Lecture Notes for “Algebra–2”, “Linear Algebra” and “Vectors and Matrices”

Gavin Brown and Oleg Kozlovski and Dmitriy Rumynin

Version March 17, 2023 for Term 2, 2022-23

Contents

1 Vector Spaces

You have met the idea of vectors in 2 or 3 dimensions. We can add them and multiply them by numbers. The same is true for collections of functions or matrices. It turns out to be useful to identify carefully what properties these collections have in common so that we can apply the intuition we get from ordinary space to these less geometric collections. Among other things we want to have a clear understanding of dimension.

1.1 Foreword

The three modules MA148 Vectors and Matrices, MA149 Linear Algebra and MA150 Algebra–2 will start with the same lecture notes that were written for the start of the term. They will develop in slightly different directions. So follow your module.

In particular, these notes are too long: some of the material in them will be cut. The cut material may differ in different modules.

1.2 Scalars

1.2.1 Field

From now on, let $\mathbb{F}$ be a field. A field is a non-zero commutative ring, where every non-zero element has a multiplicative inverse. You will get full marks in this module, if you know only the following two fields.

The real numbers $\mathbb{R}$. These are the numbers which can be expressed as decimals.

The complex numbers $\mathbb{C} = \{x + iy \mid x, y \in \mathbb{R}\}$, where $i^2 = -1$.

1.2.2 Axioms for number systems

For completeness, we give the complete list of the axioms of a field. A *field* is a set $\mathbb{F}$ together with two special elements $0 \neq 1 \in \mathbb{F}$, and two binary operations $\mathbb{F} \times \mathbb{F} \to \mathbb{F}$, called *addition* $(x, y) \mapsto x + y$ and *multiplication* $(x, y) \mapsto xy$, satisfying the following axioms.

A1. Commutativity of Addition: $\alpha + \beta = \beta + \alpha$ for all $\alpha, \beta \in \mathbb{F}$.

A2. Associativity of Addition: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ for all $\alpha, \beta, \gamma \in F$.

A3. Additive Unity: There exists $0 \in \mathbb{F}$ such that $\alpha + 0 = 0 + \alpha = \alpha$ for all $\alpha \in \mathbb{F}$.

A4. Additive Inverse: For each $\alpha \in \mathbb{F}$ there exists $-\alpha \in \mathbb{F}$ such that $\alpha + (-\alpha) = (-\alpha) + \alpha = 0$.

M1. Commutativity of Multiplication: $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbb{F}$.

M2. Associativity of Multiplication: $(\alpha\beta)\gamma = \alpha(\beta\gamma)$ for all $\alpha, \beta, \gamma \in \mathbb{F}$.

M3. Multiplicative Unity: There is $1 \in \mathbb{F}$ such that $\alpha 1 = 1\alpha = \alpha$ for all $\alpha \in \mathbb{F}$.

M4. Multiplicative Inverse: For each number $\alpha \in \mathbb{F}$ with $\alpha \neq 0$, there exists $\alpha^{-1} \in \mathbb{F}$ such that $\alpha\alpha^{-1} = \alpha^{-1}\alpha = 1$.

D. Distributivity: $(\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma$ for all $\alpha, \beta, \gamma \in \mathbb{F}$.

1.2.3 Field of real numbers

We will use $\mathbb{F}$ and $\mathbb{R}$ (and once $\mathbb{C}$) in these notes. We use $\mathbb{F}$ for the results that hold over any field, although in all our examples (and the exam) $\mathbb{F} = \mathbb{R}$. We use $\mathbb{R}$ when we need to use the special properties of the real numbers:

- The order $\alpha \geq \beta$ such that $\alpha^2 \geq 0$ for all $\alpha \in \mathbb{R}$.
- Each $\alpha \in \mathbb{R}$ such that $\alpha \geq 0$ admits a square root $\sqrt{\alpha} \in \mathbb{R}$ with $\sqrt{\alpha} \geq 0$.
- Each polynomial of positive degree $f(x)$ with coefficients in $\mathbb{R}$ admits a real root $\alpha \in \mathbb{R}$ or a pair of complex roots $\alpha, \bar{\alpha} \in \mathbb{C} \setminus \mathbb{R}$.

1.3 Vector spaces

Recall that the abelian group structure on V is a binary operation of addition that satisfies axioms A1, A2, A3 and A4.

Definition 1.3.1. A *vector space* is an abelian group V with an additional binary operation $\mathbb{F} \times V \to V$, called *scalar multiplication* $(\alpha, \mathbf{v}) \mapsto \alpha\mathbf{v}$, satisfying the following axioms:

1. $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$,
2. $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$,
3. $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$,
4. $1\mathbf{v} = \mathbf{v}$.

Elements of the field $\mathbb{F}$ will be called *scalars*. We will use the greek letters for them. We will use boldface letters like $\mathbf{v}$ to denote vectors, elements of V. The zero vector in V will be written as $\mathbf{0}_V$, or usually just $\mathbf{0}$. This is different from the zero scalar $0 = 0_{\mathbb{F}} \in \mathbb{F}$.

1.3.1 Examples of vector spaces

1. The standard vector space is the space of column vectors $\mathbb{F}^n = \{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} \mid \alpha_i \in \mathbb{F}\}$. Addition and scalar multiplication are defined by the obvious rules:

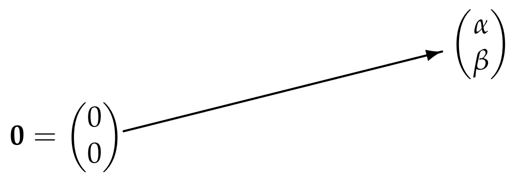
$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \alpha_1 + \beta_1 \\ \vdots \\ \alpha_n + \beta_n \end{pmatrix} , \quad \lambda \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \lambda\alpha_1 \\ \vdots \\ \lambda\alpha_n \end{pmatrix} .$$

The most familiar examples are

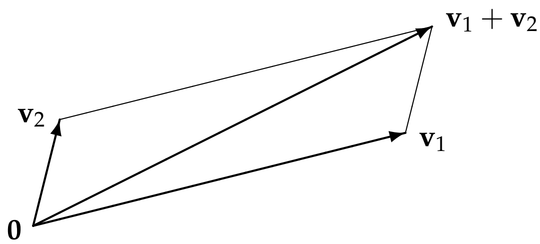
$$\mathbb{R}^2 = \{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mid \alpha, \beta \in \mathbb{R}\} \text{ and } \mathbb{R}^3 = \{ \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \mid \alpha, \beta, \gamma \in \mathbb{R}\},$$

which are the points in an ordinary 2- and 3-dimensional space, equipped with a coordinate system.

Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ can also be thought of as directed lines joining the origin to the points with the corresponding coordinates:



Addition of vectors is then given by the parallelogram law.



Note that $\mathbb{F}^1$ is essentially the same as $\mathbb{F}$ itself and $\mathbb{F}^1$ is the zero vector spaces $\{\mathbf{0}\}$.

2. Let $\mathbb{F}[x]$ be the set of polynomials in an indeterminate x with coefficients in the field $\mathbb{F}$. That is,

$$\mathbb{F}[x] = \{\alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n \mid \alpha_i \in \mathbb{F}\}.$$

Then $\mathbb{F}[x]$ is a vector space over $\mathbb{F}$.

3. Fix $n \geq 0$. Let $\mathbb{F}[x]_{\leq n}$ be the set of polynomials over $\mathbb{F}$ of degree at most n (where we agree that the degree of the zero polynomial is -1). Then $\mathbb{F}[x]_{\leq n}$ is also a vector space over $\mathbb{F}$; in fact it is a *subspace* of $\mathbb{F}[x]$ (see Definition 1.3.3).

Note that the polynomials of degree exactly n do not form a vector space. (Why?)

4. Let $\mathbb{F} = \mathbb{R}$ and let V be the set of n-times differentiable functions $f : \mathbb{R} \to \mathbb{R}$ which are solutions of the differential equation

$$\lambda_0 \frac{d^n f}{dx^n} + \lambda_1 \frac{d^{n-1} f}{dx^{n-1}} + \cdots + \lambda_{n-1} \frac{df}{dx} + \lambda_n f = 0.$$

for fixed $\lambda_0, \lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Then V is a vector space over $\mathbb{R}$, for if $f(x)$ and $g(x)$ are both solutions of this equation, then so are $f(x) + g(x)$ and $\alpha f(x)$ for all $\alpha \in \mathbb{R}$.

5. The previous example is a space of functions. There are many such examples that are important in Analysis. For example, the set $C^k((0,1), \mathbb{R})$ (of all functions $f : (0,1) \to \mathbb{R}$ such that the k-th derivative $f^{(k)}$ exists and is continuous) is a vector space over $\mathbb{R}$ with the usual pointwise definitions of addition and scalar multiplication of functions.

1.3.2 Axiomatic approach

Facing such a variety of vector spaces, a mathematician wants to derive useful methods of handling all these vector spaces. If we do it on a single example, say $\mathbb{R}^8$, how can we be certain that our methods are correct? It is only possible with the *axiomatic approach* to developing mathematics. We must use only arguments based on the vector space axioms. We have to avoid making any other assumptions. This ensures that everything we prove is valid for all vector spaces, not just the familiar ones like $\mathbb{R}^3$.

Try deducing the following easy properties from the axioms.

Lemma 1.3.2. 1. *$\alpha \mathbf{0} = \mathbf{0}$ for all $\alpha \in \mathbb{F}$,*

2. *$0\mathbf{v} = \mathbf{0}$ and $(-1)\mathbf{v} = -\mathbf{v}$ for all $\mathbf{v} \in V$.*

3. *$-(\alpha\mathbf{v}) = (-\alpha)\mathbf{v} = \alpha(-\mathbf{v})$, for all $\alpha \in \mathbb{F}$ and $\mathbf{v} \in V$.*

1.3.3 Subspaces

Let V be a vector space over the field $\mathbb{F}$.

Definition 1.3.3. A subspace of V is a non-empty subset $W \subseteq V$ such that

$$\mathbf{u}, \mathbf{v} \in W \Rightarrow \mathbf{u} + \mathbf{v} \in W \qquad \text{and} \qquad \mathbf{v} \in W,\ \alpha \in \mathbb{F} \Rightarrow \alpha\mathbf{v} \in W.$$

These two conditions can be replaced with a single condition

$$\mathbf{u}, \mathbf{v} \in W, \alpha, \beta \in \mathbb{F} \Rightarrow \alpha\mathbf{u} + \beta\mathbf{v} \in W.$$

A subspace W is itself a vector space over $\mathbb{F}$ under the operations of vector addition and scalar multiplication in V. Notice that all vector space axioms of W hold automatically. (They are inherited from V.)

For any vector space V, V is always a subspace of itself. Subspaces other than V are sometimes called *proper* subspaces. We also always have a subspace $\{\mathbf{0}\}$ consisting of the zero vector alone. This is called the *trivial* subspace.

Proposition 1.3.4. *If W_1 and W_2 are subspaces of V then so is $W_1 \cap W_2$.*

Proof. Let $\mathbf{u}, \mathbf{v} \in W_1 \cap W_2$ and $\alpha \in \mathbb{F}$. Then $\mathbf{u} + \mathbf{v} \in W_1$ (because W_1 is a subspace) and $\mathbf{u} + \mathbf{v} \in W_2$ (because W_2 is a subspace). Hence $\mathbf{u} + \mathbf{v} \in W_1 \cap W_2$. Similarly, we get $\alpha\mathbf{v} \in W_1 \cap W_2$, so $W_1 \cap W_2$ is a subspace of V. $\square$

Warning! It is **not** necessarily true that $W_1 \cup W_2$ is a subspace.

Example. Let $V = \mathbb{R}^2$, let $W_1 = \{\begin{pmatrix}\alpha \\ 0\end{pmatrix} \mid \alpha \in \mathbb{R}\}$ and $W_2 = \{\begin{pmatrix}0 \\ \alpha\end{pmatrix} \mid \alpha \in \mathbb{R}\}$. Then W_1, W_2 are subspaces of V, but $W_1 \cup W_2$ is not a subspace, because $\begin{pmatrix}1 \\ 0\end{pmatrix}, \begin{pmatrix}0 \\ 1\end{pmatrix} \in W_1 \cup W_2$, but $\begin{pmatrix}1 \\ 0\end{pmatrix} + \begin{pmatrix}0 \\ 1\end{pmatrix} = \begin{pmatrix}1 \\ 1\end{pmatrix} \notin W_1 \cup W_2$.

Note that any subspace of V that contains W_1 and W_2 has to contain all vectors of the form $\mathbf{u} + \mathbf{v}$ for $\mathbf{u} \in W_1,\ \mathbf{v} \in W_2$.

Definition 1.3.5. Let W_1, W_2 be subspaces of the vector space V. Then $W_1 + W_2$ is defined to be the set of vectors $\mathbf{v} \in V$ such that $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$ for some $\mathbf{w}_1 \in W_1,\ \mathbf{w}_2 \in W_2$. Or, if you prefer, $W_1 + W_2 = \{\mathbf{w}_1 + \mathbf{w}_2 \mid \mathbf{w}_1 \in W_1, \mathbf{w}_2 \in W_2\}$.

Do not confuse $W_1 + W_2$ with $W_1 \cup W_2$!

Proposition 1.3.6. *If W_1, W_2 are subspaces of V then so is $W_1 + W_2$. In fact, it is the smallest (with respect to the order $\subseteq$) subspace that contains both W_1 and W_2.*

Proof. Let $\mathbf{u}, \mathbf{v} \in W_1 + W_2$. Then $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$ for some $\mathbf{u}_1 \in W_1,\ \mathbf{u}_2 \in W_2$ and $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ for some $\mathbf{v}_1 \in W_1,\ \mathbf{v}_2 \in W_2$. Then $\mathbf{u} + \mathbf{v} = (\mathbf{u}_1 + \mathbf{v}_1) + (\mathbf{u}_2 + \mathbf{v}_2) \in W_1 + W_2$. Similarly, if $\alpha \in \mathbb{F}$ then $\alpha\mathbf{v} = \alpha\mathbf{v}_1 + \alpha\mathbf{v}_2 \in W_1 + W_2$. Thus $W_1 + W_2$ is a subspace of V.

Any subspace of V that contains both W_1 and W_2 must contain $W_1 + W_2$, so it is the smallest such subspace. $\square$

Examples. 1. As above, let $V = \mathbb{R}^2$, $W_1 = \{\begin{pmatrix}\alpha \\ 0\end{pmatrix} \mid \alpha \in \mathbb{R}\}$ and $W_2 = \{\begin{pmatrix}0 \\ \alpha\end{pmatrix} \mid \alpha \in \mathbb{R}\}$. Then $W_1 + W_2 = V$.

2. Let $V = \mathbb{R}^2$, let $W_1 = \{\begin{pmatrix}\alpha \\ \alpha\end{pmatrix} \mid \alpha \in \mathbb{R}\}$ and $W_2 = \{\begin{pmatrix}-\alpha \\ \alpha\end{pmatrix} \mid \alpha \in \mathbb{R}\}$. Then $W_1 + W_2 = V$.

1.4 Linear independence, spanning and bases

By *a vector sequence* we understand a finite sequence $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ of elements of a vector space V. Vectors of the form $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$ for $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$ are called *linear combinations* of $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$.

1.4.1 Linear dependence and independence

Definition 1.4.1. Let V be a vector space over the field $\mathbb{F}$. The vector sequence $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ is called *linearly dependent* if there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$, not all zero, such that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}.$$

The sequence $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$is called *linearly independent* if they are not linearly dependent. In other words, it is linearly independent if the only scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$ that satisfy the above equation are $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_n = 0$.

Examples. 1. Let $V = \mathbb{R}^2$, $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$.
Then $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = (\alpha_1 + 2\alpha_2, 3\alpha_1 + 5\alpha_2)$, which is equal to $\mathbf{0} = (0, 0)$ if and only if $\alpha_1 + 2\alpha_2 = 0$ and $3\alpha_1 + 5\alpha_2 = 0$. Thus, we have a pair of simultaneous equations in α_1, α_2 and the only solution is $\alpha_1 = \alpha_2 = 0$, so $\mathbf{v}_1$, $\mathbf{v}_2$ is linearly independent.

2. Let $V = \mathbb{R}^2$, $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 6 \end{pmatrix}$.
This time the equations are $\alpha_1 + 2\alpha_2 = 0$ and $3\alpha_1 + 6\alpha_2 = 0$, and there are non-zero solutions, such as $\alpha_1 = -2$, $\alpha_2 = 1$, and so the vector sequence $\mathbf{v}_1$, $\mathbf{v}_2$ is linearly dependent.

Lemma 1.4.2. *The following statements about a vector sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ are equivalent.*

1. *It is linearly dependent.*
2. *$\mathbf{v}_1 = \mathbf{0}$ or, for some r, $\mathbf{v}_r$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}$.*

Proof. If $\mathbf{v}_1 = \mathbf{0}$ then by putting $\alpha_1 = 1$ and $\alpha_i = 0$ for $i > 1$ we get $\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$, so $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ is linearly dependent.

If $\mathbf{v}_r$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}$, then $\mathbf{v}_r = \alpha_1\mathbf{v}_1 + \cdots + \alpha_{r-1}\mathbf{v}_{r-1}$ for some $\alpha_1, \ldots, \alpha_{r-1} \in \mathbb{F}$ and so we get $\alpha_1\mathbf{v}_1 + \cdots + \alpha_{r-1}\mathbf{v}_{r-1} - 1 \cdot \mathbf{v}_r = \mathbf{0}$ and again $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ is linearly dependent.

Conversely, suppose that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ is linearly dependent, and α_i are scalars, not all zero, satisfying $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$. Let r be maximal with $\alpha_r \neq 0$; then $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r = \mathbf{0}$. If $r = 1$ then $\alpha_1\mathbf{v}_1 = \mathbf{0}$ which is only possible if $\mathbf{v}_1 = \mathbf{0}$. Otherwise, we get

$$\mathbf{v}_r = -\frac{\alpha_1}{\alpha_r}\mathbf{v}_1 - \cdots - \frac{\alpha_{r-1}}{\alpha_r}\mathbf{v}_{r-1}.$$

In other words, $\mathbf{v}_r$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}$. $\square$

1.4.2 Spanning vectors

The proof of the next fact is routine. Try it yourself!

Proposition 1.4.3. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a vector sequence. Then the set of all linear combinations $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$ of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ forms a subspace of V.*

The subspace in this proposition is known as the *span* of the sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Definition 1.4.4. The sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *spans* V if the span of the sequence is V.

In other words, this means that every vector $\mathbf{v} \in V$ is a linear combination $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$ of $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

1.4.3 Bases of vector spaces

Definition 1.4.5. The vector sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in V forms a *basis* of V if it is linearly independent and spans V.

Proposition 1.4.6. *The vector sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ forms a basis of V if and only if every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$; that is, the coefficients $\alpha_1, \ldots, \alpha_n$ are uniquely determined by the vector $\mathbf{v}$.*

Proof. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ forms a basis of V. Then it spans V, so certainly every $\mathbf{v} \in V$ can be written as $\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$. Suppose that we also had $\mathbf{v} = \beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \cdots + \beta_n\mathbf{v}_n$ for some other scalars $\beta_i \in \mathbb{F}$. Then we have

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = (\alpha_1 - \beta_1)\mathbf{v}_1 + (\alpha_2 - \beta_2)\mathbf{v}_2 + \cdots + (\alpha_n - \beta_n)\mathbf{v}_n$$

and so

$$(\alpha_1 - \beta_1) = (\alpha_2 - \beta_2) = \cdots = (\alpha_n - \beta_n) = 0$$

by linear independence of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Hence $\alpha_i = \beta_i$ for all i, which means that the α_i are uniquely determined.

Conversely, suppose that every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$. Then $\mathbf{v}_1, \ldots, \mathbf{v}_n$ certainly spans V. If $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$, then

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n$$

and so the uniqueness assumption implies that $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent. Hence the sequence forms a basis of V. $\square$

Examples. 1. $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is a basis of $\mathbb{F}^2$. This follows from Proposition 1.4.6, because each element $\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \in \mathbb{F}^2$ can be written as $\alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2$, and this expression is clearly unique.

2. More generally, $\mathbf{e}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$, $\mathbf{e}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}$, $\mathbf{e}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}$ form a basis of $\mathbb{F}^3$, $\mathbf{e}_1 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}$, $\mathbf{e}_2 = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}$, $\mathbf{e}_3 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}$, $\mathbf{e}_4 = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$ form a basis of $\mathbb{F}^4$ and so on. This is called the *standard basis* of $\mathbb{F}^n$ for $n \in \mathbb{N}$.

(To be precise, the standard basis of $\mathbb{F}^n$ is $\mathbf{e}_1, \dots, \mathbf{e}_n$, where $\mathbf{e}_i$ is the vector with a 1 in the i-th position and a 0 in all other positions.)

3. There are many other bases of $\mathbb{F}^n$. For example, $\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}$ form a basis of $\mathbb{F}^2$, because $\begin{pmatrix}\alpha_1\\\alpha_2\end{pmatrix} = (\alpha_1 - \alpha_2)\begin{pmatrix}1\\0\end{pmatrix} + \alpha_2\begin{pmatrix}1\\1\end{pmatrix}$, and this expression is unique. In fact, any two non-zero vectors such that one is not a scalar multiple of the other form a basis for $\mathbb{F}^2$.

4. Since we defined a basis as a finite vector sequence (with additional properties, of course), not every vector space has a basis. For example, let $\mathbb{F}[x]$ be the space of polynomials in x with coefficients in $\mathbb{F}$. Let $p_1(x), p_2(x), \dots, p_n(x)$ be a finite sequence of polynomials in $\mathbb{F}[x]$. Then if d is the maximum degree of $p_1(x), p_2(x), \dots, p_n(x)$, any linear combination of $p_1(x), p_2(x), \dots, p_n(x)$ has degree at most d, and so $p_1(x), p_2(x), \dots, p_n(x)$ cannot span $\mathbb{F}[x]$.

It is customary in Maths to allow infinite bases as well: then the infinite sequence of vectors $1, x, x^2, x^3, \dots, x^n, \dots$ is a basis of $\mathbb{F}[x]$. A vector space with a finite basis is called *finite dimensional*. In fact, all of this course will be about finite-dimensional spaces, but it is important to remember that these are not the only examples. The spaces of functions mentioned in Example 5 of Section 1.3 typically have uncountably infinite dimension: so they are even less well-behaved than $\mathbb{F}[x]$ as heir bases are not even sequences.

Theorem 1.4.7. *(The basis theorem.) Suppose that $\mathbf{v}_1, \dots, \mathbf{v}_m$ and $\mathbf{w}_1, \dots, \mathbf{w}_n$ are both finite bases of the vector space V. Then $m = n$. In other words, all finite bases of V contain the same number of vectors.*

The proof of this theorem requires *sifting* and will be done in the next section.

Definition 1.4.8. The number n of vectors in a basis of the finite-dimensional vector space V is called the *dimension* of V and we write $\dim(V) = n$.

Thus, as we might expect, $\mathbb{F}^n$ has dimension n. $\mathbb{F}[x]$ is infinite-dimensional, but the space $\mathbb{F}[x]_{\leq n}$ of polynomials of degree at most n has basis $1, x, x^2, \dots, x^n$, so its dimension is $n + 1$ (not n).

1.4.4 Sifting

Lemma 1.4.9. *Suppose that the vector sequence* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{w}$ *spans* V *and that* $\mathbf{w}$ *is a linear combination of* $\mathbf{v}_1, \ldots, \mathbf{v}_n$*. Then* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *spans* V*.*

Proof. Since $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{w}$ spans V, any vector $\mathbf{v} \in V$ can be written as

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n + \beta \mathbf{w},$$

But $\mathbf{w}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$, so $\mathbf{w} = \gamma_1 \mathbf{v}_1 + \cdots + \gamma_n \mathbf{v}_n$ for some scalars γ_i, and hence

$$\mathbf{v} = (\alpha_1 + \beta\gamma_1)\mathbf{v}_1 + \cdots + (\alpha_n + \beta\gamma_n)\mathbf{v}_n$$

is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$, which therefore spans V. $\square$

There is an important process, which we shall call *sifting*, which can be applied to any sequence of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ in a vector space V. We consider each vector $\mathbf{v}_i$ in turn. If it is zero, or a linear combination of the preceding vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}$, then we remove it from the list. The output of the sifting is a new linearly independent vectors sequence with the same span as the original one.

Example. Let us sift the following sequence of vectors in $\mathbb{R}^3$.

$$\mathbf{v}_1 = \begin{pmatrix}0\\0\\0\end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix}1\\1\\1\end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix}2\\2\\2\end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \mathbf{v}_5 = \begin{pmatrix}3\\2\\2\end{pmatrix}, \mathbf{v}_6 = \begin{pmatrix}0\\0\\0\end{pmatrix}, \mathbf{v}_7 = \begin{pmatrix}1\\1\\0\end{pmatrix}, \mathbf{v}_8 = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

$\mathbf{v}_1 = \mathbf{0}$, so we remove it. $\mathbf{v}_2$ is non-zero so it stays. $\mathbf{v}_3 = 2\mathbf{v}_2$ so it is removed. $\mathbf{v}_4$ is clearly not a linear combination of $\mathbf{v}_2$, so it stays.

We have to decide next whether $\mathbf{v}_5$ is a linear combination of $\mathbf{v}_2, \mathbf{v}_4$. If so, then $\begin{pmatrix}3\\2\\2\end{pmatrix} = \alpha_1 \begin{pmatrix}1\\1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}1\\0\\0\end{pmatrix}$, which (fairly obviously) has the solution $\alpha_1 = 2, \alpha_2 = 1$, so remove $\mathbf{v}_5$. Then $\mathbf{v}_6 = \mathbf{0}$ so that is removed too.

Next we try $\mathbf{v}_7 = \begin{pmatrix}1\\1\\0\end{pmatrix} = \alpha_1 \begin{pmatrix}1\\1\\1\end{pmatrix} + \alpha_2 \begin{pmatrix}1\\0\\0\end{pmatrix}$, and looking at the three components, this reduces to the three equations

$$1 = \alpha_1 + \alpha_2; \qquad 1 = \alpha_1; \qquad 0 = \alpha_1.$$

The second and third of these equations contradict each other, and so there is no solution. Hence $\mathbf{v}_7$ is not a linear combination of $\mathbf{v}_2, \mathbf{v}_4$, and it stays.

Finally, we need to try

$$\mathbf{v}_8 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

leading to the three equations

$$0 = \alpha_1 + \alpha_2 + \alpha_3 \qquad 0 = \alpha_1 + \alpha_3; \qquad 1 = \alpha_1$$

and solving these in the normal way, we find a solution $\alpha_1 = 1, \alpha_2 = 0, \alpha_3 = -1$. Thus we delete $\mathbf{v}_8$ and we are left with just $\mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_7$.

Of course, the vectors that are removed during the sifting process depend very much on the order of the list of vectors. For example, if $\mathbf{v}_8$ had come at the beginning of the list rather than at the end, then we would have kept it.

Applying sifting, we can now prove the Basis Theorem 1.4.7. It follows from the next proposition.

Proposition 1.4.10. *(Exchange Lemma) Suppose that vector sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ spans V and that the vector sequence $\mathbf{w}_1, \ldots, \mathbf{w}_m \in V$ is linearly independent. Then $m \leq n$.*

Proof. The idea is to place the $\mathbf{w}_i$ one by one in front of the sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$, sifting each time.

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ spans V, $\mathbf{w}_1, \mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly dependent, so when we sift, at least one $\mathbf{v}_j$ is deleted. We then place $\mathbf{w}_2$ in front of the resulting sequence and sift again. Then we put $\mathbf{w}_3$ in from of the result, and sift again, and carry on doing this for each $\mathbf{w}_i$ in turn. Since $\mathbf{w}_1, \ldots, \mathbf{w}_m$ are linearly independent none of them are ever deleted. Each time we place a vector in front of a sequence which spans V, and so the extended sequence is linearly dependent, and hence at least one $\mathbf{v}_j$ gets eliminated each time.

But in total, we append m vectors $\mathbf{w}_i$, and each time at least one $\mathbf{v}_j$ is eliminated, so we must have $m \leq n$. □

1.4.5 Existence of a basis

Let us address the fundamental question. Does a vector space admit a basis?

Theorem 1.4.11. *Suppose that the vector sequence $\mathbf{v}_1, \ldots, \mathbf{v}_r$ spans the vector space V. Then there is a subsequence of $\mathbf{v}_1, \ldots, \mathbf{v}_r$ which forms a basis of V.*

Proof. We sift the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$. The vectors that we remove are linear combinations of the preceding vectors, and so by Lemma 1.4.9, the remaining vectors still span V. After sifting, no vector is zero or a linear combination of the preceding vectors (or it would have been removed), so by Lemma 1.4.2, the remaining vector sequence is linearly independent. Hence, it is a basis of V. □

Corollary 1.4.12. *If a vector space V is spanned by a finite sequence, then it admits a basis.*

In fact, if you allow infinite bases, *any vector space V admits a basis*. A proof of this requires would lead us too deep into axiomatic Set Theory: it is carried out in year 3 in both *Set Theory* and *Rings and Modules*.

Corollary 1.4.13. *Let V be a vector space of dimension n over $\mathbb{F}$. Then any sequence of n vectors which spans V is a basis of V, and no $n-1$ vectors can span V.*

Proof. After sifting a spanning sequence, the remaining vectors form a basis, so by Theorem 1.4.7, there must be precisely $n = \dim(V)$ vectors remaining. The result is now clear. □

Theorem 1.4.11 is not flexible enough for future proofs. We will need the next theorem.

Theorem 1.4.14. *Let V be a finite-dimensional vector space over $\mathbb{F}$, and suppose that the vector sequence $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is linearly independent in V. Then we can extend the sequence to a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, where $n \geq r$.*

Proof. Suppose that $\dim(V) = n$ and let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be any basis of V. We sift the combined sequence

$$\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}_1, \ldots, \mathbf{w}_n.$$

Since $\mathbf{w}_1, \ldots, \mathbf{w}_n$ spans V, the result is a basis of V by Theorem 1.4.11. Since $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is linearly independent, none of them can be a linear combination of the preceding vectors, and hence none of the $\mathbf{v}_i$ are deleted in the sifting process. Thus the resulting basis contains $\mathbf{v}_1, \ldots, \mathbf{v}_r$. □

Corollary 1.4.15. *Let V be a vector space of dimension n over $\mathbb{F}$. Then any n linearly independent vectors form a basis of V and no $n+1$ vectors can be linearly independent.*

(**Remark for the observant reader:** Notice that this corollary shows that a vector space V cannot have both a finite and an infinite basis.)

Example. The vectors $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix}$ are linearly independent in $\mathbb{R}^4$. Let us extend them to a basis of $\mathbb{R}^4$. The easiest thing is to append the standard basis of $\mathbb{R}^4$, giving the combined vector sequence

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{e}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

which we shall sift. We find that $\begin{pmatrix}1\\0\\0\\0\end{pmatrix} = \alpha_1 \begin{pmatrix}1\\2\\0\\2\end{pmatrix} + \alpha_2 \begin{pmatrix}0\\1\\0\\2\end{pmatrix}$ has no solution, so $\mathbf{e}_1$ stays. However, $\mathbf{e}_2 = \mathbf{v}_1 - \mathbf{v}_2 - \mathbf{e}_1$ so $\mathbf{e}_2$ is deleted. It is clear that $\mathbf{e}_3$ is not a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{e}_1$, because all of those have a 0 in their third component. Hence $\mathbf{e}_3$ remains. Now we have four linearly independent vectors, so must have a basis at this stage, and we can stop the sifting early. The resulting basis is

$$\mathbf{v}_1 = \begin{pmatrix}1\\2\\0\\2\end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix}0\\1\\0\\2\end{pmatrix}, \mathbf{e}_1 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}.$$

2 Linear Transformations

When you study sets the notion of function is extremely important. There is little to say about a single isolated set while functions allow you to link different sets. Similarly in Linear Algebra, a single isolated vector space is not the end of the story. We have to connect different vector spaces by functions. However, a function having little regard to the vector space operations may be of little value.

2.1 Basic properties

2.1.1 Definition

Definition 2.1.1. Let U, V be two vector spaces over the same field $\mathbb{F}$. A *linear transformation* or *linear map* T from U to V is a function $T : U \to V$ such that

(i) $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$ for all $\mathbf{u}_1, \mathbf{u}_2 \in U$;

(ii) $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$ for all $\alpha \in \mathbb{F}$ and $\mathbf{u} \in U$.

We shall usually call these linear maps (because this is shorter), although linear transformation is the standard name. Notice that the two conditions for linearity are equivalent to a single condition

$$T(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) = \alpha T(\mathbf{u}_1) + \beta T(\mathbf{u}_2) \text{ for all } \mathbf{u}_1, \mathbf{u}_2 \in U, \alpha, \beta \in \mathbb{F}.$$

First a couple of trivial consequences of the definition:

Lemma 2.1.2. *Let $T : U \to V$ be a linear map. Then*

(i) $T(\mathbf{0}_U) = \mathbf{0}_V$;

(ii) $T(-\mathbf{u}) = -T(\mathbf{u})$ *for all* $\mathbf{u} \in U$.

Proof. (i) $T(\mathbf{0}_U) = T(\mathbf{0}_U + \mathbf{0}_U) = T(\mathbf{0}_U) + T(\mathbf{0}_U)$, so $T(\mathbf{0}_U) = \mathbf{0}_V$.
(ii) Just put $\alpha = -1$ in the definition of linear map. $\square$

The key property is that linear maps are uniquely determined by their action on a basis.

Proposition 2.1.3. *Let U, V be vector spaces over $\mathbb{F}$, let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be a basis of U and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be any sequence of n vectors in V. Then there is a unique linear map $T : U \to V$ with $T(\mathbf{u}_i) = \mathbf{v}_i$ for $1 \leq i \leq n$.*

Proof. Let $\mathbf{u} \in U$. Then, since $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is a basis of U, by Proposition 1.4.6, there exist uniquely determined $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ with $\mathbf{u} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n$. Hence, if T exists at all, then we must have

$$T(\mathbf{u}) = T(\alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n) = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n,$$

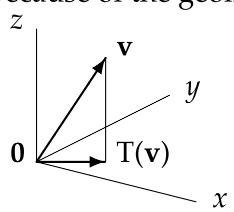
and so T is uniquely determined.

On the other hand, it is routine to check that the map $T : U \to V$ defined by the above equation is indeed a linear map, so T does exist and is unique. $\square$

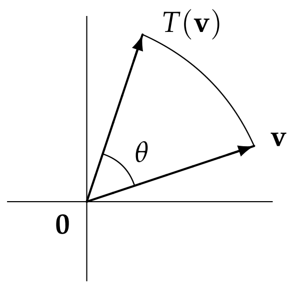
2.1.2 Examples

Many familiar geometrical transformations, such as projections, rotations, reflections and magnifications are linear maps, and the first three examples below are of this kind. Note, however, that a nontrivial translation is not a linear map, because it does not satisfy $T(\mathbf{0}_U) = \mathbf{0}_V$.

1. Let $U = \mathbb{R}^3$, $V = \mathbb{R}^2$ and define $T : U \to V$ by $T\begin{pmatrix}\alpha\\\beta\\\gamma\end{pmatrix} = \begin{pmatrix}\alpha\\\beta\end{pmatrix}$. Then T is a linear map. This type of map is known as a *projection*, because of the geometrical interpretation.



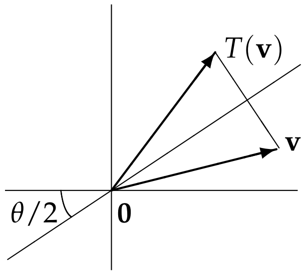
2. Let $U = V = \mathbb{R}^2$. We interpret $\mathbf{v}$ in $\mathbb{R}^2$ as a directed line vector from $\mathbf{0}$ to $\mathbf{v}$ (see the examples in Section 1.3), and let $T(\mathbf{v})$ be the vector obtained by rotating $\mathbf{v}$ through an angle θ anti-clockwise about the origin.



It is easy to see geometrically that $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$ and $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$ (because everything is simply rotated about the origin), and so T is a linear map. By considering the unit vectors, we have

$$T\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix},\ T\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-\sin\theta\\\cos\theta\end{pmatrix} \text{ and so } T\begin{pmatrix}\alpha\\\beta\end{pmatrix} = \alpha T\begin{pmatrix}1\\0\end{pmatrix} + \beta T\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}\alpha\cos\theta - \beta\sin\theta\\\alpha\sin\theta + \beta\cos\theta\end{pmatrix}.$$

3. Let $U = V = \mathbb{R}^2$ again. Now let $T(\mathbf{v})$ be the vector resulting from reflecting $\mathbf{v}$ through a line through the origin that makes an angle $\theta/2$ with the x-axis.



This is again a linear map. We find that

$$T\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix},\ T\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}\sin\theta\\ -\cos\theta\end{pmatrix}\text{ and so } T\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\alpha T\begin{pmatrix}1\\0\end{pmatrix}+\beta T\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}\alpha\cos\theta+\beta\sin\theta\\ \alpha\sin\theta-\beta\cos\theta\end{pmatrix}.$$

4. Let $U=V=\mathbb{R}[x]$, the set of polynomials over $\mathbb{R}$, and let T be differentiation, i.e., $T(p(x))=p'(x)$ for $p\in\mathbb{R}[x]$. This is easily seen to be a linear map.

5. Let $U=\mathbb{F}[x]$, the set of polynomials over $\mathbb{F}$. Every $\alpha\in\mathbb{F}$ gives rise to two linear maps, shift $S_\alpha : U\to U,\ S_\alpha(f(x))=f(x-\alpha)$ and evaluation $E_\alpha : U\to\mathbb{F},\ E_\alpha(f(x))=f(\alpha)$.

The next two examples seem dull but are important!

6. For any vector space V, we define the identity map $I_V : V\to V$ by $I_V(\mathbf{v})=\mathbf{v}$ for all $\mathbf{v}\in V$. This is a linear map.

7. For any vector spaces U, V over the field $\mathbb{F}$, we define the zero map $\mathbf{0}_{U,V} : U\to V$ by $\mathbf{0}_{U,V}(\mathbf{u})=\mathbf{0}_V$ for all $\mathbf{u}\in U$. This is also a linear map.

2.1.3 Operations on linear maps

We define the operations of *addition*, *scalar multiplication*, and *composition* on linear maps.

Let $T_1 : U\to V$ and $T_2 : U\to V$ be two linear maps, and let $\alpha\in\mathbb{F}$ be a scalar.

Addition: We define a map $T_1+T_2 : U\to V$ by the rule $(T_1+T_2)(\mathbf{u})=T_1(\mathbf{u})+T_2(\mathbf{u})$ for $\mathbf{u}\in U$.

Scalar multiplication: We define a map $\alpha T_1 : U\to V$ by the rule $(\alpha T_1)(\mathbf{u})=\alpha T_1(\mathbf{u})$ for $\mathbf{u}\in U$.

Now let $T_1 : U\to V$ and $T_2 : V\to W$ be two linear maps.

Composition: We define a map $T_2T_1 : U\to W$ by $(T_2T_1)(\mathbf{u})=T_2(T_1(\mathbf{u}))$ for $\mathbf{u}\in U$. In particular, we define $T^2=TT$ and $T^{i+1}=T^iT$ for $i>2$.

It is routine to check that T_1+T_2, αT_1 and T_2T_1 are themselves all linear maps.

Furthermore, for fixed vector spaces U and V over $\mathbb{F}$, the operations of addition and scalar multiplication on the set $\mathrm{Hom}_{\mathbb{F}}(U,V)$ of all linear maps from U to V makes $\mathrm{Hom}_{\mathbb{F}}(U,V)$ into a vector space over $\mathbb{F}$.

2.2 Dimension properties of linear maps

2.2.1 Kernels and images

Definition 2.2.1. Let $T : U \to V$ be a linear map. The *image* of T, written as $\mathrm{im}(T)$ is defined to be the set of vectors $\mathbf{v} \in V$ such that $\mathbf{v} = T(\mathbf{u})$ for some $\mathbf{u} \in U$.

The *kernel* of T, written as $\ker(T)$ is defined to be the set of vectors $\mathbf{u} \in U$ such that $T(\mathbf{u}) = \mathbf{0}_V$. Or, if you prefer:

$$\mathrm{im}(T) = \{T(\mathbf{u}) \mid \mathbf{u} \in U\}; \qquad \ker(T) = \{\mathbf{u} \in U \mid T(\mathbf{u}) = \mathbf{0}_V\}.$$

Examples. Let us consider the examples from Section 2.1.2.

In 1., $\ker(T) = \{\begin{pmatrix} 0 \\ 0 \\ \gamma \end{pmatrix} \mid \gamma \in \mathbb{R}\}$, and $\mathrm{im}(T) = \mathbb{R}^2$.
In 2. and 3., $\ker(T) = \{\mathbf{0}\}$ and $\mathrm{im}(T) = \mathbb{R}^2$.
In 4., $\ker(T)$ is the set of all constant polynomials (i.e. those of degree 0), and $\mathrm{im}(T) = \mathbb{R}[x]$.
In 5., $\ker(S_\alpha) = \{\mathbf{0}\}$, and $\mathrm{im}(S_\alpha) = \mathbb{F}[x]$, while $\ker(E_\alpha)$ is the set of all polynomials divisible by $x - \alpha$, and $\mathrm{im}(E_\alpha) = \mathbb{F}$.
In 6., $\ker(I_V) = \{\mathbf{0}\}$ and $\mathrm{im}(T) = V$.
In 7., $\ker(\mathbf{0}_{U,V}) = U$ and $\mathrm{im}(\mathbf{0}_{U,V}) = \{\mathbf{0}\}$.

Proposition 2.2.2. *(i)* $\mathrm{im}(T)$ *is a subspace of* V*; (ii)* $\ker(T)$ *is a subspace of* U*; (iii)* T *is injective if and only if* $\ker(T) = \{\mathbf{0}\}$

Proof. (i) We must show that $\mathrm{im}(T)$ is closed under addition and scalar multiplication. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathrm{im}(T)$. Then $\mathbf{v}_1 = T(\mathbf{u}_1)$, $\mathbf{v}_2 = T(\mathbf{u}_2)$ for some $\mathbf{u}_1, \mathbf{u}_2 \in U$. Then

$$\mathbf{v}_1 + \mathbf{v}_2 = T(\mathbf{u}_1) + T(\mathbf{u}_2) = T(\mathbf{u}_1 + \mathbf{u}_2) \in \mathrm{im}(T); \qquad \alpha\mathbf{v}_1 = \alpha T(\mathbf{u}_1) = T(\alpha\mathbf{u}_1) \in \mathrm{im}(T),$$

so $\mathrm{im}(T)$ is a subspace of V.

(ii) Similarly, we must show that $\ker(T)$ is closed under addition and scalar multiplication. Let $\mathbf{u}_1, \mathbf{u}_2 \in \ker(T)$. Then

$$T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{0}_U + \mathbf{0}_U) = T(\mathbf{0}_U) = \mathbf{0}_V; \qquad T(\alpha\mathbf{u}_1) = \alpha T(\mathbf{u}_1) = \alpha\mathbf{0}_V = \mathbf{0}_V,$$

so $\mathbf{u}_1 + \mathbf{u}_2$, $\alpha\mathbf{u}_1 \in \ker(T)$ and $\ker(T)$ is a subspace of U.

(iii) The "only if" is obvious since $\ker(T) = T^{-1}(\mathbf{0})$. To prove the "if", suppose $\ker(T) = \{\mathbf{0}\}$ and $T(\mathbf{u}) = T(\mathbf{v})$. Then $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{0}$, so $\mathbf{u} - \mathbf{v} \in \ker(T)$, then $\mathbf{u} - \mathbf{v} = \mathbf{0}$ and $\mathbf{u} = \mathbf{v}$. □

2.2.2 Dimension formula

We go back to the study of subspaces as this helps the understanding of kernels and images.

Theorem 2.2.3. *(Dimension Formula) Let V be a finite-dimensional vector space, and let W_1, W_2 be subspaces of V. Then*

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

Proof. First note that any subspace W of V is finite-dimensional. This follows from Corollary 1.4.15, because a largest linearly independent sequence W contains at most $\dim(V)$ vectors, and such a sequence must be a basis of W.

Let $\dim(W_1 \cap W_2) = r$ and let $\mathbf{e}_1, \ldots, \mathbf{e}_r$ be a basis of $W_1 \cap W_2$. Then $\mathbf{e}_1, \ldots, \mathbf{e}_r$ is a linearly independent sequence of vectors, so by Theorem 1.4.14 it can be extended to a basis $\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_s$ of W_1 where $\dim(W_1) = r + s$, and it can also be extended to a basis $\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{g}_1, \ldots, \mathbf{g}_t$ of W_2, where $\dim(W_2) = r + t$.

To prove the theorem, we need to show that $\dim(W_1 + W_2) = r + s + t$, and to do this, we shall show that

$$\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_s, \mathbf{g}_1, \ldots, \mathbf{g}_t$$

is a basis of $W_1 + W_2$. Certainly, they all lie in $W_1 + W_2$.

First we show that they span $W_1 + W_2$. Any $\mathbf{v} \in W_1 + W_2$ is equal to $\mathbf{w}_1 + \mathbf{w}_2$ for some $\mathbf{w}_1 \in W_1$, $\mathbf{w}_2 \in W_2$. So we can write

$$\mathbf{w}_1 = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_r \mathbf{e}_r + \beta_1 \mathbf{f}_1 + \cdots + \beta_s \mathbf{f}_s$$

for some scalars $\alpha_i, \beta_j \in \mathbb{F}$, and

$$\mathbf{w}_2 = \gamma_1 \mathbf{e}_1 + \cdots + \gamma_r \mathbf{e}_r + \delta_1 \mathbf{g}_1 + \cdots + \delta_t \mathbf{g}_t$$

for some scalars $\gamma_i, \delta_j \in \mathbb{F}$. Then

$$\mathbf{v} = (\alpha_1 + \gamma_1)\mathbf{e}_1 + \cdots + (\alpha_r + \gamma_r)\mathbf{e}_r + \beta_1 \mathbf{f}_1 + \cdots + \beta_s \mathbf{f}_s + \delta_1 \mathbf{g}_1 + \cdots + \delta_t \mathbf{g}_t$$

and so the sequence $\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_s, \mathbf{g}_1, \ldots, \mathbf{g}_t$ spans $W_1 + W_2$.

Finally, we have to show that $\mathbf{e}_1, \ldots, \mathbf{e}_r, \mathbf{f}_1, \ldots, \mathbf{f}_s, \mathbf{g}_1, \ldots, \mathbf{g}_t$ is linearly independent. Suppose that

$$\alpha_1 \mathbf{e}_1 + \cdots + \alpha_r \mathbf{e}_r + \beta_1 \mathbf{f}_1 + \cdots + \beta_s \mathbf{f}_s + \delta_1 \mathbf{g}_1 + \cdots + \delta_t \mathbf{g}_t = \mathbf{0}$$

for some scalars $\alpha_i, \beta_j, \delta_k \in \mathbb{F}$. Then

$$\alpha_1 \mathbf{e}_1 + \cdots + \alpha_r \mathbf{e}_r + \beta_1 \mathbf{f}_1 + \cdots + \beta_s \mathbf{f}_s = -\delta_1 \mathbf{g}_1 - \cdots - \delta_t \mathbf{g}_t \qquad (*)$$

The left-hand-side of this equation lies in W_1 and the right-hand-side of this equation lies in W_2. Since the two sides are equal, both must in fact lie in $W_1 \cap W_2$. Since $\mathbf{e}_1, \ldots, \mathbf{e}_r$ is a basis of $W_1 \cap W_2$, we can write

$$-\delta_1 \mathbf{g}_1 - \cdots - \delta_t \mathbf{g}_t = \gamma_1 \mathbf{e}_1 + \cdots + \gamma_r \mathbf{e}_r$$

for some $\gamma_i \in \mathbb{F}$, and so

$$\gamma_1 \mathbf{e}_1 + \cdots + \gamma_r \mathbf{e}_r + \delta_1 \mathbf{g}_1 + \cdots + \delta_t \mathbf{g}_t = \mathbf{0}.$$

But $\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{g}_1, \dots, \mathbf{g}_t$ form a basis of W_2, so they are linearly independent, and hence $\gamma_i = 0$ for $1 \le i \le r$ and $\delta_i = 0$ for $1 \le i \le t$. But now, from the equation $(*)$ above, we get

$$\alpha_1 \mathbf{e}_1 + \dots + \alpha_r \mathbf{e}_r + \beta_1 \mathbf{f}_1 + \dots + \beta_s \mathbf{f}_s = \mathbf{0}.$$

Now $\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{f}_1, \dots, \mathbf{f}_s$ form a basis of W_1, so they are linearly independent, and hence $\alpha_i = 0$ for $1 \le i \le r$ and $\beta_i = 0$ for $1 \le i \le s$. Thus $\mathbf{e}_1, \dots, \mathbf{e}_r, \mathbf{f}_1, \dots, \mathbf{f}_s, \mathbf{g}_1, \dots, \mathbf{g}_t$ are linearly independent, which completes the proof that they form a basis of $W_1 + W_2$.

Hence

$$\dim(W_1 + W_2) = r + s + t = (r + s) + (r + t) - r = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2).$$

□

Definition 2.2.4. Two subspaces W_1, W_2 of V are called *complementary* if $W_1 \cap W_2 = \{\mathbf{0}\}$ and $W_1 + W_2 = V$. In this case, we say that V is a direct sum of the subspaces W_1 and W_2 and we denote it $V = W_1 \oplus W_2$.

Corollary 2.2.5. *If $V = W_1 \oplus W_2$ is a finite-dimensional vector space, then* $\dim(V) = \dim(W_1) + \dim(W_2)$.

2.2.3 Rank and nullity

Definition 2.2.6. (i) $\dim(\operatorname{im}(T))$ is called the *rank* of T;

(ii) $\dim(\ker(T))$ is called the *nullity* of T.

Theorem 2.2.7. *(Rank-Nullity Formula) Let U, V be vector spaces over $\mathbb{F}$ with U finite-dimensional, and let $T : U \to V$ be a linear map. Then*

$$\dim(\operatorname{im}(T)) + \dim(\ker(T)) = \dim(U); \quad \textit{i.e.,} \quad \operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(U).$$

Proof. Choose a subspace of U, say W such that $U = W \oplus \ker(T)$. This can be done by Theorem 1.4.14: let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be a basis of $\ker(T)$; extend it to a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of V; let W be the span of $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$.

Consider the linear map $T' : W \to \operatorname{im}(T)$ defined by $T'(\mathbf{v}) = T(\mathbf{v})$. The map T' is injective by Proposition 2.2.2 because $\ker(T') = W \cap \ker(T) = \{\mathbf{0}\}$. The map T' is surjective: pick $\mathbf{x} \in \operatorname{im}(T)$, then $\mathbf{x} = T(\mathbf{u})$ for some $\mathbf{u} \in U$, but $\mathbf{u} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{v} \in W$ and $\mathbf{w} \in \ker(T)$, so that $\mathbf{x} = T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}) = T(\mathbf{v}) = T'(\mathbf{v})$. Hence, T' is bijective.

Clearly, if $\mathbf{u}_1, \dots, \mathbf{u}_k$ is a basis of W, then $T'(\mathbf{u}_1), \dots, T'(\mathbf{u}_k)$ is a basis of $\operatorname{im}(T)$. It follows that $\dim(W) = \dim(\operatorname{im}(T))$. By Corollary 2.2.5, $\dim(U) = \dim(W) + \dim(\ker(T)) = \dim(\operatorname{im}(T)) + \dim(\ker(T))$. □

Examples. Once again, we consider the examples from Section 2.1.2. To deal with finite-dimensional spaces we restrict to an $n+1$-dimensional space $\mathbb{F}[x]_{\le n}$ in Examples 4 and 5, that is, we consider $T : \mathbb{R}[x]_{\le n} \to \mathbb{R}[x]_{\le n}$, $S_\alpha : \mathbb{F}[x]_{\le n} \to \mathbb{F}[x]_{\le n}$, and $E_\alpha : \mathbb{F}[x]_{\le n} \to \mathbb{F}$ correspondingly. Let $n = \dim(U) = \dim(V)$ in Examples 6 and 7.

Example	$\mathrm{rank}(T)$	$\mathrm{nullity}(T)$	$\dim(U)$
1.	2	1	3
2.	2	0	2
3.	2	0	2
4.	n	1	$n+1$

Example	$\mathrm{rank}(T)$	$\mathrm{nullity}(T)$	$\dim(U)$
5. S_α	$n+1$	0	$n+1$
5. E_α	1	n	$n+1$
6.	n	0	n
7.	0	n	n

Corollary 2.2.8. *Let $T : U \to V$ be a linear map, where $\dim(U) = \dim(V) = n$. Then the following properties of T are equivalent:*

(i) *T is surjective;*

(ii) $\mathrm{rank}(T) = n$;

(iii) $\mathrm{nullity}(T) = 0$;

(iv) *T is injective;*

(v) *T is bijective;*

Proof. T is surjective $\Leftrightarrow$ $\mathrm{im}(T) = V$, so clearly (i) $\Rightarrow$ (ii). But if $\mathrm{rank}(T) = n$, then $\dim(\mathrm{im}(T)) = \dim(V)$ so (by Corollary 1.4.15) a basis of $\mathrm{im}(T)$ is a basis of V, and hence $\mathrm{im}(T) = V$. Thus (i) $\Leftrightarrow$ (ii).

(ii) $\Leftrightarrow$ (iii) follows directly from Theorem 2.2.7.

(iii) $\Leftrightarrow$ (iv) is part (iii) of Proposition 2.2.2.

Finally, (v) is equivalent to (i) and (iv), which are equivalent to each other. $\square$

Definition 2.2.9. If the conditions in the above corollary are met, then T is called a *non-singular* linear map. Otherwise, T is called *singular*.

(But normally the terms singular and non-singular are only used for linear maps $T : U \to V$ for which U and V have the same dimension.)

2.3 Matrices

Let $\mathbb{F}$ be a field and $m, n \in \mathbb{N}$. An $m \times n$ matrix A over $\mathbb{F}$ is an $m \times n$ rectangular array of numbers (i.e., scalars) in $\mathbb{F}$. The entry in row i and column j is usually written α_{ij}. (We use the corresponding Greek letter.) We write $A = (\alpha_{ij})$ to make things clear.

For example, we could take

$$\mathbb{F} = \mathbb{R}, \quad m = 3, \ n = 4, \quad A = \begin{pmatrix} 2 & -1 & -\pi & 0 \\ 3 & -3/2 & 0 & 6 \\ -1.23 & 0 & 10^{10} & 0 \end{pmatrix},$$

and then $\alpha_{13} = -\pi$, $\alpha_{33} = 10^{10}$, $\alpha_{34} = 0$, etc.

2.3.1 Matrices as linear transformations

A matrix $A \in \mathbb{F}^{m,n}$ yields a linear map between the standard vector spaces

$$L_A : \mathbb{F}^n \to \mathbb{F}^m, \ \ \mathbf{v} \mapsto L_A(\mathbf{v}) = A\mathbf{v}\,. \tag{1}$$

Let us see how this plays out on the first three examples from Section 2.1.2.

1. The matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ yields the projection $T : \mathbb{R}^3 \to \mathbb{R}^2$.

2. The matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ yields the rotation $T : \mathbb{R}^2 \to \mathbb{R}^2$.

3. The matrix $\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$ yields the reflection $T : \mathbb{R}^2 \to \mathbb{R}^2$.

2.3.2 Matrix operations

The material in this section must be familiar to you already.

Addition of matrices. Let $A = (\alpha_{ij})$ and $B = (\beta_{ij})$ be two $m \times n$ matrices over $\mathbb{F}$. We define $A + B$ to be the $m \times n$ matrix $C = (\gamma_{ij})$, where $\gamma_{ij} = \alpha_{ij} + \beta_{ij}$ for all i, j. For example,

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

Scalar multiplication. Let $A = (\alpha_{ij})$ be an $m \times n$ matrix over $\mathbb{F}$ and let $\beta \in \mathbb{F}$ be a scalar. We define the scalar multiple βA to be the $m \times n$ matrix $C = (\gamma_{ij})$, where $\gamma_{ij} = \beta\alpha_{ij}$ for all i, j.

Multiplication of matrices. Let $A = (\alpha_{ij})$ be an $l \times m$ matrix over $\mathbb{F}$ and let $B = (\beta_{ij})$ be an $m \times n$ matrix over $\mathbb{F}$. The product AB is an $l \times n$ matrix $C = (\gamma_{ij})$ where, for $1 \leq i \leq l$ and $1 \leq j \leq n$,

$$\gamma_{ij} = \sum_{k=1}^{m} \alpha_{ik}\beta_{kj} = \alpha_{i1}\beta_{1j} + \alpha_{i2}\beta_{2j} + \cdots + \alpha_{im}\beta_{mj}.$$

It is essential that the number m of columns of A is equal to the number of rows of B. Otherwise AB makes no sense. If you are familiar with scalar products of vectors, note also that γ_{ij} is the scalar product of the i-th row of A with the j-th column of B.

For example, let

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 6 \\ 3 & 2 \\ 1 & 9 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 2\times 2+3\times 3+4\times 1 & 2\times 6+3\times 2+4\times 9 \\ 1\times 2+6\times 3+2\times 1 & 1\times 6+6\times 2+2\times 9 \end{pmatrix} = \begin{pmatrix} 17 & 54 \\ 22 & 36 \end{pmatrix},$$

$$BA = \begin{pmatrix} 10 & 42 & 20 \\ 8 & 21 & 16 \\ 11 & 57 & 22 \end{pmatrix}.$$

Let $C = \begin{pmatrix} 2 & 3 & 1 \\ 6 & 2 & 9 \end{pmatrix}$. Then AC and CA are not defined.

Let $D = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. Then AD is not defined, but $DA = \begin{pmatrix} 4 & 15 & 8 \\ 1 & 6 & 2 \end{pmatrix}$.

Proposition 2.3.1. *Matrices satisfy the following laws whenever the sums and products involved are defined:*

(i) $A + B = B + A$;

(ii) $(A + B)C = AC + BC$;

(iii) $C(A + B) = CA + CB$;

(iv) $(\lambda A)B = \lambda(AB) = A(\lambda B)$;

(v) $(AB)C = A(BC)$.

Proof. These are all routine checks that the entries of the left-hand-sides are equal to the corresponding entries on the right-hand-side. Let us do (v) as an example.

Let A, B and C be $l \times m$, $m \times n$ and $n \times p$ matrices, respectively. Then $AB = D = (\delta_{ij})$ is an $l \times n$ matrix with $\delta_{ij} = \sum_{s=1}^{m} \alpha_{is}\beta_{sj}$, and $BC = E = (\varepsilon_{ij})$ is an $m \times p$ matrix with $\varepsilon_{ij} = \sum_{t=1}^{n} \beta_{it}\gamma_{tj}$. Then $(AB)C = DC$ and $A(BC) = AE$ are both $l \times p$ matrices, and we have to show that their coefficients are equal. The (i, j)-coefficient of DC is

$$\sum_{t=1}^{n} \delta_{it}\gamma_{tj} = \sum_{t=1}^{n}(\sum_{s=1}^{m} \alpha_{is}\beta_{st})\gamma_{tj} = \sum_{s=1}^{m} \alpha_{is}(\sum_{t=1}^{n} \beta_{st}\gamma_{tj}) = \sum_{s=1}^{m} \alpha_{is}\varepsilon_{sj}$$

which is the (i, j)-coefficient of AE. Hence $(AB)C = A(BC)$. $\square$

The zero and identity matrices. The $m \times n$ *zero matrix* $\mathbf{0}_{mn}$ over any field $\mathbb{F}$ has all of its entries equal to 0.

The $n \times n$ *identity matrix* $I_n = (\alpha_{ij})$ over any field $\mathbb{F}$ has $\alpha_{ii} = 1$ for $1 \leq i \leq n$, but $\alpha_{ij} = 0$ when $i \neq j$. For example,

$$I_1 = (1), \quad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that $I_n A = A$ for any $n \times m$ matrix A and $AI_n = A$ for any $m \times n$ matrix A.

Transposition. Let $A = (\alpha_{ij}) \in \mathbb{F}^{n,k}$. Its transposed matrix is $A^{\mathrm{T}} = (\alpha_{ji}) \in \mathbb{F}^{k,n}$. For example, if

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 1 & 6 & 8 \end{pmatrix} \text{ then } A^{\mathrm{T}} = \begin{pmatrix} 2 & 1 \\ 3 & 6 \\ 4 & 8 \end{pmatrix}.$$

The transposition is a linear map $\mathbb{F}^{n,k} \to \mathbb{F}^{k,n}$ such that $(XY)^{\mathrm{T}} = Y^{\mathrm{T}} X^{\mathrm{T}}$ (check this).

The inverse. Consider $A = (\alpha_{ij}) \in \mathbb{F}^{n,n}$. If the linear operator L_A from (1), satisfies Corollary 2.2.8, we say that A is invertible and it admits a unique inverse matrix A^{-1}, representing the inverse bijection L_A^{-1}.

Row and column vectors. The set of all $m \times n$ matrices over $\mathbb{F}$ will be denoted by $\mathbb{F}^{m,n}$. Note that $\mathbb{F}^{m,n}$ is itself a vector space over $\mathbb{F}$ using the operations of addition and scalar multiplication defined above, and it has dimension mn. (This should be obvious - is it?)

A $1 \times n$ matrix is called a *row vector*.

A $n \times 1$ matrix is called a *column vector*. We regard $\mathbb{F}^{n,1}$ as being the same as $\mathbb{F}^n$.

3 From Linear Transformations to Matrices

In this section we learn to represent a vector by a column, then a linear map by a matrix. Let V be a vector space with a basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$. For each $\mathbf{v} \in V$ there exist unique $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1 \mathbf{f}_1 + \ldots + \alpha_n \mathbf{f}_n$. Then the row

$$\underline{\mathbf{v}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{F}^n$$

is the coordinate vector of $\mathbf{v}$. We will this underlined notation consistently: $\underline{\mathbf{v}}$ will always be the coordinate vector of $\mathbf{v}$ in some basis.

We first address how to find the coordinate expressions in practice.

Example. Let $\mathbf{v} = x^3 - 1 \in \mathbb{F}[x]_{\leq 3}$ and $\mathbf{f}_1 = x^3 - x, \mathbf{f}_2 = (x+1)^2, \mathbf{f}_3 = x + 3, \mathbf{f}_4 = x + 1$ be a basis. Then

$$x^3 - 1 = (x^3 - x) + 2(x+1) - (x+3) = 1 \cdot \mathbf{f}_1 + 0 \cdot \mathbf{f}_2 + (-1) \cdot \mathbf{f}_3 + 2 \cdot \mathbf{f}_4$$

$$\text{so that} \qquad \underline{\mathbf{v}} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix} \in \mathbb{F}^n.$$

Remark for the inquisitive reader: Many online sources and books have a slightly different definition of a basis. Instead of vector sequences, they talk about subsets. One disatvantage they have is that they need to order their "basis" first to be able to write coordinates and matrices.

3.1 Case of standard vector space

Consider an example $V = \mathbb{R}^2$ with a basis $\mathbf{f}_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \mathbf{f}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. To write an arbitrary vector $\mathbf{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ as a linear combination of the basis, we need to solve the system of linear equations

$$\mathbf{v} = x\mathbf{f}_1 + y\mathbf{f}_2 \quad \text{or, equivalently,} \quad \begin{cases} -1 \cdot x + 0 \cdot y & = \alpha \\ 2 \cdot x + 1 \cdot y & = \beta \end{cases}$$

that we can easily solve: $x = -\alpha$ and $y = \beta - 2x = \beta + 2\alpha$. It follows that

$$\underline{\mathbf{v}} = \begin{pmatrix} -\alpha \\ 2\alpha + \beta \end{pmatrix} = P\mathbf{v} \quad \text{where} \quad P = \begin{pmatrix} -1 & 0 \\ 2 & 1 \end{pmatrix}.$$

The latest matrix P is called *the change of basis matrix*. See also Section 4.

3.1.1 Change of bases matrix

Definition 3.1.1. Let $\mathbf{f}_1, \ldots, \mathbf{f}_n$ be a basis of $\mathbb{F}^n$. Write each element of the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in this basis. The change of basis matrix is the matrix $P = (\underline{\mathbf{e}_1}, \ldots, \underline{\mathbf{e}_n}) \in \mathbb{F}^{n,n}$.

Example. Let $U = \mathbb{R}^3$, and $\mathbf{f}_1 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$, $\mathbf{f}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\mathbf{f}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Then $\mathbf{e}_1 = \mathbf{f}_3$, $\mathbf{e}_2 = \mathbf{f}_2 - \mathbf{f}_3$ and $\mathbf{e}_3 = \mathbf{f}_1 - 2\mathbf{f}_2 + 2\mathbf{f}_3$ so that

$$\underline{\mathbf{e}_1} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \ \underline{\mathbf{e}_2} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \ \underline{\mathbf{f}_3} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & -2 \\ 1 & -1 & 2 \end{pmatrix}.$$

Proposition 3.1.2. *With the above notation, let $\mathbf{v} \in \mathbb{F}^n$. Then $P\mathbf{v} = \underline{\mathbf{v}}$.*

Proof. Let us set the notation for the main protagonists as

$$P = (\sigma_{ij}), \qquad \mathbf{v} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} \qquad \text{and} \qquad \underline{\mathbf{v}} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix},$$

so that $\mathbf{v} = \sum_{i=1}^n \lambda_i \mathbf{e}_i$ and $\mathbf{v} = \sum_{j=1}^n \alpha_j \mathbf{f}_j$. By definition of P, $\mathbf{e}_j = \sum_{i=1}^n \sigma_{ij}\mathbf{f}_i$ for $1 \leq j \leq n$. Using this, we get

$$\mathbf{v} = \sum_{j=1}^n \lambda_j \mathbf{e}_j = \sum_{j=1}^n \lambda_j \sum_{i=1}^n \sigma_{ij}\mathbf{f}_i = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}\lambda_j \mathbf{f}_i$$

and then comparing coefficients of $\mathbf{f}_i$ in the expansions for $\mathbf{v}$ gives $\alpha_i = \sum_{j=1}^n \sigma_{ij}\lambda_j$ for $1 \leq i \leq n$. That is, $\underline{\mathbf{v}} = P\mathbf{v}$. $\square$

In the example above, if $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, then $\underline{\mathbf{v}} = P\mathbf{v} = \begin{pmatrix} z \\ y - 2z \\ x - y + 2z \end{pmatrix}$.

3.2 Case of euclidean spaces

We are used to thinking about $\mathbb{R}^2$ in the realm of Euclidean Geometry. Let us lie the foundation of such geometry on a general vector space.

Throughout Section 3.2, V is a vector space over $\mathbb{R}$.

3.2.1 Euclidean space

Definition 3.2.1. A *euclidean form* on V is a map $\tau : V \times V \to \mathbb{R}$ such that

(i) $\tau(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2, \mathbf{w}) = \alpha_1\tau(\mathbf{v}_1, \mathbf{w}) + \alpha_2\tau(\mathbf{v}_2, \mathbf{w})$ for all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V$ and $\alpha_1, \alpha_2 \in \mathbb{R}$,

(ii) $\tau(\mathbf{v}, \mathbf{w}) = \tau(\mathbf{w}, \mathbf{v})$ for all $\mathbf{v}, \mathbf{w} \in V$,

(iii) $\tau(\mathbf{v}, \mathbf{v}) > 0$ for all $\mathbf{v} \in V \setminus \{\mathbf{0}\}$.

Notice that τ is not a linear map. However, if you fix $\mathbf{v} \in V$, then the map $\mathbf{w} \mapsto \tau(\mathbf{v}, \mathbf{w})$ is linear. Such maps τ are called "bilinear".

Definition 3.2.2. A *euclidean space* is a pair (V, τ) where V is a vector space over $\mathbb{R}$ and τ is a euclidean form in V.

Examples. 1. Let $V = \mathbb{R}^n$, τ – the usual dot-product. In other words,

$$\tau(\,(\alpha_i)\,,\,(\beta_i)\,) = (\alpha_i) \bullet (\beta_i) = \sum_{j=1}^{n} \alpha_j\beta_j\,.$$

We call this *the standard euclidean space* and denote it $\mathbb{R}^n$ rather than $(\mathbb{R}^n, \bullet)$.

2. Let $V = \mathbb{R}^2$. Suppose $f(x) = ax^2 + 2bx + c \in \mathbb{R}[x]$ is a quadratic polynomial such that $f(x) > 0$ for all $x \in \mathbb{R}$. This holds if $a > 0$ and the discriminant is negative. It defines a euclidean form

$$\tau(\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}, \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}) = a\alpha_1\beta_1 + b\alpha_1\beta_2 + b\alpha_2\beta_1 + c\alpha_2\beta_2\,.$$

3. Let $V = \mathbb{R}[x]_{\leq n}$ and $a < b \in \mathbb{R}$. Suppose $f(x) \in \mathbb{R}[x]$ is a non-zero polynomial such that $f(x) \geq 0$ for all $x \in (a, b)$. It defines a euclidean form

$$\tau(g(x), h(x)) = \int_a^b f(x)g(x)h(x)dx\,.$$

3.2.2 Length and angle in euclidean space

Let (V, τ) be a euclidean space. For $\mathbf{v} \in V$, we define its length by $\|\mathbf{v}\| = \sqrt{\tau(\mathbf{v}, \mathbf{v})}$.

Proposition 3.2.3. *(Cauchy-Schwarz Inequality) Suppose* $\mathbf{v}, \mathbf{w} \in (V, \tau)$. *Then*

$$|\tau(\mathbf{v}, \mathbf{w})| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|\,.$$

Proof. Fix $\mathbf{v}$, $\mathbf{w}$ and consider the function $f(x) : \mathbb{R} \to \mathbb{R}$, given by $f(x) = \|\mathbf{v} + x\mathbf{w}\|^2$. Notice that it is a quadratic polynomial

$$f(x) = \tau(\mathbf{v} + x\mathbf{w}, \mathbf{v} + x\mathbf{w}) = \tau(\mathbf{w}, \mathbf{w})x^2 + 2\tau(\mathbf{v}, \mathbf{w})x + \tau(\mathbf{v}, \mathbf{v})\,.$$

It follows from part (iii) of Definition 3.2.1 that $f(x) \geq 0$. Thus, its discriminant is not positive:

$$(2\tau(\mathbf{v}, \mathbf{w}))^2 - 4\tau(\mathbf{w}, \mathbf{w})\tau(\mathbf{v}, \mathbf{v}) \leq 0\,.$$

It follows that

$$\tau(\mathbf{v},\mathbf{w})^2 \leq \tau(\mathbf{w},\mathbf{w})\tau(\mathbf{v},\mathbf{v}) \text{ and } |\tau(\mathbf{v},\mathbf{w})| = \sqrt{\tau(\mathbf{v},\mathbf{w})^2} \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|\,.$$

□

Proposition 3.2.3 allows use to define the angle φ between any two non-zero vectors $\mathbf{v}$ and $\mathbf{w}$ by

$$\tau(\mathbf{v},\mathbf{w}) = \|\mathbf{v}\| \cdot \|\mathbf{w}\| \cdot \cos\varphi \quad \text{or} \quad \varphi = \arccos \frac{\tau(\mathbf{v},\mathbf{w})}{\|\mathbf{v}\| \cdot \|\mathbf{w}\|}\,.$$

3.2.3 Orthonormal basis

Definition 3.2.4. A vector sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of a euclidean space (V,τ) is called *orthonormal*, if $\|\mathbf{v}_i\| = 1$ for all i and the angle between each $\mathbf{v}_i$ and $\mathbf{v}_j$ with $i \neq j$ is equal to $\pi/2$. *An orthonormal basis* is a basis, which is an orthonormal sequence.

In other words, the sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is orthonormal if and only if

$$\tau(\mathbf{v}_i,\mathbf{v}_j) = \delta_{ij}, \quad \text{where} \quad \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The orthonormal bases are easier to deal with because of the following property.

Lemma 3.2.5. *Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an orthonormal sequence in a euclidean space (V,τ).*

1. *If $\mathbf{v} = \alpha_1\mathbf{v}_1 + \ldots + \alpha_n\mathbf{v}_n$, then $\alpha_i = \tau(\mathbf{v},\mathbf{v}_i)$.*
2. *The sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly independent.*

Proof. If $\mathbf{v} = \alpha_1\mathbf{v}_1 + \ldots + \alpha_n\mathbf{v}_n$, then

$$\tau(\mathbf{v},\mathbf{v}_i) = \tau(\alpha_1\mathbf{v}_1 + \ldots + \alpha_n\mathbf{v}_n,\mathbf{v}_i) = \alpha_1\tau(\mathbf{v}_1,\mathbf{v}_i) + \ldots + \alpha_n\tau(\mathbf{v}_n,\mathbf{v}_i) = \alpha_i\tau(\mathbf{v}_i,\mathbf{v}_i) = \alpha_i\,.$$

The second statement follows immediately: if $\alpha_1\mathbf{v}_1 + \ldots + \alpha_n\mathbf{v}_n = \mathbf{0}$ is a linear dependency, then all $\alpha_i = 0$ by the first statement. □

Examples. 1. On the standard euclidean space $V = \mathbb{R}^2$, let us rotate the standard basis by the angle θ: $\mathbf{v}_1 = \begin{pmatrix}\cos\theta \\ \sin\theta\end{pmatrix}$, $\mathbf{v}_2 = \begin{pmatrix}-\sin\theta \\ \cos\theta\end{pmatrix}$ is an orthonormal basis. If $\mathbf{v} = \begin{pmatrix}x \\ y\end{pmatrix}$, then

$$\underline{\mathbf{v}} = \begin{pmatrix}\mathbf{v} \bullet \mathbf{v}_1 \\ \mathbf{v} \bullet \mathbf{v}_2\end{pmatrix} = \begin{pmatrix}x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta\end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{pmatrix}\mathbf{v}\,.$$

It is instructive to check that

$$(\mathbf{v}_1 \bullet \mathbf{v})\mathbf{v}_1 + (\mathbf{v}_2 \bullet \mathbf{v})\mathbf{v}_2 = (x\cos\theta + y\sin\theta)\begin{pmatrix}\cos\theta \\ \sin\theta\end{pmatrix} + (-x\sin\theta + y\cos\theta)\begin{pmatrix}-\sin\theta \\ \cos\theta\end{pmatrix} =$$

$$= \begin{pmatrix} (x\cos\theta + y\sin\theta)\cos\theta + (x\sin\theta - y\cos\theta)\sin\theta \\ (x\cos\theta + y\sin\theta)\sin\theta - (x\sin\theta - y\cos\theta)\cos\theta \end{pmatrix} = \begin{pmatrix} x(\cos^2\theta + \sin^2\theta) \\ y(\sin^2\theta + \cos^2\theta) \end{pmatrix} = \mathbf{v}\,.$$

2. On the standard euclidean space $V = \mathbb{R}^4$, the vector sequence

$$\mathbf{v}_1 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{v}_2 = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \ \mathbf{v}_3 = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \ \mathbf{v}_4 = \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

is an orthonormal basis. Hence,

$$\mathbf{v} = \frac{1}{2}\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} \quad \Rightarrow \quad \underline{\mathbf{v}} = \begin{pmatrix} \mathbf{v}\bullet\mathbf{v}_1 \\ \mathbf{v}\bullet\mathbf{v}_2 \\ \mathbf{v}\bullet\mathbf{v}_3 \\ \mathbf{v}\bullet\mathbf{v}_4 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} w+x+y+z \\ w-x-y+z \\ w+x-y-z \\ w-x+y-z \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{pmatrix}\mathbf{v}\,.$$

3.2.4 Existence of orthonormal basis

Theorem 3.2.6 (Gram-Schmidt process)**.** *Let V be a euclidean space of dimension n. Suppose that, for some r with $0 \le r \le n$, $\mathbf{f}_1, \ldots, \mathbf{f}_r$ is an orthonormal sequence. Then $\mathbf{f}_1, \ldots, \mathbf{f}_r$ can be extended to an orthonormal basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$ of V.*

Proof. The proof of the theorem will be by induction on $n - r$. We can start the induction with the case $n - r = 0$, when $r = n$, and there is nothing to prove.

Assume that $n - r > 0$. Then $r < n$. By Theorem 1.4.14, we can extend the sequence to a basis $\mathbf{f}_1, \ldots, \mathbf{f}_r, \mathbf{g}_{r+1}, \ldots, \mathbf{g}_n$ of V, containing the $\mathbf{f}_i$. The Gram-Schmidt process trick is to define

$$\mathbf{f}'_{r+1} = \mathbf{g}_{r+1} - \sum_{i=1}^{r} \tau(\mathbf{f}_i, \mathbf{g}_{r+1})\mathbf{f}_i.$$

Apply the scalar product of $\tau(\mathbf{f}_j, \ )$ to this equation for some $1 \le j \le r$ and use orthogonality:

$$\tau(\mathbf{f}_j, \mathbf{f}'_{r+1}) = \tau(\mathbf{f}_j, \mathbf{g}_{r+1}) - \sum_{i=1}^{r} \tau(\mathbf{f}_i, \mathbf{g}_{r+1})\tau(\mathbf{f}_j, \mathbf{f}_i)\tau(\mathbf{f}_j, \mathbf{g}_{r+1}) - \tau(\mathbf{f}_j, \mathbf{g}_{r+1}) = \mathbf{0}.$$

The vector $\mathbf{f}'_{r+1}$ is non-zero by linear independence of the basis, and if we define $\mathbf{f}_{r+1} = \mathbf{f}'_{r+1}/\|\mathbf{f}'_{r+1}\|$, then we still have $\tau(\mathbf{f}_j, \mathbf{f}_{r+1}) = 0$ for $1 \le j \le r$, and we also have $\|\mathbf{f}_{r+1}\| = 1$. Hence $\mathbf{f}_1, \ldots, \mathbf{f}_{r+1}$ is an orthonrmal sequence. The result follows by inductive hypothesis. □

3.3 Correspondence between linear transformations and matrices

We shall see in this section that, for fixed choice of bases, there is a very natural one-one correspondence between linear maps and matrices, such that the operations on linear maps and matrices defined in Chapters 2 and 2.3 also correspond to each other. This is perhaps the most important idea in linear algebra, because it enables us to use matrices to compute with linear maps.

3.3.1 Setting up the correspondence

Let $T : U \to V$ be a linear map, where $\dim(U) = n$, $\dim(V) = m$. Choose a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of U and a basis $\mathbf{f}_1, \ldots, \mathbf{f}_m$ of V. For each $1 \le j \le n$, $T(\mathbf{e}_j) \in V$, so $T(\mathbf{e}_j)$ can be written uniquely as a linear combination of $\mathbf{f}_1, \ldots, \mathbf{f}_m$. Let

$$\begin{array}{rcl} T(\mathbf{e}_1) & = & \alpha_{11}\mathbf{f}_1 + \alpha_{21}\mathbf{f}_2 + \cdots + \alpha_{m1}\mathbf{f}_m \\ T(\mathbf{e}_2) & = & \alpha_{12}\mathbf{f}_1 + \alpha_{22}\mathbf{f}_2 + \cdots + \alpha_{m2}\mathbf{f}_m \\ \cdots & & \\ T(\mathbf{e}_n) & = & \alpha_{1n}\mathbf{f}_1 + \alpha_{2n}\mathbf{f}_2 + \cdots + \alpha_{mn}\mathbf{f}_m \end{array}$$

where the coefficients $\alpha_{ij} \in \mathbb{F}$ (for $1 \le i \le m$, $1 \le j \le n$) are uniquely determined. In other words,

$$T(\mathbf{e}_j) = \sum_{i=1}^{m} \alpha_{ij}\mathbf{f}_i \text{ for } 1 \le j \le n.$$

The coefficients α_{ij} form an $m \times n$ matrix

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ & & \ldots & \\ \alpha_{m1} & \alpha_{m2} & \ldots & \alpha_{mn} \end{pmatrix}$$

over $\mathbb{F}$. Then A is called the matrix of the linear map T with respect to the chosen bases of U and V. In general, different choice of bases gives different matrices.

Proposition 3.3.1. *Let $T : U \to V$ be a linear map with matrix $A = (\alpha_{ij})$. Then $T(\mathbf{u}) = \mathbf{v}$ if and only if $A\underline{\mathbf{u}} = \underline{\mathbf{v}}$.*

Proof. We have

$$T(\mathbf{u}) = T(\sum_{j=1}^{n} \lambda_j \mathbf{e}_j) = \sum_{j=1}^{n} \lambda_j T(\mathbf{e}_j) = \sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{m} \alpha_{ij}\mathbf{f}_i) = \sum_{i=1}^{m} (\sum_{j=1}^{n} \alpha_{ij}\lambda_j)\mathbf{f}_i = \sum_{i=1}^{m} \mu_i \mathbf{f}_i,$$

where $\mu_i = \sum_{j=1}^{n} \alpha_{ij}\lambda_j$ is the entry in the i-th row of the column vector $A\underline{\mathbf{u}}$. This proves the result. $\square$

In particular, notice the role of the individual columns in A. The j-th column of A is $\underline{T(\mathbf{e}_j)}$ (in the basis $\mathbf{f}_1, \ldots, \mathbf{f}_m$).

3.3.2 Examples

Once again, we consider our examples from Section 2.

1. $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$. Usually, we choose the standard bases of $\mathbb{F}^m$ and $\mathbb{F}^n$, which in this case are

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

We have $T(\mathbf{e}_1) = \mathbf{f}_1$, $T(\mathbf{e}_2) = \mathbf{f}_2$, $T(\mathbf{e}_3) = \mathbf{0}$, and the matrix is

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

But suppose we choose a different basis, say

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

Then we have

$$T(\mathbf{e}'_1) = \begin{pmatrix}1\\1\end{pmatrix} = \mathbf{f}'_2,\ T(\mathbf{e}'_2) = \begin{pmatrix}0\\1\end{pmatrix} = \mathbf{f}'_1,\ T(\mathbf{e}'_3) = \begin{pmatrix}1\\0\end{pmatrix} = -\mathbf{f}'_1 + \mathbf{f}'_2$$

and the matrix of T in these basis is

$$B = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix}.$$

2. $T : \mathbb{R}^2 \to \mathbb{R}^2$, T is a rotation by θ anti-clockwise about the origin. We saw that in the standard bases

$$T(\mathbf{e}_1) = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix},\ T(\mathbf{e}_2) = \begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix} \text{ and } A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

3. $T : \mathbb{R}^2 \to \mathbb{R}^2$, T is a reflection through the line through the origin making an angle $\theta/2$ with the x-axis. We saw that in the standard bases

$$T(\mathbf{e}_1) = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix},\ T(\mathbf{e}_2) = \begin{pmatrix}\sin\theta\\ -\cos\theta\end{pmatrix} \text{ and } A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Let us do something unusual and change the basis in the range to $\mathbf{f}_1 = T(\mathbf{e}_1)$, $\mathbf{f}_2 = T(\mathbf{e}_1)$. Then the matrix of T in the basis $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{f}_1, \mathbf{f}_2$ is

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

See Section 3.3.3 for more.

4. This time we take the differentiation map T from $\mathbb{R}[x]_{\leq n}$ to $\mathbb{R}[x]_{\leq n-1}$. Then, with respect to the bases $1, x, x^2, \ldots, x^n$ and $1, x, x^2, \ldots, x^{n-1}$ of $\mathbb{R}[x]_{\leq n}$ and $\mathbb{R}[x]_{\leq n-1}$, respectively, the matrix of T is

$$\begin{pmatrix} 0 & 1 & 0 & & \ldots & & 0 \\ 0 & 0 & 2 & & \ldots & & 0 \\ 0 & 0 & 0 & 3 & \ldots & & 0 \\ 0 & 0 & 0 & & \ldots & n-1 & 0 \\ 0 & 0 & 0 & & \ldots & & n \end{pmatrix}.$$

5. Let $S_\alpha : \mathbb{F}[x]_{\leq n} \to \mathbb{F}[x]_{\leq n}$ be the shift. With respect to the basis $1, x, x^2, \ldots, x^n$ of $\mathbb{F}[x]_{\leq n}$, we calculate $S_\alpha(x^n) = (x - \alpha)^n$. The binomial formula gives the matrix of S_α,

$$\begin{pmatrix} 1 & -\alpha & \alpha^2 & & \ldots & & (-1)^n \alpha^n \\ 0 & 1 & -2\alpha & & \ldots & & (-1)^{n-1} n \alpha^{n-1} \\ 0 & 0 & 1 & 3 & \ldots & & (-1)^{n-2} \frac{n(n-1)}{2} \alpha^{n-2} \\ \vdots & & & & & & \vdots \\ 0 & 0 & 0 & & \ldots & 1 & -n\alpha \\ 0 & 0 & 0 & & \ldots & & 1 \end{pmatrix}.$$

In the same basis of $\mathbb{F}[x]_{\leq n}$ and the basis 1 of $\mathbb{F}$, $E_\alpha(x^n) = \alpha^n$. The matrix of E_α is

$$\begin{pmatrix} 1 \\ \alpha \\ \alpha^2 \\ \vdots \\ \alpha^{n-1} \\ \alpha^n \end{pmatrix}.$$

6. $T : V \to V$ is the identity map. Notice that $U = V$ in this example, and in that case we can (but we do not have to – See Section 3.3.3) choose the same basis for U and V. The matrix of T is the $n \times n$ identity matrix I_n.

7. $T : U \to V$ is the zero map. The matrix of T is the $m \times n$ zero matrix $\mathbf{0}_{mn}$.

3.3.3 Maps, operators and four key problems

Given a linear map $T : U \to V$ between finite dimensional vector spaces, can we choose the bases, where the matrix of T looks "the best"? A more scientific term for "the best" is "the normal form". This problem has 4 different variations and "the normal form" will be different in each of them.

LT: Linear Transformation Problem: Choose bases of U and V such that the matrix of T looks "the best".

If $U = V$, the linear map T is also called *a linear operator on the space U*. In this case, we have good reasons just to choose one basis (see examples 3 an 6 in Section 3.3.2).

LO: Linear Operator Problem: Suppose $U = V$. Choose a basis of U such that the matrix of T looks "the best".

If U is a euclidean space, then not all bases are "equal". As we saw already, the orthonormal bases are special. We often insist on using orthonormal bases.

ET: Euclidean Transformation Problem. Suppose U and V are euclidean spaces. Choose orthonormal bases of U and V such that the matrix of T looks "the best".

EO: Euclidean Operator Problem. Suppose $U = V$ is a euclidean space. Choose an orthonormal basis of U such that the matrix of T looks "the best".

3.4 Properties of the correspondence

Let U, V and W be vector spaces over the same field $\mathbb{F}$, let $\dim(U) = n$, $\dim(V) = m$, $\dim(W) = l$, and choose fixed bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of U and $\mathbf{f}_1, \ldots, \mathbf{f}_m$ of V, and $\mathbf{g}_1, \ldots, \mathbf{g}_l$ of W. All matrices of linear maps between these spaces will be written with respect to these bases.

Bijection.

Theorem 3.4.1. *There is a one-one correspondence between the set* $\mathrm{Hom}_{\mathbb{F}}(U, V)$ *of linear maps* $U \to V$ *and the set* $\mathbb{F}^{m,n}$ *of* $m \times n$ *matrices over* $\mathbb{F}$.

Proof. As we saw above, any linear map $T : U \to V$ determines an $m \times n$ matrix A over $\mathbb{F}$.

Conversely, let $A = (\alpha_{ij})$ be an $m \times n$ matrix over $\mathbb{F}$. Then, by Proposition 2.1.3, there is just one linear $T : U \to V$ with $T(\mathbf{e}_j) = \sum_{i=1}^{m} \alpha_{ij}\mathbf{f}_i$ for $1 \leq j \leq n$, so we have a one-one correspondence. $\square$

Addition. Let $T_1, T_2 : U \to V$ be linear maps with $m \times n$ matrices A, B respectively. Then it is routine to check that the matrix of $T_1 + T_2$ is $A + B$.

Scalar multiplication. Let $T : U \to V$ be a linear map with $m \times n$ matrices A and let $\lambda \in \mathbb{F}$ be a scalar. Then again it is routine to check that the matrix of λT is λA.

Note that the above two properties imply that the correspondence between linear maps and matrices in Theorem 3.4.1 is actually itself a linear map from $\mathrm{Hom}_{\mathbb{F}}(U, V)$ to $\mathbb{F}^{m,n}$.

Composition of linear maps and matrix multiplication. This time the correspondence is less obvious, and we state it as a theorem.

Theorem 3.4.2. *Let $T_1 : V \to W$ be a linear map with $l \times m$ matrix $A = (\alpha_{ij})$ and let $T_2 : U \to V$ be a linear map with $m \times n$ matrix $B = (\beta_{ij})$. Then the matrix of the composite map $T_1T_2 : U \to W$ is AB.*

Proof. Let AB be the $l \times n$ matrix (γ_{ij}). Then by the definition of matrix multiplication, we have $\gamma_{ik} = \sum_{j=1}^{m} \alpha_{ij}\beta_{jk}$ for $1 \le i \le l, 1 \le k \le n$.

Let us calculate the matrix of T_1T_2. We have

$$T_1T_2(\mathbf{e}_k) = T_1(\sum_{j=1}^{m} \beta_{jk}\mathbf{f}_j) = \sum_{j=1}^{m} \beta_{jk}T_1(\mathbf{f}_j) = \sum_{j=1}^{m} \beta_{jk} \sum_{i=1}^{l} \alpha_{ij}\mathbf{g}_i = \sum_{i=1}^{l}(\sum_{j=1}^{m} \alpha_{ij}\beta_{jk})\mathbf{g}_i = \sum_{i=1}^{l} \gamma_{ik}\mathbf{g}_i,$$

so the matrix of T_1T_2 is $(\gamma_{ik}) = AB$ as claimed. □

Examples.

1. Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation through an angle θ anti-clockwise about the origin. We have seen that the matrix of R_θ (using the standard basis) is $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Now clearly R_θ followed by R_ϕ is equal to $R_{\theta+\phi}$. In other words, $R_\phi R_\theta = R_{\theta+\phi}$. This can be checked by a matrix calculation:

$$\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\phi\cos\theta - \sin\phi\sin\theta & -\cos\phi\sin\theta - \sin\phi\cos\theta \\ \sin\phi\cos\theta + \cos\phi\sin\theta & -\sin\phi\sin\theta + \cos\phi\cos\theta \end{pmatrix}$$
$$= \begin{pmatrix} \cos(\phi+\theta) & -\sin(\phi+\theta) \\ \sin(\phi+\theta) & \cos(\phi+\theta) \end{pmatrix}.$$

2. Let R_θ be as in Example 1, and let $M_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be a reflection through a line through the origin at an angle $\theta/2$ to the x-axis. We have seen that the matrix of M_θ is $\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$. What is the effect of doing first R_θ and then M_ϕ? In this case, it might be easier (for some people) to work it out using the matrix multiplication! We have

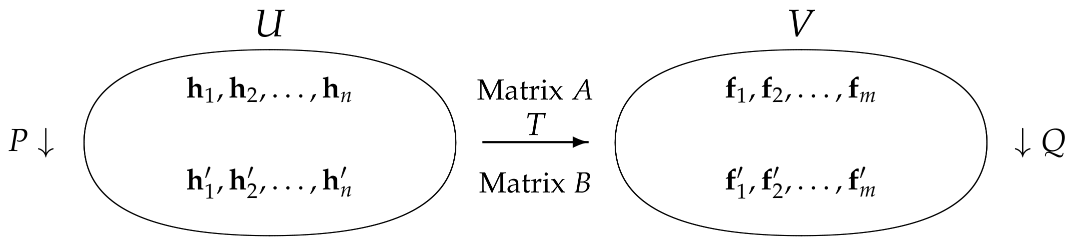
$$\begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\phi\cos\theta + \sin\phi\sin\theta & -\cos\phi\sin\theta + \sin\phi\cos\theta \\ \sin\phi\cos\theta - \cos\phi\sin\theta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

which is the matrix of $M_{\phi-\theta}$. Thus, $M_\phi R_\theta = M_{\phi-\theta}$.

We get a different result if we do first M_ϕ and then R_θ. Check yourself that $R_\theta M_\phi = M_{\phi+\theta}$ and $M_\phi M_\theta = R_{\theta+\phi}$.

4 Change of Bases

Let $T : U \to V$ be a linear map represented by two matrices A and B in the two possibly different pairs of bases.



The goal of this chapter is to learn how to handle this situation.

4.1 General theory

4.1.1 Change of bases matrix, revisited

This is a slight generalisation of Section 3.1.1. We consider a vector space V with two bases: the "old" basis $\mathbf{h}_1, \dots, \mathbf{h}_n$ and the "new" basis $\mathbf{f}_1, \dots, \mathbf{f}_n$. Recall the Theorem 1.4.7 ensures that they have the same number of elements.

Definition 4.1.1. Write each element of the old basis $\mathbf{h}_1, \dots, \mathbf{h}_n$ in the new basis. The change of basis matrix is the matrix $P = (\underline{\mathbf{h}_1}, \dots, \underline{\mathbf{h}_n}) \in \mathbb{F}^{n,n}$.

The following fact is really Proposition 3.1.2, extended to this situation.

Proposition 4.1.2. *Let $\mathbf{v} \in V$, $\underset{\sim}{\mathbf{v}} \in \mathbb{F}^n$ its coordinate vector in the old basis, $\underline{\mathbf{v}} \in \mathbb{F}^n$ its coordinate vector in the new basis. Then $P\underset{\sim}{\mathbf{v}} = \underline{\mathbf{v}}$.*

Proof. The proof is verbatim to the proof of Proposition 3.1.2. We have

$$P = (\sigma_{ij}), \qquad \underset{\sim}{\mathbf{v}} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \qquad \text{and} \qquad \underline{\mathbf{v}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix},$$

so that $\mathbf{v} = \sum_{i=1}^n \lambda_i \mathbf{h}_i = \sum_{j=1}^n \alpha_j \mathbf{f}_j$. By definition of P, $\mathbf{h}_j = \sum_{i=1}^n \sigma_{ij} \mathbf{f}_i$ for $1 \leq j \leq n$. Using this, we get

$$\mathbf{v} = \sum_{j=1}^n \lambda_j \mathbf{h}_j = \sum_{j=1}^n \lambda_j \sum_{i=1}^n \sigma_{ij} \mathbf{f}_i = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} \lambda_j \mathbf{f}_i$$

and then comparing coefficients of $\mathbf{f}_i$ in the expansions for $\mathbf{v}$ gives $\alpha_i = \sum_{j=1}^n \sigma_{ij} \lambda_j$ for $1 \leq i \leq n$. That is, $\underline{\mathbf{v}} = P\underset{\sim}{\mathbf{v}}$. $\square$

Corollary 4.1.3. *The change of basis matrix is invertible. More precisely, if P is the change of basis matrix from the basis of $\mathbf{h}_i$-s to the basis of $\mathbf{f}_i$-s and Q is the change of basis matrix from the basis of $\mathbf{f}_i$-s to the basis of $\mathbf{h}_i$-s then $P = Q^{-1}$.*

Proof. By Proposition 4.1.2, $P\underset{\sim}{\mathbf{v}} = \underline{\mathbf{v}}$ and $Q\underline{\mathbf{v}} = \underset{\sim}{\mathbf{v}}$ for all $v \in V$. It follows that $PQ\mathbf{u} = \mathbf{u}$ and $QP\mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in \mathbb{F}^n$. Hence, $I_n = QP = PQ$. $\square$

Example. Let $U = \mathbb{R}^3$, $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ be the standard basis and

$$\mathbf{f}_1 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \ \mathbf{f}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{f}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \ \text{so that } \mathbf{e}_1 = \mathbf{f}_3, \ \mathbf{e}_2 = \mathbf{f}_2 - \mathbf{f}_3, \ \mathbf{e}_3 = \mathbf{f}_1 - 2\mathbf{f}_2 + 2\mathbf{f}_3 \, .$$

The latter calculation gives Q, while the matrix P is formed by the vectors $\mathbf{f}_i$. Check yourself that $PQ = I_3$ where

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

4.1.2 Main theorem

Let us set up all notation to state the main theorem. Let $T : U \to V$ be a linear map, where $\dim(U) = n$, $\dim(V) = m$. Choose an "old" basis $\mathbf{h}_1, \ldots, \mathbf{h}_n$ of U and an "old" basis $\mathbf{f}_1, \ldots, \mathbf{f}_m$ of V. Let A be the matrix of T in these bases.

Now choose new bases $\mathbf{h}'_1, \ldots, \mathbf{h}'_n$ of U and $\mathbf{f}'_1, \ldots, \mathbf{f}'_m$ of V. Let B be the matrix of T in these bases.

Finally, let the $n \times n$ matrix P be the basis change matrix from $\{\mathbf{h}_i\}$ to $\{\mathbf{h}'_i\}$, and let the $m \times m$ matrix Q be the basis change matrix from $\{\mathbf{f}_i\}$ to $\{\mathbf{f}'_i\}$.

Theorem 4.1.4. *With the above notation, we have $QA = BP$, or equivalently $B = QAP^{-1}$.*

Proof. Fix $\mathbf{u} \in U$. Let, $\mathbf{v} = T(\mathbf{u})$. By Proposition 3.3.1, we have $A\underset{\sim}{\mathbf{u}} = \underset{\sim}{\mathbf{v}}$ and $B\underline{\mathbf{u}} = \underline{\mathbf{v}}$. By Proposition 4.1.2, $P\underset{\sim}{\mathbf{u}} = \underline{\mathbf{u}}$ and $Q\underset{\sim}{\mathbf{v}} = \underline{\mathbf{v}}$. Hence,

$$QA\underset{\sim}{\mathbf{u}} = Q\underset{\sim}{\mathbf{v}} = \underline{\mathbf{v}} = B\underline{\mathbf{u}} = BP\underset{\sim}{\mathbf{u}}.$$

Since this is true for all column vectors $\underset{\sim}{\mathbf{u}} \in \mathbb{F}^{n,1}$, this implies that $QA = BP$. We know that P is invertible from Corollary 4.1.3, so multiplying on the right by P^{-1} gives $B = QAP^{-1}$. $\square$

Example. Let us re-examine Example 1 in Section 3.3.2. To compute Q, we need to express the old basis in the new one:

$$\mathbf{f}_1 = -\mathbf{f}'_1 + \mathbf{f}'_2, \ \mathbf{f}_2 = \mathbf{f}'_1 \ \text{ so that } \ Q = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The matrix P^{-1} is formed by the $\mathbf{h}_i'$:

$$P^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Theorem 4.1.4 tells us that $B = QAP^{-1}$, does it not?

$$QAP^{-1} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix}.$$

4.1.3 Commutative squares

A commutative square is a collection of 4 vector spaces and 4 linear maps forming a diagram

$$\begin{tikzcd} U_1 \arrow[r, "T"] \arrow[d, "S"'] & V_1 \arrow[d, "Q"] \\ U_2 \arrow[r, "P"] & V_2 \end{tikzcd}$$

such that $QT = PS$. This may seem like a gimmick at this point but it is a rather deep, useful notion. We use it visualise all the change of basis phenomena. Notice that a basis $\mathbf{h}_1, \dots \mathbf{h}_n$ of a vector space U determines a linear map $R = R(\mathbf{h}_1, \dots \mathbf{h}_n) : U \to \mathbb{F}^n$ via $R(\mathbf{u}) = \underline{\mathbf{u}}$. Now Proposition 3.3.1 is saying that

the square $$\begin{tikzcd} U \arrow[r, "T"] \arrow[d, "R"'] & V \arrow[d, "R"] \\ \mathbb{F}^n \arrow[r, "A"] & \mathbb{F}^m \end{tikzcd}$$ is commutative.

Similarly Theorem 3.4.2 can be interpreted as a row concatenation of such squares for T_1 and T_2 to obtain the square for $T_1 T_2$.

$$\begin{tikzcd} U \arrow[r, "T_2"] \arrow[d, "R"'] & V \arrow[r, "T_1"] \arrow[d, "R"] & W \arrow[d, "R"] \\ \mathbb{F}^n \arrow[r, "T_B"] & \mathbb{F}^m \arrow[r, "T_A"] & \mathbb{F}^l \end{tikzcd} \rightsquigarrow \begin{tikzcd} U \arrow[r, "T_2 T_1"] \arrow[d, "R"'] & W \arrow[d, "R"] \\ \mathbb{F}^n \arrow[r, "BA"] & \mathbb{F}^l \end{tikzcd}$$

This is not merely an interpretation: it is a complete proof!

The two bases determine two linear maps $R_{old} = R(\mathbf{h}_1, \dots \mathbf{h}_n), R_{new} = R(\mathbf{h}_1', \dots \mathbf{h}_n') : U \to \mathbb{F}^n$. Employing the language of commutative squares, Proposition 3.1.2 is saying that the following is a commutative square.

the squares $$\begin{tikzcd} U \arrow[r, "I_U"] \arrow[d, "R_{new}"] & U \arrow[d, "R_{old}"] \\ \mathbb{F}^n \arrow[r, "P^{-1}"] & \mathbb{F}^n \end{tikzcd}$$ and $$\begin{tikzcd} V \arrow[r, "I_V"] \arrow[d, "R_{old}"] & V \arrow[d, "R_{new}"] \\ \mathbb{F}^m \arrow[r, "Q"] & \mathbb{F}^m \end{tikzcd}$$ are commutative.

In this language, Theorem 3.4.2 can be proved (or just interpreted) by observing that the concatenation of the three commutative squares is still a commutative square

$$\begin{tikzcd} U \arrow[r, "I_U"] \arrow[d, "R_{new}"] & U \arrow[r, "T"] \arrow[d, "R_{old}"] & V \arrow[r, "I_V"] \arrow[d, "R_{old}"] & V \arrow[d, "R_{new}"'] \\ \mathbb{F}^n \arrow[r, "P^{-1}"] & \mathbb{F}^n \arrow[r, "A"] & \mathbb{F}^n \arrow[r, "Q"] & \mathbb{F}^n \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} U \arrow[r, "T"] \arrow[d, "R_{new}"] & V \arrow[d, "R_{new}"] \\ \mathbb{F}^n \arrow[r, "QAP^{-1}"] & \mathbb{F}^m \end{tikzcd}$$

4.2 Orthogonal change of basis

4.2.1 Orthogonal operator

If we're working with a euclidean space V, we know what the "length" of a vector in V means, and what the "angle" between vectors is; so we might want to consider transformations from V to itself that preserve lengths and angles – they preserve the geometry of the space.

Definition 4.2.1. A linear operator $T: V \to V$ is said to be *orthogonal* if it preserves the scalar product on V. That is, if $\tau(T(\mathbf{v}), T(\mathbf{w})) = \tau(\mathbf{v}, \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$.

Since length and angle can be defined in terms of the scalar product, an orthogonal linear map preserves distance and angle, so geometrically it is a rigid map. In $\mathbb{R}^2$, for example, an orthogonal map is either rotation about the origin, or a reflection about a line through the origin. Let us prove the following two easy statements.

Proposition 4.2.2. *An orthogonal linear operator T is invertible.*

Proof. By Corollary 2.2.8, it suffices to show that $\ker(T) = \mathbf{0}$. Pick $\mathbf{v} \in \ker(T)$. Then $\|\mathbf{v}\| = \|T(\mathbf{v})\| = \|\mathbf{0}\| = 0$ and $\mathbf{v} = \mathbf{0}$. □

Proposition 4.2.3. *Let $\mathbf{h}_1, \ldots, \mathbf{h}_n$ be an orthonormal basis of V. A linear operator T is orthogonal if and only if $T(\mathbf{h}_1), \ldots, T(\mathbf{h}_n)$ is an orthonormal basis of V.*

Proof. If T is orthogonal, then $\tau(T(\mathbf{h}_i), T(\mathbf{h}_j)) = \tau(\mathbf{h}_i, \mathbf{h}_j) = \delta_{ij}$ for all i and j. Thus, $T(\mathbf{h}_1), \ldots, T(\mathbf{h}_n)$ is an orthonormal sequence of length $\dim(V)$. Thus, it must be a basis.

Assume that $T(\mathbf{h}_1), \ldots, T(\mathbf{h}_n)$ is an orthonormal basis of V. Pick $\mathbf{v}, \mathbf{w} \in V$. Write them $\mathbf{v} = \sum_i \alpha_i \mathbf{h}_i$ and $\mathbf{w} = \sum_i \beta_i \mathbf{h}_i$. Then $\tau(\mathbf{v}, \mathbf{w}) = \sum_{i,j} \alpha_i \beta_j \tau(\mathbf{h}_i, \mathbf{h}_j) = \sum_i \alpha_i \beta_i$ and so is $\tau(T(\mathbf{v}), T(\mathbf{w})) = \sum_{i,j} \alpha_i \beta_j \tau(T(\mathbf{h}_i), T(\mathbf{h}_j)) = \sum_i \alpha_i \beta_i$. □

4.2.2 Orthogonal matrix

The next definition makes sense over any field $\mathbb{F}$ but we will use it only over $\mathbb{R}$.

Definition 4.2.4. A matrix $A \in \mathbb{F}^{n,n}$ is called *orthogonal* if $A^{\mathrm{T}}A = I_n$.

Example. For any $\theta \in \mathbb{R}$, let $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ represent a counterclockwise rotation by an angle θ. It is easily checked that $A^{\mathrm{T}}A = AA^{\mathrm{T}} = I_2$.

Proposition 4.2.5. *A linear operator $T : V \to V$ on a euclidean space is orthogonal if and only if its matrix A (with respect to an orthonormal basis of V) is orthogonal.*

Proof. Let $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ be the columns of the matrix A. Then $\mathbf{c}_1^T, \ldots, \mathbf{c}_n^T$ are the rows of A^T. Hence, the (i,j)-th entry of $A^{\mathrm{T}}A$ is $\mathbf{c}_i^{\mathrm{T}}\mathbf{c}_j = \mathbf{c}_i \bullet \mathbf{c}_j$.

Let $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n$ be an orthonormal basis. The $\mathbf{c}_i = \underline{T(\mathbf{h}_i)}$ and everything follows from Proposition 4.2.3: T is orthogonal if and only if $T(\mathbf{h}_1), \ldots, T(\mathbf{h}_n)$ is an orthonormal sequence if and only if $\mathbf{c}_1, \ldots, \mathbf{c}_n$ is an orthonormal sequence if and only if A is orthogonal. □

Let $\mathbf{h}_1, \ldots, \mathbf{h}_n$ be an orthonormal basis of V, $\mathbf{f}_1, \ldots, \mathbf{f}_n$ another basis.

Proposition 4.2.6. *The following conditions are equivalent.*

1. *The basis $\mathbf{f}_1, \ldots, \mathbf{f}_n$ is orthonormal.*
2. *The change of basis matrix P from $\mathbf{f}_1, \ldots, \mathbf{f}_n$ to $\mathbf{h}_1, \ldots, \mathbf{h}_n$ is orthogonal.*
3. *The change of basis matrix Q from $\mathbf{h}_1, \ldots, \mathbf{h}_n$ to $\mathbf{f}_1, \ldots, \mathbf{f}_n$ is orthogonal.*

Proof. The first and the third statements are equivalent because the columns of Q are $\underline{\mathbf{f}_i}$, written in the orthonormal basis $\mathbf{h}_i$. Indeed, since the basis $\mathbf{h}_i$-s is orthonormal, in this basis we have

$$\tau(\mathbf{x}, \mathbf{y}) = \underline{\mathbf{x}} \bullet \underline{\mathbf{y}} \quad \text{for all} \quad \mathbf{x}, \mathbf{y} \in V.$$

Thus, the basis $\mathbf{f}_i$-s is orthonormal if and only if the basis $\underline{\mathbf{f}_i}$-s of the standard euclidean space $\mathbb{R}^n$ is orthonormal if and only if the matrix Q is orthogonal.

Note that $(P^{-1})^T = (P^T)^{-1}$. Thus, $P^TP = I$ if and only if $P^{-1}(P^{-1})^T = (P^TP)^{-1} = I$. Thus, P is orthogonal if and only if P^{-1} is orthogonal. Since $Q = P^{-1}$, the last two statements are equivalent. □

4.2.3 Four key problems, revisited

It is time to revisit Section 3.3.3. Let us represent a linear map $T : U \to V$ by a matrix A. We can make the following observations.

LT: Linear Transformation Problem: $A, B \in \mathbb{F}^{n,k}$ represent the same linear map (written in different bases) if and only if there exist invertible $Q \in \mathbb{F}^{n,n}$ and $P \in \mathbb{F}^{k,k}$ such that $B = QAP^{-1}$. We call such A and B *equivalent*.

LO: Linear Operator Problem: $A, B \in \mathbb{F}^{n,n}$ represent the same linear operator (written in different bases) if and only if there exists invertible $P \in \mathbb{F}^{n,n}$ such that $B = PAP^{-1}$. We call such A and B *similar*.

ET: Euclidean Transformation Problem. $A, B \in \mathbb{R}^{n,k}$ represent the same linear map between euclidean spaces (written in different orthonormal bases) if and only if there exist orthogonal $Q \in \mathbb{R}^{n,n}$ and $P \in \mathbb{R}^{k,k}$ such that $B = QAP^{-1}$ (note that $B = QAP^T$ since $P^{-1} = P^T$). We call such A and B *orthogonally equivalent*.

EO: Euclidean Operator Problem. $A, B \in \mathbb{R}^{n,n}$ represent the same linear operator on a euclidean space (written in different orthonormal bases) if and only if there exists orthogonal $P \in \mathbb{R}^{n,n}$ such that $B = PAP^{-1}$ (or $B = PAP^T$). We call such A and B *orthogonally similar*.

4.3 Elementary change of basis

Let $\mathbf{h}_1, \ldots, \mathbf{h}_n$ be a basis of V. The following three bases changes are your friends! Our goal is to observe their effect on the matrix A of a linear map, which we represent using columns $A = (\mathbf{c}_1, \mathbf{c}_2, \ldots)$ and rows $A = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \end{pmatrix}$.

4.3.1 Basis multipication

Pick $i \in \{1, \ldots, n\}$ and $\alpha \in \mathbb{F} \setminus \{0\}$. The new basis is

$$\mathbf{f}_j = \mathbf{h}_j \text{ if } j \neq i, \qquad \mathbf{f}_i = \alpha \mathbf{h}_i \,.$$

Since $\mathbf{h}_i = \alpha^{-1}\mathbf{f}_i$, the change of basis matrix P and its inverse are

$$P = E(n)^3_{\alpha^{-1},i}\,, \;\; P^{-1} = E(n)^3_{\alpha,i}$$

where $E(n)^3_{\lambda,i}$ is the $n \times n$ identity matrix with its (i,i) entry replaced by λ, so called *the elementary matrix of the third kind*. The effect of this on the matrix A depends on whether V is a domain (then the operation is the elementary column operation $\mathcal{C}3$) or a range (the elementary row operation $\mathcal{R}3$). In both cases we only show what changes in the matrix A:

$$\mathcal{C}3 : A \xrightarrow{\mathbf{c}_i \to \alpha \mathbf{c}_i} AP^{-1} = (\ldots, \alpha \mathbf{c}_i, \ldots), \quad \mathcal{R}3 : A \xrightarrow{\mathbf{r}_i \to \alpha^{-1}\mathbf{r}_i} PA = \begin{pmatrix} \vdots \\ \alpha^{-1}\mathbf{r}_i \\ \vdots \end{pmatrix}$$

Example. Let $n = 3$ and $\mathbf{f}_3 = 2\mathbf{h}_3$. Then

$$P = E(3)^3_{1/2,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \text{ and } P^{-1} = E(3)^3_{2,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

and the corresponding row and column operations are

$$\mathcal{C}3: \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow{\mathbf{c}_3 \to 2\mathbf{c}_3} \begin{pmatrix} 1 & 3 & 10 \\ 6 & 7 & 16 \\ 0 & 2 & 8 \end{pmatrix}, \quad \mathcal{R}3: \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow{\mathbf{r}_3 \to \frac{1}{2}\mathbf{r}_3} \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 0 & 1 & 2 \end{pmatrix}.$$

4.3.2 Basis swap

Pick $i \neq j$. The new basis is

$$\mathbf{f}_k = \mathbf{h}_k \text{ if } k \notin \{i, j\}, \qquad \mathbf{f}_i = \mathbf{h}_j\ ,\ \mathbf{f}_j = \mathbf{h}_i\ .$$

Since $\mathbf{h}_i = \mathbf{f}_j$ and $\mathbf{h}_j = \mathbf{f}_i$, the change of basis matrix P and its inverse are

$$P = P^{-1} = E(n)^2_{i,j}$$

where $E(n)^2_{i,j}$ is the $n \times n$ identity matrix with its i-th and j-th rows interchanged, so called *the elementary matrix of the second kind*. The effect of this on the matrix A depends on whether V is a domain (then the operation is the elementary column operation $\mathcal{C}2$) or a range (the elementary row operation $\mathcal{R}2$). In both cases we assume $i < j$ and only show what changes in the matrix A:

$$\mathcal{C}2: A \xrightarrow{\mathbf{c}_i \leftrightarrow \mathbf{c}_j} AP^{-1} = (\ldots, \mathbf{c}_j, \ldots, \mathbf{c}_i, \ldots), \quad \mathcal{R}2: A \xrightarrow{\mathbf{r}_i \leftrightarrow \mathbf{r}_j} PA = \begin{pmatrix} \vdots \\ \mathbf{r}_j \\ \vdots \\ \mathbf{r}_i \\ \vdots \end{pmatrix}$$

Example. Let $n = 3$ and $\mathbf{f}_3 = \mathbf{h}_1$, $\mathbf{f}_1 = \mathbf{h}_3$. Then

$$P = P^{-1} = E(2)^3_{1,3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

and the corresponding row and column operations are

$$\mathcal{C}2: \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow{\mathbf{c}_1 \leftrightarrow \mathbf{c}_3} \begin{pmatrix} 5 & 3 & 1 \\ 8 & 7 & 6 \\ 4 & 2 & 0 \end{pmatrix}, \quad \mathcal{R}2: \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow{\mathbf{r}_1 \leftrightarrow \mathbf{r}_3} \begin{pmatrix} 0 & 2 & 4 \\ 6 & 7 & 8 \\ 1 & 3 & 5 \end{pmatrix}.$$

4.3.3 Basis addition

Pick $i \neq j$ and $\alpha \in \mathbb{F}$. The new basis is

$$\mathbf{f}_k = \mathbf{h}_k \ \text{ if } \ k \neq i, \qquad \mathbf{f}_i = \mathbf{h}_i + \alpha \mathbf{h}_j\ .$$

Since $\mathbf{h}_i = \mathbf{f}_i - \alpha \mathbf{f}_j$, the change of basis matrix P and its inverse are

$$P = E(n)^1_{-\alpha,j,i}\ ,\ \ P^{-1} = E(n)^1_{\alpha,j,i}$$

where $E(n)^1_{\lambda,i,j}$ (where $i \neq j$) is the an $n \times n$ matrix equal to the identity, but with an additional non-zero entry λ in the (i,j) position, so called *the elementary matrix of the first kind*. The effect of this on the matrix A depends on whether V is a domain (then the operation is the elementary column operation $\mathcal{C}1$) or a range (the elementary row operation $\mathcal{R}1$). In both cases we only show what changes in the matrix A:

$$\mathcal{C}1 : A \xrightarrow{\mathbf{c}_j \to \mathbf{c}_i + \alpha \mathbf{c}_j} AP^{-1} = (\ldots, \mathbf{c}_i + \alpha \mathbf{c}_j, \ldots), \quad \mathcal{R}1 : A \xrightarrow{\mathbf{r}_j \to \mathbf{r}_j - \alpha \mathbf{r}_i} PA = \begin{pmatrix} \vdots \\ \mathbf{r}_j - \alpha \mathbf{r}_i \\ \vdots \end{pmatrix}$$

Example. Let $n = 3$ and $\mathbf{f}_2 = \mathbf{h}_2 - 4\mathbf{h}_3$. Then

$$P = E(3)^1_{4,3,2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{pmatrix} \text{ and } P^{-1} = E(3)^1_{-4,3,2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{pmatrix}$$

and the corresponding row and column operations are

$$\mathcal{C}1 : \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow{\mathbf{c}_2 \to \mathbf{c}_2 - 4\mathbf{c}_3} \begin{pmatrix} 1 & -17 & 5 \\ 6 & -25 & 8 \\ 0 & -14 & 4 \end{pmatrix}, \quad \mathcal{R}1 : \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow{\mathbf{r}_3 \to \mathbf{r}_3 + 4\mathbf{r}_2} \begin{pmatrix} 1 & 3 & 5 \\ 6 & 7 & 8 \\ 24 & 30 & 36 \end{pmatrix}$$

5 Elementary Transformations and Their Applications

We have introduced the elementary row and column transformations in the last section. It is time to put them into good use.

5.1 Solving LT

Now we solve the first key problem LT from Section 3.3.3

5.1.1 Smith's normal form

Theorem 5.1.1. *By applying elementary row and column operations, a matrix $A \in \mathbb{F}^{m\times n}$ can be brought into the block form*

$$\left(\begin{array}{c|c} I_s & \mathbf{0}_{s,n-s} \\ \hline \mathbf{0}_{m-s,s} & \mathbf{0}_{m-s,n-s} \end{array}\right),$$

where, as in Section 2.3, I_s denotes the $s \times s$ identity matrix, and $\mathbf{0}_{kl}$ the $k \times l$ zero matrix.

Proof. The proof is an algorithm whose steps contain elementary row and column operations. For a completely formal proof, we have to show that

1. after termination the resulting matrix has the required form,
2. the algorithm terminates after finitely many steps.

Both these statements are clear from the nature of the algorithm. Make sure that you understand why they are clear!

At any stage of the procedure, we are looking at the entry α_{ii} in a particular position (i, i) of the matrix. (i, i) is called the *pivot* position, and α_{ii} the *pivot* entry. We start with $(i, i) = (1, 1)$ and proceed as follows.

1. If $\alpha_{ii} = 1$ and all entries below it (in its column) and to the right (in its row) are zero (i.e. if $\alpha_{ki} = \alpha_{ik} = 0$ for all $k \geq i$), then move the pivot one place to $(i + 1, i + 1)$ and go to Step 1. Terminate if out of the matrix $(i + 1 > \min(m, n))$. Go to Step 2.

2. If $\alpha_{ii} = 0$ but $\alpha_{kj} \neq 0$ for some $k \geq i$ and $k \geq j$ then apply $\mathcal{R}2$ and $\mathcal{C}2$ to move the non-zero entry into the pivot position. Go to Step 3.

3. At this stage $\alpha_{ii} \neq 0$. If $\alpha_{ii} \neq 1$, then apply $\mathcal{R}3$ or $\mathcal{C}3$ to make $\alpha_{ii} = 1$. Go to Step 4.

4. At this stage $\alpha_{ii} = 1$. Kill all entries below it (in its column) and to the right (in its row): for any $k \neq i$, $\alpha_{ki} \neq 0$ or $\alpha_{ik} \neq 0$, apply $\mathcal{R}1$ or $\mathcal{C}1$. Go to Step 1.

□

Definition 5.1.2. The matrix in Theorem 5.1.1 is said to be in *Smith normal form* (sometimes it is called *row and column echelon form*).

Example. Let us run it on the matrix $A = \begin{pmatrix} 0 & 0 & 1 & 2 & 1 \\ 2 & 4 & 2 & -4 & 2 \\ 3 & 6 & 3 & 6 & 3 \\ 1 & 2 & 3 & 3 & 3 \end{pmatrix}$ keeping track of bases. To avoid propagation of unnecessary notation, we use $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4, \mathbf{h}_5$ for the basis of the domain and $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4$ for the basis of the range at each step. It starts with $\mathbf{h}_i = \mathbf{e}_i$ and $\mathbf{f}_j = \mathbf{e}'_j$, the standard bases of both spaces. The last column records only the basis elements that changed.

Matrix	**Pivot**	**Step**	**Operation**	**Basis Update**
A	$(1,1)$	**2**	$\mathbf{c}_1 \leftrightarrow \mathbf{c}_3$	$\mathbf{h}_1 = \mathbf{e}_3, \mathbf{h}_3 = \mathbf{e}_1$
$\begin{pmatrix} 1 & 0 & 0 & 2 & 1 \\ 2 & 4 & 2 & -4 & 2 \\ 3 & 6 & 3 & 6 & 3 \\ 3 & 2 & 1 & 3 & 3 \end{pmatrix}$	$(1,1)$	**4**	$\mathbf{c}_4 \to \mathbf{c}_4 - 2\mathbf{c}_1$ $\mathbf{c}_5 \to \mathbf{c}_5 - \mathbf{c}_1$	$\mathbf{h}_4 = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \mathbf{h}_5 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 4 & 2 & -8 & 0 \\ 3 & 6 & 3 & 0 & 0 \\ 3 & 2 & 1 & -3 & 0 \end{pmatrix}$	$(1,1)$	**4**	$\mathbf{r}_2 \to \mathbf{r}_2 - 2\mathbf{r}_1$ $\mathbf{r}_3 \to \mathbf{r}_3 - 3\mathbf{r}_1$ $\mathbf{r}_4 \to \mathbf{r}_4 - 3\mathbf{r}_1$	$\mathbf{f}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & 2 & -8 & 0 \\ 0 & 6 & 3 & 0 & 0 \\ 0 & 2 & 1 & -3 & 0 \end{pmatrix}$	$(1,1)$ $\downarrow$ $(2,2)$	**1,3**	$\mathbf{r}_2 \to \frac{1}{4}\mathbf{r}_2$	$\mathbf{f}_2 = \begin{pmatrix} 0 \\ 4 \\ 0 \\ 0 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1/2 & -2 & 0 \\ 0 & 6 & 3 & 0 & 0 \\ 0 & 2 & 1 & -3 & 0 \end{pmatrix}$	$(2,2)$	**4**	$\mathbf{c}_3 \to \mathbf{c}_3 - \frac{1}{2}\mathbf{c}_2$ $\mathbf{c}_4 \to \mathbf{c}_4 + 2\mathbf{c}_2$	$\mathbf{h}_3 = \begin{pmatrix} 1 \\ -1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{h}_4 = \begin{pmatrix} 0 \\ 2 \\ -2 \\ 1 \\ 0 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 6 & 0 & 12 & 0 \\ 0 & 2 & 0 & 1 & 0 \end{pmatrix}$	$(2,2)$	**4**	$\mathbf{r}_3 \to \mathbf{r}_3 - 6\mathbf{r}_2$ $\mathbf{r}_4 \to \mathbf{r}_4 - 2\mathbf{r}_2$	$\mathbf{f}_2 = \begin{pmatrix} 0 \\ 4 \\ 6 \\ 2 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$(2,2)$ $\downarrow$ $(3,3)$	**1,2,3**	$\mathbf{c}_3 \leftrightarrow \mathbf{c}_4$ $\mathbf{r}_3 \to \frac{1}{12}\mathbf{r}_3$	$\mathbf{h}_3 = \begin{pmatrix} 0 \\ 2 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \mathbf{h}_4 = \begin{pmatrix} 1 \\ -1/2 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{f}_3 = \begin{pmatrix} 0 \\ 0 \\ 12 \\ 0 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$(3,3)$	**4**	$\mathbf{r}_4 \to \mathbf{r}_4 - \mathbf{r}_3$	$\mathbf{f}_3 = \begin{pmatrix} 0 \\ 0 \\ 12 \\ 1 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$(3,3)$ $\downarrow$ $(4,4)$ $\downarrow$ **stop**	**1**		

Let us get together all the information about new bases we have accumulated, by wrting the inverses of the change of bases matrices:

$$(\mathbf{h}_1, \dots, \mathbf{h}_5) = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 2 & -1/2 & 0 \\ 1 & 0 & -2 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \ (\mathbf{f}_1, \dots, \mathbf{f}_4) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 3 & 6 & 12 & 0 \\ 3 & 2 & 1 & 1 \end{pmatrix}.$$

Check $A\mathbf{h}_i = \mathbf{f}_i$ for $i \leq 3$ to be sure!

Notice that the proposed algorithm is optimised for clarity. If you need to do it by hand, use *the greedy algorithm* instead. This means killing as many entries as possible and then rearranging the results. For instance, you could start on the same matrix A by utilising $\mathbf{c}_3 = \mathbf{c}_5$ and $\mathbf{c}_2 = 2\mathbf{c}_1$:

Matrix	**Operation**
$A = \begin{pmatrix} 0 & 0 & 1 & 2 & 1 \\ 2 & 4 & 2 & -4 & 2 \\ 3 & 6 & 3 & 6 & 3 \\ 1 & 2 & 3 & 3 & 3 \end{pmatrix}$	$\mathbf{c}_2 \to \mathbf{c}_2 - 2\mathbf{c}_1$ $\mathbf{c}_4 \to \mathbf{c}_4 - 2\mathbf{c}_3$ $\mathbf{c}_5 \to \mathbf{c}_5 - \mathbf{c}_3$
$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 2 & -8 & 0 \\ 3 & 0 & 3 & 0 & 0 \\ 1 & 0 & 3 & -3 & 0 \end{pmatrix}$	$\mathbf{r}_2 \leftarrow \mathbf{r}_2 - 2\mathbf{r}_1$ $\mathbf{r}_3 \to \mathbf{r}_3 - 3\mathbf{r}_1$ $\mathbf{r}_4 \to \mathbf{r}_4 - 3\mathbf{r}_1$ $\mathbf{c}_1 \leftrightarrow \mathbf{c}_3$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -8 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -3 & 0 \end{pmatrix}$	$\mathbf{c}_2 \leftrightarrow \mathbf{c}_3$ $\mathbf{r}_2 \to \frac{1}{2}\mathbf{r}_2$ $\mathbf{r}_3 \to \mathbf{r}_3 - 3\mathbf{r}_2$ $\mathbf{r}_4 \to \mathbf{r}_4 - \mathbf{r}_2$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -4 & 0 \\ 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$\mathbf{r}_2 \to \mathbf{r}_2 + 4\mathbf{r}_4$ $\mathbf{r}_3 \to \mathbf{r}_3 - 12\mathbf{r}_4$ $\mathbf{c}_3 \leftrightarrow \mathbf{c}_4$ $\mathbf{r}_3 \leftrightarrow \mathbf{r}_4$
$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	

5.1.2 The rank of a matrix

Now we would like to discuss the number s that appears in Theorem 5.1.1. Does the initial matrix uniquely determine this number?

Let $T : U \to V$ be a linear map, where $\dim(U) = n$, $\dim(V) = m$. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of U and let $\mathbf{f}_1, \dots, \mathbf{f}_m$ be a basis of V. Recall from Section 2.2.3 that $\operatorname{rank}(T) = \dim(\operatorname{im}(T))$.

Lemma 5.1.3. $\operatorname{rank}(T)$ *is the size of the largest linearly independent subset of* $T(\mathbf{e}_1), \dots, T(\mathbf{e}_n)$.

Proof. $\mathrm{im}(T)$ is spanned by the vectors $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$, and by Theorem 1.4.11, some subset of these vectors forms a basis of $\mathrm{im}(T)$. By definition of basis, this subset has size $\dim(\mathrm{im}(T)) = \mathrm{rank}(T)$, and by Corollary 1.4.15 no larger subset of $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ can be linearly independent. □

Now let $A = \begin{pmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m \end{pmatrix} = \begin{pmatrix} \mathbf{c}_1 & \ldots & \mathbf{c}_n \end{pmatrix}$ be an $m \times n$ matrix over $\mathbb{F}$ with rows $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m \in \mathbb{F}^{1,n}$ and columns $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n \in \mathbb{F}^m$. Left and right multiplications by A are linear maps

$$L_A = A : \mathbb{F}^n \to \mathbb{F}^m, L_A(\mathbf{x}) = A\mathbf{x},\ R_A = A : \mathbb{F}^{1,m} \to \mathbb{F}^{1,n}, R_A(\mathbf{y}) = \mathbf{y}A. \tag{2}$$

Definition 5.1.4. (i) The *row-space* of A is the image of R_A: it is the subspace of $\mathbb{F}^{1,n}$ spanned by the rows $\mathbf{r}_1, \ldots, \mathbf{r}_m$ of A. The *row rank* of A is equal to

- the dimension of the row-space of A,
- or the rank of R_A
- or, by Lemma 5.1.3, the size of the largest linearly independent subset of $\mathbf{r}_1, \ldots, \mathbf{r}_m$.

(ii) The *column-space* of A is the image of L_A: it is the subspace of $\mathbb{F}^m$ spanned by the columns $\mathbf{c}_1, \ldots, \mathbf{c}_n$ of A. The *column rank* of A is equal to

- the dimension of the column-space of A,
- or the rank of L_A
- or, by Lemma 5.1.3, the size of the largest linearly independent subset of $\mathbf{c}_1, \ldots, \mathbf{c}_n$.

There is no obvious reason why there should be any particular relationship between the row and column ranks, but in fact it will turn out that they are always equal.

Example.

$$A = \begin{array}{cc} \begin{pmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & 4 & 1 & 3 & 0 \\ 4 & 8 & 0 & 4 & 4 \end{pmatrix} & \begin{array}{c} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{array} \\ \begin{array}{ccccc} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4 & \mathbf{c}_5 \end{array} & \end{array}$$

We can calculate the row and column ranks by applying the sifting process (described in Section 1.4) to the row and column vectors, respectively.

Doing rows first, $\mathbf{r}_1$ and $\mathbf{r}_2$ are linearly independent, but $\mathbf{r}_3 = 4\mathbf{r}_1$, so the row rank is 2.

Now doing columns, $\mathbf{c}_2 = 2\mathbf{c}_1$, $\mathbf{c}_4 = \mathbf{c}_1 + \mathbf{c}_3$ and $\mathbf{c}_5 = \mathbf{c}_1 - 2\mathbf{c}_3$, so the column rank is also 2.

Theorem 5.1.5. *Applying $\mathcal{R}1$, $\mathcal{R}2$ or $\mathcal{R}3$ to a matrix A does not change the row or column rank of A. The same is true for $\mathcal{C}1$, $\mathcal{C}2$ and $\mathcal{C}3$.*

Proof. The matrix A represents the linear map L_A, defined in (2), in the standard basis. Applying one of the operations yields a matrix A', representing the same linear map L_A in another basis. The column ranks of A and A' are both equal to the rank of L_A.

The row ranks of A and A' are equal to the column ranks of the transposed matrices A^T and A'^T. These matrices represent R_A in the standard and the slightly changed bases. Thus, the row ranks of A and A' are both equal to the rank of R_A. □

Corollary 5.1.6. *Let s be the number of non-zero rows in the Smith normal form of a matrix A (see Theorem 5.1.1). Then both row rank of A and column rank of A are equal to s.*

Proof. Since elementary operations preserve ranks, it suffices to find both ranks of a matrix in the Smith form. Obviously, it is s. □

In particular, Corollary 5.1.6 establishes that the row rank is always equal to the column rank. This allows us to forget this artificial distinction and always talk about *the rank of a matrix*.

The most efficient way of computing the rank of a matrix by hand is to perform just enough transformations to make it obvious. For instance, let us look at the following matrix:

Matrix	**Operation**
$B = \begin{pmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & 4 & 1 & 3 & 0 \\ 4 & 8 & 1 & 5 & 2 \end{pmatrix}$	$\mathbf{r}_2 \to \mathbf{r}_2 - 2\mathbf{r}_1$ $\mathbf{r}_3 \to \mathbf{r}_3 - 4\mathbf{r}_1$
$\begin{pmatrix} 1 & 2 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 1 & 1 & -2 \end{pmatrix}$	

Since the resulting matrix has $\mathbf{r}_2 = \mathbf{r}_3$, $\text{Rank}(B) \le 2$. Since $\mathbf{c}_1$ and $\mathbf{c}_3$ are linearly independent, $\text{Rank}(B) = 2$.

5.1.3 The answer to LT

Recall that two $m \times n$ matrices A and B are said to be *equivalent* if there exist invertible P and Q with $B = QAP$; that is, if they represent the same linear map, cf. Section 4.2.3.

It is easy to check that it is an equivalence relation on the set $\mathbb{F}^{m,n}$ of $m \times n$ matrices over $\mathbb{F}$. We shall show now that the equivalence of matrices has effective characterisations.

Theorem 5.1.7. *Let $A, B \in \mathbb{F}^{m,n}$. The following conditions on A and B are equivalent.*

(i) A and B are equivalent.

(ii) A and B represent the same linear map with respect to different bases.

(iii) A and B have the same rank.

(iv) B can be obtained from A by application of elementary row and column operations.

Proof. (i) $\Leftrightarrow$ (ii): This is true by Theorem 4.1.4.

(ii) $\Rightarrow$ (iii): Since A and B both represent the same linear map T, we have $\mathrm{rank}(A) = \mathrm{rank}(B) = \mathrm{rank}(T)$.

(iii) $\Rightarrow$ (iv): By Theorem 5.1.1, if A and B both have rank s, then they can both be brought into the form

$$E_s = \left(\begin{array}{c|c} I_s & \mathbf{0}_{s,n-s} \\ \hline \mathbf{0}_{m-s,s} & \mathbf{0}_{m-s,n-s} \end{array} \right)$$

by elementary row and column operations. Since these operations are invertible, we can first transform A to E_s and then transform E_s to B.

(iv) $\Rightarrow$ (i): We saw in Section 4.3 that applying an elementary row operation to A can be achieved by multiplying A on the left by an elementary row matrix, and similarly applying an elementary column operation can be done by multiplying A on the right by an elementary column matrix. Hence (iv) implies that there exist elementary row matrices $R_1, \ldots, R_r$ and elementary column matrices $C_1, \ldots, C_s$ with $B = R_r \ldots R_1 A C_1 \ldots C_s$. Since elementary matrices are invertible, $Q = R_r \ldots R_1$ and $P = C_1 \ldots C_s$ are invertible and $B = QAP$. $\square$

In the above proof, we also showed the following:

Proposition 5.1.8. *Any $m \times n$ matrix is equivalent to the matrix E_s defined above, where $s = \mathrm{rank}(A)$.*

The matrices E_s are *canonical forms* for the linear maps in the LT problem. This means that they are easily recognizable representatives of equivalence classes of matrices.

5.2 Solving systems of linear equations

Row transformations solve systems of linear equations. But don't ever try column transformations: it is faux pas.

5.2.1 Linear equations and the inverse image problem

The study and solution of systems of simultaneous linear equations is the main motivation behind the development of the theory of linear algebra and of matrix operations. Let us consider a system of m equations in n unknowns $x_1, x_2 \ldots x_n$, $m, n \geq 1$.

$$\left\{ \begin{array}{ccccccccc} \alpha_{11}x_1 & + & \alpha_{12}x_2 & + & \cdots & + & \alpha_{1n}x_n & = & \beta_1 \\ \alpha_{21}x_1 & + & \alpha_{22}x_2 & + & \cdots & + & \alpha_{2n}x_n & = & \beta_2 \\ \vdots & & \vdots & & & & & & \vdots \\ \alpha_{m1}x_1 & + & \alpha_{m2}x_2 & + & \cdots & + & \alpha_{mn}x_n & = & \beta_m \end{array} \right. \tag{3}$$

All coefficients α_{ij} and β_i belong to $\mathbb{F}$. Solving this system means finding all collections $x_1, x_2 \ldots x_n \in \mathbb{F}$ such that (3) holds true.

Let $A = (\alpha_{ij}) \in \mathbb{F}^{m,n}$ be the $m \times n$ matrix of coefficients. Consider the column vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \in \mathbb{F}^m.$$

This allows us to rewrite system (3) as a single equation

$$A\mathbf{x} = \mathbf{b} \ \text{ or } \ L_A(\mathbf{x}) = \mathbf{b} \tag{4}$$

where the coefficient A is a matrix, the right-hand side $\mathbf{b}$ is a given vector in $\mathbb{F}^m$, the unknown $\mathbf{x}$ is a vector in $\mathbb{F}^n$ and L_A is the linear map from (2).

Using the new terms of linear maps, we have just reduced solving a system of linear equations to the *inverse image* problem. That is, given a linear map $T : U \to V$, and a fixed vector $\mathbf{b} \in V$, find $T^{-1}(\mathbf{b}) := \{\mathbf{x} \in U \mid T(\mathbf{x}) = \mathbf{b}\}$.

In fact, these two problems are equivalent! In the opposite direction, just choose bases in U and V and denote a matrix of T in these bases by A. Proposition 3.3.1 says that $T(\mathbf{x}) = \mathbf{b}$ if and only if $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$. This reduces the inverse image problem to solving a system of linear equations.

Let us make several observations that follow from the properties of the linear maps.

The case when $\mathbf{b} = \mathbf{0}$, or equivalently when $\beta_i = 0$ for $1 \leq i \leq m$, is called the *homogeneous* case. Here the set of solutions is

$$\ker(L_A) = \{\mathbf{x} \in \mathbb{F}^n \mid L_A(\mathbf{x}) = \mathbf{0}\} = \{\mathbf{x} \in \mathbb{F}^n \mid A\mathbf{x} = \mathbf{0}\},$$

which is sometimes called the *nullspace* of the matrix A.

In general, it is easy to see that if $\mathbf{x}_0$ is one solution to a system of equations, then the complete set of solutions is equal to

$$\mathbf{x}_0 + \ker(L_A) = \{\mathbf{x}_0 + \mathbf{y} \mid \mathbf{y} \in \ker(L_A)\}.$$

It is possible that there are no solutions at all; this occurs when $\mathbf{b} \notin \operatorname{im}(L_A)$. If there are solutions, then there is a unique solution precisely when $\ker(L_A) = \{\mathbf{0}\}$. If the field $\mathbb{F}$ is infinite and there are solutions but $\ker(L_A) \neq \{\mathbf{0}\}$, then there are infinitely many solutions.

5.2.2 Gauss method and elementary row transformations

There are two standard high school methods for solving linear systems: *substitution method* (where you consequently express variables through other variables and substitute the result in remaining equations) and *elimination method* (sometimes called *Gauss method*). The latter is usually more effective, so we would like to contemplate its nature.

In fact, Gauss method are exactly elementary row transformations! We saw in Section 4.3 that a row transformation changes A to PA where P is one of the elementary matrices. Since P is invertible,

$$A\mathbf{x} = \mathbf{b} \ \Leftrightarrow PA\mathbf{x} = P\mathbf{b}$$

that allows us to improve the equation until we can see the solution. Notice that the column transformation changes A to AP and it is just destroying the equation.

Example. Consider the following three systems.

$$\begin{cases} 2x+y &= 1 \\ 4x+2y &= 1 \end{cases} \qquad \begin{cases} 2x+y &= 1 \\ 4x+y &= 1 \end{cases} \qquad \begin{cases} 2x+y &= 1 \\ 4x+2y &= 2 \end{cases}$$

The Gauss method solves them by subtracting twice the first equation from the second equation:

$$\begin{cases} 2x+y &= 1 \\ 0 &= -1 \end{cases} \qquad \begin{cases} 2x+y &= 1 \\ -y &= -1 \end{cases} \qquad \begin{cases} 2x+y &= 1 \\ 0 &= 0 \end{cases}$$

The first equation has no solutions. The second equation has a single solution $(x, y) = (0, 1)$. The third equation has infinitely many solutions $(x, y) = (t, 1-2t)$.

In terms of the elementary row transformations, we write the systems in matrix form:

$$\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Then we perform $\mathbf{r}_2 \leftarrow \mathbf{r}_2 - 2\mathbf{r}_1$ on all three of them. Equivalently, we multiply by $P = E(2)^1_{-2,2,1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$:

$$\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

5.2.3 Augmented matrix

We would like to make the process of solving more mechanical by forgetting about the variable names w, x, y, z, etc. and doing the whole operation as a matrix calculation. For this, we use the *augmented matrix* of the system of equations, which for the system $A\mathbf{x} = \mathbf{b}$ of m equations in n unknowns, where A is the $m \times n$ matrix (α_{ij}) is defined to be the $m \times (n+1)$ matrix

$$\overline{A} = (\ A \mid \mathbf{b}\ ) = \left(\begin{array}{cccc|c} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} & \beta_1 \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} & \beta_2 \\ \vdots & & \vdots & & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn} & \beta_m \end{array}\right).$$

The vertical line in the matrix is put there just to remind us that the rightmost column is different from the others, and arises from the constants on the right-hand side of the equations.

Let us look at the following system of linear equations over $\mathbb{R}$, that is, we want to find all $x, y, z, w \in \mathbb{R}$ satisfying the equations.

$$\left\{\begin{array}{rcrcrcrcr} 2w & - & x & + & 4y & - & z & = & 1 \\ w & + & 2x & + & y & + & z & = & 2 \\ w & - & 3x & + & 3y & - & 2z & = & -1 \\ -3w & - & x & - & 5y & & & = & -3 \end{array}\right.$$

Elementary row operations on $\overline{A}$ are precisely Gauss transformations of the corresponding linear system. Thus, the solution can be carried out mechanically as follows:

Matrix	**Operation**
$\left(\begin{array}{rrrr\|r} 2 & -1 & 4 & -1 & 1 \\ 1 & 2 & 1 & 1 & 2 \\ 1 & -3 & 3 & -2 & -1 \\ -3 & -1 & -5 & 0 & -3 \end{array}\right)$	$\mathbf{r}_1 \to \mathbf{r}_1/2$
$\left(\begin{array}{rrrr\|r} 1 & -1/2 & 2 & -1/2 & 1/2 \\ 1 & 2 & 1 & 1 & 2 \\ 1 & -3 & 3 & -2 & -1 \\ -3 & -1 & -5 & 0 & -3 \end{array}\right)$	$\mathbf{r}_2 \to \mathbf{r}_2 - \mathbf{r}_1,\ \mathbf{r}_3 \to \mathbf{r}_3 - \mathbf{r}_1,\ \mathbf{r}_4 \to \mathbf{r}_4 + 3\mathbf{r}_1$
$\left(\begin{array}{rrrr\|r} 1 & -1/2 & 2 & -1/2 & 1/2 \\ 0 & 5/2 & -1 & 3/2 & 3/2 \\ 0 & -5/2 & 1 & -3/2 & -3/2 \\ 0 & -5/2 & 1 & -3/2 & -3/2 \end{array}\right)$	$\mathbf{r}_3 \to \mathbf{r}_3 + \mathbf{r}_2,\ \mathbf{r}_4 \to \mathbf{r}_4 + \mathbf{r}_2$
$\left(\begin{array}{rrrr\|r} 1 & -1/2 & 2 & -1/2 & 1/2 \\ 0 & 5/2 & -1 & 3/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$	$\mathbf{r}_2 \to 2\mathbf{r}_2/5$
$\left(\begin{array}{rrrr\|r} 1 & -1/2 & 2 & -1/2 & 1/2 \\ 0 & 1 & -2/5 & 3/5 & 3/5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$	$\mathbf{r}_1 \to \mathbf{r}_1 + \mathbf{r}_2/2$
$\left(\begin{array}{rrrr\|r} 1 & 0 & 9/5 & -1/5 & 4/5 \\ 0 & 1 & -2/5 & 3/5 & 3/5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$	

The original system has been transformed to the following equivalent system, that is, both systems have the same solutions.

$$\left\{\begin{array}{rcrcrcr} w & + & 9/5y & - & 1/5z & = & 4/5 \\ x & - & 2/5y & + & 3/5z & = & 3/5 \end{array}\right.$$

In the latter system, variables y and z can take arbitrary values in $\mathbb{R}$ in the solution set; say $y = s$, $z = t$. Then equations tell us that $x = 2s/5 - 3t/5 + 3/5$ and $w = -9s/5 + t/5 + 4/5$

(be careful to get the signs right!), and so the complete set of solutions is

$$\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -9s/5 + t/5 + 4/5 \\ 2s/5 - 3t/5 + 3/5 \\ s \\ t \end{pmatrix} = \begin{pmatrix} 4/5 \\ 3/5 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -9/5 \\ 2/5 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1/5 \\ -3/5 \\ 0 \\ 1 \end{pmatrix}, \; s, t \in \mathbb{R}.$$

5.2.4 Row echelonisation a matrix

Let $A = (\alpha_{ij})$ be an $m \times n$ matrix over the field $\mathbb{F}$. For i-th row let $\alpha_{i,c(i)}$ be the first (leftmost) non-zero entry in this row. In other words, $\alpha_{i,c(i)} \neq 0$ while $\alpha_{ij} = 0$ for all $j < c(i)$. By convention, $c(i) = \infty$ if $\alpha_{ij} = 0$ for all j.

After applying this procedure, the resulting matrix $A = (\alpha_{ij})$ has the following properties.

(i) All zero rows are below all non-zero rows.

(ii) Let $\mathbf{r}_1, \ldots, \mathbf{r}_s$ be the non-zero rows. Then each $\mathbf{r}_i$ with $1 \leq i \leq s$ has 1 as its first non-zero entry. (In other words, $\alpha_{i,c(i)} = 1$.)

(iii) $c(1) < c(2) < \cdots < c(s)$.

(iv) $\alpha_{k,c(i)} = 0$ for all $k > i$.

Definition 5.2.1. A matrix satisfying these properties is said to be in *row echelon form*.

There is a stronger version of the last property

(v) $\alpha_{k,c(i)} = 0$ for all $k \neq i$.

Definition 5.2.2. A matrix satisfying properties (i), (ii), (iii), and (v) is said to be in *row reduced echelon form*.

Here is the intuition behind these forms:

- The number of non-zero rows in a row echelon form of A is the rank of A (prove it yourself).
- The row reduced echelon form of A (it is not just my misuse of articles: this form is, indeed, unique) solves the system of linear equations.

In this light, the following theorem says that every system of linear equations can be solved by Gauss method.

Theorem 5.2.3. *Every matrix can be brought to the row reduced echelon form by elementary row transformations.*

Proof. This proof is similar to the proof of Theorem 5.1.1, yet somewhat more technically complex. Again we describe an algorithm whose steps contain elementary row operations. We have to show that

1. after termination the resulting matrix has the row reduced echelon form,
2. the algorithm terminates after finitely many steps.

Both these statements are clear from the nature of the algorithm. Make sure that you understand why they are clear!

At any stage of the procedure, we are looking at the entry α_{ij} in a particular position (i, j) of the matrix. (i, j) is called the *pivot* position, and α_{ij} the *pivot* entry. We start with $(i, j) = (1, 1)$ and proceed as follows.

1. If α_{ij} and all entries below it in its columns are zero (i.e. if $\alpha_{kj} = 0$ for all $k \geq i$), then move the pivot one place to the right, to $(i, j + 1)$ and repeat Step 1, or terminate if $j = n$.

2. If $\alpha_{ij} = 0$ but $\alpha_{kj} \neq 0$ for some $k > j$ then apply $\mathcal{R}2$ and interchange $\mathbf{r}_i$ and $\mathbf{r}_k$.

3. At this stage $\alpha_{ij} \neq 0$. If $\alpha_{ij} \neq 1$, then apply $\mathcal{R}3$ and multiply $\mathbf{r}_i$ by α_{ij}^{-1}.

4. At this stage $\alpha_{ij} = 1$. If, for any $k \neq i$, $\alpha_{kj} \neq 0$, then apply $\mathcal{R}1$, and subtract α_{kj} times $\mathbf{r}_i$ from $\mathbf{r}_k$.

5. At this stage, $\alpha_{kj} = 0$ for all $k \neq i$. If $i = m$ or $j = n$ then terminate. Otherwise, move the pivot diagonally down to the right to $(i + 1, j + 1)$, and go back to Step 1.

□

If one needs only a row echelon form, this can done faster by replacing steps 4 and 5 with weaker and faster steps as follows.

4a. At this stage $\alpha_{ij} = 1$. If, for any $k > i$, $\alpha_{kj} \neq 0$, then apply $\mathcal{R}1$, and subtract α_{kj} times $\mathbf{r}_i$ from $\mathbf{r}_k$.

5a. At this stage, $\alpha_{kj} = 0$ for all $k > i$. If $i = m$ or $j = n$ then terminate. Otherwise, move the pivot diagonally down to the right to $(i + 1, j + 1)$, and go back to Step 1.

In the example below, we find a row echelon form of a matrix by applying the faster algorithm. The number in the 'Step' column refers to the number of the step applied in the description of the procedure above.

Example. $A = \begin{pmatrix} 0 & 0 & 1 & 2 & 1 \\ 2 & 4 & 2 & -4 & 2 \\ 3 & 6 & 3 & -6 & 3 \\ 1 & 2 & 3 & 3 & 3 \end{pmatrix}$.

Matrix	Pivot	Step	Operation
$\begin{pmatrix} 0 & 0 & 1 & 2 & 1 \\ 2 & 4 & 2 & -4 & 2 \\ 3 & 6 & 3 & -6 & 3 \\ 1 & 2 & 3 & 3 & 3 \end{pmatrix}$	$(1,1)$	**2**	$\mathbf{r}_1 \leftrightarrow \mathbf{r}_2$
$\begin{pmatrix} 2 & 4 & 2 & -4 & 2 \\ 0 & 0 & 1 & 2 & 1 \\ 3 & 6 & 3 & -6 & 3 \\ 1 & 2 & 3 & 3 & 3 \end{pmatrix}$	$(1,1)$	**3**	$\mathbf{r}_1 \to \mathbf{r}_1/2$
$\begin{pmatrix} 1 & 2 & 1 & -2 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 3 & 6 & 3 & -6 & 3 \\ 1 & 2 & 3 & 3 & 3 \end{pmatrix}$	$(1,1)$	**4**	$\mathbf{r}_3 \to \mathbf{r}_3 - 3\mathbf{r}_1$ $\mathbf{r}_4 \to \mathbf{r}_4 - \mathbf{r}_1$
$\begin{pmatrix} 1 & 2 & 1 & -2 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 5 & 2 \end{pmatrix}$	$(1,1) \to (2,2) \to (2,3)$	**5, 1**	
$\begin{pmatrix} 1 & 2 & 1 & -2 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 5 & 2 \end{pmatrix}$	$(2,3)$	**4**	$\mathbf{r}_4 \to \mathbf{r}_4 - 2\mathbf{r}_2$
$\begin{pmatrix} 1 & 2 & 1 & -2 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$(2,3) \to (3,4)$	**5, 2**	$\mathbf{r}_3 \leftrightarrow \mathbf{r}_4$
$\begin{pmatrix} 1 & 2 & 1 & -2 & 1 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$(3,4) \to (4,5) \to$ **stop**	**5, 1**	

Note that the row reduced echelon form of A can be obtained from a row echelon form. In this example, three further row transformations are needed:

$$\begin{array}{c} \mathbf{r}_1 \to \mathbf{r}_1 + 2\mathbf{r}_3 \\ \mathbf{r}_2 \to \mathbf{r}_2 - 2\mathbf{r}_3 \\ \mathbf{r}_1 \to \mathbf{r}_1 - \mathbf{r}_2 \end{array} \qquad \begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

5.2.5 Rank criterion

The following criterion has mostly theoretical value. The proof may appear in your example sheets. If not try proving it on your own.

Proposition 5.2.4. *Let A be the $n \times m$ matrix of a linear system, $\overline{A}$ its augmented $n \times (m+1)$ matrix. The system of linear equations has a solution if and only if* $\operatorname{rank}(A) = \operatorname{rank}(\overline{A})$.

5.3 The inverse of a matrix

The row reduction is an efficient practical method for finding the inverse matrix

5.3.1 Definition and basic properties

Let $A \in \mathbb{F}^{m,n}$. Recall (cf. Section 2.3.2) that A is called invertible, if there exists the inverse matrix $A^{-1} \in \mathbb{F}^{m,n}$ such that $AA^{-1} = I_m$ and $A^{-1}A = I_n$.

Example. Let $A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -2 \\ 0 & 1 \\ -2 & 5 \end{pmatrix}$. Then $AB = I_2$, but $BA \neq I_3$, so a non-square matrix can have "a right inverse" which is not "a left inverse". Such misfortunes do not happen with square matrices.

Lemma 5.3.1. *If A and B are $n \times n$ matrices such that $AB = I_n$, then A and B are invertible, and $A^{-1} = B$, $B^{-1} = A$.*

Proof. If $B\mathbf{x} = \mathbf{0}$ then $\mathbf{x} = I_n\mathbf{x} = A(B\mathbf{x}) = \mathbf{0}$. Thus, $L_B : \mathbb{F}^n \to \mathbb{F}^n$ is injective. By Theorem 2.2.7, $\operatorname{rank}(L_B) = n - \operatorname{nullity}(L_B) = n - 0 = n$ so that L_B is surjective and bijective. The inverse is a linear map. Thus, B is invertible and B^{-1} exists. Finally, $A = ABB^{-1} = B^{-1}$ and $A^{-1} = (B^{-1})^{-1} = B$. $\square$

Lemma 5.3.2. *If A and B are invertible $n \times n$ matrices, then AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. This is clear, because $ABB^{-1}A^{-1} = B^{-1}A^{-1}AB = I_n$. $\square$

5.3.2 Row reduced form of an invertible matrix

Proposition 5.3.3. *The row reduced echelon form of an invertible $n \times n$ matrix A is I_n.*

Proof. First note that if an $n \times n$ matrix A is invertible, then it has rank n. Consider the row reduced echelon form $B = (\beta_{ij})$ of A. As we saw in Section 5.1.2, we have $\beta_{ic(i)} = 1$ for $1 \leq i \leq n$ (since $\operatorname{rank}(A) = \operatorname{rank}(B) = n$), where $c(1) < c(2) < \cdots < c(n)$, and clearly this is only possible if $c(i) = i$ for $1 \leq i \leq n$. Then, since all other entries in column $c(i)$ are zero, we have $B = I_n$. $\square$

Corollary 5.3.4. *An invertible matrix is a product of elementary matrices.*

Proof. The sequence of row operations in the proof of Proposition 5.3.3 can be written as

$$A \leftarrow E_1 A \leftarrow E_2 E_1 A \leftarrow \ldots \leftarrow E_r E_{r-1} \ldots E_1 A = I_n.$$

Since elementary matrices are invertible and their inverses are also elementary,

$$A = (E_r E_{r-1} \ldots E_1)^{-1} = E_1^{-1} E_2^{-1} \ldots E_r^{-1}.$$

□

5.3.3 Algorithm for matrix inversion

Corollary 5.3.4 yields an explicit formula for the inverse matrix

$$A^{-1} = E_r E_{r-1} \ldots E_1 = E_r E_{r-1} \ldots E_1 I_n$$

that can be turned into an algorithm. Just reduce A to its row reduced echelon form I_n, using elementary row operations, while simultaneously applying the same row operations to the identity matrix I_n. These operations transform I_n to A^{-1}.

In practice, we might not know whether or not A is invertible before we start, but we will find out while carrying out this procedure because, if A is not invertible, then its rank will be less than n, and it will not row reduce to I_n.

Example.

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 4 & 1 & 3 \\ 2 & 1 & 6 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{r}_1 \to \mathbf{r}_1/3$$

$$\begin{pmatrix} 1 & 2/3 & 1/3 \\ 4 & 1 & 3 \\ 2 & 1 & 6 \end{pmatrix} \qquad \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{r}_2 \to \mathbf{r}_2 - 4\mathbf{r}_1$$
$$\mathbf{r}_3 \to \mathbf{r}_3 - 2\mathbf{r}_1$$

$$\begin{pmatrix} 1 & 2/3 & 1/3 \\ 0 & -5/3 & 5/3 \\ 0 & -1/3 & 16/3 \end{pmatrix} \qquad \begin{pmatrix} 1/3 & 0 & 0 \\ -4/3 & 1 & 0 \\ -2/3 & 0 & 1 \end{pmatrix}$$

$$\mathbf{r}_2 \to -3\mathbf{r}_2/5$$

$$\begin{pmatrix} 1 & 2/3 & 1/3 \\ 0 & 1 & -1 \\ 0 & -1/3 & 16/3 \end{pmatrix} \qquad \begin{pmatrix} 1/3 & 0 & 0 \\ 4/5 & -3/5 & 0 \\ -2/3 & 0 & 1 \end{pmatrix}$$

$$\mathbf{r}_1 \to \mathbf{r}_1 - 2\mathbf{r}_2/3$$
$$\mathbf{r}_3 \to \mathbf{r}_3 + \mathbf{r}_2/3$$

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

$$\mathbf{r}_3 \to \mathbf{r}_3/5$$

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -1/5 & 2/5 & 0 \\ 4/5 & -3/5 & 0 \\ -2/25 & -1/25 & 1/5 \end{pmatrix}$$

$$\mathbf{r}_1 \to \mathbf{r}_1 - \mathbf{r}_3$$
$$\mathbf{r}_2 \to \mathbf{r}_2 + \mathbf{r}_3$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -3/25 & 11/25 & -1/5 \\ 18/25 & -16/25 & 1/5 \\ -2/25 & -1/25 & 1/5 \end{pmatrix}$$

So

$$A^{-1} = \begin{pmatrix} -3/25 & 11/25 & -1/5 \\ 18/25 & -16/25 & 1/5 \\ -2/25 & -1/25 & 1/5 \end{pmatrix}.$$

It is always a good idea to check the result afterwards. This is easier if we remove the common denominator 25, and we can then easily check that

$$\begin{pmatrix} 3 & 2 & 1 \\ 4 & 1 & 3 \\ 2 & 1 & 6 \end{pmatrix} \begin{pmatrix} -3 & 11 & -5 \\ 18 & -16 & 5 \\ -2 & -1 & 5 \end{pmatrix} = \begin{pmatrix} -3 & 11 & -5 \\ 18 & -16 & 5 \\ -2 & -1 & 5 \end{pmatrix} \begin{pmatrix} 3 & 2 & 1 \\ 4 & 1 & 3 \\ 2 & 1 & 6 \end{pmatrix} = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 25 \end{pmatrix}$$

which confirms the result!

6 The Determinant: Methods of Calculation

This is a breather chapter, roughly in line with the breather week. The only goal of this chapter is to introduce the determinant and to learn how to compute it.

6.1 From first principles

Let A be an $n \times n$ matrix over the field $\mathbb{F}$. The *determinant* of A, which is written as $\det(A)$ or sometimes as $|A|$, is a certain scalar that is defined from A in a rather complicated way.

6.1.1 Definition of the determinant

The definition for $n \leq 3$ is already familiar to you:

$$\det(a) = a, \ \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \ \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{matrix} a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - \\ -a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + \\ +a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{matrix}$$

Now we turn to the general definition for $n \times n$ matrices. Suppose that we take the product of n entries from the matrix, where we take exactly one entry from each row and one from each column. Such a product is called an *elementary product*.

Recall the symmetric group S_n for Algebra-1 (or Sets and Numbers) last term. The elements of S_n are permutations of the set $X_n = \{1, 2, 3, \ldots, n\}$, is simply a bijection from X_n to itself. There are $n!$ permutations altogether, so $|S_n| = n!$.

An elementary product contains one entry from each row of A, so let the entry in the product from the ith row be $a_{i\phi(i)}$, where ϕ is some as-yet unknown function from X_n to X_n. Since the product also contains exactly one entry from each column, each integer $j \in X_n$ must occur exactly once as $\phi(i)$. But this is just saying that $\phi \colon X_n \to X_n$ is a bijection; that is $\phi \in S_n$. So an elementary product has the general form $a_{1\phi(1)}a_{2\phi(2)} \ldots a_{n\phi(n)}$ for some $\phi \in S_n$, and there are $n!$ elementary products altogether. We want to define

$$\det(A) = \sum_{\phi \in S_n} \pm a_{1\phi(1)}a_{2\phi(2)} \ldots a_{n\phi(n)}$$

but we still have to decide which of the elementary products has a plus sign and which has a minus sign. In fact this depends on the sign of the permutation ϕ, which we must now define.

Recall that a transposition is a permutation of X_n that interchanges two numbers i and j in X_n and leaves all other numbers fixed. It is written as (i, j). The two key facts are

- every $\phi \in S_n$ is a product of permutations,
- for each fixed ϕ the number of permutations in this product is always odd or always even.

Definition 6.1.1. A permutation $\phi \in S_n$ is said to be even, and to have sign $+1$, if ϕ is a composition of an even number of transpositions; and ϕ is said to be odd, and to have sign -1, if ϕ is a composition of an odd number of transpositions.

Example. The permutation $\phi \in S_5$ defined by

$$\phi(1) = 4,\ \phi(2) = 5,\ \phi(3) = 3,\ \phi(4) = 2,\ \phi(5) = 1$$

is equal to $(1,4) \cdot (2,4) \cdot (2,5)$ (Remember that permutations are functions $X_n \to X_n$. Their product is the composition starting from the right, so this means first apply the function $(2,5)$ (which interchanges 2 and 5) then apply $(2,4)$ and finally apply $(1,4)$.) or $(2,3) \cdot (3,4) \cdot (2,5)) \cdot (3,5) \cdot (1,5)$. Hence, $\text{sign}(\phi) = -1$.

Now we can give the general definition of the determinant. It is optimal for calculation only if $n \leq 2$. Already for $n = 3$, you should use other methods.

Definition 6.1.2. The determinant of a $n \times n$ matrix $A = (a_{ij})$ is the scalar quantity

$$\det(A) = \sum_{\phi \in S_n} \text{sign}(\phi) a_{1\phi(1)} a_{2\phi(2)} \dots a_{n\phi(n)}.$$

6.1.2 Characteristic polynomial

Definition 6.1.3. For an $n \times n$ matrix A, the determinant $\det(A - x) := \det(A - xI_n)$ is called the characteristic polynomial of A.

Note that x is an indeterminant variable so that $A - xI_n$ is a matrix with coefficients in the field of rational functions $\mathbb{F}(x)$. Its elements are $f(x)/h(x)$ where $f(x), h(x) \in \mathbb{F}[x]$ and $h \neq 0$ under the usual addition and multiplications rules. All the theorems and computation methods for the determinants apply equally to characteristic functions.

Example. Let $A = \begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix}$. Then

$$\det(A - x) = \begin{vmatrix} 1 - x & 2 \\ 5 & 4 - x \end{vmatrix} = (1 - x)(4 - x) - 10 = x^2 - 5x - 6 = (x - 6)(x + 1).$$

6.1.3 Staircase matrices

An upper staircase matrix (or lower staircase) is a matrix of the form

$$A_{upper} = \begin{pmatrix} [\ B_1\ ] & * & \dots & * \\ \mathbf{0} & [\ B_2\ ] & \dots & * \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & [\ B_n\ ] \end{pmatrix} \quad (\text{or } A_{lower} = \begin{pmatrix} [\ B_1\ ] & \mathbf{0} & \dots & \mathbf{0} \\ * & [\ B_2\ ] & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ * & * & \dots & [\ B_n\ ] \end{pmatrix})$$

with n square blocks on the diagonal. Their determinant is

$$\det(A_{upper}) = \det(A_{lower}) = \det(B_1) \cdot \det(B_2) \cdots \det(B_n)\,. \tag{5}$$

Example. The following matrix is un an upper staircase with three 2×2 blocks.

$$\det \begin{pmatrix} 3 & 0 & 1 & 2 & 3 & 0 \\ 1 & 1 & 5 & 2 & 0 & 3 \\ 0 & 0 & 1 & 1 & 1 & 3 \\ 0 & 0 & 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{vmatrix} 3 & 0 \\ 1 & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \cdot \begin{vmatrix} 1 & 3 \\ 1 & 1 \end{vmatrix} = 3 \cdot (2-1) \cdot (1-3) = -6\,.$$

Triangular and diagonal matrices are subsets of the set of staircase matrices. An upper staircase matrix is called *upper triangular* if all the diagonal blocks have size 1. Equivalently, all of its entries below the main diagonal are zero, that is, (a_{ij}) is upper triangular if $a_{ij} = 0$ for all $i > j$.

The matrix is called *diagonal* if all entries not on the main diagonal are zero; that is, $a_{ij} = 0$ for $i \neq j$. The following is a partial case of Equation(5) but we give an independent simpler proof

Proposition 6.1.4. *If $A = (a_{ij})$ is upper (or lower) triangular, then $\det(A) = a_{11}a_{22} \cdots a_{nn}$ is the product of the entries on the main diagonal of A.*

Proof. If $\phi \in S_n$ is not the identity permutation, there exists i such that $\phi(i) < i$. The corresponding elementary product contain $a_{i\phi(i)}$ but $a_{ij} = 0$ when $i > j$. So the only non-zero elementary product in the sum occurs when ϕ is the identity permutation. Hence $\det(A) = a_{11}a_{22} \ldots a_{nn} = 1$. $\square$

Example. $A = \begin{pmatrix} 3 & 0 & -\frac{1}{2} \\ 0 & -1 & -11 \\ 0 & 0 & \frac{2}{5} \end{pmatrix}$ is upper triangular, and $B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 17 & 0 \\ 0 & 0 & -3 \end{pmatrix}$ is diagonal. By Proposition 6.1.4, $\det(A - x) = (3 - x)(-1 - x)(\frac{2}{5} - x)$ and $\det(B - x) = -x(17 - x)(-3 - x)$.

6.2 Gaussian transformations

An efficient way to compute the determinants is to use row and column transformations, at least for $n \geq 3$. For $n = 2$, it is easiest to do it straight from the definition.

6.2.1 The effect of matrix operations on the determinant

We will prove this theorem in the next section:

Theorem 6.2.1. *Elementary row and column operations affect the determinant of a matrix as follows.*

(i) $\det(I_n) = 1$.

(ii) *Let B result from A by applying $\mathcal{R}2$ or $\mathcal{C}2$ (swap). Then* $\det(B) = -\det(A)$.

(iii) *If A has two equal rows or columns, then* $\det(A) = 0$.

(iv) *Let B result from A by applying $\mathcal{R}1$ or $\mathcal{C}1$ (addition). Then* $\det(B) = \det(A)$.

(v) *Let B result from A by applying $\mathcal{R}3$ or $\mathcal{C}3$ (multiplication). Then* $\det(B) = \lambda \det(A)$.

6.2.2 Algorithm

Let us turn Theorem 6.2.1 into an algorithm for a rational matrix $A \in \mathbb{Q}^{n,n}$. Over more general fields, just skip the first step.

1. Use $\mathcal{R}3$ and $\mathcal{C}3$ to make all the coefficients integer.
2. Use $\mathcal{R}1$, $\mathcal{C}1$, $\mathcal{R}2$ and $\mathcal{C}2$ to reduce A to upper or lower staircase form with blocks of size at most 2.
3. Apply Formula (5)

Example. We just proceed with the algorithm.

$$\begin{vmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 \\ 2 & 1 & 3 & 1 \\ 1/2 & 1 & 2 & 1 \end{vmatrix} \overset{\mathbf{r}_4 \to 2\mathbf{r}_4}{=} \frac{1}{2} \begin{vmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 \\ 2 & 1 & 3 & 1 \\ 1 & 2 & 4 & 2 \end{vmatrix} \overset{\substack{\mathbf{r}_4 \to \mathbf{r}_4 - \mathbf{r}_2 \\ \mathbf{r}_3 \to \mathbf{r}_3 - 2\mathbf{r}_2}}{=} \frac{1}{2} \begin{vmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 \\ 0 & -3 & 1 & -1 \\ 0 & 0 & 3 & 1 \end{vmatrix} \overset{\mathbf{r}_3 \to \mathbf{r}_3 + 3\mathbf{r}_1}{=}$$

$$= \frac{1}{2} \begin{vmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 \\ 0 & 0 & 4 & 5 \\ 0 & 0 & 3 & 1 \end{vmatrix} \overset{(5)}{=} \frac{1}{2} \begin{vmatrix} 0 & 1 \\ 1 & 2 \end{vmatrix} \cdot \begin{vmatrix} 4 & 5 \\ 3 & 1 \end{vmatrix} = \frac{1}{2} \cdot (-1) \cdot (-11) = \frac{11}{2}$$

Notice that the algorithm would not work so nicely for the characteristic polynomial $\det(A - x)$ because we will be getting higher and higher powers of x with each transformation.

6.3 Cofactor expansion

That is the final method, optimal for computing characteristic polynomials. It may be also useful for matrices of large size. Occasionally, combining different methods could optimal.

6.3.1 Minors and cofactors

Let $A = (a_{ij})$ be an $n \times n$ matrix. Let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i-th row and the j-th column of A. Now let $M_{ij} = \det(A_{ij})$. Then M_{ij} is called the (i, j)-th *minor* of A.

Example. If $A = \begin{pmatrix} 2 & 1 & 0 \\ 3 & -1 & 2 \\ 5 & -2 & 0 \end{pmatrix}$, then $A_{12} = \begin{pmatrix} 3 & 2 \\ 5 & 0 \end{pmatrix}$ and $A_{31} = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}$, and so $M_{12} = -10$ and $M_{31} = 2$.

We define the (i,j)-th *cofactor* of A as

$$c_{ij} = (-1)^{i+j} M_{ij} = (-1)^{i+j} \det(A_{ij}).$$

In other words, c_{ij} is equal to M_{ij} if $i+j$ is even, and to $-M_{ij}$ if $i+j$ is odd. In the example above,

$$\begin{array}{llllll}
c_{11} = \begin{vmatrix} -1 & 2 \\ -2 & 0 \end{vmatrix} = 4, & c_{12} = -\begin{vmatrix} 3 & 2 \\ 5 & 0 \end{vmatrix} = 10, & c_{13} = \begin{vmatrix} 3 & -1 \\ 5 & -2 \end{vmatrix} = -1 \\
c_{21} = -\begin{vmatrix} 1 & 0 \\ -2 & 0 \end{vmatrix} = 0, & c_{22} = \begin{vmatrix} 2 & 0 \\ 5 & 0 \end{vmatrix} = 0, & c_{23} = -\begin{vmatrix} 2 & 1 \\ 5 & -2 \end{vmatrix} = 9 \\
c_{31} = \begin{vmatrix} 1 & 0 \\ -1 & 2 \end{vmatrix} = 2, & c_{32} = -\begin{vmatrix} 2 & 0 \\ 3 & 2 \end{vmatrix} = -4, & c_{33} = \begin{vmatrix} 2 & 1 \\ 3 & -1 \end{vmatrix} = -5.
\end{array}$$

Theorem 6.3.1. *Let A be an $n \times n$ matrix.*

(i) (Expansion of a determinant by the i-th row.) For any i with $1 \le i \le n$, we have

$$\det(A) = a_{i1}c_{i1} + a_{i2}c_{i2} + \cdots + a_{in}c_{in} = \sum_{j=1}^{n} a_{ij}c_{ij}.$$

(ii) (Expansion of a determinant by the j-th column.) For any j with $1 \le j \le n$, we have

$$\det(A) = a_{1j}c_{1j} + a_{2j}c_{2j} + \cdots a_{nj}c_{nj} = \sum_{i=1}^{n} a_{ij}c_{ij}.$$

For example, expanding the determinant of the matrix A above by the first row, the third row, and the second column give respectively:

$$\begin{array}{lcccc}
\det(A) & = & 2 \cdot 4 + 1 \cdot 10 + 0 \cdot (-1) & = & 18 \\
\det(A) & = & 5 \cdot 2 + (-2) \cdot (-4) + 0 \cdot (-5) & = & 18 \\
\det(A) & = & 1 \cdot 10 + (-1) \cdot 0 + (-2) \cdot (-4) & = & 18
\end{array}$$

6.3.2 Algorithm

Just expand by a row or a column to reduce to smaller size matrices. This is very quick if the matrix contains a lot of zeroes.

Example. The following matrix is custom-made for the method:

$$A = \begin{pmatrix} 9 & 0 & 2 & 6 \\ 1 & 2 & 9 & -3 \\ 0 & 0 & -2 & 0 \\ -1 & 0 & -5 & 2 \end{pmatrix}.$$

Expanding by the third row, we get $\det(A) = -2 \begin{vmatrix} 9 & 0 & 6 \\ 1 & 2 & -3 \\ -1 & 0 & 2 \end{vmatrix}$, and then expanding by the second column, $\det(A) = -2 \cdot 2 \begin{vmatrix} 9 & 6 \\ -1 & 2 \end{vmatrix} = -4 \cdot (18+6) = -96$.

Example. Let us compute the characteristic polynomial of

$$A = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 \\ 2 & 1 & 3 & 1 \\ 3 & 2 & 4 & 2 \end{pmatrix}$$

by combining both methods. We use $\mathbf{c}_1$ because the $(1,1)$-entry looks the simplest but we avoid clearing up $\mathbf{c}_2$ to avoid higher powers of x:

$$\det(A-x) = \begin{vmatrix} -x & 1 & 1 & 2 \\ 1 & 2-x & 1 & 1 \\ 2 & 1 & 3-x & 1 \\ 3 & 2 & 4 & 2-x \end{vmatrix} \overset{\substack{\mathbf{c}_4 \to \mathbf{c}_4 - \mathbf{c}_1 \\ \mathbf{c}_3 \to \mathbf{c}_3 - \mathbf{c}_1}}{=} \begin{vmatrix} -x & 1 & 1+x & 2+x \\ 1 & 2-x & 0 & 0 \\ 2 & 1 & 1-x & -1 \\ 3 & 2 & 1 & -1-x \end{vmatrix} =$$

Now we expand by the second row:

$$= - \begin{vmatrix} 1 & 1+x & 2+x \\ 1 & 1-x & -1 \\ 2 & 1 & -1-x \end{vmatrix} + (2-x) \begin{vmatrix} -x & 1+x & 2+x \\ 2 & 1-x & -1 \\ 3 & 1 & -1-x \end{vmatrix} =$$

Now apply $\mathbf{r}_3 \to \mathbf{r}_3 - 2\mathbf{r}_1$, $\mathbf{r}_2 \to \mathbf{r}_2 - \mathbf{r}_1$ on the first matrix and $\mathbf{c}_1 \to \mathbf{c}_1 - 3\mathbf{c}_2$ on the second:

$$= - \begin{vmatrix} 1 & 1+x & 2+x \\ 0 & -2x & -3-x \\ 0 & -1-2x & -5-3x \end{vmatrix} + (2-x) \begin{vmatrix} -3-4x & 1+x & 2+x \\ -1+3x & 1-x & -1 \\ 0 & 1 & -1-x \end{vmatrix}$$

Now expand by the first columns in both matrices

$$- \begin{vmatrix} -2x & -3-x \\ -1-2x & -5-3x \end{vmatrix} + (2-x)(-3-4x) \begin{vmatrix} 1-x & -1 \\ 1 & -1-x \end{vmatrix} - (2-x)(-1+3x) \begin{vmatrix} 1+x & 2+x \\ 1 & -1-x \end{vmatrix} =$$

and compute each 2×2 determinant separately in larger brackets

$$= -\Big(-2x(-5-3x) - (-1-2x)(-3-x) \Big) + \Big((4x^2-5x-6)((1-x)(-1-x)+1) \Big) +$$

$$+ \Big((3x^2-7x+2)((1+x)(-1-x)-(2+x)) \Big) = \Big(-4x^2+3x+3 \Big) +$$

$$+ \Big(4x^4 - 5x^3 - 6x^2 \Big) + \Big(-3x^4 - 2x^3 + 10x^2 + 15x - 6 \Big) = x^4 - 7x^3 + 12x - 3.$$

6.3.3 Vandermonde matrix and its determinant

Let $a_1, a_2, \ldots, a_n$ be elements of the field $\mathbb{F}$. *The Vandermonde matrix*

$$V := \begin{pmatrix} 1 & a_1 & a_1^2 & \ldots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \ldots & a_2^{n-1} \\ \vdots & & \vdots & & \vdots \\ 1 & a_n & a_n^2 & \ldots & a_n^{n-1} \end{pmatrix} \in \mathbb{F}^{n,n}$$

is an important matrix for many applications. Let us compute its determinant:

$$\det(V) = \prod_{1 \le i < j \le n} (a_j - a_i)\,. \tag{6}$$

We proceed by induction. For $n = 1$ the formula is trivial and for $n = 2$ we have

$$\begin{vmatrix} 1 & a_1 \\ 1 & a_2 \end{vmatrix} = a_2 - a_1\,.$$

Suppose we know the formula for $n - 1$. Let us do the elementary column operations on V

$$\mathbf{c}_n \to \mathbf{c}_n - a_1\mathbf{c}_{n-1}, \;\; \mathbf{c}_{n-1} \to \mathbf{c}_{n-1} - a_1\mathbf{c}_{n-2}, \;\; \ldots \;\; \mathbf{c}_2 \to \mathbf{c}_2 - a_1\mathbf{c}_1$$

in that particular order. Thus,

$$\det(V) = \begin{vmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 1 & a_2 - a_1 & a_2(a_2 - a_1) & a_2^2(a_2 - a_1) & \ldots & a_2^{n-2}(a_2 - a_1) \\ & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & a_n - a_1 & a_n(a_n - a_1) & a_n^2(a_n - a_1) & \ldots & a_n^{n-2}(a_n - a_1) \end{vmatrix}.$$

Finally expanding by the first row (or using the lower staircase formula)

$$\det(V) = (a_2 - a_1)\ldots(a_n - a_1) \begin{vmatrix} 1 & a_2 & a_2^2 & \ldots & a_2^{n-2} \\ 1 & a_3 & a_3^2 & \ldots & a_3^{n-2} \\ & \ldots & \ldots & \ldots & \ldots \\ 1 & a_n & a_n^2 & \ldots & a_n^{n-2} \end{vmatrix}$$

and the result follows by the inductive hypothesis.

6.3.4 The inverse of a matrix using determinants

Let $A \in \mathbb{F}^{n,n}$ be an $n \times n$ matrix. We define the *adjoint* matrix $\operatorname{adj}(A)$ of A to be the $n \times n$ matrix of which the (i, j)-th element is the cofactor c_{ji}. In other words, it is the transpose of the matrix of cofactors. For instance, in the example at the start of Section 6.3.1,

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 3 & -1 & 2 \\ 5 & -2 & 0 \end{pmatrix}, \qquad \operatorname{adj}(A) = \begin{pmatrix} 4 & 0 & 2 \\ 10 & 0 & -4 \\ -1 & 9 & -5 \end{pmatrix}.$$

We will not prove the next theorem but check that $A\operatorname{adj}(A) = \operatorname{adj}(A)A = 18I_3$ in the last example.

Theorem 6.3.2. $A\,\text{adj}(A) = \det(A) I_n = \text{adj}(A) A$

Corollary 6.3.3. *If* $\det(A) \neq 0$, *then* $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$.

This formula for finding inverses should be used for 2×2 matrices and **sometimes** for 3×3 matrices. For larger matrices, the row reduction method described in Section 5.3 is quicker.

In the example above,

$$\begin{pmatrix} 2 & 1 & 0 \\ 3 & -1 & 2 \\ 5 & -2 & 0 \end{pmatrix}^{-1} = \frac{1}{18} \begin{pmatrix} 4 & 0 & 2 \\ 10 & 0 & -4 \\ -1 & 9 & -5 \end{pmatrix},$$

and in the example in Section 5.3.3,

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 4 & 1 & 3 \\ 2 & 1 & 6 \end{pmatrix}, \ \text{adj}(A) = \begin{pmatrix} 3 & -11 & 5 \\ -18 & 16 & -5 \\ 2 & 1 & -5 \end{pmatrix}, \ \det(A) = -25, \ A^{-1} = \frac{-1}{25} \text{adj}(A).$$

7 The Determinant: Algebraic and Geometric Properties

We will catch up with some of the algebraic proofs, missing in the last chapter. Then we will give the geometric meaning of the determinant over $\mathbb{R}$.

7.1 Algebraic Properties

7.1.1 Determinant of transposed matrix

Recall that the transposed matrix A^{T} of $A = (a_{ij})$ is defined in Section 2.3.2. For example, $\begin{pmatrix} 1 & 3 & 5 \\ -2 & 0 & 6 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} 1 & -2 \\ 3 & 0 \\ 5 & 6 \end{pmatrix}$.

Theorem 7.1.1. *Let* $A = (a_{ij})$ *be an* $n \times n$ *matrix. Then* $\det(A^{\mathrm{T}}) = \det(A)$.

Proof. Let $A^{\mathrm{T}} = (b_{ij})$ where $b_{ij} = a_{ji}$. Then

$$\begin{aligned} \det(A^{\mathrm{T}}) &= \sum_{\phi \in S_n} \operatorname{sign}(\phi) b_{1\phi(1)} b_{2\phi(2)} \dots b_{n\phi(n)} \\ &= \sum_{\phi \in S_n} \operatorname{sign}(\phi) a_{\phi(1)1} a_{\phi(2)2} \dots a_{\phi(n)n}. \end{aligned}$$

Now, by rearranging the terms in the elementary product, we have

$$a_{\phi(1)1} a_{\phi(2)2} \dots a_{\phi(n)n} = a_{1\phi^{-1}(1)} a_{2\phi^{-1}(2)} \dots a_{n\phi^{-1}(n)},$$

where ϕ^{-1} is the *inverse* permutation to ϕ. Notice also that if ϕ is a composition $\tau_1 \circ \tau_2 \circ \cdots \circ \tau_r$ of transpositions τ_i, then $\phi^{-1} = \tau_r \circ \cdots \circ \tau_2 \circ \tau_1$ (because each $\tau_i \circ \tau_i$ is the identity permutation). Hence $\operatorname{sign}(\phi) = \operatorname{sign}(\phi^{-1})$. Also, summing over all $\phi \in S_n$ is the same as summing over all $\phi^{-1} \in S_n$, so we have

$$\begin{aligned} \det(A^{\mathrm{T}}) &= \sum_{\phi \in S_n} \operatorname{sign}(\phi) a_{1\phi^{-1}(1)} a_{2\phi^{-1}(2)} \dots a_{n\phi^{-1}(n)} \\ &= \sum_{\phi^{-1} \in S_n} \operatorname{sign}(\phi^{-1}) a_{1\phi^{-1}(1)} a_{2\phi^{-1}(2)} \dots a_{n\phi^{-1}(n)} = \det(A). \end{aligned}$$

□

If you find proofs like the above, where we manipulate sums of products, hard to follow, then it might be helpful to write it out in full in a small case, such as $n = 3$. Then

$$\begin{aligned}\det(A^{\mathrm{T}}) &= b_{11}b_{22}b_{33} - b_{11}b_{23}b_{32} - b_{12}b_{21}b_{33} \\ &\qquad + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} \\ &\qquad + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \\ &\qquad + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \\ &= \det(A).\end{aligned}$$

7.1.2 Proof of Theorem 6.2.1

Proof. (i) This is Proposition 6.1.4.

In the rest of the proof, we can supply only proofs for the rows. Indeed, by Theorem 7.1.1, we can apply column operations to A by transposing it, applying the corresponding row operations, and then re-transposing it. Thus, the proofs for columns follow from the proofs for rows.

(ii) To keep the notation simple, we shall suppose that we interchange the first two rows, but the same argument works for interchanging any pair of rows. Then if $B = (b_{ij})$, we have $b_{1j} = a_{2j}$ and $b_{2j} = a_{1j}$ for all j. Hence

$$\begin{aligned}\det(B) &= \sum_{\phi \in S_n} \operatorname{sign}(\phi) b_{1\phi(1)} b_{2\phi(2)} \dots b_{n\phi(n)} \\ &= \sum_{\phi \in S_n} \operatorname{sign}(\phi) a_{1\phi(2)} a_{2\phi(1)} a_{3\phi(3)} \dots a_{n\phi(n)}.\end{aligned}$$

For $\phi \in S_n$, let $\psi = \phi \circ (1, 2)$, so $\phi(1) = \psi(2)$ and $\phi(2) = \psi(1)$, and $\operatorname{sign}(\psi) = -\operatorname{sign}(\phi)$. Now, as ϕ runs through all permutations in S_n, so does ψ (but in a different order), so summing over all $\phi \in S_n$ is the same as summing over all $\psi \in S_n$. Hence

$$\begin{aligned}\det(B) &= \sum_{\phi \in S_n} -\operatorname{sign}(\psi) a_{1\psi(1)} a_{2\psi(2)} \dots a_{n\psi(n)} \\ &= \sum_{\psi \in S_n} -\operatorname{sign}(\psi) a_{1\psi(1)} a_{2\psi(2)} \dots a_{n\psi(n)} = -\det(A).\end{aligned}$$

(iii) Again to keep notation simple, assume that the equal rows are the first two. Using the same notation as in (ii), namely $\psi = \phi \circ (1, 2)$, the two elementary products:

$$a_{1\psi(1)} a_{2\psi(2)} \dots a_{n\psi(n)} \quad \text{and} \quad a_{1\phi(1)} a_{2\phi(2)} \dots a_{n\phi(n)}$$

are equal. This is because $a_{1\psi(1)} = a_{2\psi(1)}$ (first two rows equal) and $a_{2\psi(1)} = a_{2\phi(2)}$ (because $\phi(2) = \psi(1)$); hence $a_{1\psi(1)} = a_{2\phi(2)}$. Similarly $a_{2\psi(2)} = a_{1\phi(1)}$ and the two products differ by interchanging their first two terms. But $\operatorname{sign}(\psi) = -\operatorname{sign}(\phi)$ so the two corresponding signed products cancel each other out. Thus each signed product in $\det(A)$ cancels with another and the sum is zero.

(iv) Again, to simplify notation, suppose that we replace the second row $\mathbf{r}_2$ by $\mathbf{r}_2 + \lambda\mathbf{r}_1$ for some $\lambda \in K$. Then

$$\begin{aligned}\det(B) &= \sum_{\phi\in S_n} \operatorname{sign}(\phi) a_{1\phi(1)}(a_{2\phi(2)} + \lambda a_{1\phi(2)}) a_{3\phi(3)} \dots a_{n\phi(n)} \\ &= \sum_{\phi\in S_n} \operatorname{sign}(\phi) a_{1\phi(1)} a_{2\phi(2)} \dots a_{n\phi(n)} \\ &\quad + \lambda \sum_{\phi\in S_n} \operatorname{sign}(\phi) a_{1\phi(1)} a_{1\phi(2)} \dots a_{n\phi(n)}.\end{aligned}$$

Now the first term in this sum is $\det(A)$, and the second is $\lambda\det(C)$, where C is a matrix in which the first two rows are equal. Hence $\det(C) = 0$ by (iii), and $\det(B) = \det(A)$.

(v) Easy. Note that this holds even when the scalar $\lambda = 0$.

□

7.1.3 The determinant of a product

Let us start with an example:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} \text{ and } B = \begin{pmatrix} -1 & -1 \\ 2 & 0 \end{pmatrix}.$$

Note that $\det(A) = -4$, $\det(B) = 2$ and

$$AB = \begin{pmatrix} 3 & -1 \\ 1 & -3 \end{pmatrix} \text{ and } \det(AB) = -8 = \det(A)\det(B).$$

Bingo! Let us prove that this simple relationship holds in general.

Remember that there is no simple relationship between $\det(A + B)$ and $\det(A), \det(B)$. In our example,

$$A + B = \begin{pmatrix} 0 & 1 \\ 5 & 2 \end{pmatrix} \text{ and } \det(A + B) = -5 \neq \det(A) + \det(B).$$

Lemma 7.1.2. *If E is an $n \times n$ elementary matrix, and B is any $n \times n$ matrix, then* $\det(EB) = \det(E)\det(B)$.

Proof. E is one of the three types $E(n)^1_{\lambda,ij}$, $E(n)^2_{ij}$ or $E(n)^3_{\lambda,i}$, and multiplying B on the left by E has the effect of applying $\mathcal{R}1$, $\mathcal{R}2$ or $\mathcal{R}3$ to B, respectively. Hence, by Theorem 6.2.1, $\det(EB) = \det(B), -\det(B)$, or $\lambda\det(B)$, respectively. But by considering the special case $B = I_n$, we see that $\det(E) = 1, -1$ or λ, respectively, and so $\det(EB) = \det(E)\det(B)$ in all three cases. □

Recall Definition 2.2.9 of the singular map. An $n \times n$ matrix A is called *singular* if the map $L_A : \mathbf{x} \mapsto A\mathbf{x}$ is singular. By Corollary 2.2.8, this is equivalent to $\operatorname{rank}(A) < n$.

Theorem 7.1.3. *For an $n \times n$ matrix A, $\det(A) = 0$ if and only if A is singular.*

Proof. A can be reduced to Smith normal form by using elementary row and column operations. By Theorem 5.1.5, none of these operations affect the rank of A, and so they do not affect whether A is singular or not singluar. By Theorem 6.2.1, they do not affect whether $\det(A) = 0$ or $\det(A) \neq 0$. So we can assume that A is in Smith normal form.

Then $\operatorname{rank}(A)$ is the number of non-zero rows of A and, by Proposition 6.1.4, $\det(A) = a_{11}a_{22}\cdots a_{nn}$. Thus, $\det(A) = 0$ if and only if $a_{nn} = 0$ if and only if $\operatorname{rank}(A) < n$. □

Theorem 7.1.4. *For any two $n \times n$ matrices A and B, we have $\det(AB) = \det(A)\det(B)$.*

Proof. Suppose first that $\det(A) = 0$. Then we have $\operatorname{rank}(A) < n$ by Theorem 7.1.3. Let $T_1, T_2 : V \to V$ be linear maps corresponding to A and B, where $\dim(V) = n$. Then AB corresponds to T_1T_2 (by Theorem 3.4.2). By Corollary 2.2.8, $\operatorname{rank}(A) = \operatorname{rank}(T_1) < n$ implies that T_1 is not surjective. But then T_1T_2 cannot be surjective, so $\operatorname{rank}(T_1T_2) = \operatorname{rank}(AB) < n$. Hence $\det(AB) = 0$ so $\det(AB) = \det(A)\det(B)$.

On the other hand, if $\det(A) \neq 0$, then A is nonsingular, and hence invertible, so by Theorem 5.3.4, A is a product $E_1E_2\ldots E_r$ of elementary matrices E_i. Hence $\det(AB) = \det(E_1E_2\ldots E_rB)$. Now the result follows from the above lemma, because

$$\det(AB) = \det(E_1)\det(E_2\ldots E_rB) = \det(E_1)\det(E_2)\det(E_3\ldots E_rB) =$$
$$\det(E_1)\det(E_2)\ldots\det(E_r)\det(B) = \det(E_1E_2\ldots E_r)\det(B) = \det(A)\det(B).$$ □

Let us derive some immediate consequences.

Corollary 7.1.5. *If $A \in \mathbb{F}^{n,n}$ is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.*

Proof. $AA^{-1} = I_n$ implies that $1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$. □

Corollary 7.1.6. *If $A \in \mathbb{R}^{n,n}$ is orthogonal, then $\det(A) = \pm 1$.*

Proof. Since $\det(A) = \det(A^{\mathrm{T}})$ and $A^{\mathrm{T}} = A^{-1}$, $\det(A) = \det(A^{-1}) = \det(A)^{-1}$. Hence, $\det(A)^2 = 1$ and $\det(A) = \pm 1$. □

Corollary 7.1.7. *If $A, B \in \mathbb{F}^{n,n}$ are similar, then $\det(A) = \det(B)$ and, furthermore, $\det(A - x) = \det(B - x)$.*

Proof. As in Section 4.2.3, there exists an invertible matrix P with $B = PAP^{-1}$. Then

$$\det(B - xI_n) = \det(PAP^{-1} - xI_n) = \det(A - xI_n)P^{-1}) \overset{Th.\ 7.1.3}{=}$$
$$= \det(P)\det(A - xI_n)\det(P^{-1}) = \det(P)\det(P)^{-1}\det(A - xI_n) = \det(A - xI_n).$$

The proof for det is verbatim. □

7.1.4 Trace

An alternative proof of the first statement of Corollary 7.1.7 follows from the fact that $\det(A)$ is the free term of $\det(A-x)$. Indeed, since $\det(A-x)=\det(B-x)$, their free term are equal.

But there is another significant coefficient: *the trace* of the matrix A is the sum of its diagonal coefficients:

$$\mathrm{Tr}(a_{ij}) = \sum_i a_{ii}\,.$$

For example,

$$\mathrm{Tr}\,\begin{pmatrix}7\end{pmatrix} = 7,\ \ \mathrm{Tr}\begin{pmatrix}11 & 2\\ 3 & -3\end{pmatrix} = 11+(-3) = 8,\ \ \mathrm{Tr}\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{pmatrix} = 1+5+9 = 15.$$

Proposition 7.1.8. *If $A \in \mathbb{F}^{n,n}$ then*

$$\det(A-x) = \det(A) + ?x + \ldots + ?x^{n-2} + (-1)^{n-1}\,\mathrm{Tr}(A)x^{n-1} + (-1)^n x^n\,.$$

Proof. Note that we use ? for the remaining coefficient, about which we make no statement.

To understand the two highest terms, observe that any off-diagonal elementary product for computing $\det(A-x)$ has at most $n-2$ terms from the main diagonal: thus, it can contribute only to the coefficients of x^k with $k \le n-2$. Hence, the top two coefficients come from the diagonal elementary product:

$$(a_{11}-x)(a_{22}-x)\cdots(a_{nn}-x) = ? + \ldots + ?x^{n-2} + (a_{11}+\ldots+a_{nn})(-x)^{n-1} + (-x)^n =$$

$$= ? + \ldots + ?x^{n-2} + (-1)^{n-1}\,\mathrm{Tr}(A)x^{n-1} + (-1)^n x^n.$$

Finally, one gets the free term by setting $x=0$ and $\det(A-0)=\det(A)$. □

Corollary 7.1.9. *If $A, B \in \mathbb{F}^{n,n}$ are similar, then $\mathrm{Tr}(A) = \mathrm{Tr}(B)$.*

7.2 Geometric Properties

You will find the proofs in this section slightly different. We have to be less precise and more intuitive discussing the n-dimensional volume Vol_n. Given n vectors $\mathbf{v}_1, \ldots \mathbf{v}_n \in \mathbb{R}^n$, let us consider the parallelepiped spanned by them

$$P(\mathbf{v}_1,\ldots,\mathbf{v}_n) := \{\sum_{i=1}^n \beta_i \mathbf{v}_i \mid \beta_i \in [0,1]\}.$$

Our goal is to understand and to sketch a proof for the key formula for its volume

$$\mathrm{Vol}_n(P(\mathbf{v}_1,\ldots,\mathbf{v}_n)) = |\det(\mathbf{v}_1,\ldots,\mathbf{v}_n)|\,. \tag{7}$$

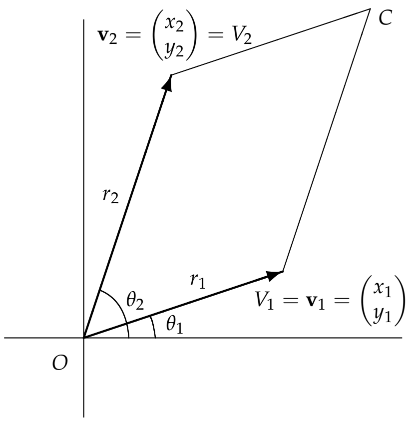
7.2.1 Volume in small dimensions

Let us use our intuitive understand of the volume in the dimensions $n \leq 3$ to prove Equation(7). By A we denote the matrix $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ with columns $\mathbf{v}_1, \ldots \mathbf{v}_n$.

If $n = 1$, then $\mathbf{v}_1 = (a)$ for some real number. Depending on whether a is positive or not, $P(\mathbf{v}_1) = [0, a]$ or $P(\mathbf{v}_1) = [-a, 0]$. Since $A = (a)$, $\mathrm{Vol}_1(P(\mathbf{v}_1)) = |a| = |\det(A)|$.

For $n = 2$, change the order of $\mathbf{v}_1, \mathbf{v}_2$ if necessary, so that $\det(A) \geq 0$. Now consider the positions of the vectors $V_1 = \mathbf{v}_1$, $V_2 = \mathbf{v}_2$ in the plane. Then, in the diagram below, $P(\mathbf{v}_1, \mathbf{v}_2)$ is the parallelogram OV_1CV_2. Its area is

$$r_1 r_2 \sin(\theta_2 - \theta_1) = r_1 r_2 (\sin\theta_2 \cos\theta_1 - \sin\theta_1 \cos\theta_2) = x_1 y_2 - x_2 y_1 = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}$$



Finally, consider $n = 3$. Let $A = (\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Let us expand by the first column

$$\det(A) = a_{11}c_{11} + a_{21}c_{21} + a_{31}c_{31} = \mathbf{v}_1 \bullet \begin{pmatrix} c_{11} \\ c_{21} \\ c_{31} \end{pmatrix}$$

where we are using the dot product. Note that the vector (c_{i1}) is the cross product $\mathbf{v}_2 \times \mathbf{v}_3$:

$$c_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = a_{22}a_{33} - a_{32}a_{23}\,,\ c_{21} = -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} = a_{32}a_{13} - a_{12}a_{33}\,,$$

$$c_{31} = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = a_{12}a_{23} - a_{22}a_{13}\,.$$

Thus, the determinant can be written as

$$\det(A) = \mathbf{v}_1 \bullet (\mathbf{v}_2 \times \mathbf{v}_3)\,,$$

a well-known (in either *MMM-2* or *Differential Equations*) formula for the 3D-volume.

7.2.2 Arbitrary dimension

To prove Equation (7), we must define Vol_n first. Instead, we can just use Equation (7) to define Vol_n. Why does it make sense?

Consider the case $\det A = 0$. Then $\mathbf{v}_1 \ldots \mathbf{v}_n$ span a proper subspace of $\mathbb{R}^n$. Thus, $P(\mathbf{v}_1 \ldots \mathbf{v}_n)$ collapses to a parallelepiped of smaller dimension. We can agree that $\mathrm{Vol}_n(P(\mathbf{v}_1 \ldots \mathbf{v}_n)) = 0$.

Consider the case $\det A \neq 0$. In this case we can use only column operations in Theorem 6.2.1 to compute the determinant. Let us just make certain that all the rules will be agreeable in terms of the nD volume:

(i) $\det(I_n) = 1$ means that the unit cube has volume 1.

(ii) If B results from A by applying $\mathcal{C}2$, then $\det(B) = -\det(A)$. This means swapping $\mathbf{v}_i \leftrightarrow \mathbf{v}_j$ does not change $P(\mathbf{v}_1 \ldots \mathbf{v}_n)$: its volume does not change.

(iv) If B results from A by applying $\mathcal{C}1$, then $\det(B) = \det(A)$. This means $\mathbf{v}_i \to \mathbf{v}_i + \lambda \mathbf{v}_j$ slides $P(\mathbf{v}_1 \ldots \mathbf{v}_n)$ in one direction: its volume does not change.

(v) If B results from A by applying $\mathcal{C}3$, then $\det(B) = \lambda \det(A)$. This means $\mathbf{v}_i \to \lambda \mathbf{v}_i$ stretches $P(\mathbf{v}_1 \ldots \mathbf{v}_n)$ in one direction: its volume gets multiplied by $|\lambda|$.

(iii) If A has two equal columns, then $\det(A) = 0$. This means that $P(\mathbf{v}_1 \ldots \mathbf{v}_n)$ has dimension less than n, then its volume is zero: indeed, column reduce; the last column is zero; add the first column to the last column; observe a parallelepiped of the same volume and this volume must be zero.

7.2.3 Orientation

Does Equation (7) work without the absolute value signs? Yes, but you need to make sense of the negative volume!

Consider the change of basis matrix P from a basis $\mathbf{v}_i$ to $\mathbf{f}_i$ in a vector space V over $\mathbb{R}$. We say that the bases have *the same orientation* if $\det(P) > 0$ and have *the opposite orientations* if $\det(P) < 0$.

Having the same orientation is the equivalence relation on the set of all bases in V. It has two equivalence classes. *An orientation of V* is the choice of one of these two orientation classes. The standard vector space $\mathbb{R}^n$ has a standard basis, hence comes with a standard orientation. The bases with the standard orientation are called *positively oriented*. The bases with the opposite to the standard orientation are called *negatively oriented*.

Now by an oriented nD parallelepiped, we understand a pair $(P(\mathbf{v}_1, \ldots, \mathbf{v}_n), \epsilon)$ where $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis and $\epsilon = \pm 1$ depending on whether $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is positively or negatively oriented. This allows us to define the oriented volume and have the following version of Equation (7):

$$\det(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \mathrm{Vol}_n^{or}(P(\mathbf{v}_1, \ldots, \mathbf{v}_n), \epsilon) := \epsilon \, \mathrm{Vol}_n(P(\mathbf{v}_1, \ldots, \mathbf{v}_n)) \,. \tag{8}$$

8 Eigenvalues and Eigenvectors

We are ready to make some progress on LO, cf. Section 4.2.3. Unlike LT, we are not going to solve it. In Multilinear Algebra next year, you will see a solution for the case $\mathbb{F} = \mathbb{C}$: cf. *Jordan Normal Form* on internet.

In this course, we shall settle the question which matrices are similar to a diagonal matrix. Such matrices are called *diagonalisable*.

8.1 Eigenvectors and eigenvalues

8.1.1 The definition of an eigenvector and an eigenvalue

In some books, eigenvectors and eigenvalues are called *characteristic vectors* and *characteristic roots*. We will not use these terms.

Definition 8.1.1. Let $T : V \to V$ be a linear operator, where V is a vector space over $\mathbb{F}$. Suppose that for some non-zero vector $\mathbf{v} \in V$ and some scalar $\lambda \in \mathbb{F}$, we have $T(\mathbf{v}) = \lambda\mathbf{v}$. Then $\mathbf{v}$ is called an *eigenvector* of T, and λ is called the *eigenvalue* of T corresponding to $\mathbf{v}$.

Note that the zero vector is **not** an eigenvector. (This would not be a good idea, because $T\mathbf{0} = \lambda\mathbf{0}$ for all λ.) However, the zero scalar $0_{\mathbb{F}}$ may sometimes be an eigenvalue.

Example. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 2a_1 \\ 0 \end{pmatrix}.$$

Then $T(\mathbf{e}_1) = 2\mathbf{e}_1$, so 2 is an eigenvalue and $\mathbf{e}_1$ is an eigenvector. Also $T(\mathbf{e}_2) = \mathbf{0} = 0\mathbf{e}_2$, so 0 is an eigenvalue and $\mathbf{e}_2$ is an eigenvector.

In this example, notice that in fact $\alpha\mathbf{e}_1$ and $\alpha\mathbf{e}_2$ are eigenvectors for any $\alpha \neq 0$. In fact, in general, it is easy to see that if $\mathbf{v}$ is an eigenvector of T, then so is $\alpha\mathbf{v}$ for any non-zero scalar α.

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be a basis of V, and let $A = (a_{ij})$ be the matrix of T with respect to this basis. As in Section 3.3.1, to each vector $\mathbf{v} \in V$, we associate the column vector $\underline{\mathbf{v}} \in \mathbb{F}^n$ such that by Proposition 3.3.1, for $\mathbf{u}, \mathbf{v} \in V$, we have $T(\mathbf{u}) = \mathbf{v}$ if and only if $A\underline{\mathbf{u}} = \underline{\mathbf{v}}$. In particular,

$$T(\mathbf{v}) = \lambda\mathbf{v} \iff A\underline{\mathbf{v}} = \lambda\underline{\mathbf{v}}$$

that leads to the next definition:

Definition 8.1.2. Let A be an $n \times n$ matrix over $\mathbb{F}$. Suppose that, for some non-zero column vector $\mathbf{x} \in \mathbb{F}^n$ and some scalar $\lambda \in \mathbb{F}$, we have $A\mathbf{x} = \lambda\mathbf{x}$. Then $\mathbf{x}$ is called an *eigenvector* of A, and λ is called the *eigenvalue* of A corresponding to $\mathbf{x}$.

It follows from Proposition 3.3.1 that if the matrix A corresponds to the linear map T, then λ is an eigenvalue of T if and only if it is an eigenvalue of A.

8.1.2 Role of characteristic polynomial

Theorem 8.1.3. *Let A be an $n \times n$ matrix. Then λ is an eigenvalue of A if and only if λ is a root of the characteristic polynomial* $\det(A - x)$.

Proof. Suppose that λ is an eigenvalue of A. Then $A\mathbf{x} = \lambda\mathbf{x}$ for some non-zero $\mathbf{x} \in \mathbb{F}^n$. This is equivalent to $A\mathbf{x} = \lambda I_n\mathbf{x}$, or $(A - \lambda I_n)\mathbf{x} = \underline{\mathbf{0}}$. But this says exactly that $\mathbf{x}$ is in the kernel of L_A. By Corollary 2.2.8, the matrix $A - \lambda I_n$ is singular. By Theorem 7.1.3,

$$0 = \det(A - \lambda I_n) = \det(A - x)|_{x=\lambda}.$$

Conversely, if λ is a root of $\det(A - x)$ then $\det(A - \lambda I_n) = 0$ and $A - \lambda I_n$ is singular. By Corollary 2.2.8, there exists a non-zero $\mathbf{x} \in \mathbb{F}^n$ with $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$, which is equivalent to $A\mathbf{x} = \lambda I_n\mathbf{x} = \lambda\mathbf{x}$, and so λ is an eigenvalue of A. $\square$

The above theorem gives us a method of calculating eigenvalues of a matrix. Once the eigenvalues are known, it is then straightforward to compute the corresponding eigenvectors.

Example 1. Let $A = \begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix}$. Then

$$\det(A - x) = \begin{vmatrix} 1-x & 2 \\ 5 & 4-x \end{vmatrix} = (1-x)(4-x) - 10 = x^2 - 5x - 6 = (x-6)(x+1).$$

Hence the eigenvalues of A are the roots of $(x-6)(x+1) = 0$; that is 6 and -1.

Let us now find the eigenvectors corresponding to the eigenvalue 6. We seek a non-zero column vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 6\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \quad \text{that is} \quad \begin{pmatrix} -5 & 2 \\ 5 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

We can take $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$ to be our eigenvector; or indeed any non-zero multiple of $\begin{pmatrix} 2 \\ 5 \end{pmatrix}$.

Similarly, for the eigenvalue -1, we want a non-zero column vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \quad \text{that is} \quad \begin{pmatrix} 2 & 2 \\ 5 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and we can take $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ to be our eigenvector.

Example 2. This example shows that the eigenvalues can depend on the field $\mathbb{F}$. Let

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad \text{Then} \quad \det(A - x) = \begin{vmatrix} -x & -1 \\ 1 & -x \end{vmatrix} = x^2 + 1,$$

so the characteristic equation is $x^2 + 1 = 0$. If $\mathbb{F} = \mathbb{R}$ (the real numbers) then this equation has no solutions, so there are no eigenvalues or eigenvectors. However, if $\mathbb{F} = \mathbb{C}$ (the complex numbers), then there are two eigenvalues i and $-i$. By a calculation similar to the last example, $\begin{pmatrix} -1 \\ i \end{pmatrix}$ and $\begin{pmatrix} 1 \\ i \end{pmatrix}$ are eigenvectors corresponding to i and $-i$ respectively.

By Corollary 7.1.7, similar matrices have the same characteristic equation. There are two important consequences of this. First, similar matrices have the same eigenvalues. Second, since different matrices corresponding to a linear operator T are all similar, they all have the same characteristic equation, so we can unambiguously refer to $\det(T)$, $\mathrm{Tr}(T)$ and $\det(T - x)$: they can be computed in any basis.

8.1.3 Diagonalisation

The linear operators and the matrices satisfying the next theorem are called *diagonalisable*.

Theorem 8.1.4. *Let $T : V \to V$ be a linear map. Then the matrix of T is diagonal with respect to some basis of V if and only if V has a basis consisting of eigenvectors of T.*

Equivalently, let A be an $n \times n$ matrix over $\mathbb{F}$. Then A is similar to a diagonal matrix if and only if the space $\mathbb{F}^n$ has a basis of eigenvectors of A.

Proof. The equivalence of the two statements follows directly from the correspondence between linear maps and matrices, and the corresponding definitions of eigenvectors and eigenvalues.

Suppose that the matrix $B = (a_{ij})$ of T is diagonal with respect to the basis $\mathbf{f}_1, \dots, \mathbf{f}_n$ of V. Recall from Section 3.3.1 that the images of the i-th basis vector of V is represented by the i-th column of B. But since B is diagonal, this column has the single non-zero entry a_{ii}. Hence $T(\mathbf{f}_i) = a_{ii}\mathbf{f}_i$, and so each basis vector $\mathbf{f}_i$ is an eigenvector of B.

Conversely, suppose that $\mathbf{f}_1, \dots, \mathbf{f}_n$ is a basis of V consisting entirely of eigenvectors of T. Then, for each i, we have $T(\mathbf{f}_i) = \lambda_i \mathbf{f}_i$ for some $\lambda_i \in \mathbb{F}$. But then the matrix of T with respect to this basis is the diagonal matrix $B = (a_{ij})$ with $a_{ii} = \lambda_i$ for each i. $\square$

There is one case where the eigenvalues can be written down immediately.

Proposition 8.1.5. *Suppose that the matrix A is upper (or lower) triangular. Then the eigenvalues of A are just the diagonal entries a_{ii} of A.*

Proof. In this case, $A - x$ is also upper triangular. By Corollary 6.1.4,

$$\det(A - x) = \prod_{i=1}^{n} (a_{ii} - x)$$

so the eigenvalues are the a_{ii}. $\square$

Example. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then A is upper triangular, so its only eigenvalue is 1. We can now see that A is not diagonalisable. Otherwise, we would have a basis of eigenvectors. Let P be the change of basis matrix. Since both eigenvalues are 1, $PAP^{-1} = I_2$ and $A = P^{-1}(PAP^{-1})P = P^{-1}I_2P = I_2$, a contradiction.

8.1.4 The case of distinct eigenvalues

This is the case where we can solve LO at this point.

Theorem 8.1.6. *Let $\lambda_1, \dots, \lambda_r$ be distinct eigenvalues of $T : V \to V$, and let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be corresponding eigenvectors. (So $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ for $1 \le i \le r$.) Then $\mathbf{v}_1, \dots, \mathbf{v}_r$ are linearly independent.*

Proof. Suppose that for some $a_1, \dots, a_r \in \mathbb{F}$ we have

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_r\mathbf{v}_r = \mathbf{0}.$$

Then, applying T to this equation gives

$$a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + \dots + a_r\lambda_r\mathbf{v}_r = \mathbf{0}.$$

Furthermore, applying T repeatedly to this equation gives

$$a_1\lambda_1^k\mathbf{v}_1 + a_2\lambda_2^k\mathbf{v}_2 + \dots + a_r\lambda_r^k\mathbf{v}_r = \mathbf{0}$$

for all k. Organising the first r such equations together into a matrix form yields

$$A \cdot \begin{pmatrix} a_1\mathbf{v}_1 \\ \vdots \\ a_r\mathbf{v}_r \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} \quad \text{where} \quad A = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_r \\ \vdots & & \vdots & & \vdots \\ \lambda_1^{r-1} & \lambda_2^{r-1} & \lambda_3^{r-1} & \dots & \lambda_r^{r-1} \end{pmatrix}$$

is the transpose of the Vandermonde matrix from Section 6.3.3. By Equation (6), it admits an inverse A^{-1}. Multiply by it to get

$$\begin{pmatrix} a_1\mathbf{v}_1 \\ \vdots \\ a_r\mathbf{v}_r \end{pmatrix} = A^{-1}A \cdot \begin{pmatrix} a_1\mathbf{v}_1 \\ \vdots \\ a_r\mathbf{v}_r \end{pmatrix} = A^{-1} \cdot \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}.$$

It follows that all a_i are equal to zero. □

Notice that the columns in the last proof are unusual: their entries are elements of V. It does not cause any issues with the proof!

Corollary 8.1.7. *If the linear map $T : V \to V$ (or equivalently the $n \times n$ matrix A) has n distinct eigenvalues, where $n = \dim(V)$, then T (or A) is diagonalisable.*

Proof. Under the hypothesis, there are n linearly independent eigenvectors, which form a basis of V by Corollary 1.4.15. The result follows from Theorem 8.1.4. □

Corollary 8.1.8. *Let $A, B \in \mathbb{F}^{n,n}$ and A has n distinct eigenvalues. Then A and B are similar if and only if they have the same eigenvalues.*

Example. Let

$$A = \begin{pmatrix} 4 & 5 & 2 \\ -6 & -9 & -4 \\ 6 & 9 & 4 \end{pmatrix} \qquad \text{so that} \qquad |A - x| = \begin{vmatrix} 4-x & 5 & 2 \\ -6 & -9-x & -4 \\ 6 & 9 & 4-x \end{vmatrix}.$$

To help evaluate this determinant, apply first the row operation $\mathbf{r}_3 \to \mathbf{r}_3 + \mathbf{r}_2$ and then the column operation $\mathbf{c}_2 \to \mathbf{c}_2 - \mathbf{c}_3$, giving

$$|A - x| = \begin{vmatrix} 4-x & 5 & 2 \\ -6 & -9-x & -4 \\ 0 & -x & -x \end{vmatrix} = \begin{vmatrix} 4-x & 3 & 2 \\ -6 & -5-x & -4 \\ 0 & 0 & -x \end{vmatrix},$$

and then expanding by the third row we get

$$-x\big((4-x)(-5-x) + 18\big) = -x(x^2 + x - 2) = -x(x+2)(x-1)$$

so the eigenvalues are 0, 1 and -2. Since these are distinct, we know from the above corollary that A can be diagonalised. In fact, the eigenvectors will be the new basis for the diagonal matrix, so we will calculate these.

In the following calculations, we will denote eigenvectors $\mathbf{v}_1$, etc. by $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, where x_1, x_2, x_3 need to be calculated by solving simultaneous equations.

For the eigenvalue $\lambda = 0$, an eigenvector $\mathbf{v}_1$ satisfies $A\mathbf{v}_1 = \mathbf{0}$, which gives the three equations:

$$4x_1 + 5x_2 + 2x_3 = 0; \qquad -6x_1 - 9x_2 - 4x_3 = 0; \qquad 6x_1 + 9x_2 + 4x_3 = 0.$$

The third is clearly redundant, and adding twice the first to the second gives $2x_1 + x_2 = 0$ and then we see that one solution is $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$.

For $\lambda = 1$, we want an eigenvector $\mathbf{v}_2$ with $A\mathbf{v}_2 = \mathbf{v}_2$, which gives the equations

$$4x_1 + 5x_2 + 2x_3 = x_1; \qquad -6x_1 - 9x_2 - 4x_3 = x_2; \qquad 6x_1 + 9x_2 + 4x_3 = x_3; \qquad \Longrightarrow$$

$$3x_1 + 5x_2 + 2x_3 = 0; \qquad -6x_1 - 10x_2 - 4x_3 = 0; \qquad 6x_1 + 9x_2 + 3x_3 = 0.$$

Adding the second and third equations gives $x_2 + x_3 = 0$ and then we see that a solution is $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.

Finally, for $\lambda = -2$, $A\mathbf{v}_3 = -2\mathbf{v}_3$ gives the equations

$$6x_1 + 5x_2 + 2x_3 = 0; \qquad -6x_1 - 7x_2 - 4x_3 = 0; \qquad 6x_1 + 9x_2 + 6x_3 = 0,$$

of which one solution is $\mathbf{v}_3 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$.

Now, if we change basis to $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, we should get the diagonal matrix with the eigenvalues $0, 1, -2$ on the diagonal. We can check this by direct calculation. Since the inverse of the basis change matrix P has the new basis vectors as its columns,

$$P^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ -2 & -1 & -2 \\ 3 & 1 & 2 \end{pmatrix} \Rightarrow P = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & 0 \\ -1 & -2 & -1 \end{pmatrix}, \; PAP^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

The last equality follows from Theorem 4.1.4. Check it directly by hand!

Warning. The converse of Corollary 8.1.7 is not true. If it turns out that there do not exist n distinct eigenvalues, then you cannot conclude from this that the matrix is not be diagonalisable. This is rather obvious, because the identity matrix has only a single eigenvalue, but it is diagonal already. Even so, this is one of the most common mistakes that students make.

If there are fewer than n distinct eigenvalues, then the matrix may or may not be diagonalisable. See the repeated real eigenvalue case in Section 8.2.3.

8.2 Applications

Applications of the eigenvalues are numerous. We only consider two.

8.2.1 Fibonacci and Lucas numbers

The famous Fibonacci numbers $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$ are defined as

$$F_0 = 0, F_1 = 1; \; F_m = F_{m-1} + F_{m-2}, \; m \geq 2.$$

The Lucas numbers $2, 1, 3, 4, 7, 11, 18, \ldots$ have the same recursive formula but a different start:

$$L_0 = 2, L_1 = 1; \; L_m = L_{m-1} + L_{m-2}, \; m \geq 2.$$

It helps to write both recursions in the vector form, using the vectors and the matrix

$$\mathbf{f}_n = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}, \; \mathbf{s}_n = \begin{pmatrix} L_{n+1} \\ L_n \end{pmatrix}, \; A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

The definitions become

$$\mathbf{f}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; \mathbf{f}_{n+1} = A\mathbf{f}_n \text{ and } \mathbf{s}_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{s}_{n+1} = A\mathbf{s}_n.$$

This gives us semi-useful answers such as $\mathbf{f}_n = A^n \mathbf{f}_0$ and a method to prove some of the identities. For instance,

$$\begin{pmatrix} L_{n+1} \\ L_n \end{pmatrix} = \mathbf{s}_n = A^n \begin{pmatrix} 2 \\ 1 \end{pmatrix} = A^n(\mathbf{f}_0 + \mathbf{f}_1) = A^n\mathbf{f}_0 + A^n\mathbf{f}_1 = \mathbf{f}_n + \mathbf{f}_{n+1} = \begin{pmatrix} F_{n+1} + F_{n+2} \\ F_n + F_{n+1} \end{pmatrix}$$

proves that $L_n = F_n + F_{n+1}$. Yet we can do much better by diagonalising the matrix A. Note that

$$\det(A - x) = \begin{pmatrix} 1-x & 1 \\ 1 & -x \end{pmatrix} = x^2 - x - 1 = (x - \lambda)(x - \lambda_\flat)$$

where the roots are the golden ratio and its companion

$$\lambda_1 = \lambda = \frac{1+\sqrt{5}}{2}, \;\; \lambda_2 = \lambda_\flat = (1 - \lambda) = \frac{1-\sqrt{5}}{2}.$$

Solving two linear systems $(A - \lambda_i)\mathbf{x} = \mathbf{0}$ gives eigenvectors

$$\mathbf{w}_1 = \begin{pmatrix} 1+\sqrt{5} \\ 2 \end{pmatrix}, \; \mathbf{w}_2 = \begin{pmatrix} 1-\sqrt{5} \\ 2 \end{pmatrix}.$$

Now express the starting vectors in the basis of eigenvalues

$$\mathbf{f}_0 = \frac{1}{2\sqrt{5}}(\mathbf{w}_1 - \mathbf{w}_2), \;\; \mathbf{s}_0 = \frac{1}{2}(\mathbf{w}_1 + \mathbf{w}_2)$$

and we get the general formulas

$$\mathbf{f}_n = \frac{1}{2\sqrt{5}}(A^n\mathbf{w}_1 - A^n\mathbf{w}_2) = \frac{1}{2\sqrt{5}}(\lambda^n\mathbf{w}_1 - \lambda_\flat^n\mathbf{w}_2), \, \mathbf{s}_0 = \frac{1}{2}(A^n\mathbf{w}_1 + A^n\mathbf{w}_2) = \frac{1}{2}(\lambda^n\mathbf{w}_1 + \lambda_\flat^n\mathbf{w}_2)\,.$$

Looking at the lower entry of the vectors, we arrive at Binet's formula:

$$F_n = \frac{1}{2\sqrt{5}}(2\lambda^n - 2\lambda_\flat^n) = \frac{\lambda^n - \lambda_\flat^n}{\sqrt{5}}, \;\; L_n = \frac{1}{2}(2\lambda^n + 2\lambda_\flat^n) = \lambda^n + \lambda_\flat^n\,.$$

8.2.2 Cayley-Hamilton theorem

This theorem will be proved for general n in Multilinear Algebra next year.

Theorem 8.2.1. *If $A \in \mathbb{F}^{2,2}$, then $A^2 = \mathrm{Tr}(A)A - \det(A)I_2$.*

Proof. Let us write $A = (a_{ij})$ and compute the right-hand side:

$$(a_{11} + a_{22})\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} - \begin{pmatrix} \det(A) & 0 \\ 0 & \det(A) \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{21}a_{12} & (a_{11} + a_{22})a_{12} \\ (a_{11} + a_{22})a_{21} & a_{22}^2 + a_{21}a_{12} \end{pmatrix} = A^2.$$

$\square$

This gives the following way to solve the LO for the 2×2 over any field $\mathbb{F}$.

Corollary 8.2.2. *If $A \in \mathbb{F}^{2,2}$ and $A \neq \alpha I_2$, then A is similar to $B = \begin{pmatrix} 0 & -\det(A) \\ 1 & \operatorname{Tr}(A) \end{pmatrix}$.*

Proof. Since $A \neq \alpha I_2$, we can find a nonzero vector $\mathbf{x} \in \mathbb{F}^2$ which is not an eigenvector. Then $A\mathbf{x}$ and $\mathbf{x}$ are linearly independent. Consider the basis $\mathbf{f}_1 = \mathbf{x}$, $\mathbf{f}_2 = A\mathbf{x}$. In this basis,

$$A\mathbf{f}_1 = A\mathbf{x} = \mathbf{f}_2$$

and, by Theorem 8.2.1,

$$A\mathbf{f}_2 = A(A\mathbf{x}) = A^2\mathbf{x} = (\operatorname{Tr}(A)A - \det(A)I_2)\mathbf{x} = \operatorname{Tr}(A)A\mathbf{x} - \det(A)\mathbf{x} = tr(A)\mathbf{f}_2 - \det(A)\mathbf{f}_1\,.$$

Thus, the matrix of L_A in this basis is B. □

8.2.3 All 2x2 real matrices

Let us solve LO in $\mathbb{R}^{2,2}$, using Corollary 8.2.2. Write $\det(A - x) = (\lambda_1 - x)(\lambda_2 - x)$. The polynomial $\det(A - x)$ is real, so we have the following three possibilities.

Repeated real eigenvalue $\lambda = \lambda_1 = \lambda_2 \in \mathbb{R}$. By Corollary 8.2.2, we have two similarity classes: scalar and non-scalar matrices. Again by Corollary 8.2.2, we can choose any matrix with the same trace and determinant as a representative. Good representatives are

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \text{ and } \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

For instance, let

$$A = \begin{pmatrix} 1 & 4 \\ -1 & -3 \end{pmatrix} \text{ so that } \det(A - x) = x^2 + 2x + 1 = (x+1)^2$$

so there is a single eigenvalue -1 with multiplicity 2. Since A is not scalar, A is similar to $B = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$. To find the change of basis explicitly, observe that Theorem 8.2.1 tells us that $(A + I_2)^2 = 0$. It follows that any vector $(A + I_2)\mathbf{x}$ will be either zero, or an eigenvector. Since the first column of $A + I_2$ is non-zero, we can choose

$$\mathbf{v}_2 := \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \mathbf{v}_1 := (A + I_2)\mathbf{v}_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \text{ so that } P^{-1} = (\mathbf{v}_1\ \mathbf{v}_2) = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}.$$

By Theorem 4.1.4, $PAP^{-1} = B$. Check it by hand!

Two distinct real eigenvalues $\lambda_1 \neq \lambda_2 \in \mathbb{R}$, By Corollary 8.2.2, we have a single similarity class and we can choose any matrix with the same trace and determinant as a representative. The good representative is

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

For instance, let

$$A = \begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix} \quad \text{so that} \quad \det(A - x) = x^2 - 2x - 3 = (x-3)(x+1)$$

so there are two distinct eigenvalues, 3 and -1. Associated eigenvectors are $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$, so we put $P^{-1} = \begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix}$ and get $PAP^{-1} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$ from Theorem 4.1.4. Check it by hand!

Two complex, nonreal eigenvalues $\lambda_1 = ae^{\phi i}$ and $\lambda_2 = ae^{-\phi i}$ with $a, \phi \in \mathbb{R}$, $a > 0$ and $\phi \notin 2\pi\mathbb{Z}$. By Corollary 8.2.2, we have a single similarity class and we can choose any matrix with the same trace and determinant as a representative. The good representative is

$$\begin{pmatrix} a\cos\phi & -a\sin\phi \\ a\sin\phi & a\cos\phi \end{pmatrix} = a\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}.$$

For instance, let

$$A = \begin{pmatrix} 1 & -1 \\ 7 & -3 \end{pmatrix} \quad \text{so that} \quad \det(A - x) = x^2 + 2x + 4 = (x - 2e^{2\pi i/3})(x - 2e^{-2\pi i/3})$$

so that A is similar to $B = \begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}$ (we choose $\phi = -2\pi/3$ for this). To choose the corresponding change of basis, we choose any $\mathbf{v}_1$, say $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The first column of B forces our hand for the second vector

$$A\mathbf{v}_1 = -\mathbf{v}_1 - \sqrt{3}\mathbf{v}_2 \;\Rightarrow\; \mathbf{v}_2 = \frac{-1}{\sqrt{3}}(A + I_2)\mathbf{v}_1 = \frac{-1}{\sqrt{3}}\begin{pmatrix} 2 & -1 \\ 7 & -2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{-1}{\sqrt{3}}\begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

It follows that $P^{-1} = \begin{pmatrix} 1 & -2/\sqrt{3} \\ 0 & -7/\sqrt{3} \end{pmatrix}$ and $PAP^{-1} = B$. Check it by hand!

9 Spectral Theorem and Singular Value Decomposition

In the final section we consider ET and EO, cf. Section 4.2.3. First we will solve EO for self-adjoint operators. Then we will solve ET in general.

9.1 Spectral theorem

We will study EO now. Hence, we will work with a euclidean space $V = (V, \tau)$ throughout.

9.1.1 Adjoint map

Proposition 9.1.1. *Let $T : V \to V$ be a linear operator. There exists a unique linear operator S such that for all $\mathbf{v}, \mathbf{w} \in V$*

$$\tau(T(\mathbf{v}), \mathbf{w}) = \tau(\mathbf{v}, S(\mathbf{w})) . \tag{9}$$

Proof. Let $\mathbf{f}_1, \dots, \mathbf{f}_n$ be an orthonormal basis. Equation (9) implies that

$$S(\mathbf{w}) = \sum_i \tau(\mathbf{f}_i, S(\mathbf{w}))\mathbf{f}_i = \sum_i \tau(T(\mathbf{f}_i), \mathbf{w})\mathbf{f}_i . \tag{10}$$

This implies that S, if exists, is unique.

The existence of S is clear too: equation (10) defines S. □

Definition 9.1.2. The unique linear map S in Proposition 9.1.1 is called the *adjoint* of T. We write this as T^*. We say T is *self-adjoint* if $T^* = T$.

Orthonormal bases are handy for computing adjoint maps, as clear from the next fact.

Proposition 9.1.3. *Fix an orthonormal basis $\mathbf{f}_1, \dots, \mathbf{f}_n$ of V. Let A be the matrix of a linear operator T in this basis. Then A^{T} is the matrix of the adjoint operator T^* in the same basis.*

Proof. Let $A = (a_{ij})$. Then $T(\mathbf{f}_i) = \sum_k a_{ki}\mathbf{f}_k$. Equation (10) implies that

$$T^*(\mathbf{f}_j) = \sum_i \tau(T(\mathbf{f}_i), \mathbf{f}_j)\mathbf{f}_i = \sum_{i,k} \tau(a_{ki}\mathbf{f}_k, \mathbf{f}_j)\mathbf{f}_i = \sum_i a_{ji}\mathbf{f}_i$$

which proves that A^{T} is the matrix of T^*. □

Example. Consider the standard euclidean space $\mathbb{R}^2$ and the linear map T given by $A = \begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix}$ in the standard basis. By Proposition 9.1.3, T^* given by A^{T}. Let us change the basis to

$$\mathbf{f}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \ \mathbf{f}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ P^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \ P = (P^{-1})^{-1} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix}.$$

In the new basis T and T^* are given by

$$P\begin{pmatrix} 1 & 2 \\ 5 & 4 \end{pmatrix}P^{-1} = \begin{pmatrix} 2 & 6 \\ 2 & 3 \end{pmatrix} \text{ and } P\begin{pmatrix} 1 & 5 \\ 2 & 4 \end{pmatrix}P^{-1} = \begin{pmatrix} -1 & 0 \\ 5 & 6 \end{pmatrix}$$

The matrices are no longer easily relatable because the basis is not orthonormal.

Corollary 9.1.4. *In an orthonormal basis, symmetric matrices correspond to self-adjoint operators.*

9.1.2 Eigenvalues and eigenvectors of self-adjoint operators

First, we'll warm up by proving a proposition which we'll need in the main proof.

Proposition 9.1.5. *Let A be an $n \times n$ real symmetric matrix. Then A has an eigenvalue in $\mathbb{R}$, and all complex eigenvalues of A lie in $\mathbb{R}$.*

Proof. The characteristic equation $\det(A - x) = 0$ is a polynomial equation of degree n in x. Since $\mathbb{C}$ is an algebraically closed field, it certainly has a root $\lambda \in \mathbb{C}$, which is an eigenvalue for A if we regard A as a matrix over $\mathbb{C}$. It remains to prove that any such λ lies in $\mathbb{R}$.

For a column vector $\mathbf{v}$ or matrix B over $\mathbb{C}$, we denote by $\overline{\mathbf{v}}$ or $\overline{B}$ the result of replacing all entries of $\mathbf{v}$ or B by their complex conjugates. Since the entries of A lie in $\mathbb{R}$, we have $\overline{A} = A$.

Let $\mathbf{v} \in \mathbb{C}^{n,1}$ be a complex eigenvector associated with λ. Then

$$A\mathbf{v} = \lambda\mathbf{v} \tag{11}$$

so, taking complex conjugates and using $\overline{A} = A$, we get

$$A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}. \tag{12}$$

Transposing (11) and using $A^{\mathrm{T}} = A$ gives

$$\mathbf{v}^{\mathrm{T}}A = \lambda\mathbf{v}^{\mathrm{T}}, \tag{13}$$

so by equations (12) and (13), we have

$$\lambda\mathbf{v}^{\mathrm{T}}\overline{\mathbf{v}} = \mathbf{v}^{\mathrm{T}}A\overline{\mathbf{v}} = \overline{\lambda}\mathbf{v}^{\mathrm{T}}\overline{\mathbf{v}}.$$

Note that $\mathbf{v} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$ is non-zero, since eigenvectors are non-zero. Hence,

$$\mathbf{v}^{\mathrm{T}}\overline{\mathbf{v}} = a_1\overline{a_1} + \ldots + a_n\overline{a_n} = |a_1|^2 + \ldots + |a_n|^2$$

is a positive real number. Thus, $\lambda = \overline{\lambda}$ and $\lambda \in \mathbb{R}$. □

We are ready to prove the spectral theorem. It solves both LO and EO for self-adjoint operators and symmetric real matrices.

Theorem 9.1.6. *Let V be a euclidean space of dimension n. If $T : V \to V$ is a selfadjoint linear operator, there is an orthonormal basis $\mathbf{f}_1, \dots, \mathbf{f}_n$ of V consisting of eigenvectors of T.*

An $n \times n$ real symmetric matrix is orthogonally similar to a diagonal matrix.

Proof. By Corollary 9.1.4, the statements are equivalent. Let us prove the first statement.

We proceed by induction on $n = \dim V$. If $n = 0$ there is nothing to prove, so let us assume the proposition holds for $n - 1$.

By proposition 9.1.5, T has an eigenvalue in $\lambda_1 \in \mathbb{R}$. Let $\mathbf{v} \in V$ be a corresponding eigenvector in V. Then $\mathbf{f}_1 = \mathbf{v}/|\mathbf{v}|$ is also an eigenvector, and $|\mathbf{f}_1| = 1$.

We consider the space $W = \mathbf{v}^\perp := \{\mathbf{w} \in V \mid \tau(\mathbf{w}, \mathbf{f}_1) = 0\}$. Since W is the kernel of a surjective linear map

$$V \longrightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \tau(\mathbf{x}, \mathbf{v}),$$

it is a subspace of V of dimension $n - 1$. We claim that T maps W into itself. Indeed, suppose $\mathbf{w} \in W$; we want to show that $T(\mathbf{w}) \in W$ also. This follows from T being self-adjoint:

$$\tau(T(\mathbf{w}), \mathbf{v}) = \tau(\mathbf{w}, T(\mathbf{v})) = \tau(\mathbf{w}, \alpha_1 \mathbf{v}) = \alpha_1 \tau(\mathbf{w}, \mathbf{v}) = 0.$$

Now we know that T maps W into itself. Moreover, W is a euclidean space of dimension $n - 1$, so we may apply the induction hypothesis to the restriction of T to W. This gives us an orthonormal basis $\mathbf{f}_2, \dots, \mathbf{f}_n$ of W consisting of eigenvectors of T. Then the set $\mathbf{f}_1, \dots, \mathbf{f}_n$ is an orthonormal basis of V, which also consists of eigenvectors of T. $\square$

9.1.3 Orthogonality of eigenvectors

The next property, implicit in the proof of the spectral theorem 9.1.6, is useful when calculating examples. It helps us to write down more vectors in the final orthonormal basis immediately.

Proposition 9.1.7. *Let A be a real symmetric matrix, and let λ_1, λ_2 be two distinct eigenvalues of A, with corresponding eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$. Then $\mathbf{v}_1 \bullet \mathbf{v}_2 = 0$.*

Proof. We will use the equality $\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1^\mathrm{T} \mathbf{v}_2$. We have

$$A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2 \mathbf{v}_2. \tag{14}$$

The trick is now to look at the expression $\mathbf{v}_1^\mathrm{T} A \mathbf{v}_2$. On the one hand, by (14) we have

$$\mathbf{v}_1^\mathrm{T} A \mathbf{v}_2 = \mathbf{v}_1 \bullet (A\mathbf{v}_2) = \mathbf{v}_1 \bullet (\lambda_2 \mathbf{v}_2) = \lambda_2 (\mathbf{v}_1 \bullet \mathbf{v}_2). \tag{15}$$

On the other hand, $A^\mathrm{T} = A$, so $\mathbf{v}_1^\mathrm{T} A = \mathbf{v}_1^\mathrm{T} A^\mathrm{T} = (A\mathbf{v}_1)^\mathrm{T}$, so using (14) again we have

$$\mathbf{v}_1^\mathrm{T} A \mathbf{v}_2 = (A\mathbf{v}_1)^\mathrm{T} \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1^\mathrm{T}) \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1) \bullet \mathbf{v}_2) = \lambda_1 (\mathbf{v}_1 \bullet \mathbf{v}_2). \tag{16}$$

Comparing (15) with (16), we have $(\lambda_2 - \lambda_1)(\mathbf{v}_1 \bullet \mathbf{v}_2) = 0$. Since $\lambda_2 - \lambda_1 \neq 0$ by assumption, we have $\mathbf{v}_1 \bullet \mathbf{v}_2 = 0$. $\square$

Corollary 9.1.8. *Let λ_1, λ_2 be two distinct eigenvalues of a self-adjoint linear operator T, with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2$. Then $\tau(\mathbf{v}_1, \mathbf{v}_2) = 0$.*

Example. Let $n = 2$ and let

$$A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \Rightarrow \det(A - x) = (1-x)^2 - 9 = x^2 - 2x - 8 = (x-4)(x+2)\,.$$

Hence, the eigenvalues of A are 4 and -2. Solving $A\mathbf{v} = \lambda\mathbf{v}$ for $\lambda = 4$ and -2, we find corresponding eigenvectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Proposition 9.1.7 tells us that these vectors are orthogonal to each other (which we can of course check directly), so if we divide them by their lengths to give vectors of length 1, giving

$$\mathbf{f}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},\ \mathbf{f}_2 = \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix},\ P^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},\ P = (P^{-1})^{\mathrm{T}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},$$

an orthonormal basis consisting of eigenvectors of A, which is what we want. We could compute the inverse by transposition because P is orthogonal. Check that $PAP^{-1} = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}$.

Example. Let $n = 3$ and

$$A = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 6 & -2 \\ 1 & -2 & 3 \end{pmatrix}.$$

Then, expanding by the first row,

$$\begin{aligned} \det(A - x) &= (3-x)(6-x)(3-x) - 4(3-x) - 4(3-x) + 4 + 4 - (6-x) \\ &= -x^3 + 12x^2 - 36x + 32 = (2-x)(x-8)(x-2), \end{aligned}$$

so the eigenvalues are 2 (repeated) and 8. For the eigenvalue 8, if we solve $A\mathbf{v} = 8\mathbf{v}$ then we find a solution $\mathbf{x} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$. Since 2 is a repeated eigenvalue, we need two corresponding eigenvectors, which must be orthogonal to each other. By Proposition 9.1.7, they are orthogonal to $\mathbf{x}$, which allows us to write an acceptable solution:

$$\mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},\ \mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Check by hand that these are eigenvectors with eigenvalues 2! To get an orthonormal basis, we just need to divide by their lengths:

$$\mathbf{f}_1 = \frac{1}{\sqrt{6}}\mathbf{x} = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix},\ \mathbf{f}_2 = \frac{1}{\sqrt{2}}\mathbf{y} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{pmatrix},\ \mathbf{f}_3 = \frac{1}{\sqrt{3}}\mathbf{x} = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix},\quad P^{-1} = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

Orthogonality gives us $P = (P^{-1})^{\mathrm{T}}$ and we know that $PA^{-1}P$ is the diagonal matrix with entries 8, 2, 2. Check it by hand!

9.2 Singular value decomposition

To address ET, consider a linear map $T : V \to W$ between euclidean spaces. We already know that we can choose bases in V and W such that the matrix of T is in the Smith normal form $\left(\begin{array}{c|c} I_n & 0 \\ \hline 0 & 0 \end{array}\right)$ where $n = \mathrm{rank}(T)$. Its refinement is known as *the singular value decomposition*.

9.2.1 Motivational example

We want to solve ET for a linear operator

$$T = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 .$$

Its eigenvalues are 1 and 2. Let us call the corresponding eigenvectors $\mathbf{f}_1$ and $\mathbf{f}_2$. We will also need its inverse:

$$\mathbf{f}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \mathbf{f}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \ T^{-1} = \begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix} .$$

We start by computing the image $T(\mathcal{S}_2)$ of the unit sphere $\mathcal{S}_2$:

$$T(\mathcal{S}_2) \ = \ \{\mathbf{x} \in \mathbb{R}^2 \ | \ \|T^{-1}(\mathbf{x})\| = 1\} \ = \ \{\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \ | \ \| \begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \| = 1\} .$$

The length of the vector in the last expression is

$$(x + \frac{1}{2}y)^2 + (\frac{1}{2}y)^2 = x^2 + xy + \frac{1}{2}y^2$$

so that $T(\mathcal{S}_2)$ is an ellipse given by the equation $x^2 + xy + \frac{1}{2}y^2 = 1$. Let us parametrise it, using polar coordinates:

$$T(\mathcal{S}_2) \ = \ \{T \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \in \mathbb{R}^2 \ | \ \theta \in \mathbb{R}\} \ = \ \{\begin{pmatrix} \cos\theta - \sin\theta \\ 2\sin\theta \end{pmatrix} \in \mathbb{R}^2 \ | \ \theta \in \mathbb{R}\} .$$

Its semiaxes are directions of the largest and the smallest vectors on the ellipse. We can find them by elementary trigonometry, using $\alpha = \arcsin(1/\sqrt{5})$ (and $\cos\alpha = 2/\sqrt{5}$):

$$\begin{aligned}
(\cos\theta - \sin\theta)^2 + (2\sin\theta)^2 = \ & \cos^2\theta - 2\cos\theta\sin\theta + 5\sin^2\theta = \\
= 1 - \sin(2\theta) + 4\sin^2\theta = & \ 1 - \sin(2\theta) + 2(1 - \cos(2\theta)) = \\
= 3 - (\sin(2\theta) + 2\cos(2\theta)) = & \ 3 - \sqrt{5}(\frac{1}{\sqrt{5}}\sin(2\theta) + \frac{2}{\sqrt{5}}\cos(2\theta)) = \\
= 3 - \sqrt{5}(\sin\alpha\sin(2\theta) + \cos\alpha\cos(2\theta)) = & \ 3 - \sqrt{5}\cos(2\theta - \alpha) .
\end{aligned}$$

Thus, we get the largest vector when $\cos(2\theta - \alpha) = -1$ and it has length $\sqrt{3 + \sqrt{5}}$. We can pick $2\theta - \alpha = \pi$, which means $\theta = (\pi + \alpha)/2$. The key vectors are

$$\mathbf{v}_1 = \begin{pmatrix} \cos\frac{\pi+\alpha}{2} \\ \sin\frac{\pi+\alpha}{2} \end{pmatrix} = \begin{pmatrix} -\sin\frac{\alpha}{2} \\ \cos\frac{\alpha}{2} \end{pmatrix} = \begin{pmatrix} -\sqrt{\frac{\sqrt{5}-2}{2\sqrt{5}}} \\ \sqrt{\frac{2+\sqrt{5}}{2\sqrt{5}}} \end{pmatrix}$$

and

$$\mathbf{u}_1 = \frac{1}{T(\mathbf{v}_1)} T(\mathbf{v}_1) = \frac{1}{\sqrt{3+\sqrt{5}}} \begin{pmatrix} \cos\frac{\pi+\alpha}{2} - \sin\frac{\pi+\alpha}{2} \\ 2\sin\frac{\pi+\alpha}{2} \end{pmatrix} = \ldots \text{Good Luck!}$$

Similarly, we get the smallest vector when $\sin(2\theta+\alpha) = 1$ and it has length $\sqrt{3-\sqrt{5}}$. We can repeat this calculation but we know for certain that the corresponding vectors $\mathbf{v}_2$ and $\mathbf{u}_2$ are obtained by rotating $\mathbf{v}_1$ and $\mathbf{u}_1$ by $\pi/2$ clockwise:

$$\mathbf{v}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mathbf{v}_1 = \begin{pmatrix} -\sqrt{\frac{2+\sqrt{5}}{2\sqrt{5}}} \\ -\sqrt{\frac{\sqrt{5}-2}{2\sqrt{5}}} \end{pmatrix} \quad \text{and} \quad \mathbf{u}_2 = \frac{1}{\sqrt{3-\sqrt{5}}} T(\mathbf{v}_2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mathbf{v}_2 \,.$$

The upshot of all this is not that the precise final forms of the vectors look hard (this tend to happen with SVD) but that $\mathbf{v}_1, \mathbf{v}_2$ is a good orthonormal basis of the domain of T and $\mathbf{u}_1, \mathbf{u}_2$ is a good orthonormal basis of the range of T. In these bases, the matrix of T is

$$\begin{pmatrix} \sqrt{3+\sqrt{5}} & 0 \\ 0 & \sqrt{3-\sqrt{5}} \end{pmatrix}.$$

9.2.2 Main theorem

Theorem 9.2.1. *(SVD for linear maps) Suppose $T : (V, \tau_V) \to (W, \tau_W)$ is a linear map of rank n between euclidean spaces. Then there exist unique positive numbers $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_n > 0$, called the singular values of T, and orthonormal bases of V and W such that the matrix of T with respect to these bases is*

$$\left(\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right) \quad \textit{where} \quad D = \begin{pmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_n \end{pmatrix}.$$

Proof. Choose orthonormal bases of V and W. Write T as a matrix A in these bases. Consider an operator $S : V \to V$ given by the matrix $A^{\mathrm{T}}A$ in the same basis:

$$\mathbf{v} \mapsto S(\mathbf{v}) \;\text{ where }\; \underline{S(\mathbf{v})} = A^{\mathrm{T}}A\underline{\mathbf{v}}\,.$$

Since $(A^{\mathrm{T}}A)^{\mathrm{T}} = A^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}} = A^{\mathrm{T}}A$, the matrix $A^{\mathrm{T}}A$ is symmetric and the operator S is selfadjoint. By Theorem 9.1.6, we can choose an orthonormal basis $\mathbf{f}_1, \ldots, \mathbf{f}_k$ of V, consisting of eigenvectors of S. Then $S(\mathbf{f}_i) = a_i\mathbf{f}_i$ where a_i are the eigenvalues of A.

Observe that the eigenvalues are non-negative:

$$a_i = \tau_V(\mathbf{f}_i, S\mathbf{f}_i) = \underline{\mathbf{f}_i}^{\mathrm{T}}(A^{\mathrm{T}}A\underline{\mathbf{f}_i}) = (A\underline{\mathbf{f}_i})^{\mathrm{T}}(A\underline{\mathbf{f}_i}) = \tau_W(T\mathbf{f}_i, T\mathbf{f}_i) \geq 0\,.$$

In fact, we have computed the length of the image vectors: $|T\mathbf{f}_i| = \sqrt{a_i}$. A similar trick shows that $T(\mathbf{f}_i)$ is orthogonal to $T(\mathbf{f}_j)$ for $i \neq j$:

$$\tau_W(T\mathbf{f}_i, T\mathbf{f}_j) = (A\underline{\mathbf{f}_i})^{\mathrm{T}}(A\underline{\mathbf{f}_j}) = \underline{\mathbf{f}_i}^{\mathrm{T}}(A^{\mathrm{T}}A\underline{\mathbf{f}_j}) = \tau_V(\mathbf{f}_i, S\mathbf{f}_j) = a_j\tau_V(\mathbf{f}_i, \mathbf{f}_j) = 0\,.$$

Now reorder the orthonormal basis $\mathbf{f}_1, \ldots, \mathbf{f}_k$ so that

$$a_1 \geq \ldots \geq a_t > 0,\ a_{t+1} = \ldots = a_k = 0\,,$$

define $\gamma_i = \sqrt{a_i}$ for $i \leq t$ and define an orthonormal basis $\mathbf{h}_1, \ldots, \mathbf{h}_m$ of W by

$$\mathbf{h}_i := \frac{1}{\gamma_i} T(\mathbf{f}_i) \ \text{ for all } \ i \leq t$$

and picking some $\mathbf{h}_{t+1}, \ldots, \mathbf{h}_m$, which can be done by the Gram-Schmidt process (Theorem 3.2.6).

Since $T(\mathbf{f}_i) = \gamma_i \mathbf{h}_i$ for $i \leq t$ and $T(\mathbf{f}_j) = 0$ for $j > t$. Thus, $t = \mathrm{rank}(T) = n$ and the matrix of T with respect to these bases has the required form.

It remains to prove uniqueness of the singular values. Suppose we have orthonormal bases $\mathbf{f}'_1, \ldots, \mathbf{f}'_k$ of V and $\mathbf{h}'_1, \ldots, \mathbf{h}'_m$ of W, in which T is represented by a matrix

$$\left(\begin{array}{c|c} B & 0 \\ \hline 0 & 0 \end{array}\right), \text{ where } B = \begin{pmatrix} \beta_1 & & \\ & \ddots & \\ & & \beta_t \end{pmatrix}, \ \beta_1 \geq \ldots \geq \ldots \beta_t > 0.$$

First, $t = \mathrm{rank}(T) = n$. Then β_i^2 is an eigenvalue of S (we write in the original basis in the next calculation):

$$S(\mathbf{f}'_i) = \sum_j \tau_V(\mathbf{f}'_j, S(\mathbf{f}'_i))\mathbf{f}'_j = \sum_j \tau_V(\mathbf{f}'_j, S(\mathbf{f}'_i))\mathbf{f}'_j = \sum_j \underline{\mathbf{f}'_j}^{\mathrm{T}}(A^{\mathrm{T}} A \underline{\mathbf{f}'_i})\mathbf{f}'_j =$$

$$= \sum_j (A\underline{\mathbf{f}'_j})^{\mathrm{T}}(A\underline{\mathbf{f}'_i})\mathbf{f}'_j = \sum_j \tau_W(T(\mathbf{f}'_j), T(\mathbf{f}'_i))\mathbf{f}'_j = \sum_j \tau_W(\beta_j \mathbf{h}'_j, \beta_i \mathbf{h}'_i)\mathbf{f}'_j = \beta_i^2 \mathbf{f}'_i\,.$$

By the uniqueness of eigenvalues, $\gamma_i = \beta_i$. □

Before we proceed with some examples, all on the standard euclidean spaces $\mathbb{R}^n$, let us restate the SVD for matrices:

Corollary 9.2.2. *(SVD for matrices) Given any real $k \times m$ matrix A, there exist unique singular values $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_n > 0$ and (non-unique) orthogonal matrices P and Q such that*

$$PAQ^{-1} = PAQ^{\mathrm{T}} = \widetilde{D} \quad \textit{where} \quad \widetilde{D} = \left(\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array}\right) \quad \textit{and} \quad D = \begin{pmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_n \end{pmatrix}.$$

Note that terminologically SVD needs to **decompose** A. Thus, by the SVD people often understand the presentation of A as

$$A = P^{\mathrm{T}} \widetilde{D} Q\,.$$

This solves EO: $\widetilde{D}$ is orthogonally equivalent to A. Furthermore, matrices of the form like $\widetilde{D}$ are normal forms of the orthogonal equivalence classes.

Example. Consider the linear map $\mathbb{R}^2 \to \mathbb{R}^2$, given by the symmetric matrix $A = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$, in the example after Corollary 9.1.8. There we found the orthogonal matrix

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \text{ with } PAP^{\mathrm{T}} = \begin{pmatrix} 4 & 0 \\ 0 & -2 \end{pmatrix}.$$

This is not the SVD of A because the diagonal matrix contains a negative entry. To get to the SVD we just need to pick different bases for the domain and the range: the columns $\mathbf{f}_1$, $\mathbf{f}_2$ of P^{-1} can still be a basis of the domain, while the basis of the range could become $\mathbf{f}_1$, $-\mathbf{f}_2$. This is the SVD:

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}, \; Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \; PAQ^{\mathrm{T}} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

The same method works for any symmetric matrix: the SVD is just orthogonal diagonalisation with additional care needed for signs. If the matrix is not symmetric, we need to follow the proof of Theorem 9.2.1 during the calculation.

Example. Consider a linear map $T : \mathbb{R}^3 \to \mathbb{R}^2$, given by $A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$. The matrix of S in the standard basis is

$$A^T A = \begin{pmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{pmatrix} \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

The eigenvalues of this matrix are 360, 90 and 0. Hence the singular values of A are

$$\gamma_1 = \sqrt{360} = 6\sqrt{10} \geq \gamma_2 = \sqrt{90} = 3\sqrt{10}.$$

At this stage we are assured of the existence of orthogonal matrices P and Q such that

$$PAQ^{-1} = PAQ^{\mathrm{T}} = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}.$$

To find the orthogonal matrices we need to find eigenvectors of $A^T A$:

$$\mathbf{e}_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}, \; \mathbf{e}_2 = \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}, \; \mathbf{e}_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}$$

and then their images under A:

$$\mathbf{f}_1 = \frac{1}{6\sqrt{10}} A\mathbf{e}_1 = \begin{pmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}, \; \mathbf{f}_2 = \frac{1}{3\sqrt{10}} A\mathbf{e}_2 = \begin{pmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{pmatrix}.$$

Hence, the orthogonal matrices are

$$P^{-1} = \begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix}, \; Q^{-1} = \begin{pmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{pmatrix}.$$

9.2.3 Geometric intuition

Let us summarize the geometric intuition about linear maps on euclidean spaces. First, in a euclidean space we always choose an orthonormal basis. Then we write all the matrices in an orthonormal basis or two orthonormal bases.

The first type of useful matrices are orthogonal matrices $P \in \mathbb{R}^{n,n}$. They are cool algebraically because calculation of their inverses: $P^{-1} = P^{\mathrm{T}}$. They have two geometric meanings:

- they represent orthogonal linear operators, those operators that preserve distances and angles,
- they represent the basis changes between orthonormal bases, the only bases changes that we allow ourselves.

The second type of useful matrices are symmetric matrices $S \in \mathbb{R}^{n,n}$. They are easy to recognize: $S = S^{\mathrm{T}}$. They have two geometric meanings as well:

- they represent self-adjoint linear operators,
- they represent the operators that can be written as diagonal operators in an orthonormal basis.

Finally, to understand SVD, we need two figures in $\mathbb{R}^n$: the unit sphere $\mathcal{S}$ and an ellipsoid $\mathcal{E}$

$$\mathcal{S}_n := \{(x_i) \mid x_1^2 + \ldots + x_n^2 = 1\},\ \mathcal{E}_n := \{(x_i) \mid a_1 x_1^2 + \ldots + a_n x_n^2 = 1\}$$

where we require all a_i to be positive. We always write these equations in coordinates in some orthonormal basis. There is only one sphere: any basis will produce the same sphere. There are infinitely many ellipsoids: we can choose different bases and different constants a_i.

Let us now consider the geometric meaning of SVD of a linear operator $T : V \to W$:

- T takes the unit sphere $\mathcal{S}_n$ to an ellipsoid $\mathcal{E}_k$ where $n = \dim V$ and $k = \mathrm{rank}(T)$.
- If $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_k > 0$ are the singular values of T, the ellipsoid $\mathcal{E}_k$ can be given by the equation
$$\frac{1}{\gamma_1^2} x_1^2 + \frac{1}{\gamma_2^2} x_2^2 + \ldots + \frac{1}{\gamma_k^2} x_k^2 = 1$$
inside the image of T.
- The largest singular value γ_1 is the half-girth of the ellipsoid $\mathcal{E}_k$, the length of its major semi-axis.
- The largest singular value γ_1 can be thought of a measure of how "large" T is. More precisely, it is *the operator norm* of T, which you will study next year.
- The last statement can be expressed as the following useful formula:

$$\gamma_1 = \sup \left\{ \frac{|T(\mathbf{x})|}{|\mathbf{x}|} \mid \mathbf{x} \in V \setminus \{0\} \right\} = \max\{|T(\mathbf{x})| \mid \mathbf{x} \in \mathcal{S}_n \subset V\}.$$