ordered basis $\mathcal{B}$.
We could argue that its roots are the eigenvalues of L and these depend only on L (and the characteristic polynomial is monic so there is only one possible choice of polynomial with this set of roots).
Alternatively, in Theorem 6.27 we saw that if $\mathcal{A}$ is another ordered basis of V, then the

matrices $M_{\mathcal{A}}(L)$ and $M_{\mathcal{B}}(L)$ are similar (since there exists an invertible matrix $P = C_{\mathcal{B}}^{\mathcal{A}}$ such that $M_{\mathcal{B}}(L) = P^{-1}M_{\mathcal{A}}(L)P$), and we saw in Lemma 2.17 that similar matrices have the same characteristic polynomial. □

An important point here is that we can use any choice of ordered basis to calculate eigenvalues and eigenvectors of L. The matrix A will look different for different choices, but the eigenvalues will not change. In particular, we are welcome to use a basis with respect to which A is easy to work with (e.g. diagonal).

If V is a finite dimensional complex vector space then any linear map $L : V \to V$ must have at least one eigenvector (this follows from the Fundamental Theorem of Algebra, which in this case says that the characteristic polynomial (which is by definition non-constant) must have a root over $\mathbb{C}$). This is not necessarily true over other fields (and we have seen examples of linear maps on e.g. $\mathbb{R}^2$ with no eigenvectors).

When V is not finite dimensional things become more complicated since we can no longer rely on matrices. Here are some examples.

Example 6.33.

(i) Let $V = \mathbb{R}[x]$, the vector space of polynomials in x of any degree with real coefficients. For $p(x) \in V$ define $L(p) = \int_0^x p(t)dt$. Note that $L(p)$ is a polynomial in x, so $L(p) \in V$, and integration is linear, so $L : V \to V$ is a linear map. We claim that this map has no eigenvectors. Indeed, if $p(x)$ is a solution to the integral equation $L(p(x)) = \int_0^x p(t)dt = \lambda p(x)$ then by differentiation it must solve $p(x) = \lambda p'(x)$.
If $\lambda = 0$ then $p(x) = \mathbf{0}(x)$ (the zero polynomial), and the zero vector of any space is by definition not an eigenvector, so there is no eigenvector for $\lambda = 0$.
If $\lambda \neq 0$ then we can divide by it to get $\frac{1}{p(x)}p'(x) = \frac{1}{\lambda}$, which is a separable first order differential equation with solution $p(x) = \alpha \exp(\frac{x}{\lambda})$ for some constant $\alpha \in \mathbb{R}$. The exponential function is not a polynomial, so this solution is not in V, hence there are no eigenvectors for $\lambda \neq 0$ either.
Hence the spectrum of L on V is the empty set in this example.

(ii) Let $V = C^\infty([a,b],\mathbb{R})$ be the vector space of all real-valued functions with continuous derivatives of all orders (such functions are also said to be *infinitely differentiable*) and let $L : V \to V$ be given by $L(f) = f'$. Note that we indeed have $L(f) \in V$ since the derivative of an infinitely differentiable function is infinitely differentiable, so this map is well-defined, and differentiation is linear so this is a linear map. Note that $e^{\lambda x}$ for $\lambda \in \mathbb{R}$ is an infinitely differentiable function, and we have $L(e^{\lambda x}) = \lambda e^{\lambda x}$ for every $\lambda \in \mathbb{R}$, so $e^{\lambda x}$ is an eigenvector for every λ and the spectrum of L is all of $\mathbb{R}$.

In general the theory of eigenvalues for operators on infinite dimensional vector spaces is less about linear algebra than functional analysis, so we will not do more on this in these notes.

Given a linear map $L : V \to V$ on an n-dimensional vector space over $\mathbb{F}$, we know that every choice of ordered basis $\mathcal{B} = (v_1, \dots, v_n)$ for V gives us a corresponding $n \times n$ matrix $M_{\mathcal{B}}(L)$. But is there a best choice of basis? The easiest matrices to work with are diagonal matrices, so we want to know when it is possible to choose $\mathcal{B}$ so that

$$M_{\mathcal{B}}(L) = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}, \quad \lambda_1, \dots \lambda_n \in \mathbb{F}.$$

Definition 6.34. *For V a finite dimensional vector space, we say a linear map $L : V \to V$ is **diagonalisable** when V admits a basis $\mathcal{B}$ for which the matrix $M_{\mathcal{B}}(L)$ representing it is diagonal.*

Equivalently, L is diagonalisable if and only if some (and in fact, any) matrix $M_{\mathcal{B}}(L)$ representing it is diagonalisable.

Recall that an $n \times n$ matrix A is diagonalisable when there exists an invertible matrix P for which $P^{-1}AP$ is diagonal. We saw that this happens precisely when $\mathbb{F}^n$ has a basis consisting of eigenvectors for A.

Using the isomorphism $\psi_{\mathcal{B}} : V \to \mathbb{F}^n$, a linear map $L : V \to V$ is diagonalisable if and only if V admits a basis $\mathcal{B} = (v_1, \dots, v_n)$ where each v_j is an eigenvector of L.

As before, the matrix P is formed using the eigenvectors of L (written as coordinate vectors) as columns. The diagonal matrix we obtain will then have the corresponding eigenvalues of L on the diagonal.

Example 6.35. The **Legendre equation** from mathematical physics is the differential equation

$$(1 - x^2)y'' - 2xy' - \lambda y = 0, \quad \lambda \in \mathbb{R},$$

for a function y of x.

This is a second order differential equation, but unlike the ones you have seen earlier, it does not have constant coefficients, so it is harder to solve.

Define $L(y) = (1 - x^2)y'' - 2xy'$. Note that, by properties of differentiation, this map is linear. We could view L as a linear operator on the We could define this on the infinite dimensional vector space $C^\infty([a, b], \mathbb{R})$ as above. Then the Legendre equation becomes $L(y) = \lambda y$, an eigenvector problem.

However, notice that if y is a polynomial of degree n then so is $(1 - x^2)y''$ and $-2xy'$, so we can restrict L to $\mathbb{R}_n[x]$, a finite dimensional vector space. Then we can represent L

by an $(n+1)\times(n+1)$ matrix.
For example, for $n = 2$ and the standard ordered basis $\mathcal{B} = (1, x, x^2)$ for $V = \mathbb{R}_2[x]$ we have

$$L(1) = (1-x^2)1'' - 2x1' = 0 = 0\cdot 1 + 0\cdot x + 0\cdot x^2$$
$$L(x) = (1-x^2)x'' - 2x\cdot x' = -2x = 0\cdot 1 + (-2)\cdot x + 0\cdot x^2$$
$$L(x^2) = (1-x^2)(x^2)'' - 2x(x^2)' = (1-x^2)\cdot 2 - 2x\cdot 2x = 2 - 6x^2 = 2\cdot 1 + 0\cdot x + (-6)\cdot x^2.$$

Therefore, with respect to $\mathcal{B}$, L is represented by the matrix

$$A = M_{\mathcal{B}}(L) = \begin{pmatrix} 0 & 0 & 2 \\ 0 & -2 & 0 \\ 0 & 0 & -6 \end{pmatrix}.$$

The matrix A is upper triangular hence its eigenvalues are the diagonal entries $\lambda_1 = 0$, $\lambda_2 = -2$, $\lambda_3 = -6$. We find corresponding eigenvectors as usual, e.g. for $\lambda_3 = -6$ we compute

$$\begin{pmatrix} 0-(-6) & 0 & 2 \\ 0 & -2-(-6) & 0 \\ 0 & 0 & -6-(-6) \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{cases} 6x_1 + 2x_3 = 0 \\ 4x_2 = 0 \end{cases} \implies \begin{cases} x_3 = -3x_1 \\ x_2 = 0 \end{cases}$$

so an eigenvector of A for $\lambda_3 = -6$ is $v_3 = \left(\begin{smallmatrix} 1 \\ 0 \\ -3 \end{smallmatrix}\right)$. Similarly, for $\lambda_1 = 0$ and $\lambda_2 = -2$ we find the following corresponding eigenvectors $v_1 = \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right)$.
To get the corresponding eigenvectors of L we use these coordinates as the coefficients of the polynomials (i.e. we apply $\psi_{\mathcal{B}}^{-1}$):

$$\psi_{\mathcal{B}}^{-1}(v_1) = \psi_{\mathcal{B}}^{-1}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) = 1\cdot 1 + 0\cdot x + 0\cdot x^2 = 1$$
$$\psi_{\mathcal{B}}^{-1}(v_2) = \psi_{\mathcal{B}}^{-1}\left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right) = 0\cdot 1 + 1\cdot x + 0\cdot x^2 = x$$
$$\psi_{\mathcal{B}}^{-1}(v_3) = \psi_{\mathcal{B}}^{-1}\left(\begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}\right) = 1\cdot 1 + 0\cdot x + (-3)\cdot x^2 = 1 - 3x^2.$$

So we get three eigenvectors:

$$p_1(x) = 1, \quad p_2(x) = x, \quad p_3(x) = 1 - 3x^2.$$

We often use the term **eigenfunction** to describe an eigenvector for a linear operator acting on a space of functions.
These polynomials are known as the first three **Legendre polynomials** (though it is

customary to use a different scaling for the chosen eigenvectors, e.g. $\frac{1}{2}(3x^2 - 1)$ instead of $1 - 3x^2$).

The differential operator L on $\mathbb{R}_2[x]$ is diagonalisable because it is represented by a diagonalisable matrix A (since it is a 3×3 matrix with 3 distinct eigenvalues). Equally, it is straightforward to check that the eigenfunctions p_1, p_2, p_3 of L are linearly independent, hence provide a basis for $\mathbb{R}_2[x]$.

It is also easy to explicitly check that if we let P be a matrix whose columns are the eigenvectors of A, then we get that $P^{-1}AP$ is a diagonal matrix with the corresponding eigenvalues on the diagonal:

$$P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix}, \qquad P^{-1}AP = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -6 \end{pmatrix}.$$

Note that if we had picked a larger n, then we would still get the same eigenvalues and eigenfunctions, plus additional ones, e.g. for $n = 3$ we could additionally get the eigenvalue -12 with a corresponding eigenvector of A being $v_4 = \left(\begin{smallmatrix} 0 \\ -3 \\ 0 \\ 5 \end{smallmatrix}\right)$, hence a corresponding eigenfunction of $p_4(x) = -3x + 5x^3$ (or $\frac{1}{2}(5x^3 - 3x)$ in the customary scaling). So by increasing n, we can find infinitely many Legendre polynomials, i.e. infinitely many solutions to the Legendre equation.

When we study inner product spaces in the next chapter we will see that Legendre polynomials are examples of **orthogonal polynomials**, where "orthogonal" means orthogonal with respect to a certain inner product (i.e. generalisation of the dot product). Orthogonal polynomials often appear as the solutions to other differential equations in physics.

6.5. **Dual vector spaces and maps.** For any vector space V over a field $\mathbb{F}$ there is a natural partner vector space called the *dual space* of V. If V is finite dimensional then this space is isomorphic to V.

Definition 6.36. *For a vector space V over a field $\mathbb{F}$, a* ***linear functional*** *is a linear map $L : V \to \mathbb{F}$, i.e. an element of $\mathrm{Hom}(V, \mathbb{F})$.*
The space $\mathrm{Hom}(V, \mathbb{F})$ of linear functionals forms a vector space over $\mathbb{F}$ called the ***dual space*** *of V, denoted V^*. The operations are the usual (pointwise) addition and scalar multiplication of functions.*

Example 6.37. As a motivating example, if $V = \mathbb{R}^3$, viewed as column vectors, then its dual space V^* could be interpreted as row vectors, since row vectors "act" on column vectors through matrix multiplication, the action is linear since matrix multiplication is, and the output is a real number, so this indeed gives us a linear map taking vectors in

$V = \mathbb{R}^3$ to elements of $\mathbb{F} = \mathbb{R}$:

$$\begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = y_1 x_1 + y_2 x_2 + y_3 x_3 \in \mathbb{R}.$$

Row vectors can obviously be scaled and added so they form a vector space, and there is clearly an isomorphism between row vectors and column vectors.
In the result below we will see that if V has a basis $(v_1, \dots, v_n)$, then V^* has a corresponding basis $(v_1^*, \dots, v_n^*)$ called the *dual basis*, where the rule connecting the two is $v_j^*(v_k) = \delta_{jk}$, where δ_{jk} is the *Kronecker delta* function

$$\delta_{jk} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} .$$

So in the case of $\mathbb{R}^3$ and row vectors acting on column vectors, we have a standard basis for $V = \mathbb{R}^3$:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

and we have a corresponding dual basis for V^* given by

$$f_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \quad f_3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

Note that we indeed have $f_j(e_k) = \delta_{jk}$, since for example

$$f_1(e_1) = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1$$

since the indices are equal but

$$f_1(e_2) = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 0$$

since the indices are different.
The next proposition repeats this example in the general case of any finite dimensional vector space.

Proposition 6.38. *Let V be a finite dimensional vector space over a field $\mathbb{F}$, with a basis $v_1, \dots, v_n$.*
Then V^* has a basis given by the linear functionals $v_j^* : V \to \mathbb{F}$, $j = 1, \dots, n$ which satisfy $v_j^*(v_k) = \delta_{jk}$.
Hence $\dim(V) = \dim(V^*)$ and V is isomorphic to V^*.

An isomorphism is given by the linear map $L : V \to V^*$ characterized by $L(v_j) = v_j^*$ for $j = 1, \dots, n$.

Proof. To show that the linear functionals $v_1^*, \dots v_j^*$ are a basis for V^* we need to show that they span V^* and are linearly independent. We begin by showing that they span V^*. So we need to show that every element of V^*, i.e. every linear functional $f : V \to \mathbb{F}$, is a linear combination of $v_1^*, \dots, v_n^*$. To see this first note that, like every linear map, f is completely determined by its values on the basis $v_1, \dots, v_n$. Define $\gamma_j = f(v_j)$ for $j = 1, \dots, n$. Note that, by definition of f, we have $\gamma_j \in \mathbb{F}$. We claim that

$$f = \sum_{j=1}^{n} \gamma_j v_j^*.$$

This is because

$$(\sum_{j=1}^{n} \gamma_j v_j^*)(v_k) = \sum_{j=1}^{n} \gamma_j (v_j^*(v_k)) = \sum_{j=1}^{n} \gamma_j \delta_{jk} = \gamma_k = f(v_k).$$

Hence the linear functional $\sum_{j=1}^{n} \gamma_j v_j^*$ takes the same values on the basis $v_1, \dots, v_n$ as f does, so it is the same function. Thus every $f \in V^*$ is a linear combination of the elements $v_1^*, \dots, v_n^*$, so they span V^*.

Now we have to show that $v_1^*, \dots, v_n^*$ are linearly independent. Suppose there exist scalars $\alpha_1, \dots, \alpha_n$ for which

$$\sum_{j=1}^{n} \alpha_j v_j^* = \mathbf{0},$$

where $\mathbf{0}$ is the zero linear functional, i.e. the map $V \to \mathbb{F}$ such that $\mathbf{0}(v) = 0$ for all $v \in V$. Then applying this map to the basis vector v_k we have $(\sum_{j=1}^{n} \alpha_j v_j^*)(v_k) = \mathbf{0}(k) = 0$. But using the definition of the dual basis we also have $(\sum_{j=1}^{n} \alpha_j v_j^*)(v_k) = \sum_{j=1}^{n} \alpha_j (v_j^*(v_k)) = \sum_{j=1}^{n} \alpha_j \delta_{jk} = \alpha_k$. So we have $\alpha_k = 0$. This is true for all basis vectors v_k, hence all the coefficients α_k are 0 and $v_1^*, \dots, v_n^*$ are linearly independent.

We have already shown that these elements span V^*, hence they form a basis for V^*.

We have $\dim(V^*) = n = \dim(V)$, since both vector spaces have bases with n elements. By Corollary 6.21, any two vector spaces of the same finite dimension and over the same field are isomorphic, hence V and V^* are isomorphic.

If $\mathcal{B} = (v_1, \dots, v_n)$ and $\mathcal{B}^* = (v_1^*, \dots, v_n^*)$, then we know we have isomorphisms (coordinate maps) defined by

$$\psi_{\mathcal{B}} : V \to \mathbb{F}^n, \quad v_j \mapsto \mathbf{e}_j$$
$$\psi_{\mathcal{B}^*} : V \to \mathbb{F}^n, \quad v_j^* \mapsto \mathbf{e}_j.$$

The inverse of a vector space isomorphism is a vector space isomorphism, and the composite of two vector space isomorphisms is a vector space isomorphism, so we get an isomorphism $L : V \to V^*$ given by the composite

$$L : V \xrightarrow{\psi_{\mathcal{B}}} \mathbb{F}^n \xrightarrow{(\psi_{\mathcal{B}^*})^{-1}} V$$
$$v_j \mapsto \mathbf{e}_j \mapsto v_j^*.$$

□

Note that the isomorphism given in the proposition above depends on a choice of basis for V. In that sense it is not "natural", since there is no obvious natural choice of basis for an arbitrary vector space.

We may also look at the **double dual** of V, i.e. the dual space of the dual space, $(V^*)^*$. Elements of $(V^*)^*$ are linear maps from V^* to $\mathbb{F}$, i.e. they are linear maps that take a linear functional $f : V \to \mathbb{F}$ as input, and output an element of $\mathbb{F}$. When V is finite dimensional, it is possible to show that $(V^*)^*$ is isomorphic to V, in a way that does not depend on a choice of basis, so there is a *natural isomorphism* between the two vector spaces.

Using dual spaces, to every linear map we can associate a corresponding dual map (sometimes also called the *transpose map*) between the dual spaces.

Definition 6.39. *If $L : V \to W$ is a linear map then the **dual map** of L is the linear map $L^* : W^* \to V^*$ given by $L^*(f) = f \circ L$ for $f \in W^*$.*

In the example of $V = \mathbb{R}^n$, viewed as column vectors, with V^* being interpreted as row vectors, and $W = \mathbb{R}^m$ (as row vectors), a linear map $L : V \to W$ can be represented by an $m \times n$ matrix A, and it can be shown that the the dual map $L^* : W^* \to V^*$ is then represented by the $n \times m$ transpose matrix $A^\top$.

We will see an explicit example of a dual map in the next section, when we look at inner products and bilinear forms.

7. Inner product spaces

In this section we consider how to attribute a meaningful concept of 'length' to a vector, i.e. how to combine two vectors to get a scalar, in a way that behaves like it does in $\mathbb{R}^n$. We would also like to generalise the concept of vectors being orthogonal (or perpendicular).

In $\mathbb{R}^n$, a length is a non-negative real number, and the angle between two vectors is described using its cosine, which is again a real number. We know how to compute lengths and (cosines of) angles for vectors in $\mathbb{R}^n$: we use the dot (or scalar) product. For example, in $\mathbb{R}^3$, for column vectors

$$u = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad v = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

we have

$$u \cdot v = u^\top v = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

By definition the length of u is $||u|| = \sqrt{u \cdot u}$ (note that we always take $\sqrt{\ }$ to mean the non-negative square root) and the cosine of the angle between vectors u, v is

$$\cos(\theta) = \frac{u \cdot v}{||u|| \, ||v||}.$$

The dot product satisfies the properties $u \cdot v = v \cdot u$, $(\alpha u + \beta v) \cdot w = \alpha(u \cdot w) + \beta(v \cdot w)$, $||u|| \geq 0$, and $||u|| = 0$ if and only if $u = 0$.

We would like to consider other products with the same properties, such as “weighted sums” like

$$x_1^2 + 3x_2^2 + \tfrac{1}{4}x_3^2,$$

where the coordinates contribute different amounts to the ‘length’ measurement. We would also like to consider similar products on abstract real and complex vector spaces, including infinite dimensional spaces.

Note that the dot product is not helpful for defining ‘length’ and ‘orthogonal’ on a complex vector space like $\mathbb{C}^n$, since the quantity $u \cdot u$ will typically just be a complex number (not a real number), which does not give a useful measure of length. Even when this quantity is real, it can be negative, or 0 for non-zero vectors. For example, if

$$u = \begin{pmatrix} i \\ 0 \\ 0 \end{pmatrix}, \quad v = \begin{pmatrix} i \\ 1 \\ 0 \end{pmatrix}$$

then

$$\begin{aligned} u \cdot u &= i^2 + 0^2 + 0^2 = -1 \\ v \cdot v &= i^2 + 1^2 + 0^2 = 0, \text{ but } v \neq \mathbf{0}. \end{aligned}$$

We can rectify this by using that fact that for $z \in \mathbb{C}$ its modulus $|z| = \sqrt{z\bar{z}}$ is real and non-negative. This allows us to define a **Hermitian (or complex) inner product** on $\mathbb{C}^3$

by

$$\langle u, v\rangle := \bar{u}^\top v = \bar{x}_1\, y_1 + \bar{x}_2 y_2 + \bar{x}_3 y_3.$$

With this definition we get

$$\langle u, u\rangle = |x_1|^2 + |x_2|^2 + |x_3|^2 = \sum_{j=1}^{3} |x_j|^2.$$

This is real, non-negative and $\langle u, u\rangle = 0$ if and only if $u = \mathbf{0}$.
We will see that this product satisfies similar properties to the dot product, but not quite the same (for example since, in general, $\langle u, v\rangle \neq \langle v, u\rangle$), so the theory over $\mathbb{R}$ and $\mathbb{C}$ has some differences.

To proceed, we study vector spaces over $\mathbb{R}$ and $\mathbb{C}$ separately.

7.1. **Inner products on real vector spaces.**

Definition 7.1. *Let V be a vector space over $\mathbb{R}$. An **inner product** on V is a function*

$$\langle\ ,\ \rangle : V \times V \to \mathbb{R}$$

such that for any $u, v, w \in V$ and $\alpha \in \mathbb{R}$ we have

(i) $\langle u, v\rangle = \langle v, u\rangle$
(ii) $\langle u + v\,,\, w\rangle = \langle u, w\rangle + \langle v, w\rangle$
(iii) $\langle \alpha u, v\rangle = \alpha\, \langle u, v\rangle$
(iv) $\langle u, u\rangle \geq 0$, *with equality if and only if $u = \mathbf{0}$.*

*A vector space over $\mathbb{R}$ with an inner product is called a **real inner product space.** Sometimes it is also called a **Euclidean space**.*

*Property (i) says that an inner product is **symmetric**. Properties (ii) and (iii) say that it is linear in the first variable, which together with (i) means it is linear in both variables, i.e. it is **bilinear**. (iv) says that an inner product is **positive definite**.*

Note that (ii) and (iii) in the definition above imply that in any inner product space, for $u, v, w \in V$ and $\alpha, \beta \in \mathbb{R}$, we have

$$\langle \alpha u + \beta v\,,\, w\rangle = \alpha\, \langle u, w\rangle + \beta\, \langle v, w\rangle,$$

and combined with symmetry (i) we get

$$\langle u\,,\, \alpha v + \beta w\rangle = \langle \alpha v + \beta w\,,\, u\rangle = \alpha\, \langle v, u\rangle + \beta\, \langle w, u\rangle = \alpha\, \langle u, v\rangle + \beta\, \langle u, w\rangle.$$

So $\langle\ ,\ \rangle$ is indeed linear in both variables, i.e. bilinear.

Example 7.2. The following vector spaces are real inner product spaces.

(i) $\mathbb{R}^n$ for any natural number n, with the inner product being the usual dot product:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

(ii) The space $V = C([0,1],\mathbb{R})$ of continuous real-valued functions defined on the interval $[0,1]$, together with the inner product given by

$$\langle f, g\rangle := \int_0^1 f(t)g(t)\,dt.$$

Let us check that this satisfies the definition of an inner product. We know that continuous functions are integrable, and the product of two continuous functions is continuous, so this definition outputs a real number for all $f, g \in V$, so this is indeed a function from $V \times V$ to $\mathbb{R}$ hence the product is well-defined.
Now let $f, g, h \in V$ and $\alpha \in \mathbb{R}$.
Symmetry:

$$\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt = \int_0^1 g(t)f(t)\,dt = \langle g, f\rangle.$$

Linearity in the first variable:

$$\begin{aligned} \langle f+g\,,\, h\rangle &= \int_0^1 (f+g)(t)\,h(t)\,dt \\ &= \int_0^1 (f(t)+g(t))\,h(t)\,dt \\ &= \int_0^1 (f(t)h(t)+g(t)h(t))\,dt \\ &= \int_0^1 f(t)h(t)\,dt + \int_0^1 g(t)h(t)\,dt \\ &= \langle f, h\rangle + \langle g, h\rangle. \end{aligned}$$

$$\langle \alpha f, g\rangle = \int_0^1 (\alpha f)(t)g(t)\,dt = \int_0^1 \alpha\,(f(t)g(t))\,dt = \alpha \int_0^1 f(t)g(t)\,dt = \alpha\,\langle f, g\rangle.$$

Positive definiteness:

$$\langle f, f\rangle = \int_0^1 f(t)f(t)\,dt = \int_0^1 (f(t))^2\,dt \geq 0,$$

since the integrand $(f(t))^2 \geq 0$ for all $t \in [0,1]$, since it is a square of a real number. And if $\langle f, f\rangle = \int_0^1 (f(t))^2\,dt = 0$ then, since $(f(t))^2 \geq 0$ for all $t \in [0,1]$ so

the graph of $(f(t))^2$ never goes below the x-axis, we must have $f(t) = 0$ for all $t \in [0, 1]$, i.e. f must be the zero function, which is the zero vector in V.
We have now checked all parts of the definition, so this is indeed an inner product on V. Recall that V is infinite dimensional, so this is an example of an inner product on an infinite dimensional space.
Note that the proof above did not rely on the interval being $[0, 1]$, all we used was the fact that $1 > 0$, making the integral of $(f(t))^2$ non-negative. So we also have an inner product on $C([a, b], \mathbb{R})$ given by $\langle f, g\rangle := \int_a^b f(t)g(t)\,dt$ (as long as $a < b$).

(iii) $V = \mathbb{R}^2$ with the inner product given by

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle := 3x_1y_1 + 2x_2y_2,$$

for $x_1, x_2, y_1, y_2 \in \mathbb{R}$. Let us check that this satisfies the conditions. The expression on the right is a real number, so we have a map $V \times V \to \mathbb{R}$ hence the product is well-defined.
Now let $z_1, z_2, \alpha \in \mathbb{R}$.
Symmetry:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = 3x_1y_1 + 2x_2y_2 = 3y_1x_1 + 2y_2x_2 = \left\langle \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\rangle.$$

Linearity in the first variable:

$$\begin{aligned} \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}, \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\rangle \\ &= 3(x_1 + y_1)z_1 + 2(x_2 + y_2)z_2 \\ &= 3x_1z_1 + 2x_2z_2 + 3y_1z_1 + 2y_2z_2 \\ &= \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\rangle + \left\langle \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\rangle. \end{aligned}$$

$$\begin{aligned} \left\langle \alpha \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle \\ &= 3(\alpha x_1)y_1 + 2(\alpha x_2)y_2 \\ &= \alpha(3x_1y_1 + 2x_2y_2) \\ &= \alpha \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle. \end{aligned}$$

Positive definiteness:

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\rangle = 3x_1^2 + 2x_2^2 \geq 0$$

for all possible inputs, since these are squares of real numbers. If we have $\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\rangle = 3x_1^2 + 2x_2^2 = 0$ then, since $3x_1^2 \geq 0$ and $2x_2^2 \geq 0$, we must have $x_1 = 0$ and $x_2 = 0$, so $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = \mathbf{0}$.
So all the conditions are satisfied hence this is an inner product on V.

Using the inner product, we can now define the length, or *norm* of vectors.

Definition 7.3. *In an inner product space V the **norm** (or length) of a vector v is*

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

*We refer to vectors of norm 1 as **unit vectors**.*

If v is any nonzero vector in V, then the vector $\frac{v}{\|v\|}$ has norm 1, so is a unit vector.
Since inner products are by definition positive definite, $\langle v, v \rangle \geq 0$ for all $v \in V$, so taking the square root gives us a real number. Recall that we always take $\sqrt{}$ to mean the non-negative square root, so the norm of a vector is always non-negative.
Note that the norm is a map $\|\cdot\| : V \to \mathbb{R}$, but this map is not linear. For example, for $\alpha \in \mathbb{R}$ we have $\|\alpha v\| = \sqrt{\langle \alpha v, \alpha v \rangle} = \sqrt{\alpha^2 \langle v, v \rangle} = |\alpha| \sqrt{\langle v, v \rangle} = |\alpha| \, \|v\|$, which is not necessarily equal to $\alpha \|v\|$, as we need for linearity.

Example 7.4. For $V = C([0,1], \mathbb{R})$, we saw that $\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt$ is an inner product. Let us find the norm of the vector $f(x) = x$ with respect to this inner product. We have $\|f\| = \sqrt{\langle f, f \rangle}$, where

$$\langle f, f \rangle = \int_0^1 f(t)f(t)\,dt = \int_0^1 t^2\,dt = \left[\frac{t^3}{3}\right]_0^1 = \frac{1}{3}.$$

So $\|f\| = \sqrt{\frac{1}{3}} = \frac{1}{\sqrt{3}}$.
Hence $\frac{f(x)}{\|f(x)\|} = \frac{x}{1/\sqrt{3}} = \sqrt{3}x$ is a unit vector with respect to this inner product.
For $V = \mathbb{R}^2$, we saw that $\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = 3x_1y_1 + 2x_2y_2$ is an inner product. Let us find the norm of the vector $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ with respect to this inner product. We have

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{\langle \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) \rangle} = \sqrt{3 \cdot 0 \cdot 0 + 2 \cdot 1 \cdot 1} = \sqrt{2}.$$

So $\frac{v}{\|v\|} = \frac{v}{\sqrt{2}} = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$ is a unit vector with respect to this inner product.

A real inner product is an example of a more general construction called a **bilinear form**, which we state as a definition.

Definition 7.5. *Let V, W, U be three vector spaces over the same field $\mathbb{F}$. A map $f : V \times W \to U$ is said to be a* ***bilinear map*** *if it is linear in each of its arguments. This means that for each fixed $v \in V$, the map $f(v, -) : W \to U$ (taking $w \in W$ to $f(v, w) \in U$) is linear, and for each fixed $w \in W$ the map $f(-, w) : V \to U$ is linear.*

In the special case where $W = V$ and $U = \mathbb{F}$ (viewed as a 1-dimensional vector space over itself), a bilinear map $f : V \times V \to \mathbb{F}$ is called a ***bilinear form****.*
In detail, a bilinear form is a map $f : V \times V \to \mathbb{F}$ that satisfies, for all $u, v, w \in V$ and $\alpha \in \mathbb{F}$:

(i) $f(u + v\,,\, w) = f(u, w) + f(v, w)$
(ii) $f(\alpha u, v) = \alpha f(u, v)$
(iii) $f(u\,,\, v + w) = f(u, v) + f(u, w)$
(iv) $f(u, \alpha v) = \alpha f(u, v)$.

Note that a bilinear form is not necessarily symmetric, but, as we saw for inner products, if it is symmetric then properties (iii) and (iv) follow from properties (i) and (ii).

Example 7.6. On $V = \mathbb{R}^2$, an example of a bilinear form that is not an inner product is given by the map

$$f : V \times V \to \mathbb{R}, \quad f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = 3x_1y_1 + x_1y_2 + 2x_2y_2.$$

This map is bilinear (the proof is similar to the previous example of an inner product on $\mathbb{R}^2$) and but fails to be an inner product since it is not symmetric (since swapping the order of the inputs would give us a y_1x_2 term, and, in general, $x_1y_2 \neq y_1x_2$).
Another example is the map

$$g : V \times V \to \mathbb{R}, \quad g\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = -3x_1y_1 + 2x_2y_2,$$

which is symmetric but fails the positive definite criterion, since it is possible for $g\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$ to be negative.

We will now see how to use matrices to check when a map on $\mathbb{R}^n \times \mathbb{R}^n$ is a bilinear map, and when it is an inner product.

Proposition 7.7. *A map $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a bilinear form if and only if there is a matrix $A \in M_{n\times n}(\mathbb{R})$ such that*
$$f(u,v) = u^\top Av$$
for all $u, v \in \mathbb{R}^n$ (thought of as column vectors). The entries A_{jk} of the matrix are given by $A_{jk} = f(\mathbf{e}_j, \mathbf{e}_k)$. The matrix A is known as the ***matrix representing the bilinear form*** *f.*

Proof. If a matrix A satisfying the condition $f(u,v) = u^\top Av$ for all $u, v \in \mathbb{R}^n$ exists, then linearity in u and in v follows from properties of the transpose, linearity of vector addition and scalar multiplication, and linearity of matrix multiplication. For example,

$$\begin{aligned} f(u_1+u_2,v) &= (u_1+u_2)^\top Av = (u_1^\top + u_2^\top)Av = u_1^\top Av + u_2^\top Av = f(u_1,v) + f(u_2,v) \\ f(\alpha u, v) &= (\alpha u)^\top Av = \alpha(u^\top Av) = \alpha f(u,v), \end{aligned}$$

which proves conditions (i) and (ii) (linearity in the first variable). The two other conditions (linearity in the second variable) are proved similarly, so we see that f is bilinear. Note that A is an $n \times n$ matrix, v is an n-dimensional column vector and $u^\top$ is an n-dimensional row vector, so Av is an n-dimensional column vector, and $u^\top Av$ is a 1×1 matrix, or equivalently, an element of $\mathbb{R}$. So f is bilinear and it does indeed go from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$, hence it is a bilinear form.
Conversely, let $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear form and define $A_{jk} = f(\mathbf{e}_j, \mathbf{e}_k)$. By writing u and v as a linear combination of the standard basis, and using bilinearity, we can show that $f(u,v) = u^\top Av$. In detail, let

$$u = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n \quad \text{and} \quad v = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \ldots + y_n\mathbf{e}_n.$$

Then

$$
\begin{aligned}
f(u,v) &= f\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}\right) \\
&= f(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n \ , \ y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \ldots + y_n\mathbf{e}_n) \\
&= f(x_1\mathbf{e}_1 \ , \ y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \ldots + y_n\mathbf{e}_n) + \ldots + f(x_n\mathbf{e}_n \ , \ y_1\mathbf{e}_1 + y_2\mathbf{e}_2 + \ldots + y_n\mathbf{e}_n) \\
&= f(x_1\mathbf{e}_1, y_1\mathbf{e}_1) + \ldots + f(x_1\mathbf{e}_1, y_n\mathbf{e}_n) + \ldots + f(x_n\mathbf{e}_n, y_1\mathbf{e}_1) + \ldots + f(x_n\mathbf{e}_n, y_n\mathbf{e}_n) \\
&= x_1y_1f(\mathbf{e}_1, \mathbf{e}_1) + \ldots + x_1y_nf(\mathbf{e}_1, \mathbf{e}_n) + \ldots + x_ny_1f(\mathbf{e}_n, \mathbf{e}_1) + \ldots + x_ny_nf(\mathbf{e}_n, \mathbf{e}_n) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} x_jy_k \, f(\mathbf{e}_j, \mathbf{e}_k) \qquad (\star) \\
&= \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} f(\mathbf{e}_1,\mathbf{e}_1)y_1 + f(\mathbf{e}_1,\mathbf{e}_2)y_2 + \ldots f(\mathbf{e}_1,\mathbf{e}_n)y_n \\ f(\mathbf{e}_2,\mathbf{e}_1)y_1 + f(\mathbf{e}_2,\mathbf{e}_2)y_2 + \ldots f(\mathbf{e}_2,\mathbf{e}_n)y_n \\ \vdots \\ f(\mathbf{e}_n,\mathbf{e}_1)y_1 + f(\mathbf{e}_n,\mathbf{e}_2)y_2 + \ldots f(\mathbf{e}_n,\mathbf{e}_n)y_n \end{pmatrix} \\
&= \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} f(\mathbf{e}_1,\mathbf{e}_1) & f(\mathbf{e}_1,\mathbf{e}_2) & \ldots & f(\mathbf{e}_1,\mathbf{e}_n) \\ f(\mathbf{e}_2,\mathbf{e}_1) & f(\mathbf{e}_2,\mathbf{e}_2) & \ldots & f(\mathbf{e}_2,\mathbf{e}_n) \\ \vdots & \vdots & \ddots & \vdots \\ f(\mathbf{e}_n,\mathbf{e}_1) & f(\mathbf{e}_n,\mathbf{e}_2) & \ldots & f(\mathbf{e}_n,\mathbf{e}_n) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\
&= \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\
&= u^\top Av.
\end{aligned}
$$

We have used bilinearity several times in this calculation. Note that line ($\star$) tells us that all bilinear forms on $\mathbb{R}^n$ have the form of some linear combination of terms of the form x_jy_k, and that the A_{jk} entry in the matrix of the bilinear map is exactly the coefficient of x_jy_k, since by definition we have $A_{jk} = f(\mathbf{e}_j, \mathbf{e}_k)$.

□

The moral of the result and proof above is that, just like a linear map is *uniquely* determined by its action on the basis vectors, so a bilinear map is *uniquely* determined by its action on pairs of basis vectors. In both cases the matrix representing the map contains exactly this information.

Example 7.8. Consider $V = \mathbb{R}^2$. Using the formula $A_{jk} = f(\mathbf{e}_j, \mathbf{e}_k)$ from the proposition we get that the dot product is represented by the 2×2 identity matrix I_2, since, e.g. $A_{11} = f(\mathbf{e}_1, \mathbf{e}_1) = \mathbf{e}_1 \cdot \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1$ and $A_{12} = f(\mathbf{e}_1, \mathbf{e}_2) = \mathbf{e}_1 \cdot \mathbf{e}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$.

Returning to the maps in the previous example, we see that the matrix representing the bilinear form

$$f : V \times V \to \mathbb{R}, \quad f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = 3x_1y_1 + x_1y_2 + 2x_2y_2$$

is

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}.$$

We can see this in two ways. One way is to use the formula $A_{jk} = f(\mathbf{e}_j, \mathbf{e}_k)$ to work out the entries of the matrix:

$$\begin{aligned}
A_{11} = f(\mathbf{e}_1, \mathbf{e}_1) = f\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = 3 \cdot 1 \cdot 1 + 1 \cdot 0 + 2 \cdot 0 \cdot 0 = 3 \\
A_{12} = f(\mathbf{e}_1, \mathbf{e}_2) = f\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = 3 \cdot 1 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 \cdot 1 = 1 \\
A_{21} = f(\mathbf{e}_2, \mathbf{e}_1) = f\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = 3 \cdot 0 \cdot 1 + 0 \cdot 0 + 2 \cdot 1 \cdot 0 = 0 \\
A_{22} = f(\mathbf{e}_2, \mathbf{e}_2) = f\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = 3 \cdot 0 \cdot 0 + 0 \cdot 1 + 2 \cdot 1 \cdot 1 = 2.
\end{aligned}$$

Alternatively, we can use the reasoning from the end of the last proof to note that A_{jk} is the coefficient of x_jy_k in the expansion of the bilinear form, so we can simply record these coefficients as the entries of the matrix A.
Let us check that we do indeed have $f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} A \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$:

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} A \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3y_1 + y_2 \\ 2y_2 \end{pmatrix} = 3x_1y_1 + x_1y_2 + 2x_2y_2 = f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right),$$

as expected.

Similarly, we see that the matrix representing the bilinear form

$$g : V \times V \to \mathbb{R}, \quad g\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = -3x_1y_1 + 2x_2y_2,$$

is

$$B = \begin{pmatrix} -3 & 0 \\ 0 & 2 \end{pmatrix}.$$

We now consider which matrices lead to inner products. Recall that an $n \times n$ matrix A is said to be **symmetric** if $A^\top = A$, where $A^\top$ denotes the transpose of A.
A real symmetric $n \times n$ matrix is said to be **positive definite** if $v^\top A v \geq 0$ for all column vectors $v \in \mathbb{R}^n$, with equality if and only if $v = \mathbf{0}$.
For any $n \times n$ matrix A, a **leading principal minor** of A is the determinant of the sub-matrix formed by taking the top left $k \times k$ submatrix of A, for any $1 \leq k \leq n$.

We now give a useful criterion for determining when a real symmetric matrix is positive definite.

Lemma 7.9. *Let A be a real symmetric $n \times n$ matrix. Then the following are equivalent:*

(i) *A is positive definite.*
(ii) *All the eigenvalues of A are positive.*
(iii) *(Sylvester's Criterion) All the leading principal minors of A are positive.*

We can now say exactly when a given bilinear form is a real inner product, using matrices:

Proposition 7.10. *A bilinear form $\langle\ ,\ \rangle$ on $\mathbb{R}^n \times \mathbb{R}^n$ is a real inner product if and only if the matrix representing representing $\langle\ ,\ \rangle$ is a real, symmetric, positive definite matrix.*

Example 7.11. Consider $V = \mathbb{R}^2$ again. We saw that the matrix representing the dot product is the identity matrix I_2. The dot product is an inner product, and the identity matrix is real, symmetric and positive definite (since all of its eigenvalues are 1, so positive).
We saw that the bilinear form

$$f : V \times V \to \mathbb{R}, \quad f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = 3x_1y_1 + x_1y_2 + 2x_2y_2$$

is represented by the matrix

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}.$$

This bilinear form was not an inner product on V since it was not symmetric, and this corresponds to the matrix A not being symmetric.
We also saw that the matrix representing the bilinear form

$$g : V \times V \to \mathbb{R}, \quad g\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) = -3x_1y_1 + 2x_2y_2$$

is
$$B = \begin{pmatrix} -3 & 0 \\ 0 & 2 \end{pmatrix}.$$
This bilinear form was not an inner product on V since it was not positive definite, and this corresponds to the matrix B not being positive definite (since it has a negative eigenvalue, -3).
Now let $V = \mathbb{R}^3$ and consider the following map:
$$h : V \times V \to \mathbb{R}, \quad h\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}\right) = 2x_1y_1 + x_2y_2 - 2x_2y_3 - 2x_3y_2 + kx_3y_3, \quad \text{for some } k \in \mathbb{R}.$$
This is a bilinear form on V since we can represent it by the matrix
$$C = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & k \end{pmatrix}$$
(using the method described in the proof of Proposition 7.7, so the (j, k)-entry of the matrix is given by the coefficient of x_jy_k). This matrix is clearly real and symmetric. To check when it is positive definite, we use Sylvester's Criterion, so we compute the determinants of the top left 1×1, 2×2 and 3×3 submatrices of C:
$$\begin{vmatrix} 2 \end{vmatrix} = 2 > 0$$
$$\begin{vmatrix} 2 & 0 \\ 0 & 1 \end{vmatrix} = 2 \cdot 1 - 0 \cdot 0 = 2 > 0$$
$$\begin{vmatrix} 2 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & -2 & k \end{vmatrix} = 2 \cdot (1 \cdot k - (-2) \cdot (-2)) = 2(k - 4).$$
We have $2(k - 4) > 0 \iff k > 4$, so we see that the matrix C is positive definite if and only if all the leading principal minors of C are positive, which happens if and only if $k > 4$. Hence h is an inner product on V if and only if $k > 4$.

We can repeat this in more generality, to determine when a bilinear form on a finite dimensional real vector space V is an inner product.

Theorem 7.12. *Let V be a finite dimensional vector space over $\mathbb{R}$, and let $\mathcal{B} = (v_1, \ldots, v_n)$ be any ordered basis for V.*
For $u, v \in V$, let $\mathbf{u} = \psi_{\mathcal{B}}(u)$ and $\mathbf{v} = \psi_{\mathcal{B}}(v)$ (so $\mathbf{u}$ and $\mathbf{v}$ are the coordinate (column) vectors of u and v with respect to the basis $\mathcal{B}$).
Then a map $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$ is a bilinear form if and only if there is a matrix

$A \in M_{n\times n}(\mathbb{R})$ such that

$$\langle u, v\rangle = \mathbf{u}^\top A\mathbf{v}$$

for all $u, v \in V$. The entries A_{jk} of the matrix are given by $A_{jk} = \langle v_j, v_k\rangle$. The matrix A is known as the ***matrix representing the bilinear form $\langle\ ,\ \rangle$ with respect to the basis $\mathcal{B}$****.*
The bilinear form $\langle\ ,\ \rangle$ is a real inner product on V if and only if the matrix A is a (real) symmetric, positive definite matrix.

Example 7.13. We saw that $\langle f, g\rangle := \int_0^1 f(t)g(t)\,dt$ defines an inner product on the space $C([0,1],\mathbb{R})$ of continuous real-valued functions defined on the interval $[0,1]$.
Let $V = \mathbb{R}_1[x]$. Then $V \subseteq C([0,1],\mathbb{R})$, so this formula also defines an inner product on V. Let us find the matrix A of this inner product with respect to the standard basis $\mathcal{B} = (v_1, v_2) = (1, x)$ of V:

$$\begin{aligned}
A_{11} &= \langle v_1, v_1\rangle = \langle 1, 1\rangle = \int_0^1 1^2\,dt = \Big[t\Big]_0^1 = 1\\
A_{12} &= \langle v_1, v_2\rangle = \langle 1, x\rangle = \int_0^1 1\cdot t\,dt = \left[\frac{t^2}{2}\right]_0^1 = \frac{1}{2}\\
A_{21} &= \langle v_2, v_1\rangle = \langle x, 1\rangle = \int_0^1 t\cdot 1\,dt = \left[\frac{t^2}{2}\right]_0^1 = \frac{1}{2} = A_{12}\\
A_{22} &= \langle v_2, v_2\rangle = \langle x, x\rangle = \int_0^1 t^2\,dt = \left[\frac{t^3}{3}\right]_0^1 = \frac{1}{3}.
\end{aligned}$$

So we get the matrix

$$C = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}.$$

This matrix is real and symmetric, and we can check that it is positive definite by using Sylvester's Criterion:

$$\begin{aligned}
\begin{vmatrix} 1 \end{vmatrix} &= 1 > 0\\
\begin{vmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{vmatrix} &= 1\cdot\frac{1}{3} - \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{12} > 0.
\end{aligned}$$

Let us also check that this matrix does indeed represent this inner product. A polynomial in $\mathbb{R}_1[x]$ has the form $\alpha_0 + \alpha_1 x$ for some $\alpha_0, \alpha_1 \in \mathbb{R}$. The corresponding coordinate vector in $\mathbb{R}^2$ is given by

$$\psi_{\mathcal{B}}(\alpha_0 + \alpha_1 x) = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}.$$

So for two vectors $\alpha_0 + \alpha_1 x, \beta_0 + \beta_1 x \in V$ we have

$$\begin{aligned}
\langle \alpha_0 + \alpha_1 x\,,\, \beta_0 + \beta_1 x\rangle &= \int_0^1 (\alpha_0 + \alpha_1 t)(\beta_0 + \beta_1 t)\,dt \\
&= \int_0^1 (\alpha_0\beta_0 + \alpha_0\beta_1 t + \alpha_1\beta_0 t + \alpha_1\beta_1 t^2)\,dt \\
&= \left[\alpha_0\beta_0 t + \alpha_0\beta_1\frac{t^2}{2} + \alpha_1\beta_0\frac{t^2}{2} + \alpha_1\beta_1\frac{t^3}{3}\right]_0^1 \\
&= \alpha_0\beta_0 + \frac{\alpha_0\beta_1}{2} + \frac{\alpha_1\beta_2}{2} + \frac{\alpha_1\beta_1}{3} \\
&= \begin{pmatrix}\alpha_0 & \alpha_1\end{pmatrix}\begin{pmatrix}1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3}\end{pmatrix}\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix} \\
&= (\psi_{\mathcal{B}}(\alpha_0 + \alpha_1 x))^\top A\,\psi_{\mathcal{B}}(\beta_0 + \beta_1 x),
\end{aligned}$$

as expected.

Note that the above result says that a matrix representing a real inner product, *with respect to any basis*, needs to be symmetric and positive definite. How are the matrices representing the same bilinear form with respect to different bases related?

Proposition 7.14. *Let V be a finite dimensional vector space over $\mathbb{R}$, and let $f : V \times V \to \mathbb{R}$ be a bilinear form. Let $\mathcal{A}, \mathcal{B}$ be two ordered bases for V. Let B be the matrix representing f with respect to $\mathcal{B}$ and A be the matrix representing f with respect to $\mathcal{A}$. Then there is an invertible matrix P such that*

$$B = P^\top A P.$$

In fact, we have $P = C_{\mathcal{B}}^{\mathcal{A}}$, the change of basis matrix from $\mathcal{B}$ to $\mathcal{A}$.

Proof. For vectors $u, v \in V$, let $\mathbf{u}_{\mathcal{A}} = \psi_{\mathcal{A}}(u)$, $\mathbf{v}_{\mathcal{A}} = \psi_{\mathcal{A}}(v)$ be the coordinate vectors of u and v with respect to the ordered basis $\mathcal{A}$ for V, and similarly let let $\mathbf{u}_{\mathcal{B}} = \psi_{\mathcal{B}}(u)$, $\mathbf{v}_{\mathcal{B}} = \psi_{\mathcal{B}}(v)$ be the coordinate vectors with respect to the ordered basis $\mathcal{B}$.
Recall that we then have

$$\mathbf{u}_{\mathcal{A}} = \psi_{\mathcal{A}}(u) = (\psi_{\mathcal{A}} \circ (\psi_{\mathcal{B}})^{-1} \circ \psi_{\mathcal{B}})(u) = C_{\mathcal{B}}^{\mathcal{A}}\psi_{\mathcal{B}}(u) = C_{\mathcal{B}}^{\mathcal{A}}\mathbf{u}_{\mathcal{B}},$$

where $C^{\mathcal{A}}_{\mathcal{B}}$ is the change of basis matrix from $\mathcal{B}$ to $\mathcal{A}$, representing the linear map $\psi_{\mathcal{A}} \circ (\psi_{\mathcal{B}})^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. Then the inner product of u and v is given by

$$\begin{aligned}\langle u, v\rangle &= \mathbf{u}_{\mathcal{A}}^\top A\,\mathbf{v}_{\mathcal{A}}\\ &= (C^{\mathcal{A}}_{\mathcal{B}}\mathbf{u}_{\mathcal{B}})^\top A\,(C^{\mathcal{A}}_{\mathcal{B}}\mathbf{v}_{\mathcal{B}})\\ &= \mathbf{u}_{\mathcal{B}}^\top (C^{\mathcal{A}}_{\mathcal{B}})^\top A\,C^{\mathcal{A}}_{\mathcal{B}}\mathbf{v}_{\mathcal{B}}\\ &= \mathbf{u}_{\mathcal{B}}^\top B\,\mathbf{v}_{\mathcal{B}},\end{aligned}$$

so we see that the two matrices A and B, representing $\langle\,,\rangle$ with respect to the two ordered bases $\mathcal{A}$ and $\mathcal{B}$, are related by

$$B = (C^{\mathcal{A}}_{\mathcal{B}})^\top A\,C^{\mathcal{A}}_{\mathcal{B}}.$$

□

Matrices A, B that satisfy the condition $B = P^\top AP$ for some invertible matrix P are called **congruent matrices**.

Congruence is an equivalence relation on the set of square matrices (check this). It is straightforward to show that if A and B are congruent, then A is symmetric if and only if B is symmetric, and A is positive definite if and only if B is positive definite.

We now move on to consider inner products on complex vector spaces, and then we will return to real inner product spaces to discuss orthogonality and properties of the norm that hold for both real and complex vector spaces.

7.2. **Inner products on complex vector spaces.** For complex vectors we require slightly different properties to ensure that lengths are real and non-negative.

For any $z = x + iy \in \mathbb{C}$ with $x, y \in \mathbb{R}$ let $\overline{z} = x - iy$ be the complex conjugate of z.

Definition 7.15. *Let V be a vector space over $\mathbb{C}$. A* ***(Hermitian) inner product*** *on V is a function*

$$\langle\,,\,\rangle : V \times V \to \mathbb{C}$$

such that for any $u, v, w \in V$ and $\alpha \in \mathbb{C}$ we have

(i) $\langle u, v\rangle = \overline{\langle v, u\rangle}$
(ii) $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$
(iii) $\langle u, \alpha v\rangle = \alpha\langle u, v\rangle$ *for all* $\alpha \in \mathbb{C}$
(iv) $\langle u, u\rangle \geq 0$ *(in particular $\langle u, u\rangle \in \mathbb{R}$), with equality if and only if $u = \mathbf{0}$.*

A vector space over $\mathbb{C}$ with an inner product is called a ***complex inner product space.***

Property (i) is called ***conjugate (or Hermitian) symmetry****. Properties (ii) and (iii) are linearity in the second argument. The last property (iv) is the same* ***positive-definiteness*** *as for real inner products. Because of (i), (ii) and (iii) are equivalent to*

$$\langle \alpha u + \beta v, w\rangle = \overline{\alpha}\langle u, w\rangle + \overline{\beta}\langle v, w\rangle \quad \forall\, u, v, w \in V,\ \alpha, \beta \in \mathbb{C}$$

This is referred to as ***conjugate linearity in the first argument****.*

Let us see why we get conjugate linearity in the first argument, considering addition and scalar multiplication separately. For any $u, v, w \in V$ and $\alpha \in \mathbb{C}$ we have

$$\begin{aligned}
\langle u + v, w\rangle &= \overline{\langle w, u + v\rangle} && \text{by (i)}\\
&= \overline{\langle w, u\rangle + \langle w, v\rangle} && \text{by (ii)}\\
&= \overline{\langle w, u\rangle} + \overline{\langle w, v\rangle} && \text{since } \overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}\ \ \forall\, z_1, z_2 \in \mathbb{C}\\
&= \langle u, w\rangle + \langle v, w\rangle && \text{by (i),}\\
\langle \alpha u, v\rangle &= \overline{\langle v, \alpha u\rangle} && \text{by (i)}\\
&= \overline{\alpha\langle v, u\rangle} && \text{by (iii)}\\
&= \overline{\alpha}\,\overline{\langle v, u\rangle} && \text{since } \overline{z_1 z_2} = \overline{z_1}\cdot\overline{z_2}\ \ \forall\, z_1, z_2 \in \mathbb{C}\\
&= \overline{\alpha}\langle u, v\rangle && \text{by (i).}
\end{aligned}$$

So a Hermitian inner product $\langle\ ,\ \rangle$ is linear in the second argument and conjugate linear in the first argument*.

Definition 7.16. *Let V be a complex inner product space.*
The ***norm*** *(or length) is a function from V to $\mathbb{R}_{\geq 0}$ defined by*

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Vectors of norm 1 are called ***unit vectors****.*

Note that this definition makes sense because a Hermitian inner product is positive definite, so for any $v \in V$, $\langle v, v\rangle$ is a non-negative real number. As before, the norm is *not* a linear map, and $\|\alpha u\| = |\alpha|\,\|u\|$ for $\alpha \in \mathbb{C}$, where now $|\alpha|$ means the modulus of the complex number α.

*It is a choice of convention to choose linearity for the second rather than the first slot. This convention tends to be preferred in Mathematical Physics. The other convention puts linearity in the first slot, and therefore (because (i) must still apply) conjugate linearity is in the second slot.

The following result completely classifies all the possible Hermitian inner products on $\mathbb{C}^n$. To state it we need the notion of the conjugate transpose of a matrix (or vector).

Definition 7.17. *For any $p \times n$ matrix $A = (A_{jk})$ we define its **conjugate transpose** to be $A^\dagger = (\overline{A})^\top$, i.e., $A^\dagger = (\overline{A_{kj}})$.*
*We say an $n \times n$ complex matrix A is **Hermitian** when $A^\dagger = A$.*
*An $n \times n$ Hermitian matrix is said to be **positive definite** when $u^\dagger A u \geq 0$ for all column vectors $v \in \mathbb{C}^n$, with equality if and only if $v = \mathbf{0}$.*

Note: in speaking, the symbol $A^\dagger$ is often called "A dagger".
Notice that the entries of A are all real if and only if $A^\dagger = A^\top$. In this case Hermitian becomes symmetric, and positive definiteness is consistently defined for real and complex matrices. In particular, all *real* symmetric matrices are Hermitian, and a Hermitian matrix is only symmetric if it is real.

Theorem 7.18. *Let V be a complex finite dimensional vector space, and let $\mathcal{B} = (v_1, \ldots v_j)$ be an (ordered) basis for V. An operation $\langle\ ,\ \rangle$ on $V \times V$ is an Hermitian inner product if and only if there is a Hermitian positive definite matrix $A \in M_{n\times n}(\mathbb{C})$ for which*

$$\langle u, v\rangle = \mathbf{u}^\dagger A \mathbf{v},$$

for all $u, v \in V$, where $\mathbf{u} = \psi_{\mathcal{B}}(u)$ and $\mathbf{v} = \psi_{\mathcal{B}}(v)$ are the corresponding coordinate (column) vectors in $\mathbb{C}^n$. The entries A_{jk} of the matrix are given by $A_{jk} = \langle v_j, v_k\rangle$.

If $V = \mathbb{C}^n$ and we use the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for $\mathbb{C}^n$, then the above theorem has the simpler form $\langle u, v\rangle = u^\dagger A v$, with $A_{jk} = \langle \mathbf{e}_j, \mathbf{e}_k\rangle$.
Note that the proof of Proposition 7.7 can easily be adapted to the complex case, so the matrix of a Hermitian inner product on $\mathbb{C}^n$ can be found using the same method: the A_{jk} entry is the coefficient in front of the $\overline{x_j} y_k$ term.

It is an exercise to show that the eigenvalues of a Hermitian matrix are real. Note that Lemma 7.9 is actually true not only for $n\times n$ real symmetric matrices, but for all Hermitian matrices, so we can check whether a Hermitian matrix is positive definite by checking whether its eigenvalues are all positive, or using Sylvester's criterion (i.e. checking whether its leading principal minors are all positive).

Example 7.19. The following vector spaces are complex inner product spaces.

(i) $V = \mathbb{C}$ (viewed as a 1-dimensional vector space over itself), with the inner product defined by $\langle z, w\rangle = \overline{z} w$.

(ii) $V = \mathbb{C}^n$ for any $n \in \mathbb{N}$, with the inner product defined by

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \overline{x_1}y_1 + \overline{x_2}y_2 + \cdots + \overline{x_n}y_n.$$

Recall that this is known as the **standard Hermitian inner product** (and the example given in (i) is the special case of this for $n = 1$).
Note that

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\rangle = \overline{x_1}x_1 + \overline{x_2}x_2 + \cdots + \overline{x_n}x_n = |x_1|^2 + |x_2|^2 + \ldots |x_n|^n \geq 0,$$

and this is equal to 0 only if all the x_i are 0, so this function is positive definite.
Linearity in the second argument is easily checked, the reasoning is the same as for the dot product on $\mathbb{R}^n$.
Let us finally check conjugate symmetry:

$$\begin{aligned}
\left\langle \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \right\rangle &= \overline{y_1}x_1 + \overline{y_2}x_2 + \cdots + \overline{y_n}x_n \\
&= x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n} \\
&= \overline{\overline{x_1}} \cdot \overline{y_1} + \overline{\overline{x_2}} \cdot \overline{y_2} + \cdots \overline{\overline{x_n}} \cdot \overline{y_n} \\
&= \overline{\overline{x_1}y_1 + \overline{x_2}y_2 + \cdots + \overline{x_n}y_n} \\
&= \overline{\left\langle \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \right\rangle}.
\end{aligned}$$

So this does indeed define a Hermitian inner product on $\mathbb{C}^n$.
The matrix representing $\langle , \rangle$ is the identity matrix I_n, since we have $\langle u, v\rangle = u^\dagger I_n v = u^\dagger v$. Note that I_n is Hermitian (since it is real and symmetric) and positive definite (since its eigenvalues are all $1 > 0$).

(iii) For $[0, 1] \subseteq \mathbb{R}$, the set V of continuous functions $f : [0, 1] \to \mathbb{C}$ forms a vector space over $\mathbb{C}$ (with the usual pointwise scalar multiplication and addition of functions). This becomes an inner product space if we define the inner product by

$$\langle f, g\rangle = \int_0^1 \overline{f(t)}g(t)dt$$

(check that this is indeed a Hermitian inner product on V).

(iv) $V = \mathbb{C}^2$ with the inner product given by

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = 3\overline{x_1}y_1 + 2\overline{x_2}y_2.$$

Using the method described above, the matrix representing $\langle\,,\rangle$ is $\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$.

(v) Consider

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

We have

$$A^\dagger = (\overline{A})^\top = \begin{pmatrix} \overline{2} & \overline{i} \\ \overline{-i} & \overline{2} \end{pmatrix}^\top = \begin{pmatrix} 2 & -i \\ i & 2 \end{pmatrix}^\top = \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix} = A,$$

so A is Hermitian. To check whether A is positive definite we can use Sylvester's criterion:

$$\begin{vmatrix} 2 \end{vmatrix} = 2 > 0$$
$$\begin{vmatrix} 2 & i \\ -i & 2 \end{vmatrix} = 2^2 - i \cdot (-i) = 2^2 + i^2 = 1 > 0,$$

so A is positive definite. Hence $\langle u, v\rangle = u^\dagger A v$ defines an inner product on $\mathbb{C}^2$. Explicitly, we have

$$\begin{aligned}
\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle &= \begin{pmatrix} \overline{x_1} & \overline{x_2} \end{pmatrix} \begin{pmatrix} 2 & i \\ -i & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\
&= \begin{pmatrix} \overline{x_1} & \overline{x_2} \end{pmatrix} \begin{pmatrix} 2y_1 + iy_2 \\ -iy_1 + 2y_2 \end{pmatrix} \\
&= \overline{x_1}(2y_1 + iy_2) + \overline{x_2}(-iy_1 + 2y_2) \\
&= 2\overline{x_1}y_1 + i\,\overline{x_1}y_2 - i\,\overline{x_2}y_1 + 2\overline{x_2}y_2.
\end{aligned}$$

7.3. **Cauchy-Schwarz, Triangle Inequalities, and Metric Spaces.**

We now return to the situation of a general (real or complex) inner product space, and consider two inequalities that follow from the axioms for real or Hermitian inner products. They can be very powerful inequalities, particularly in Analysis.

Theorem 7.20 (Cauchy-Schwarz inequality)**.** *If u and v are any two vectors in an inner product space V then*
$$|\langle u,v\rangle| \le \|u\|\,\|v\|\,.$$

Proof. Let us first consider the case of a real inner product space V.
For any $w \in V$ the inner product $\langle w,w\rangle$ is real and non-negative. Let $w = \alpha u + \beta v$ where $\alpha = -\langle u,v\rangle$ and $\beta = \langle u,u\rangle$. Then

$$\begin{aligned}\langle \alpha u+\beta v\,,\,\alpha u+\beta v\rangle &= \langle \alpha u,\alpha u\rangle + \langle \alpha u,\beta v\rangle + \langle \beta v,\alpha u\rangle + \langle \beta v,\beta v\rangle\\ &= \alpha^2\langle u,u\rangle + \alpha\beta\langle u,v\rangle + \beta\alpha\langle v,u\rangle + \beta^2\langle v,v\rangle\\ &= \alpha^2\langle u,u\rangle + 2\alpha\beta\langle u,v\rangle + \beta^2\langle v,v\rangle\\ &= \alpha^2\cdot\beta + 2\alpha\beta\cdot(-\alpha) + \beta^2\langle v,v\rangle\\ &= \beta\left(-\alpha^2 + \beta\langle v,v\rangle\right)\\ &= \beta(\langle u,u\rangle\langle v,v\rangle - \langle u,v\rangle^2).\end{aligned}$$

Note that $\beta = \langle u,u\rangle \ge 0$ by positive definiteness, with $\beta = \langle u,u\rangle = 0$ if and only if $u = 0$. In that case the Cauchy-Schwarz inequality is trivially true, with both sides zero. Otherwise, if $b > 0$, then we know $\langle \alpha u+\beta v\,,\,\alpha u+\beta v\rangle$ is nonnegative, hence we must have $\langle u,u\rangle\langle v,v\rangle - \langle u,v\rangle^2 \ge 0$. Rewriting this, we get $\langle u,v\rangle^2 \le \langle u,u\rangle\langle v,v\rangle$. Taking the square root of both sides, we get
$$|\langle u,v\rangle| \le \|u\|\,\|v\|\,,$$
as required.
If V is now a complex inner product space, then the argument is similar. Again for any $w \in V$ the inner product $\langle w,w\rangle$ is real and non-negative. Let $w = \alpha u + \beta v$ where $\alpha = -\langle u,v\rangle$ and $\beta = \langle u,u\rangle$, so that $\beta \in \mathbb{R}$ and $\beta \ge 0$. Then

$$\begin{aligned}\langle \alpha u+\beta v\,,\,\alpha u+\beta v\rangle &= \overline{\alpha}\alpha\langle u,u\rangle + \overline{\alpha}\beta\langle u,v\rangle + \overline{\beta}\alpha\langle v,u\rangle + \overline{\beta}\beta\langle v,v\rangle\\ &= \overline{\alpha}\alpha\langle u,u\rangle + \overline{\alpha}\beta\langle u,v\rangle + \beta\,\alpha\langle v,u\rangle + \beta^2\langle v,v\rangle\\ &= \overline{\alpha}\alpha\,\beta + \overline{\alpha}\beta\,(-\alpha) + \beta\,\alpha\,(-\overline{\alpha}) + \beta^2\langle v,v\rangle\\ &= -\beta\,\alpha\,\overline{\alpha} + \beta^2\langle v,v\rangle\\ &= \beta\,(-|\alpha|^2 + \beta\langle v,v\rangle)\\ &= \beta\,(\langle u,u\rangle\langle v,v\rangle - |\langle u,v\rangle|^2)\,.\end{aligned}$$

As before, if $\beta = \langle u,u\rangle = 0$ then $u = 0$ and the Cauchy-Schwarz inequality is trivially true with both sides zero.
Otherwise $\beta > 0$ and $\langle \alpha u+\beta v\,,\,\alpha u+\beta v\rangle$ is real and non-negative, so we have $\langle u,u\rangle\langle v,v\rangle - |\langle u,v\rangle|^2 \ge 0$. Rewriting this gives us $|\langle u,v\rangle|^2 \le \langle u,u\rangle\langle v,v\rangle$. Taking the square root of both

sides we get

$$|\langle u, v\rangle| \leq \|u\| \, \|v\| \, ,$$

as required (note that $|\langle u, v\rangle|$ is now the modulus of the complex number $\langle u, v\rangle$, whereas in the real case it was the absolute value of the real number $\langle u, v\rangle$). □

On any real inner product space, the Cauchy-Schwarz inequality can be written in the form $-\|u\| \, \|v\| \leq \langle u, v\rangle \leq \|u\| \, \|v\|$. For non-zero vectors u, v this implies

$$-1 \leq \frac{\langle u, v\rangle}{\|u\|\|v\|} \leq 1.$$

Therefore we could define the cosine of the angle between two non-zero vectors u, v by $\cos(\theta) = \langle u, v\rangle/(\|u\|\|v\|)$. Hence the idea of an angle between vectors makes sense for any real inner product.

In $\mathbb{R}^2$ we know that the sum of any two sides of a triangle is longer than the third side. This holds in any inner product space:

Theorem 7.21 (Triangle inequality)**.** *If V is an inner product space and $u, v \in V$ then*

$$\|u + v\| \leq \|u\| + \|v\| \, .$$

Proof. Let V be a real inner product space. By definition we have

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

The norm is always non-negative, so taking the square root of the above equation gives us

$$\|u + v\| \leq \|u\| + \|v\| \, ,$$

as required.
For V a complex inner product space the proof is similar:

$$\begin{aligned}
||u+v||^2 &= \langle u+v\,,\,u+v\rangle \\
&= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\
&= ||u||^2 + \langle u,v\rangle + \overline{\langle u,v\rangle} + ||v||^2 \\
&= ||u||^2 + 2\,\mathrm{Re}(\langle u,v\rangle) + ||v||^2 && \text{since } z+\overline{z} = \mathrm{Re}(z)\ \ \forall z \in \mathbb{C} \\
&\le ||u||^2 + 2|\langle u,v\rangle| + ||v||^2 && \text{since } \mathrm{Re}(z) \le |z|\ \ \forall z \in \mathbb{C} \\
&\le ||u||^2 + 2\,||u||\,||v|| + ||v||^2 && \text{by Cauchy-Schwarz inequality} \\
&= (||u|| + ||v||)^2\,.
\end{aligned}$$

Taking square roots again gives us the desired inequality. □

Definition 7.22. *In any inner product space V, the corresponding* ***metric*** *or* ***distance function*** *is a function $d : V \times V \to \mathbb{R}$ defined by*

$$d(u,v) = ||u-v||\,.$$

The metric on an inner product space V satisfies, for all $u,v,w \in V$:

(i) $d(u,v) \ge 0$, with $d(u,v) = 0 \iff u = v$
(ii) $d(u,v) = d(v,u)$
(iii) $d(u,w) \le d(u,v) + d(v,w)$.

Property (iii) is also called the **triangle inequality**. It can be proved using the version given above.

Example 7.23. For $\mathbb{R}^n$ with the usual inner product, if $\mathbf{v} = [x_1,\dots,x_n]^T$ and $\mathbf{u} = [y_1,\dots,y_n]^T$, then

$$||\mathbf{v}|| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

and

$$d(\mathbf{u},\mathbf{v}) = \sqrt{(y_1-x_1)^2 + \cdots + (y_n-x_n)^2} = \sqrt{\sum_{i=1}^{n} (y_i-x_i)^2}.$$

This is known as the **standard Euclidean metric**.
If we view $\mathbf{u}$ and $\mathbf{v}$ as the position vectors of two points in $\mathbb{R}^n$, e.g. if $\mathbf{u} = \overrightarrow{OP}$ and $\mathbf{v} = \overrightarrow{OQ}$, for two points P,Q and the origin O, then $d(\mathbf{u},\mathbf{v})$ gives us the usual (Euclidean) distance between the points P and Q.

Remark 7.24. We have defined a norm with the help of an inner product. Then we used the norm to define a metric. More generally, on any set M, one can define a metric first, to be a function $M \times M \to \mathbb{R}$ satisfying the three properties given above. This gives rise to so-called **metric spaces**.
A vector space for which a norm is defined is called a **normed vector space**. A normed vector space is also a metric space, but the converse is not true since a metric can be defined independently from a norm and the existence of a metric does not imply the existence of a norm.
Also, every inner product space is a normed vector space, but the converse is not true, because a norm can be defined independently from an inner product (in detail, a **norm** $\|\cdot\| : V \to \mathbb{R}$ is a function on a vector space V over a field $\mathbb{F}$ (where $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$), satisfying, for all $u, v \in V$, $\alpha \in \mathbb{F}$: $\|u + v\| \leq \|u\| + \|v\|$) (the triangle inequality), $\|\alpha v\| = |\alpha| \, \|v\|$, $\|v\| \geq 0$, with $\|v\| = 0$ if and only if $v = \mathbf{0}$).
The precise relations between metric, normed and inner product spaces are studied in Functional Analysis.

7.4. **Orthogonality.**

Throughout this section we will assume V is a (real or complex) inner product space.

Definition 7.25. *Two vectors u and v of an inner product space V are said to be **orthogonal** if $\langle u, v\rangle = 0$ (we also say u **is orthogonal to** v and vice versa).*
When u and v are orthogonal we sometimes write $u \perp v$.

Note that if $\langle u, v\rangle = 0$ then $\langle v, u\rangle = 0$ since for a real inner product we have $\langle v, u\rangle = \langle u, v\rangle = 0$, and for a Hermitian (complex) inner product we have $\langle v, u\rangle = \overline{\langle u, v\rangle} = \overline{0} = 0$.
Also, for any $v \in V$ we have $\langle v, \mathbf{0}\rangle = 0$ since

$$\begin{aligned} \langle v, \mathbf{0}\rangle &= \langle v, w - w\rangle && \text{for any } w \in V \\ &\left.\begin{aligned} &= \langle v, w\rangle + \langle v, -w\rangle \\ &= \langle v, w\rangle - \langle v, w\rangle \end{aligned}\right\} && \text{by linearity in 2nd argument} \\ &= 0. \end{aligned}$$

Similarly, $\langle \mathbf{0}, v\rangle = 0$ for all $v \in V$.
So the zero vector $\mathbf{0}$ is orthogonal to every vector in an inner product space.

Definition 7.26. *A set S of non-zero vectors in an inner product space is said to be **orthogonal** if $u \perp v$ for every pair u, v of distinct vectors in S.*
*If in addition every $u \in S$ is a unit vector, then S is said to be **orthonormal.***

Note the "non-zero" condition in the definition above. We will be interested in constructing orthogonal bases, and we know that a basis for a space cannot contain $\mathbf{0}$, but $\mathbf{0}$ is orthogonal to every vector in the space, so to have a useful definition of an orthogonal set it is necessarily to only consider non-zero vectors.

One of the reasons we are interested in orthogonality is the following result:

Theorem 7.27. *Any orthogonal set S in an inner product space V is linearly independent. Hence, if V has dimension n and S has size n, then S is a basis for V.*

Proof. Let V be an inner product space over the field $\mathbb{F}$ (where $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$). Suppose S is a subset of non-zero vectors of V such that $\langle u, v\rangle = 0$ for all $u, v \in S$ with $u \neq v$.
Let $v_1, \dots, v_k$ be any k distinct vectors from S, and suppose there exist $\alpha_1, \dots \alpha_n \in \mathbb{F}$ such that

$$\alpha_1 v_1 + \ldots + \alpha_k v_k = \mathbf{0}.$$

We take the inner product of both sides with v_i:

$$\begin{aligned}
0 &= \langle v_i, \mathbf{0}\rangle \\
&= \langle v_i\,,\, \alpha_1 v_1 + \ldots + \alpha_k v_k\rangle \\
&\left.\begin{aligned} &= \langle v_i, \alpha_1 v_1\rangle + \cdots + \langle v_i, \alpha_k v_k\rangle \\ &= \alpha_1\langle v_i, v_1\rangle + \ldots + \alpha_i\langle v_i, v_i\rangle + \ldots + \alpha_k\langle v_i, v_k\rangle \end{aligned}\right\} \quad \text{by linearity in 2nd argument} \\
&= \alpha_i\langle v_i, v_i\rangle \qquad \text{since distinct elements of } S \text{ are mutually orthogonal.}
\end{aligned}$$

Since v_i is non-zero, $\langle v_i, v_i\rangle \neq 0$, so we have $\alpha_i = 0$. This holds for all i, so the v_i are linearly independent. This holds for any finite set of vectors from S, so S is linearly independent. □

What is the benefit of using orthogonal and orthonormal bases? Suppose $(v_1, \dots, v_n)$ is an orthogonal basis for V. Then any $v \in V$ can be written as

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n,$$

for some $\alpha_1, \dots, \alpha_n \in \mathbb{F}$, and it is easy to find these coefficients. We have

$$\begin{aligned}
\langle v_i, v\rangle &= \langle v_i\,,\, \alpha_1 v_1 + \cdots + \alpha_n v_n\rangle \\
&= \alpha_1\langle v_i, v_1\rangle + \alpha_2\langle v_i, v_2\rangle + \cdots + \alpha_n\langle v_i, v_n\rangle \\
&= \alpha_i\langle v_i, v_i\rangle
\end{aligned}$$

so, since v_i is a basis element, hence $v_i \neq \mathbf{0}$, so $\langle v_i, v_i\rangle \neq 0$, we have

$$\alpha_i = \frac{\langle v_i, v\rangle}{\langle v_i, v_i\rangle} = \frac{\langle v_i, v\rangle}{\|v_i\|^2}.$$

When the basis is orthonormal the formula for the coefficients is even simpler, since then we have $\|v_i\| = 1$, so we get

$$\alpha_i = \langle v_i, v \rangle.$$

Example 7.28.

(i) In $\mathbb{R}^n$ (with the standard inner product, the dot product) the standard basis is orthonormal.

(ii) In $\mathbb{R}^3$ (with the standard inner product) the vectors

$$\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$$

form an orthogonal basis (check this!). Note that they do not form an orthonormal basis because they are not unit vectors, but we could get an orthonormal basis by normalising the vectors (i.e. dividing each vector by its norm):

$$\left\| \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\| = \sqrt{1^2 + (-1)^2 + 1^2} = \sqrt{3}$$
$$\left\| \begin{pmatrix} 5 \\ 4 \\ -1 \end{pmatrix} \right\| = \sqrt{5^2 + 4^2 + (-1)^2} = \sqrt{42}$$
$$\left\| \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} \right\| = \sqrt{(-1)^2 + (2)^2 + (3)^2} = \sqrt{14}.$$

So we get that

$$\begin{pmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} \frac{5}{\sqrt{42}} \\ \frac{4}{\sqrt{42}} \\ -\frac{1}{\sqrt{42}} \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \\ \frac{3}{\sqrt{14}} \end{pmatrix}$$

is an orthonormal basis for $\mathbb{R}^3$.

(iii) In $\mathbb{C}^2$ (with the standard (Hermitian) inner product) the vectors

$$\begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix}$$

form an orthonormal basis, since

$$\left\langle \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix} \right\rangle = \overline{\left(\frac{i}{\sqrt{2}}\right)}\frac{1}{\sqrt{2}} + \overline{\left(\frac{1}{\sqrt{2}}\right)}\frac{i}{\sqrt{2}} = \frac{-i}{2} + \frac{i}{2} = 0$$
$$\left\langle \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\rangle = \overline{\left(\frac{i}{\sqrt{2}}\right)}\frac{i}{\sqrt{2}} + \overline{\left(\frac{1}{\sqrt{2}}\right)}\frac{1}{\sqrt{2}} = \frac{-i \cdot i}{2} + \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1$$
$$\left\langle \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix} \right\rangle = \overline{\left(\frac{1}{\sqrt{2}}\right)}\frac{1}{\sqrt{2}} + \overline{\left(\frac{i}{\sqrt{2}}\right)}\frac{i}{\sqrt{2}} = \frac{1}{2} + \frac{1}{2} = 1.$$

We can also encode the ideas of orthonormal bases using square matrices.

Definition 7.29.
*We say a real matrix $Q \in M_{n\times n}(\mathbb{R})$ is **orthogonal** if $Q^\top Q = I_n$ (equivalently, $Q^{-1} = Q^\top$).*
*We say a complex matrix $P \in M_{n\times n}(\mathbb{C})$ is **unitary** if $P^\dagger P = I_n$ (equivalently, $P^{-1} = P^\dagger$).*

Lemma 7.30.
A basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ is orthonormal with respect to the standard real inner product if and only if these are the columns of an orthogonal matrix Q.
A basis $v_1, \ldots, v_n$ of $\mathbb{C}^n$ is orthonormal with respect to the standard Hermitian inner product if and only if these are the columns of a unitary matrix P.