Linear Algebra

Min Yan

August 8, 2017

Contents

1 Vector Space

Linear algebra describes the most basic mathematical structure. The key object in linear algebra is vector space, which is characterised by the operations of addition and scalar multiplication. The key relation between objects is linear transformation, which is characterised by preserving the two operations. The key example is Euclidean space, which is the model for all finite dimensional vector spaces.

The theory of linear algebra can be developed over any field, which is a "number system" where the usual four arithmetic operations are defined. In fact, a more general theory (of modules) can be developed over any ring, which is a system where only the addition, subtraction and multiplication (no division) are defined. Since the linear algebra of real vector spaces already reflects most of the true spirit of linear algebra, we will concentrate on real vector spaces until Chapter **??**.

1.1 Definition

Definition 1.1. A *(real) vector space* is a set V, together with the operations of addition and scalar multiplication

$$\vec{u}+\vec{v}\colon V\times V\to V,\quad a\vec{u}\colon \mathbb{R}\times V\to V,$$

such that the following are satisfied.

1. Commutativity: $\vec{u}+\vec{v}=\vec{v}+\vec{u}$.
2. Associativity for Addition: $(\vec{u}+\vec{v})+\vec{w}=\vec{u}+(\vec{v}+\vec{w})$.
3. Zero: There is an element $\vec{0}\in V$ satisfying $\vec{u}+\vec{0}=\vec{u}=\vec{0}+\vec{u}$.
4. Negative: For any $\vec{u}$, there is $\vec{v}$ (to be denoted $-\vec{u}$), such that $\vec{u}+\vec{v}=\vec{0}=\vec{v}+\vec{u}$.
5. One: $1\vec{u}=\vec{u}$.
6. Associativity for Scalar Multiplication: $(ab)\vec{u}=a(b\vec{u})$.
7. Distributivity in $\mathbb{R}$: $(a+b)\vec{u}=a\vec{u}+b\vec{u}$.
8. Distributivity in V: $a(\vec{u}+\vec{v})=a\vec{u}+a\vec{v}$.

Due to the associativity of addition, we may write $\vec{u}+\vec{v}+\vec{w}$ and even longer expressions without ambiguity.

Example 1.1. The zero vector space $\{\vec{0}\}$ consists of single element $\vec{0}$. This leaves no choice for the two operations: $\vec{0}+\vec{0}=\vec{0}$, $c\vec{0}=\vec{0}$. It can be easily verified that all eight axioms hold.

Example 1.2. The *Euclidean space* $\mathbb{R}^n$ is the set of n-tuples

$$\vec{x}=(x_1,x_2,\dots,x_n),\quad x_i\in\mathbb{R}.$$

The i-th number x_i is the i-th *coordinate* of the vector. The Euclidean space is a vector space with coordinate wise addition and scalar multiplication

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n),$$
$$a(x_1, x_2, \dots, x_n) = (ax_1, ax_2, \dots, ax_n).$$

Geometrically, we often express a vector in Euclidean space as a dot or an arrow from the origin $\vec{0} = (0, 0, \dots, 0)$ to the dot. Figure 1.1 shows that the addition is described by parallelogram, and the scalar multiplication is described by stretching and shrinking.

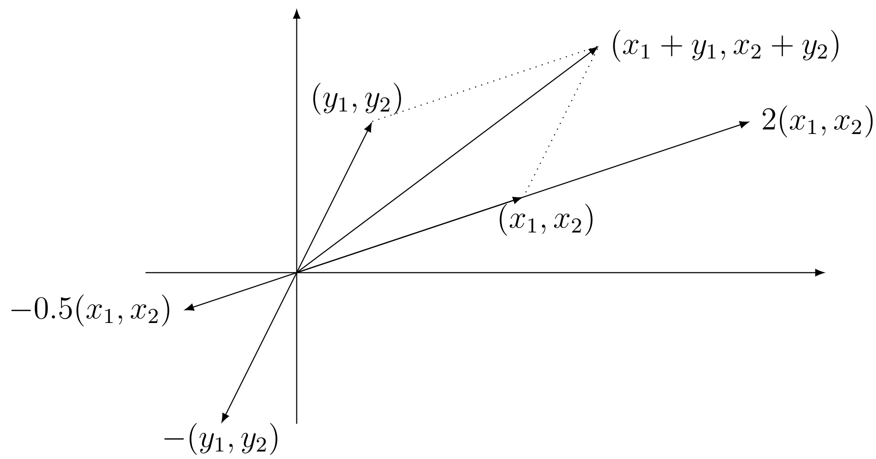

Figure 1.1: Euclidean space $\mathbb{R}^2$.

For the purpose of calculation (especially when mixed with matrices), it is more convenient to write the vector as vertical $n \times 1$ matrix, or the transpose (indicated by superscript T) of horizontal $1 \times n$ matrix

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1 \ x_2 \ \cdots \ x_n)^T.$$

We can write

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad a \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{pmatrix}.$$

Example 1.3. All polynomials of degree $\leq n$ form a vector space

$$P_n = \{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n\}.$$

The addition and scalar multiplication are the usual operations of functions. In fact, the coefficients of the polynomial provides a one-to-one correspondence

$$a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \in P_n \longleftrightarrow (a_0, a_1, a_2, \dots, a_n) \in \mathbb{R}^{n+1}.$$

Since the one-to-one correspondence preserves the addition and scalar multiplication, it identifies the polynomial space P_n with the Euclidean space $\mathbb{R}^{n+1}$, as far as the two operations are concerned. Such identifications are *isomorphisms*.

The rigorous definition of isomorphism and discussions about the concept will appear in Section 2.4.

Example 1.4. An $m \times n$ *matrix* A is mn numbers arranged in m rows and n columns. The number a_{ij} in the i-th row and j-column of A is called the ij*-entry* of A. We also indicate the matrix by $A = (a_{ij})$.

All $m \times n$ matrices form a vector space $M_{m\times n}$ with the obvious addition and scalar multiplication. For example, in $M_{2\times 3}$ we have

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix} + \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{pmatrix} = \begin{pmatrix} x_{11}+y_{11} & x_{12}+y_{12} \\ x_{21}+y_{21} & x_{22}+y_{22} \\ x_{31}+y_{31} & x_{32}+y_{32} \end{pmatrix}, \quad a\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix} = \begin{pmatrix} ax_{11} & ax_{12} \\ ax_{21} & ax_{22} \\ ax_{31} & ax_{32} \end{pmatrix}.$$

We have an isomorphism

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} \in M_{2\times 3} \longleftrightarrow (x_1, x_2, x_3, y_1, y_2, y_3) \in \mathbb{R}^6,$$

that can be used to translate linear algebra problems about matrices to linear algebra problems in Euclidean spaces. We also have the general *transpose* isomorphism that identifies $m \times n$ matrices with $n \times n$ matrices (see Example 2.32 for the general formula).

$$A = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} \in M_{2\times 3} \longleftrightarrow A^T = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} \in M_{3\times 2}.$$

Example 1.2 gives an isomorphism

$$(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \longleftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in M_{n\times 1}.$$

The transpose is also an isomorphism

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in M_{n\times 1} \longleftrightarrow \vec{x}^T = (x_1\ x_2\ \cdots\ x_n) \in M_{1\times n}.$$

The addition, scalar multiplication, and transpose of matrices are defined in the most "obvious" way. However, even simple definitions need to be justified. See Section 2.2 for the justification of addition and scalar multiplication. See Section ????? for the justification of transpose.

Example 1.5. All sequences $(x_n)_{n=1}^{\infty}$ of real numbers form a vector space V, with the addition and scalar multiplications given by

$$(x_n) + (y_n) = (x_n + y_n), \quad a(x_n) = (ax_n).$$

Example 1.6. All smooth functions form a vector space C^{∞}, with the usual addition and scalar multiplication of functions. The vector space is not isomorphic to the usual Euclidean space because it is "infinite dimensional".

Exercise 1.1. Prove that $(a+b)(\vec{x}+\vec{y}) = a\vec{x} + b\vec{y} + b\vec{x} + a\vec{y}$ in any vector space.

Exercise 1.2. Introduce the following addition and scalar multiplication in $\mathbb{R}^2$

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_2, x_2 + y_1), \quad a(x_1, x_2) = (ax_1, ax_2).$$

Check which axioms of vector space are true, and which are false.

Exercise 1.3. Introduce the following addition and scalar multiplication in $\mathbb{R}^2$

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, 0), \quad a(x_1, x_2) = (ax_1, 0).$$

Check which axioms of vector space are true, and which are false.

Exercise 1.4. Introduce the following addition and scalar multiplication in $\mathbb{R}^2$

$$(x_1, x_2) + (y_1, y_2) = (x_1 + Ay_1, x_2 + By_2), \quad a(x_1, x_2) = (ax_1, ax_2).$$

Show that this makes $\mathbb{R}^2$ into a vector space if and only if $A = B = 1$.

Exercise 1.5. Show that all convergent sequences form a vector space.

Exercise 1.6. Show that all even smooth functions form a vector space.

Exercise 1.7. Explain that the transpose of matrix satisfies $(A+B)^T = A^T + B^T$, $(aA)^T = aA^T$, $(A^T)^T = A$.

Consequence of Axiom

Now we establish some basic properties the vector space. You can directly verify these properties in Euclidean spaces. However, the proof for general vector spaces can only use the axioms.

Proposition 1.2. *The zero vector is unique.*

Proof. Suppose $\vec{0}_1$ and $\vec{0}_2$ are two zero vectors. By applying the first equality in Axiom 3 to $\vec{u} = \vec{0}_1$ and $\vec{0} = \vec{0}_2$, we get $\vec{0}_1 + \vec{0}_2 = \vec{0}_1$. By applying the second equality in Axiom 3 to $\vec{0} = \vec{0}_1$ and $\vec{u} = \vec{0}_2$, we get $\vec{0}_2 = \vec{0}_1 + \vec{0}_2$. Combining the two equalities, we get $\vec{0}_2 = \vec{0}_1 + \vec{0}_2 = \vec{0}_1$. $\square$

Proposition 1.3. *If $\vec{u}+\vec{v}=\vec{u}$, then $\vec{v}=\vec{0}$.*

By Axioms 2, we also have $\vec{v}+\vec{u}=\vec{u}$, then $\vec{v}=\vec{0}$. Both properties are the *cancelation law*.

Proof. Suppose $\vec{u}+\vec{v}=\vec{u}$. By Axiom 3, there is $\vec{w}$, such that $\vec{w}+\vec{u}=\vec{0}$. We use $\vec{w}$ instead of $\vec{v}$ in the axiom, because $\vec{v}$ is already used in the proposition. Then

$$\begin{aligned}
\vec{v} &= \vec{0}+\vec{v} && \text{(Axiom 3)}\\
&= (\vec{w}+\vec{u})+\vec{v} && \text{(choice of } \vec{w})\\
&= \vec{w}+(\vec{u}+\vec{v}) && \text{(Axiom 2)}\\
&= \vec{w}+\vec{u} && \text{(assumption)}\\
&= \vec{0}. && \text{(choice of } \vec{w})
\end{aligned}$$

□

Proposition 1.4. *$a\vec{u}=\vec{0}$ if and only if $a=0$ or $\vec{u}=\vec{0}$.*

Proof. First we prove $0\vec{u}=\vec{0}$. By Axiom 7, we have

$$0\vec{u}+0\vec{u}=(0+0)\vec{u}=0\vec{u}.$$

By Proposition 1.3, we get $0\vec{u}=\vec{0}$.

Next we prove $a\vec{0}=\vec{0}$. By Axioms 8 and 3, we have

$$a\vec{0}+a\vec{0}=a(\vec{0}+\vec{0})=a\vec{0}.$$

By Proposition 1.3, we get $a\vec{0}=\vec{0}$.

The equalities $0\vec{u}=\vec{0}$ and $a\vec{0}=\vec{0}$ give the if part of the proposition. The only if part means $a\vec{u}=\vec{0}$ implies $a=0$ or $\vec{u}=\vec{0}$. This is the same as $a\vec{u}=\vec{0}$ and $a\neq 0$ imply $\vec{u}=\vec{0}$. So we assume $a\vec{u}=\vec{0}$ and $a\neq 0$ and apply Axioms 5, 6 and $a\vec{0}=\vec{0}$ (just proved) to get

$$\vec{u}=1\vec{u}=(a^{-1}a)\vec{u}=a^{-1}(a\vec{u})=a^{-1}\vec{0}=\vec{0}.$$

□

Exercise 1.8. Directly verify Propositions 1.2, 1.3, 1.4 in $\mathbb{R}^n$.

Exercise 1.9. Prove that the vector $\vec{v}$ in Axiom 4 is unique. This justifies the notation $-\vec{u}$.

Exercise 1.10. Prove the more general version of the cancelation law: $\vec{u}+\vec{v}_1=\vec{u}+\vec{v}_2$ implies $\vec{v}_1=\vec{v}_2$.

Exercise 1.11. We use Exercise 1.9 to define $\vec{u}-\vec{v}=\vec{u}+(-\vec{v})$. Prove the following properties

$$(-1)\vec{u}=-\vec{u},\quad -(-\vec{u})=\vec{u},\quad -(\vec{u}-\vec{v})=-\vec{u}+\vec{v},\quad -(\vec{u}+\vec{v})=-\vec{u}-\vec{v}.$$

1.2 Span and Linear Independence

The repeated use of addition and scalar multiplication gives the *linear combination*

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n.$$

By using the axioms, it is easy to verify the usual properties of linear combinations such as

$$(a_1\vec{v}_1+a_2\vec{v}_2+\cdots+a_n\vec{v}_n)+(b_1\vec{v}_1+b_2\vec{v}_2+\cdots+b_n\vec{v}_n) = (a_1+b_1)\vec{v}_1+(a_2+b_2)\vec{v}_1+\cdots+(a_n+b_n)\vec{v}_n, \tag{1.1}$$

and

$$c(a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n) = ca_1\vec{v}_1 + ca_2\vec{v}_2 + \cdots + ca_n\vec{v}_n. \tag{1.2}$$

The linear combination produces many more vectors from several "seed vectors". If we start with a nonzero seed vector $\vec{u}$, then all its linear combinations $a\vec{u}$ form a straight line passing through the origin $\vec{0}$. If we start with two non-parallel vectors $\vec{u}$ and $\vec{v}$, then all their linear combinations $a\vec{u}+b\vec{v}$ form a plane passing through the origin $\vec{0}$.

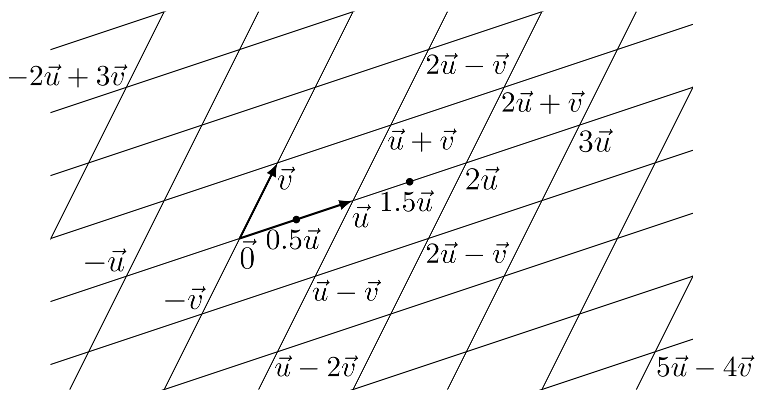

Figure 1.2: Linear combination.

Exercise 1.12. Suppose each of $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m$ is a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$. Prove that a linear combination of $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m$ is also a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$.

Definition of Span and Linear Independence

We often have mechanisms that produce new objects from existing objects. For example, the addition and multiplication produce new numbers from two existing numbers. Then linear combinations is comparable to all the new objects that can be produced by several "seed objects". For example, by using the mechanism "+1" and starting with the seed number 1, we can produce all the natural numbers $\mathbb{N}$. For another example, by using multiplication and starting with all prime numbers as "seed numbers", we can produce $\mathbb{N}$.

Two questions naturally arises. The first is whether the mechanism and the seed objects can produce all the objects. The answer is yes for the mechanism "+1" and seed 1 producing all natural numbers. The answer is no for the following

1. mechanism "+2", seed 1, producing $\mathbb{N}$.

2. mechanism "+1", seed 2, producing $\mathbb{N}$.

3. mechanism "+1", seed 1, producing $\mathbb{Q}$ (rational numbers).

The answer is also yes for the multiplication and all prime numbers producing $\mathbb{N}$.

The second question is whether the way new objects are produced is unique. For example, 12 is obtained by applying $+1$ to 1 eleven times, and this process is unique. For another example, 12 is obtained by multiplying the prime numbers $2, 2, 3$ together, and this collection of prime numbers is unique.

For the linear combination, the first question is span, and the second question is linear independence.

Definition 1.5. A set of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ *span* V if any vector in V can be expressed as a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$. The vectors are *linearly independent* if the coefficient in the linear combination is unique

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n = b_1\vec{v}_1 + b_2\vec{v}_2 + \cdots + b_n\vec{v}_n \implies a_1 = b_1,\ a_2 = b_2,\ \dots,\ a_n = b_n.$$

The vectors are *linearly dependent* if they are not linearly independent.

Proposition 1.6. *A set of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ is linearly independent if and only if the linear combination expression of the zero vector is unique*

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n = \vec{0} \implies a_1 = a_2 = \cdots = a_n = 0.$$

Proof. The property in the proposition is the special case of $b_1 = \cdots = b_n = 0$ in the definition of linear independence. Conversely, if the special case holds, then

$$\begin{aligned} & a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n = b_1\vec{v}_1 + b_2\vec{v}_2 + \cdots + b_n\vec{v}_n \\ \implies & (a_1 - b_1)\vec{v}_1 + (a_2 - b_2)\vec{v}_2 + \cdots + (a_n - b_n)\vec{v}_n = \vec{0} \\ \implies & a_1 - b_1 = a_2 - b_2 = \cdots = a_n - b_n = 0. \quad \text{(spacial case)} \end{aligned}$$

The second implication is obtained by applying the special case. □

Proposition 1.7. *A set of vectors is linearly dependent if and only if one vector is a linear combination of the other vectors.*

Proof. By Proposition 1.6, $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are linearly dependent if and only if there are $a_1, a_2, \dots, a_n$, not all 0, such that $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n = \vec{0}$. If $a_i \neq 0$, then the equality implies

$$\vec{v}_i = -\frac{a_1}{a_i}\vec{v}_1 - \cdots - \frac{a_{i-1}}{a_i}\vec{v}_{i-1} - \frac{a_{i+1}}{a_i}\vec{v}_{i+1} - \cdots - \frac{a_n}{a_i}\vec{v}_n.$$

This expresses the i-th vector as the a linear combination of the other vectors.

Conversely, if $\vec{v}_i = a_1\vec{v}_1 + \cdots + a_{i-1}\vec{v}_{i-1} + a_{i+1}\vec{v}_{i+1} + \cdots + a_n\vec{v}_n$, then

$$\vec{0} = a_1\vec{v}_1 + \cdots + a_{i-1}\vec{v}_{i-1} - 1\vec{v}_i + a_{i+1}\vec{v}_{i+1} + \cdots + a_n\vec{v}_n,$$

where the i-th coefficient is $-1 \neq 0$. By Proposition 1.6, the vectors are linearly dependent. □

Proposition 1.8. *A single vector is linearly dependent if and only if it is the zero vector. Two vectors are linearly dependent if and only if one is a scalar multiple of another.*

Proof. The zero vector $\vec{0}$ is linearly dependent because $1\vec{0} = \vec{0}$, with the coefficient $1 \neq 0$. Conversely, if $\vec{v} \neq \vec{0}$ and $a\vec{v} = \vec{0}$, then by Proposition 1.4, we have $a = 0$. This proves that the non-zero vector is linearly independent.

By Proposition 1.7, two vectors $\vec{u}$ and $\vec{v}$ are linearly dependent if and only if either $\vec{u}$ is a linear combination of $\vec{v}$, or $\vec{v}$ is a linear combination of $\vec{u}$. Since the linear combination of a single vector is simply the scalar multiplication, the proposition follows. □

Exercise 1.13. Suppose a set of vectors are linearly independent. Prove that a smaller set is still linearly independent.

Exercise 1.14. Suppose a set of vectors are linearly dependent. Prove that a bigger set is still linearly dependent.

Exercise 1.15. Suppose $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are linearly independent. Prove that $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n, \vec{v}_{n+1}$ is still linearly independent if and only if $\vec{v}_{n+1}$ is not a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$.

Exercise 1.16. Show that $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span V if and only if $\vec{v}_1, \dots, \vec{v}_j, \dots, \vec{v}_i, \dots, \vec{v}_n$ span V. Moreover, the linear independence is also equivalent.

Exercise 1.17. Show that $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span V if and only if $\vec{v}_1, \dots, c\vec{v}_i, \dots, \vec{v}_n$, $c \neq 0$, span V. Moreover, the linear independence is also equivalent.

Exercise 1.18. Show that $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span V if and only if $\vec{v}_1, \dots, \vec{v}_i + c\vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n$ span V. Moreover, the linear independence is also equivalent.

Calculation of Span and Linear Independence

Now we calculate the concepts of span and linear independence.

Example 1.7. We try to determine whether $\vec{b} = (10, 11, 12)$ is a linear combination of

$$\vec{v}_1 = (1, 2, 3), \quad \vec{v}_2 = (4, 5, 6), \quad \vec{v}_3 = (7, 8, 9).$$

This means that we can find numbers x_1, x_2, x_3, such that

$$\begin{aligned} x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 &= x_1(1,2,3) + x_2(4,5,6) + x_3(7,8,9) \\ &= (x_1 + 4x_2 + 7x_3, 2x_1 + 5x_2 + 8x_3, 3x_1 + 6x_2 + 9x_3) \\ &= (10, 11, 12) = \vec{b}. \end{aligned}$$

The equality means the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= 11, \\ 3x_1 + 6x_2 + 9x_3 &= 12. \end{aligned}$$

Therefore $\vec{b}$ is a linear combination of $\vec{v}_1, \vec{v}_2, \vec{v}_3$ if and only if the system has solution.

The system of linear equations can be simplified by the process of *Gaussian elimination*. Specifically, we may eliminate x_1 in the second and third equations by $\text{eq}_2 - 2\text{eq}_1$ (multiply the first equation by -2 and add to the second equation) and $\text{eq}_3 - 3\text{eq}_1$ (multiply the first equation by -3 and add to the third equation) to get

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ - 3x_2 - 6x_3 &= -9, \\ - 6x_2 - 12x_3 &= -18. \end{aligned}$$

Then we use $\text{eq}_3 - 2\text{eq}_2$ and $-\frac{1}{3}\text{eq}_2$ (multiplying $-\frac{1}{3}$ to the second equation) to get

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ x_2 + 2x_3 &= 3, \\ 0 &= 0. \end{aligned}$$

The third equation is trivial. We get $x_2 = 3 - 2x_3$ from the second equation. Then we substitute this into the first equation to get $x_1 = -2 + x_3$. We conclude the general solution

$$x_1 = -2 + x_3, \quad x_2 = 3 - 2x_3, \quad x_3 \text{ arbitrary}.$$

If we take $x_3 = 0$, then we get a special solution $x_1 = -2, x_2 = 3, x_3 = 0$. If we take $x_3 = 1$, then we get another special solution $x_1 = -1, x_2 = 1, x_3 = 1$. The two solutions give two linear combination expressions

$$(10, 11, 12) = -2(1,2,3) + 3(4,5,6) + 0(7,8,9) = -1(1,2,3) + 1(4,5,6) + 1(7,8,9).$$

The non-uniqueness of the expression also shows that $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly dependent.

We can use Gaussian elimination because the method does not change the solutions of the system. For example, if eq_1, eq_2, eq_3 hold, then $\text{eq}_1' = \text{eq}_1$, $\text{eq}_2' = \text{eq}_2 - 2\text{eq}_1$, $\text{eq}_3' = \text{eq}_3 - 3\text{eq}_1$ hold. Conversely, if eq_1', eq_2', eq_3' hold, then $\text{eq}_1 = \text{eq}_1'$, $\text{eq}_2 = \text{eq}_2' + 2\text{eq}_1'$, $\text{eq}_3 = \text{eq}_3' + 3\text{eq}_1'$ hold.

Example 1.8. We try to determine whether the three vectors in Example 1.7 span $\mathbb{R}^3$. This means that, for any vector $\vec{b} = (b_1, b_2, b_3)$ in $\mathbb{R}^3$, we can find numbers x_1, x_2, x_3, such that $x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{b}$. By the interpretation in the earlier example, this means that the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= b_1, \\ 2x_1 + 5x_2 + 8x_3 &= b_2, \\ 3x_1 + 6x_2 + 9x_3 &= b_3, \end{aligned}$$

has solution for all b_1, b_2, b_3.

Similar to the earlier example, we may apply $\mathrm{eq}_2 - 2\mathrm{eq}_1$ and $\mathrm{eq}_3 - 3\mathrm{eq}_1$ to eliminate x_1 in the second and third equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= b_1, & (\mathrm{eq}_1' &= \mathrm{eq}_1) \\ -3x_2 - 6x_3 &= -2b_1 + b_2, & (\mathrm{eq}_2' &= \mathrm{eq}_2 - 2\mathrm{eq}_1) \\ -6x_2 - 12x_3 &= -3b_1 + b_3, & (\mathrm{eq}_3' &= \mathrm{eq}_3 - 3\mathrm{eq}_1) \end{aligned}$$

Then we may apply $\mathrm{eq}_3' - 2\mathrm{eq}_2'$ and $-\frac{1}{3}\mathrm{eq}_2'$ to eliminate x_2 in the third equation

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= b_1, & (\mathrm{eq}_1'' &= \mathrm{eq}_1' = \mathrm{eq}_1) \\ x_2 + 2x_3 &= \frac{2}{3}b_1 - \frac{1}{3}b_2, & (\mathrm{eq}_2'' &= -\frac{1}{3}\mathrm{eq}_2' = -\frac{2}{3}\mathrm{eq}_1 + \frac{1}{3}\mathrm{eq}_2) \\ 0 &= b_1 - 2b_2 + b_3. & (\mathrm{eq}_3'' &= \mathrm{eq}_3' - 2\mathrm{eq}_2' = \mathrm{eq}_1 - 2\mathrm{eq}_2 + \mathrm{eq}_3) \end{aligned}$$

The last equation shows that there is no solution unless $b_1 - 2b_2 + b_3 = 0$. Therefore the three vectors do not span $\mathbb{R}^3$.

Now we modify the third vector to $\vec{v}_3 = (7, 8, a)$, with a to be determined. Then whether the three vectors span $\mathbb{R}^3$ is the same as whether the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= b_1, \\ 2x_1 + 5x_2 + 8x_3 &= b_2, \\ 3x_1 + 6x_2 + ax_3 &= b_3, \end{aligned}$$

has solution for all b_1, b_2, b_3. By the same elimination, the system can be simplified to

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= b_1, \\ x_2 + 2x_3 &= \frac{2}{3}b_1 - \frac{1}{3}b_2, \\ (a-9)x_3 &= b_1 - 2b_2 + b_3. \end{aligned}$$

The last equation has solution for all b_1, b_2, b_3 if and only if $a \neq 9$. In fact, the solution is $x_3 = \frac{b_1 - 2b_2 + b_3}{a-9}$. We may substitute this solution to the second equation to get $x_2 = \frac{2b_1 - b_2}{3} - 2\frac{b_1 - 2b_2 + b_3}{a-9}$. Then we may further substitute the formulae for x_2, x_3 the first equation to determines x_1. The process is called *back substitution* and shows that the three vectors span $\mathbb{R}^3$ if and only if $a \neq 9$.

Example 1.9. We try to determine whether the three vectors in Example 1.7 are linearly independent. By Proposition 1.6, this means that

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{0} \implies x_1 = x_2 = x_3 = 0.$$

By the interpretation in Examples 1.7 and 1.8 (taking $b_1 = b_2 = b_3 = 0$), the linear independence means that the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 0, \\ 2x_1 + 5x_2 + 8x_3 &= 0, \\ 3x_1 + 6x_2 + 9x_3 &= 0, \end{aligned}$$

has only the trivial solution $x_1 = x_2 = x_3 = 0$. Note that the *homogeneous* system above (with zero right side) always has the trivial solution $x_1 = x_2 = x_3 = 0$. The key here is that there is no non-trivial solution.

By the elimination in Example 1.7, the homogeneous system can be simplified to

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 0, \\ x_2 + 2x_3 &= 0, \\ 0 &= 0. \end{aligned}$$

It is easy to see that the simplfied system has non-trivial solution $x_3 = 1$, $x_2 = -2$ (from the second equation), $x_1 = -4 \cdot (-2) - 7 \cdot 1 = 1$ (back substitution). Indeed, we can verify that

$$\vec{v}_1 - 2\vec{v}_2 + \vec{v}_3 = 1\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - 2\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + 1\begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \vec{0}.$$

This explicitly shows that the three vectors are linearly independent.

Similar to Example 1.8, we modify the third vector to $\vec{v}_3 = (7, 8, a)$. Then the linear independence means that the homogeneous system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 0, \\ 2x_1 + 5x_2 + 8x_3 &= 0, \\ 3x_1 + 6x_2 + ax_3 &= 0, \end{aligned}$$

has only the trivial solution. By the same elimination, the system can be simplified to

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 0, \\ 3x_2 + 6x_3 &= 0, \\ (a - 9)x_3 &= 0. \end{aligned}$$

If $a \neq 9$, then the last equation implies $x_3 = 0$. Substituting $x_3 = 0$ into the second equation, we get $x_2 = 0$. Further substituting $x_2 = x_3 = 0$ into the first equation, we get $x_1 = 0$. This shows that the homogeneous system has only the trivial solution, and therefore the three vectors are linearly independent.

Example 1.10. We try to determine whether the polynomials $p_1 = 1+2t+3t^2$, $p_2 = 4+5t+6t^2$, $p_3 = 7+8t+9t^2$ span the vector space P_2. This means that any polynomial $b_1 + b_2t + b_3t^2$ can be expressed as a linear combination

$$\begin{aligned} b_1 + b_2t + b_3t^2 &= x_1(1+2t+3t^2) + x_2(4+5t+6t^2) + x_3(7+8t+9t^2) \\ &= (x_1+4x_2+7x_3) + (2x_1+5x_2+8x_3)t + (3x_1+6x_2+9x_3)t^2. \end{aligned}$$

The equality is the same as system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= b_1, \\ 2x_1 + 5x_2 + 8x_3 &= b_2, \\ 3x_1 + 6x_2 + 9x_3 &= b_3. \end{aligned}$$

Then $p_1(t), p_2(t), p_3(t)$ span P_2 if and only if the system has solution for all b_1, b_2, b_3. By the discussion in Example 1.8, we know the answer is no. Therefore the three polynomials do not span P_2.

Similarly, the three polynomials are linearly independent if and only if

$$\begin{aligned} 0 &= x_1(1+2t+3t^2) + x_2(4+5t+6t^2) + x_3(7+8t+9t^2) \\ &= (x_1+4x_2+7x_3) + (2x_1+5x_2+8x_3)t + (3x_1+6x_2+9x_3)t^2 \end{aligned}$$

implies $x_1 = x_2 = x_3 = 0$. This is the same as that the homogeneous system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 0, \\ 2x_1 + 5x_2 + 8x_3 &= 0, \\ 3x_1 + 6x_2 + 9x_3 &= 0, \end{aligned}$$

has only the trivial solution $x_1 = x_2 = x_3 = 0$. By the discussion in Example 1.9, we know the answer is no. Therefore the three polynomials are linearly dependent. In fact, we have $p_1(t) - 2p_2(t) + p_3(t) = 0$.

We see that the problems of span and linear independence in general vector spaces may be translated into the similar problems in the Euclidean spaces. In fact, the translation is given by the isomprhism

$$a_0 + a_1t + a_2t^2 \in P_2 \longleftrightarrow (a_0, a_1, a_2) \in \mathbb{R}^3.$$

To calculate the concepts, we further translate into problems about solutions of systems of linear equations.

Exercise 1.19. Determine whether the vectors span Euclidean space or are linearly independent. You need to first translate the problem into the problem about some systems of linear equations, and then try to solve the system.

1. $(1,2,3),(2,3,1),(3,1,2)$.
2. $(1,2,3,4),(2,3,4,5),(3,4,5,6)$.
3. $(1,2,3),(4,5,6),(7,8,9),(10,11,12)$.
4. $(1,2,3),(2,3,1),(3,1,a)$.
5. $(1,2,3,4),(2,3,4,5),(3,4,a,a)$.
6. $(1,2,3),(4,5,6),(7,8,9),(10,11,a)$.

Exercise 1.20. Translate the problem of polynomials span suitable polynomial vector space or are linearly independent into the problem about some systems of linear equations.

1. $1+2t+3t^2, 2+3t+t^2, 3+t+2t^2$.
2. $1+2t+3t^2+4t^3, 2+3t+4t^2+5t^3, 3+4t+5t^2+6t^3$.
3. $3+2t+t^2, 6+5t+4t^2, 9+8t+7t^2, 12+11t+10t^2$.
4. $1+2t+3t^2, 2+3t+t^2, 3+t+at^2$.
5. $1+2t+3t^2+4t^3, 2+3t+4t^2+5t^3, 3+4t+at^2+at^3$.
6. $3+2t+t^2, 6+5t+4t^2, a+8t+7t^2, b+11t+10t^2$.

Exercise 1.21. Translate the problem of polynomials span suitable polynomial vector space or are linearly independent into the problem about some systems of linear equations.

1. $\begin{pmatrix}1&2\\3&4\end{pmatrix}, \begin{pmatrix}3&4\\1&2\end{pmatrix}, \begin{pmatrix}2&1\\4&3\end{pmatrix}, \begin{pmatrix}4&3\\2&1\end{pmatrix}$.
2. $\begin{pmatrix}1&4\\2&3\end{pmatrix}, \begin{pmatrix}4&3\\1&2\end{pmatrix}, \begin{pmatrix}3&2\\4&1\end{pmatrix}, \begin{pmatrix}2&1\\3&4\end{pmatrix}$.
3. $\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}, \begin{pmatrix}4&5&6\\1&2&3\end{pmatrix}$.
4. $\begin{pmatrix}1&3&5\\2&4&6\end{pmatrix}, \begin{pmatrix}5&1&3\\6&2&4\end{pmatrix}, \begin{pmatrix}3&5&1\\4&6&2\end{pmatrix}$.

Another Way of Argueing Linear Independence

The examples so far are about finite dimensional vector spaces. The following shows how to argue about span or linear independence in some infinite dimensional vector spaces.

Example 1.11. To show that $\cos t, \sin t, e^t$ are linearly independent, we need to verify

$$x_1\cos t + x_2\sin t + x_3e^t = 0 \implies x_1 = x_2 = x_3 = 0.$$

If the left side holds, then by evaluating at $t = 0, \frac{\pi}{2}, -\frac{\pi}{2}$, we get

$$x_1 + x_3 = 0, \quad x_2 + x_3e^{\frac{\pi}{2}} = 0, \quad -x_2 + x_3e^{-\frac{\pi}{2}} = 0.$$

Adding the second and the third equations together, we get $x_3(e^{\frac{\pi}{2}} + e^{-\frac{\pi}{2}}) = 0$. This implies $x_3 = 0$. Substituting $x_3 = 0$ to the first and second equations, we get $x_1 = x_2 = 0$. This proves that the three functions are linearly independent.

Example 1.12. To show that $\cos t, \sin t, e^t$ do not span the space of all smooth functions, we only need to find a function that cannot be written as a linear combination of the three functions.

Suppose the constant function is a linear combination of the three functions

$$x_1 \cos t + x_2 \sin t + x_3 e^t = 1.$$

By evaluating at $t = 0, \pi, \frac{\pi}{2}, -\frac{\pi}{2}$, we get

$$x_1 + x_3 = 1, \quad -x_1 + x_3 e^{\pi} = 1, \quad x_2 + x_3 e^{\frac{\pi}{2}} = 1, \quad -x_2 + x_3 e^{-\frac{\pi}{2}} = 1.$$

Adding the first two equations together, we get $x_3(1 + e^{\pi}) = 2$. Adding the last two equations together, we get $x_3(e^{\frac{\pi}{2}} + e^{-\frac{\pi}{2}}) = 2$. Comparing the two equations about x_3, we get $1 + e^{\pi} = e^{\frac{\pi}{2}} + e^{-\frac{\pi}{2}}$. Since this is not true, we conclude that there is no x_1, x_2, x_3 satisfying $x_1 \cos t + x_2 \sin t + x_3 e^t = 1$. This proves that 1 is not in the span of the three functions.

Exercise 1.22. Determine whether the given functions are linearly independent, and whether $f(x), g(t)$ are in the span of the given functions.

1. $\cos^2 t, \sin^2 t$. $f(t) = 1$, $g(t) = t$.
2. $\cos^2 t, \sin^2 t, 1$. $f(t) = \cos 2t$, $g(t) = t$.
3. $1, t, e^t, te^t$. $f(t) = (1 + t)e^t$, $g(t) = f'(t)$.
4. $\cos^2 t, \cos 2t$. $f(t) = a$, $g(t) = a + \sin^2 t$.

Exercise 1.23. For $a \neq b$, show that e^{at} and e^{bt} are linearly independent. What about e^{at}, e^{bt} and e^{ct}?

1.3 System of Linear Equations

In Examples 1.8, 1.9, 1.10, we saw that the calculation of span and linear independence is equivalent to answering some problems about solutions of systems of linear equations. In general, a *system of linear equation* is

$$\begin{aligned} a_{11}x_1 + \ a_{12}x_2 + \cdots + \ a_{1n}x_n &= b_1, \\ a_{21}x_1 + \ a_{22}x_2 + \cdots + \ a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

The *coefficient matrix*, the right side, and the *augmented matrix* of the system are

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}, \quad (A\ \vec{b}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

The columns of A correspond to the variables. We also denote the system of linear equations by $A\vec{x} = \vec{b}$.

We call the system *homogeneous* if the right side is zero. Therefore a homogeneous system is of the form $A\vec{x} = \vec{0}$.

Exercise 1.24. How does the exchange of rows of A affect the solution of $A\vec{x} = \vec{b}$? What about multiplying a nonzero number to a row of A?

Exercise 1.25. How does the exchange of columns of A affect the solution of $A\vec{x} = \vec{b}$? What about multiplying a nonzero number to a column of A?

Exercise 1.26. Show that a system is homogeneous if and only if $\vec{x} = \vec{0}$ is a solution.

Row Echelon Form

In Example 1.7, we saw that systems of linear equations can be solved by Gaussian elimination, which cancels variables by combining equations in the system. In general, the Gaussian elimination consists of three operations.

- $\text{eq}_i \leftrightarrow \text{eq}_j$: exchange the i-th and j-th equations.
- $c\,\text{eq}_i$: multiplying a number $c \neq 0$ to the i-th equation.
- $\text{eq}_i + c\,\text{eq}_j$: add the c multiple of the j-th equation to the i-th equation.

Note that $\vec{x}$ satisfies eq_i and eq_j if and only if it satisfies $\text{eq}_i + c\ \text{eq}_j$ and eq_j. This shows that the third operation preserves solutions. Similar argument shows that the first and second operations also preserve solutions.

The Gaussian elimination in Example 1.7 is equivalent to the following *row operation* on the augmented matrix

$$(A\ \vec{b}) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \xrightarrow[R_3-3R_1]{R_2-2R_1} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix} \xrightarrow[-\frac{1}{3}R_2]{R_3-2R_2} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In general, the three elimination operations correspond to three types of row operations.

- $R_i \leftrightarrow R_j$: exchange the i-th and j-th rows.
- cR_i: multiplying a number $c \neq 0$ to the i-th row.
- $R_i + cR_j$: add the c multiple of the j-th row to the i-th row.

The goal of elimination is to simplify the system. Correspondingly, the third row operation simplifies the matrix by creating as many zero entries as possible. The second row operation simplifies the coefficients in one equation. The first row operation rearranges the order of the equations from the "longest" (most complicated) to "shortest" (simplest).

The simplest *shape* one can achieve by the three row operations is the *row echelon form.* For the system in Example 1.7, the row echelon form is

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \bullet \neq 0, \quad * \text{ arbitrary.} \tag{1.3}$$

The entries indicated by $\bullet$ are called the *pivots*. For the augmented matrix in the example, the first and second columns are the *pivot columns*, and the first and second rows are the *pivot rows*.

In general, a row echelon form has the shape of upside down staircase, and the shape is characterized by the locations of the pivots. The pivots are the leading nonzero entries in the rows. They appear in the first several rows in later and later positions, and the subsequent rows are completely zero. The following are all the 2×3 row echelon forms

$$\begin{pmatrix} \bullet & * & * \\ 0 & \bullet & * \end{pmatrix}, \begin{pmatrix} \bullet & * & * \\ 0 & 0 & \bullet \end{pmatrix}, \begin{pmatrix} \bullet & * & * \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \bullet & * \\ 0 & 0 & \bullet \end{pmatrix}, \begin{pmatrix} 0 & \bullet & * \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & \bullet \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Exercise 1.27. Why the shape in (1.3) cannot be further improved? How can you improve the following shape to the upside down staircase by using row operations?

1. $\begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \\ \bullet & * & * & * \end{pmatrix}$.
2. $\begin{pmatrix} \bullet & * & * & * \\ \bullet & * & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix}$.
3. $\begin{pmatrix} 0 & \bullet & * & * \\ 0 & \bullet & * & * \\ \bullet & * & * & * \end{pmatrix}$.
4. $\begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ \bullet & * & * & * \end{pmatrix}$.

Exercise 1.28. Display all the 2×2 row echelon forms. How about 3×2 matrices?

Exercise 1.29. How many $m \times n$ row echelon forms are there?

Example 1.13. We have interpreted the Gaussian elimination in Example 1.7 as row operations. The following is another sequence of row operations on the same augmented matrix.

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \xrightarrow[R_3-R_2]{R_2-R_1} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \xrightarrow[R_1\leftrightarrow R_2]{R_3-R_2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 7 & 10 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{R_2-R_1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 3 & 6 & 9 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix we get is different from before, but has the same shape as before.

Example 1.14. To solve the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= 11, \\ 3x_1 + 6x_2 + ax_3 &= b, \end{aligned}$$

we carry out the following row operations on the augmented matrix

$$(A\ \vec{b}) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & a & b \end{pmatrix} \xrightarrow[R_3-3R_1]{R_2-2R_1} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & a-21 & b-30 \end{pmatrix} \xrightarrow{R_3-2R_2} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & a-9 & b-12 \end{pmatrix}.$$

The row echelon form depends on the values of a and b.

If $a \neq 9$, then the result of the row operations is already a row echelon form

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \end{pmatrix}.$$

The third row is the equation $\bullet x_3 = *$, from which we get unique value $x_3 = \frac{*}{\bullet}$. The key here is that we can divide $\bullet$ because the number is nonzero. Substituting value of x_3 into the second equation $\bullet x_2 + * x_3 = *$, we get the unique value $x_2 = \frac{1}{\bullet}(* - * x_3)$. Again we can divide $\bullet$ because $\bullet \neq 0$. Substituting the values of x_2, x_3 into the first equation $\bullet x_1 + * x_2 + * x_3 = *$, we can divide $\bullet$ and get the unique value of x_1. The back substitution shows that the system has unique solution when $a \neq 9$. The uniqueness corresponds to the fact that all the columns of A (meaning ignoring the last column $\vec{b}$) are pivot columns.

If $a = 9$, then the result of the row operations is

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & b - 12 \end{pmatrix}.$$

If $b \neq 12$, then this is the row echelon form

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & \bullet \end{pmatrix}.$$

The last row is the equation $0 = \bullet$. Since this is a contradiction, we find that the system has no solution when $a = 9$ and $b \neq 12$. The non-existence of the solution corresponds to the fact that the last column $\vec{b}$) is a pivot column.

If $a = 9$ and $b = 12$, then the result of the row operations is the row echelon form

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The last row is the trivial equation $0 = 0$, which is no longer a contradiction. The second equation is $\bullet x_2 + * x_3 = *$. We can divide the non-zero number $\bullet$ to get $x_2 = \frac{1}{\bullet}(* - * x_3)$, with arbitrary x_3. Then we may substitute the formula for x_2 into the first equation $\bullet x_1 + * x_2 + * x_3 = *$ and divide $\bullet$ to get the formula of x_1 (involving the arbitrary x_3). The back substitution shows that the system has solution with free variable x_3 when $a = 9$ and $b = 12$. The free variable corresponds to the *non-pivot* column of A.

The discussion in the example can be easily generalised. For a system of linear equations $A\vec{x} = \vec{b}$, our first question should be the existence of solutions. After we know the existence, then we can ask the second question, which is the structure of solution. Usually the general solution is presented as some variables (called *non-free* variables) expressed in terms of the other variables (called *free* variables), and the other variables can take arbitrary values. We also note that variables correspond to the columns of A.

Theorem 1.9. *A system of linear equations $A\vec{x} = \vec{b}$ has solution if and only if $\vec{b}$ is not a pivot column in the augmented matrix $(A\ \vec{b})$. If the system has solution, then variables corresponding to pivot columns of A can be chosen as non-free, and variables corresponding to non-pivot columns of A can be chosen as free. In particular, the solution is unique if and only if all columns of A are pivot.*

Exercise 1.30. Solve system of linear equations. Compare systems and their solutions.

1. $\begin{aligned} x_1 + 2x_2 &= 3, \\ 4x_1 + 5x_2 &= 6. \end{aligned}$
2. $\begin{aligned} x_1 &= 3, \\ 4x_1 &= 6. \end{aligned}$
3. $\begin{aligned} x_1 &= 3, \\ 5x_2 &= 6. \end{aligned}$
4. $x_1 + 2x_2 = 3.$
5. $4x_1 + 5x_2 = 6.$
6. $\begin{aligned} x_1 + 2x_2 &= 3, \\ 4x_1 + 5x_2 &= 6, \\ 7x_1 + 8x_2 &= 9. \end{aligned}$
7. $\begin{aligned} x_1 + 2x_2 + 3x_3 &= 0, \\ 4x_1 + 5x_2 + 6x_3 &= 0. \end{aligned}$
8. $\begin{aligned} x_1 + 4x_2 &= 0, \\ 2x_1 + 5x_2 &= 0, \\ 3x_1 + 6x_2 &= 0. \end{aligned}$
9. $\begin{aligned} x_1 + 2x_2 &= 1, \\ 4x_1 + 5x_2 &= 1. \end{aligned}$
10. $x_1 + 2x_2 = 1.$
11. $\begin{aligned} x_1 + 2x_2 &= 1, \\ 4x_1 + 5x_2 &= 1, \\ 7x_1 + 8x_2 &= 1. \end{aligned}$

Exercise 1.31. Solve system of linear equations. Compare systems and their solutions.

1. $\begin{aligned} x_1 + 2x_2 &= a, \\ 4x_1 + 5x_2 &= b. \end{aligned}$
2. $\begin{aligned} x_1 &= a, \\ 4x_1 &= b. \end{aligned}$
3. $\begin{aligned} x_1 &= a, \\ 5x_2 &= b. \end{aligned}$
4. $x_1 + 2x_2 = a.$
5. $\begin{aligned} x_1 + 2x_2 &= a, \\ 4x_1 + 5x_2 &= b, \\ 7x_1 + 8x_2 &= c. \end{aligned}$
6. $\begin{aligned} x_1 + ax_2 &= 3, \\ 4x_1 + 5x_2 &= 6. \end{aligned}$
7. $\begin{aligned} x_1 + ax_2 &= 1, \\ 4x_1 + 5x_2 &= 1. \end{aligned}$
8. $\begin{aligned} x_1 + ax_2 &= b, \\ 4x_1 + 5x_2 &= 6. \end{aligned}$
9. $\begin{aligned} x_1 + ax_2 &= 3, \\ 4x_1 + bx_2 &= 6. \end{aligned}$
10. $\begin{aligned} x_1 + ax_2 &= 3, \\ 4x_1 + 5x_2 &= b. \end{aligned}$
11. $\begin{aligned} ax_1 + bx_2 &= 3, \\ 4x_1 + 5x_2 &= 6. \end{aligned}$
12. $\begin{aligned} ax_1 + 2x_2 &= b, \\ 4x_1 + 5x_2 &= 6. \end{aligned}$

Exercise 1.32. Carry out the row operation and explain what the row operation tells you about some system of linear equations.

1. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 5 & 6 & 7 & 8 \\ 7 & 8 & 9 & 10 \end{pmatrix}.$
2. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 5 & 6 & 7 & a \\ 7 & 8 & 9 & b \end{pmatrix}.$
3. $\begin{pmatrix} 1 & 2 & 4 & 3 \\ 3 & 4 & 6 & 5 \\ 5 & 6 & a & 7 \\ 7 & 8 & b & 9 \end{pmatrix}.$
4. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 5 & 6 & 7 & 8 \\ 7 & 8 & a & b \end{pmatrix}.$
5. $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \\ 4 & 1 & 2 \end{pmatrix}.$
6. $\begin{pmatrix} 1 & 2 & 3 & 1 \\ 2 & 3 & 1 & 2 \\ 3 & 1 & 2 & 3 \end{pmatrix}.$
7. $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \\ 1 & 2 & 3 \end{pmatrix}.$

Exercise 1.33. Solve system of linear equations.

1. $$\begin{aligned} a_1x_1 &= b_1, \\ a_2x_2 &= b_2, \\ &\vdots \\ a_nx_n &= b_n. \end{aligned}$$

2. $$\begin{aligned} x_1 - x_2 &= b_1, \\ x_2 - x_3 &= b_2, \\ &\vdots \\ x_{n-1} - x_n &= b_{n-1}. \end{aligned}$$

3. $$\begin{aligned} x_1 - x_2 &= b_1, \\ x_2 - x_3 &= b_2, \\ &\vdots \\ x_{n-1} - x_n &= b_{n-1}, \\ x_n - x_1 &= b_n. \end{aligned}$$

4. $$\begin{aligned} x_1 + x_2 &= b_1, \\ x_2 + x_3 &= b_2, \\ &\vdots \\ x_{n-1} + x_n &= b_{n-1}. \end{aligned}$$

5. $$\begin{aligned} x_1 + x_2 &= b_1, \\ x_2 + x_3 &= b_2, \\ &\vdots \\ x_{n-1} + x_n &= b_{n-1}, \\ x_n + x_1 &= b_n. \end{aligned}$$

6. $$\begin{aligned} x_1 + x_2 + x_3 &= b_1, \\ x_2 + x_3 + x_4 &= b_2, \\ &\vdots \\ x_{n-2} + x_{n-1} + x_n &= b_{n-2}. \end{aligned}$$

7. $$\begin{aligned} x_1 + x_2 + x_3 &= b_1, \\ x_2 + x_3 + x_4 &= b_2, \\ &\vdots \\ x_{n-2} + x_{n-1} + x_n &= b_{n-2}, \\ x_{n-1} + x_n + x_1 &= b_{n-1}, \\ x_n + x_1 + x_2 &= b_n. \end{aligned}$$

8. $$\begin{aligned} x_1 + x_2 + x_3 + \cdots + x_n &= b_1, \\ x_2 + x_3 + \cdots + x_n &= b_2, \\ &\vdots \\ x_n &= b_n. \end{aligned}$$

9. $$\begin{aligned} x_2 + x_3 + \cdots + x_{n-1} + x_n &= b_1, \\ x_1 + x_3 + \cdots + x_{n-1} + x_n &= b_2, \\ &\vdots \\ x_1 + x_2 + \cdots + x_{n-2} + x_{n-1} &= b_n. \end{aligned}$$

Reduced Row Echelon Form

For the existence, we note that last column $\vec{b}$ is pivot if and only if the corresponding row is $(0 \ \cdots \ 0 \ \bullet)$, which means the contradictory equation $0 = \bullet$. For the structure of solution (in case $\vec{b}$ is not pivot), we can see the structure more clearly by further simply the entries of the matrix. Specifically, suppose the augmented matrix is simplified to the row echelon form (1.3). Then we may use the second row operation $\frac{1}{\bullet}R_i$ to further simplify to

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

Next we may use $R_1 - cR_2$ to get cancel c

$$\begin{pmatrix} 1 & 0 & a_1 & b_1 \\ 0 & 1 & a_2 & b_2 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.4}$$

This is the simplest *matrix* (although the shape can not be further improved) one can get by row operations, and is called the *reduced row echelon form*. The reduced row echelon form is characterised by

- The pivot entries are 1.
- The entries above the pivot entries are 0.

The following are all the 2×3 reduced row echelon forms

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

The reduced row echelon form (1.4) is translated into the system of equations $x_1 + a_1x_3 = b_1$ and $x_2 + a_2x_3 = b_2$. By moving the (non-pivot) free variable x_3 to the right, the equations become the general solution

$$x_1 = b_1 - a_1x_3, \quad x_2 = b_2 - a_2x_3, \quad x_3 \text{ arbitrary.}$$

We see that the reduced echelon form is equivalent to the general solution. The conversion between the reduced echelon form and the general solution shows that pivot means non-free and non-pivot means free.

Example 1.15. We already obtained a row echelon form of the system of linear equations in Example 1.7

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= 11, \\ 3x_1 + 6x_2 + 9x_3 &= 12. \end{aligned}$$

The following is the further reduction to reduced row echelon form

$$(A\,\vec{b}) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{R_1 - 4R_2} \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then we can read the general solution

$$x_1 = -2 + x_3, \quad x_2 = 3 - 2x_3, \quad x_3 \text{ arbitrary.}$$

directly from the reduced row echelon form.

Example 1.16. For the homogeneous system

$$\begin{aligned} x_1 + 4x_2 + 7x_3 + 10x_4 &= 0, \\ 2x_1 + 5x_2 + 8x_3 + 11x_4 &= 0, \\ 3x_1 + 6x_2 + 9x_3 + 12x_4 &= 0, \end{aligned}$$

the reduced row echelon form in Example 1.15 also gives

$$\begin{pmatrix} 1 & 4 & 7 & 10 & 0 \\ 2 & 5 & 8 & 11 & 0 \\ 3 & 6 & 9 & 12 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -1 & -2 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then we can read the general solution of the homogeneous system

$$x_1 = x_3 + 2x_4, \quad x_2 = -2x_3 - 3x_4, \quad x_3, x_4 \text{ arbitrary.}$$

Exercise 1.34. Display all the 2×2, 3×2 and 3×4 reduced row echelon forms.

Exercise 1.35. Given the reduced row echelon form of the system of linear equations, find the general solution.

1. $\begin{pmatrix} 1 & 0 & a_1 & b_1 \\ 0 & 0 & 1 & b_2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

2. $\begin{pmatrix} 1 & 0 & a_1 & b_1 & 0 \\ 0 & 0 & 1 & b_2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.

3. $\begin{pmatrix} 1 & 0 & 0 & b_1 \\ 0 & 1 & 0 & b_2 \\ 0 & 0 & 1 & b_3 \end{pmatrix}$.

4. $\begin{pmatrix} 1 & a_1 & a_2 & b_1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

5. $\begin{pmatrix} 1 & a_1 & a_2 & b_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.

6. $\begin{pmatrix} 1 & a_1 & 0 & a_2 & b_1 \\ 0 & 0 & 1 & a_3 & b_2 \end{pmatrix}$.

7. $\begin{pmatrix} 1 & 0 & a_1 & a_2 & b_1 \\ 0 & 1 & a_3 & a_4 & b_2 \end{pmatrix}$.

8. $\begin{pmatrix} 1 & 0 & a_1 & b_1 \\ 0 & 1 & a_2 & b_2 \end{pmatrix}$.

9. $\begin{pmatrix} 0 & 1 & 0 & a_1 & b_1 \\ 0 & 0 & 1 & a_2 & b_2 \end{pmatrix}$.

10. $\begin{pmatrix} 1 & 0 & a_1 & b_1 \\ 0 & 1 & a_2 & b_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

11. $\begin{pmatrix} 1 & 0 & a_1 & 0 & a_2 & b_1 \\ 0 & 1 & a_3 & 0 & a_4 & b_2 \\ 0 & 0 & 0 & 1 & a_5 & b_3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.

12. $\begin{pmatrix} 1 & 0 & a_1 & 0 & a_2 & b_1 & 0 \\ 0 & 1 & a_3 & 0 & a_4 & b_2 & 0 \\ 0 & 0 & 0 & 1 & a_5 & b_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.

Exercise 1.36. Given the general solution of system of linear equations, find the reduced row echelon form. Moreover, which system is homogeneous?

1. $x_1 = -x_3$, $x_2 = 1 + x_3$; x_3 arbitrary.
2. $x_1 = -x_3$, $x_2 = 1 + x_3$; x_3, x_4 arbitrary.
3. $x_2 = -x_4$, $x_3 = 1 + x_4$; x_1, x_4 arbitrary.
4. $x_2 = -x_4$, $x_3 = x_4 - x_5$; x_1, x_4, x_5 arbitrary.
5. $x_1 = 1 - x_2 + 2x_5$, $x_3 = 1 + 2x_5$, $x_4 = -3 + x_5$; x_2, x_5 arbitrary.
6. $x_1 = 1 + 2x_2 + 3x_4$, $x_3 = 4 + 5x_4 + 6x_5$; x_2, x_4, x_5 arbitrary.
7. $x_1 = 2x_2 + 3x_4 - x_6$, $x_3 = 5x_4 + 6x_5 - 4x_6$; x_2, x_4, x_5, x_6 arbitrary.

Exercise 1.37. Show that a matrix may have several row echelon forms (same shape but different entries). However, explain that the reduced row echelon form of any matrix is unique.

1.4 Calculation in Euclidean Space

After we develop the theory of solutions of systems of linear equations, we may apply the theory to the calculation of span and linear independence in Euclidean space in Section 1.2.

In Example 1.8, we saw that the linear combination of $\vec{v}_1 = (1,2,3), \vec{v}_2 = (4,5,6), \vec{v}_3 = (7,8,9)$ is

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = x_1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + x_3 \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = \begin{pmatrix} x_1 + 4x_2 + 7x_3 \\ 2x_1 + 5x_2 + 8x_3 \\ 3x_1 + 6x_2 + 9x_3 \end{pmatrix}.$$

This is the left side of the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= b_1, \\ 2x_1 + 5x_2 + 8x_3 &= b_2, \\ 3x_1 + 6x_2 + 9x_3 &= b_3. \end{aligned}$$

In general, we use vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n \in \mathbb{R}^m$ to form columns of a matrix

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n), \quad \vec{v}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix}. \tag{1.5}$$

Then the linear combination is the left of the system of linear equations

$$\begin{aligned} x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n &= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = A\vec{x}. \end{aligned} \tag{1.6}$$

Calculation of Span

Expressing a vector $\vec{b} \in \mathbb{R}^m$ as a linear combination $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{b}$ is the same as solving the corresponding system of linear equations $A\vec{x} = \vec{b}$.

Example 1.17. Recall the row operation in Example 1.14

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & a & b \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & a-9 & b-12 \end{pmatrix}.$$

If $a \neq 9$, then the result is already a row echelon form with $1, -3, a-9$ as pivots. Since the last column is not pivot, the system has solution. This means that $(10, 11, b)$ is a linear combination of $(1,2,3), (4,5,6), (7,8,a)$.

If $a = 9$ and $b = 12$, then the result is already a row echelon form with $1, -3$ as pivots (and the last row is all 0). Again the last column is not pivot, and $(10, 11, 12)$ is a linear combination of $(1,2,3), (4,5,6), (7,8,9)$.

If $a = 9$ and $b \neq 12$, then the result is already a row echelon form with $1, -3, b-12$ as pivots. Since the last column is pivot, we find that $(10, 11, b)$ is not a linear combination of $(1,2,3), (4,5,6), (7,8,9)$.

Example 1.18. The row operation in Example 1.14 (and Example 1.17) gives the row operation

$$\begin{pmatrix} 1 & 4 & 7 & 10 & b_1 \\ 2 & 5 & 8 & 11 & b_2 \\ 3 & 6 & a & b & b_3 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 & b_1' \\ 0 & -3 & -6 & -9 & b_2' \\ 0 & 0 & a-9 & b-12 & b_3' \end{pmatrix}.$$

Here $\vec{b}' = (b_1', b_2', b_3')$ is obtained from $\vec{b} = (b_1, b_2, b_3)$ by performing the same sequence of row operations. each b_i' is a linear expression of b_1, b_2, b_3 (find the explicit formula!).

If $a \neq 9$ or $b \neq 12$, then either $a-9$ or $b-12$ is pivot. Therefore the last column is never pivot. This means that every vector $\vec{b} = (b_1, b_2, b_3)$ is a linear combination of $(1,2,3)$, $(4,5,6)$, $(7,8,a)$, $(10,11,b)$. In other words, the four vectors span $\mathbb{R}^3$.

If $a = 9$ and $b = 12$, then the last row is $(0\ 0\ 0\ 0\ b_3')$, and the last column is pivot if and only if $b_3' \neq 0$. We claim that it is always possible to find some $\vec{b}$ such that $b_3' \neq 0$. Then this $\vec{b}$ is not a linear combination of $(1,2,3)$, $(4,5,6)$, $(7,8,9)$, $(10,11,12)$. This means that the four vectors do not span $\mathbb{R}^3$.

Recall that $\vec{b}'$ is obtained from $\vec{b}$ by row operations

$$\vec{b} \xrightarrow{R_i \leftrightarrow R_j} (*) \xrightarrow{cR_i} (*) \xrightarrow{R_i + cR_j} \vec{b}'.$$

The row operations can be reversed

$$\vec{b} \xleftarrow{R_i \leftrightarrow R_j} (*) \xleftarrow{c^{-1}R_i} (*) \xleftarrow{R_i - cR_j} \vec{b}'.$$

We start with $\vec{b}' = (0,0,1)$ (so that $b_3' = 1 \neq 0$), and use the reverse row operations to get $\vec{b}$. Then this $\vec{b}$ is not a linear combination of $(1,2,3)$, $(4,5,6)$, $(7,8,9)$, $(10,11,12)$.

The discussion in Example 1.18 can be extended to the following criterion on when some vectors can span the Euclidean space.

Proposition 1.10. *Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n \in \mathbb{R}^m$ and $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$. The following are equivalent.*

1. *$\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span $\mathbb{R}^m$.*
2. *$A\vec{x} = \vec{b}$ has solution for all $\vec{b} \in \mathbb{R}^m$.*
3. *All rows of A are pivot. In other words, the row echelon form of A has no zero row $(0\ 0\ \cdots\ 0)$.*

Example 1.19. To determine whether the vectors $\vec{v}_1 = (1,2,3), \vec{v}_2 = (2,3,4), \vec{v}_3 = (3,4,1), \vec{v}_4 = (4,1,2)$ span $\mathbb{R}^3$, we carry out the row operation

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -7 \\ 0 & -2 & -8 & -10 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -7 \\ 0 & 0 & -4 & 4 \end{pmatrix}.$$

Since there is no zero row in the row echelon form, the four vectors span $\mathbb{R}^3$.

If we perform the same row operations on the first three columns only, then we get

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3) = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{pmatrix}.$$

This shows that $\vec{v}_1, \vec{v}_2, \vec{v}_3$ span $\mathbb{R}^3$. By the similar idea, $\vec{v}_1, \vec{v}_2, \vec{v}_4$ span $\mathbb{R}^3$, and $\vec{v}_1, \vec{v}_2$ do not span $\mathbb{R}^3$.

The vectors $\vec{v}_1, \vec{v}_2$ in Example 1.19 form a 3×2 matrix. Since each column has at most one pivot, the number of pivots is at most 2. Therefore among the 3 rows, at least one row is not pivot. By Proposition 1.10, we conclude that $\vec{v}_1, \vec{v}_2$ cannot span $\mathbb{R}^3$.

We have argued that two vectors cannot span $\mathbb{R}^3$, by considering the shape of the row echelon form of 3×2 matrix instead of calculating the actual row echelon form. In general, the argument gives the following result.

Proposition 1.11. *If n vectors span $\mathbb{R}^m$, then $n \geq m$.*

Equivalently, if $n < m$, then n vectors cannot span $\mathbb{R}^m$.

Example 1.20. By Proposition 1.11, we know that $(1,2,3,4), (5,6,7,8), (9,10,11,12)$ cannot span $\mathbb{R}^4$. Although it is easy to do row operations on the three vectors (see Example 1.22), there is no need to calculate.

Another example is that $(1, 0, \sqrt{2}, \pi), (\log 2, e, 100, -0.5), (\sqrt{3}, e^{-1}, \sin 1, 2.3)$ cannot span $\mathbb{R}^4$. The row operation calculation would be very complicated.

Exercise 1.38. Determine whether the vectors span the vector space. For those that do not span, add more vectors to achieve span.

1. $(1,1,0,0)$, $(-1,1,0,0)$, $(0,0,1,1)$, $(0,0,-1,1)$.
2. $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$.
4. $1 + t^2$, $1 + t - 2t^2$, $3t - t^2$, $t + 2t^2$.
5. $\vec{e}_1 - \vec{e}_2$, $\vec{e}_2 - \vec{e}_3$, $\dots$, $\vec{e}_{n-1} - \vec{e}_n$, $\vec{e}_n - \vec{e}_1$.
6. $\vec{e}_1, \vec{e}_1 + 2\vec{e}_2, \vec{e}_1 + 2\vec{e}_2 + 3\vec{e}_3, \dots, \vec{e}_1 + 2\vec{e}_2 + \cdots + n\vec{e}_n$.

Exercise 1.39. Prove that if $\vec{u}, \vec{v}, \vec{w}$ span V, then the linear combinations of vectors span V.

1. $\vec{u} + 2\vec{v} + 3\vec{w}$, $2\vec{u} + 3\vec{v} + 4\vec{w}$, $3\vec{u} + 4\vec{v} + \vec{w}$, $4\vec{u} + \vec{v} + 2\vec{w}$.

2. $\vec{u}+2\vec{v}+3\vec{w}$, $4\vec{u}+5\vec{v}+6\vec{w}$, $7\vec{u}+8\vec{v}+a\vec{w}$, $10\vec{u}+11\vec{v}+b\vec{w}$, $a\neq 9$ or $b\neq 12$.

In general, if the columns of a matrix

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \end{pmatrix}$$

span $\mathbb{R}^3$, then $a_{11}\vec{u}+a_{21}\vec{v}+a_{31}\vec{w}, a_{12}\vec{u}+a_{22}\vec{v}+a_{32}\vec{w},\dots,a_{1n}\vec{u}+a_{2n}\vec{v}+a_{3n}\vec{w}$ span V. Exercise 2.35 explains the property by using onto linear transformations.

Calculation of Linear Independence

Using $A\vec{x}=x_1\vec{v}_1+x_2\vec{v}_2+\cdots+x_n\vec{v}_n$, the definition of linear independence becomes that the solution of the system $A\vec{x}=\vec{b}$ is unique. This means that the general solution has no free variable. By Proposition 1.9, free variables correspond to the non-pivot columns of A. Therefore uniqueness means that all columns of A are pivot.

Proposition 1.12. *Let $\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n\in\mathbb{R}^m$ and $A=(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$. The following are equivalent.*

1. *$\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n$ are linearly independent.*
2. *The solution of $A\vec{x}=\vec{b}$ his unique.*
3. *All columns of A are pivot.*

Note that the criterion for the linear independence does not depend on the right side $\vec{b}$. In fact, Proposition 1.6 says that we only need to verify the unique linear combination for $\vec{b}=\vec{0}$. This is the same as the homogeneous system $A\vec{x}=\vec{0}$ has only the trivial solution $\vec{x}=\vec{0}$.

We remark that the criterion in Proposition 1.12 is independent of the right side $\vec{b}$. This implies that the solution of $A\vec{x}=\vec{b}$ is unique if and only if the solution of the *homogeneous system* $A\vec{x}=0$ is unique. The observation will be explained again by the structure of all solutions (Proposition 3.9).

Example 1.21. To determine whether the vectors $\vec{v}_1=(1,2,3),\vec{v}_2=(2,3,4),\vec{v}_3=(3,4,1),\vec{v}_4=(4,1,2)$ are linearly independent, we use the row operation in Example 1.19

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4)=\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \end{pmatrix}\to\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -7 \\ 0 & 0 & -4 & 4 \end{pmatrix}.$$

Since the last column is not pivot, the four vectors are linearly dependent.

If we remove the freedom (corresponding to $\vec{v}_4$) by considering $\vec{v}_1,\vec{v}_2,\vec{v}_3$ only, then the same row operations gives

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)=\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{pmatrix}\to\begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \end{pmatrix}.$$

Since all three columns are pivot, the three vectors are linearly independent. This is not surprising because removing freedom means no freedom, or uniqueness.

The same row operations show that $\vec{v}_1, \vec{v}_2, \vec{v}_4$ are linearly independent, and $\vec{v}_1, \vec{v}_2$ are also linearly independent.

Example 1.22. To determine the linear independence of $\vec{v}_1 = (1,2,3,4), \vec{v}_2 = (5,6,7,8), \vec{v}_3 = (9,10,11,a)$, we carry out the row oepration

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3) = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & a \end{pmatrix} \to \begin{pmatrix} 1 & 5 & 9 \\ 0 & 1 & 2 \\ 0 & 0 & a-12 \\ 0 & 0 & 0 \end{pmatrix}$$

If $a \neq 12$, then all columns are pivot, and the three vectors are linearly independent. If $a = 12$, then the third column is not pivot, and the three vectors are linearly dependent.

The four vectors in Example 1.21 form a 3×4 matrix. Since each row has at most one pivot, the number of pivots is at most 3. Therefore among the 4 columns, at least one column is not pivot. By Proposition 1.12, we conclude that the four vectors are linearly dependent. The argument can be extended to a general result.

Proposition 1.13. *If n vectors in $\mathbb{R}^m$ are linearly independent, then $n \leq m$.*

Equivalently, if $n > m$, then n vectors in $\mathbb{R}^m$ are linearly dependent.

Example 1.23. By Proposition 1.13, we know that $(1,2,3), (4,5,6), (7,8,a), (10,11,b)$ must be linearly dependent, no matter what a, b are. The corresponding row operation in Example 1.14 is not needed here.

Exercise 1.40. Determine linear independence. For linearly dependent collection, find maximal linearly independent subset.

1. $(1,1,0,0)$, $(-1,1,0,0)$, $(0,0,1,1)$, $(0,0,-1,1)$.
2. $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} -2 & -4 \\ -6 & -8 \end{pmatrix}$.
4. $1+t^2$, $1+t-2t^2$, $3t-t^2$, $t+2t^2$.
5. $\vec{e}_1 - \vec{e}_2$, $\vec{e}_2 - \vec{e}_3$, $\dots$, $\vec{e}_{n-1} - \vec{e}_n$, $\vec{e}_n - \vec{e}_1$.
6. $\vec{e}_1, \vec{e}_1 + 2\vec{e}_2, \vec{e}_1 + 2\vec{e}_2 + 3\vec{e}_3, \dots, \vec{e}_1 + 2\vec{e}_2 + \cdots + n\vec{e}_n$.

Exercise 1.41. Prove that for any $\vec{u}, \vec{v}, \vec{w}$, the linear combinations of vectors are always linearly dependent.

1. $\vec{u}+4\vec{v}+7\vec{w}$, $2\vec{u}+5\vec{v}+8\vec{w}$, $3\vec{u}+6\vec{v}+9\vec{w}$.
2. $\vec{u}+4\vec{v}+7\vec{w}$, $2\vec{u}+5\vec{v}+8\vec{w}$, $3\vec{u}+6\vec{v}+a\vec{w}$, $10\vec{u}+11\vec{v}+b\vec{w}$.

In general, if the columns of the matrix

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \end{pmatrix}$$

are linearly dependent, then $a_{11}\vec{u}+a_{21}\vec{v}+a_{31}\vec{w}, a_{12}\vec{u}+a_{22}\vec{v}+a_{32}\vec{w}, \dots, a_{1n}\vec{u}+a_{2n}\vec{v}+a_{3n}\vec{w}$ are always linearly dependent. Exercise 2.44 explains the property by using onto linear transformations.

1.5 Basis

In Section 1.4, we saw how to calculate span and linear independence in Euclidean space, by using the theory of solutions of system of linear equations in Section 1.3. In Example 1.10, we saw how to calculate span and linear independence in other vector spaces, by translating the problem to the Euclidean space. The translation is based on the following concept.

Definition 1.14. A set of vectors is a *basis* if they span the vector space and are linearly independent. In other words, any vector can be uniquely expressed as a linear combination of the vectors in the basis.

The basis is the perfect situation for a collection of vectors. Similarly, we may regard 1 as the "basis" for $\mathbb{N}$ with respect to the mechanism $+1$, and regard all prime numbers as the "basis" for $\mathbb{N}$ with respect to the multiplication.

Example 1.24. The *standard basis* of the Euclidean space $\mathbb{R}^n$ is

$$\vec{e}_1=(1,0,\dots,0),\ \vec{e}_2=(0,1,\dots,0),\ \dots,\ \vec{e}_n=(0,0,\dots,1).$$

The i-th coordinate of $\vec{e}_i$ is 1 and all the other coordinates are 0. We have

$$\begin{aligned} x_1\vec{e}_1+x_2\vec{e}_2+\cdots+x_n\vec{e}_n &= (x_1,0,\dots,0)+(0,x_2,\dots,0)+(0,0,\dots,x_n) \\ &= (x_1,x_2,\dots,x_n). \end{aligned}$$

The formula shows that any vector can be written as a linear combination of $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$, with the coordinates of the vector as the unique cooefficients.

Example 1.25. Any polynomial $p(t)=a_0+a_1t+a_2t^2+\cdots+a_nt^n$ of degree n is a linear combination of *monomials* $1, t, t^2, \dots, t^n$. Moreover, we know that two polynomials are qual if and only if their coefficients are equal. Therefore the monomials form a basis of the vector space P_n.

Consider the monomials $1, t-1, (t-1)^2$ at 1. Any quadratic polynomial

$$\begin{aligned} a_0 + a_1 t + a_2 t^2 &= a_0 + a_1[1 + (t-1)] + a_2[1 + (t-1)]^2 \\ &= (a_0 + a_1 + a_2) + (a_1 + 2a_2)(t-1) + a_2(t-1)^2 \end{aligned}$$

is a linear combination of $1, t-1, (t-1)^2$. Moreover, if two linear combinations are equal

$$a_0 + a_1(t-1) + a_2(t-1)^2 = b_0 + b_1(t-1) + b_2(t-1)^2,$$

then substituting t by $t+1$ gives the equality

$$a_0 + a_1 t + a_2 t^2 = b_0 + b_1 t + b_2 t^2.$$

This means that $a_0 = b_0$, $a_1 = b_1$, $a_2 = b_2$. We just argued that $1, t-1, (t-1)^2$ is a basis of P_2. In general, $1, t-1, (t-1)^2, \dots, (t-1)^n$ is a basis of P_n.

Example 1.26. For 3×2 matrices, we have

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix} = x_{11}\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} + x_{21}\begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} + x_{31}\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} + x_{12}\begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} + x_{22}\begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} + x_{32}\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Similar to Example 1.24, this implies that the six special matrices on the right (one entry is 1 and all other entries are 0) form a basis of $M_{3\times 2}$.

Example 1.27. To determine whether

$$\vec{v}_1 = (1, -1, 0), \quad \vec{v}_2 = (1, 0, -1), \quad \vec{v}_3 = (1, 1, 1)$$

form a basis of $\mathbb{R}^3$, we carry out the row operation

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

We find the all rows are pivot, which means that the vectors span $\mathbb{R}^3$. We also find the all columns are pivot, which means that the vectors are linearly independent. Therefore the three vectors form a basis of $\mathbb{R}^3$.

Example 1.28. The three vectors of $\mathbb{R}^3$

$$\vec{v}_1 = (1, 2, 3), \quad \vec{v}_2 = (4, 5, 6), \quad \vec{v}_3 = (7, 8, 9)$$

do not form a basis because they do not span $\mathbb{R}^3$ by Example 1.8. Another reason (one reason is enough) for not being a basis is that the three vectors are linearly dependent by Example 1.9.

Examples 1.8 and 1.9 also show that the vectors

$$\vec{v}_1 = (1, 2, 3), \quad \vec{v}_2 = (4, 5, 6), \quad \vec{v}_3 = (7, 8, a)$$

form a basis is and only if $a \neq 9$.

Exercise 1.42. Show that the collection is a basis.

1. $(1,1,0),(1,0,1),(0,1,1)$ in $\mathbb{R}^3$.
2. $(1,1,1,0),(1,1,0,1),(1,0,1,1),(0,1,1,1)$ in $\mathbb{R}^4$.
3. $1+t, 1+t^2, t+t^2$ in P_2.
4. $\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}$ in $M_{2\times 2}$.

Exercise 1.43. For which a is the collection is a basis?

1. $(1,1,0),(1,0,1),(0,1,a)$ in $\mathbb{R}^3$.
2. $(1,1,1,0),(1,1,0,1),(1,0,1,1),(0,1,1,a)$ in $\mathbb{R}^4$.
3. $1+t, 1+t^2, t+at^2$ in P_2.
4. $\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\ 1 & a\end{pmatrix}$ in $M_{2\times 2}$.

Exercise 1.44. Show that $(a,b),(c,d)$ form a basis of $\mathbb{R}^2$ if and only if $ad \neq bc$. What is the condition for a to be a basis of $\mathbb{R}^1$?

Exercise 1.45. Show that α is a basis if and only if β is a basis.

1. $\alpha = \{\vec{v}_1, \vec{v}_2\}$, $\beta = \{\vec{v}_1+\vec{v}_2, \vec{v}_1-\vec{v}_2\}$.
2. $\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, $\beta = \{\vec{v}_1+\vec{v}_2, \vec{v}_1+\vec{v}_3, \vec{v}_2+\vec{v}_3\}$.
3. $\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, $\beta = \{\vec{v}_1, \vec{v}_1+\vec{v}_2, \vec{v}_1+\vec{v}_2+\vec{v}_3\}$.
4. $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$, $\beta = \{\vec{v}_1, \vec{v}_1+\vec{v}_2, \dots, \vec{v}_1+\vec{v}_2+\cdots+\vec{v}_n\}$.

Exercise 2.50 gives a generalisation.

Exercise 1.46. Use Exercises 1.16, 1.17, 1.18 to show that $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is a basis if and only if any one of the following is a basis.

1. $\{\vec{v}_1, \dots, \vec{v}_j, \dots, \vec{v}_i, \dots, \vec{v}_n\}$.
2. $\{\vec{v}_1, \dots, c\vec{v}_i, \dots, \vec{v}_n\}$, $c \neq 0$.
3. $\{\vec{v}_1, \dots, \vec{v}_i + c\vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n\}$.

Exercise 2.50 gives a generalisation.

Coordinate

Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is an (ordered) basis of V. Then any vector $\vec{x} \in V$ can be uniquely expressed as a linear combination

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n.$$

The coefficients $x_1, x_2, \dots, x_n$ are the *coordinates* of $\vec{x}$ with respect to the basis. The unique expression means that the α-coordinate map

$$\vec{x} \in V \mapsto [\vec{x}]_\alpha = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$$

is well defined. The linear combination gives the reverse map

$$(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mapsto x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \in V.$$

The two way maps identify (called *isomorphism*) the general vector space V with the Euclidean space $\mathbb{R}^n$. Moreover, the coordinate preserves the addition and scalar multiplication.

Proposition 1.15. $[\vec{x} + \vec{y}]_\alpha = [\vec{x}]_\alpha + [\vec{y}]_\alpha$, $[a\vec{x}]_\alpha = a[\vec{x}]_\alpha$.

Proof. Let $[\vec{x}]_\alpha = (x_1, x_2, \dots, x_n)$ and $[\vec{y}]_\alpha = (y_1, y_2, \dots, y_n)$. Then by the definition of coordinates, we have

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n, \quad \vec{y} = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n.$$

Adding the two together, we have

$$\vec{x} + \vec{y} = (x_1 + y_1)\vec{v}_1 + (x_2 + y_2)\vec{v}_2 + \cdots + (x_n + y_n)\vec{v}_n.$$

By the definition of coordinates, this means

$$[\vec{x} + \vec{y}]_\alpha = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n) = (x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = [\vec{x}]_\alpha + [\vec{y}]_\alpha.$$

The proof of $[a\vec{x}]_\alpha = a[\vec{x}]_\alpha$ is similar. □

The proposition implies that the coordinate identification preserves linear combinations. Therefore we can use the coordinate to translate linear algebra problems (such as span and linear independence) in general vector space to corresponding problems in Euclidean space. The problems in Euclidean space can then be solved by further translating into problems about systems of linear equations.

Example 1.29. Denote the standard basis (in the standard order) by $\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$. Then the equality $(x_1, x_2, \dots, x_n) = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n$ means

$$[(x_1, x_2, \dots, x_n)]_\epsilon = (x_1, x_2, \dots, x_n).$$

We may write $[\vec{x}]_\epsilon = \vec{x}$ for short.

If we change the order in the standard basis, then we should also change the order of coordinates

$$\begin{aligned}
[(x_1, x_2, x_3)]_{\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}} &= (x_1, x_2, x_3), \\
[(x_1, x_2, x_3)]_{\{\vec{e}_2, \vec{e}_1, \vec{e}_3\}} &= (x_2, x_1, x_3), \\
[(x_1, x_2, x_3)]_{\{\vec{e}_3, \vec{e}_2, \vec{e}_1\}} &= (x_3, x_2, x_1).
\end{aligned}$$

Example 1.30. The coordinates of a polynomial with respect to the monomial basis is simply the coefficients in the polynomial

$$[a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n]_{\{1, t, t^2, \ldots, t^n\}} = (a_0, a_1, a_2, \ldots, a_n).$$

This is used in Example 1.10 to translate linear algebra problems for polynomials. For example, the polynomials $1 + 2t + 3t^2, 4 + 5t + 6t^2, 7 + 8t + at^2$ form a basis of P_2 if and only if their coordinates with respect to the monomial basis $\alpha = \{1, t, t^2\}$

$$[1 + 2t + 3t^2]_\alpha = (1, 2, 3), \quad [4 + 5t + 6t^2]_\alpha = (4, 5, 6), \quad [7 + 8t + at^2]_\alpha = (7, 8, a)$$

form a basis of $\mathbb{R}^3$. By Example 1.28, this happens if and only if $a \neq 9$.

Exercise 1.47. For an ordered basis $\alpha = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is of V, explain that $[\vec{v}_i]_\alpha = \vec{e}_i$.

Exercise 1.48. A *permutation* of $\{1, 2, \ldots, n\}$ is a one-to-one correspondence $\pi \colon \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. Let $\alpha = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis of V.

1. Show that $\pi(\alpha) = \{\vec{v}_{\pi(1)}, \vec{v}_{\pi(2)}, \ldots, \vec{v}_{\pi(n)}\}$ is still a basis.

2. What is the relation between the coordinates with respect to α and with respect to $\pi(\alpha)$?

Exercise 1.49. Given the coordinates with respect to basis, what are the coordinates with respect to the new bases in Exercise 1.46?

Example 1.31. We would like to calculate the coordinates with respect to the basis $\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \{(1, -1, 0), (1, 0, -1), (1, 1, 1)\}$ in Example 1.27. Finding the coordinate of $\vec{e}_1 = (1, 0, 0)$ means solving $x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{e}_1$. This is a system of linear equations. We can carry out the row operation on the augmented matrix similar to Example 1.27 and then further obtain the reduced row echelon form

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{e}_1) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 3 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{pmatrix}.$$

We can read the answer directly from the educed row echelon form and get $[\vec{e}_1]_\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

We can get the coordinates $[\vec{e}_2]_\alpha$ and $[\vec{e}_3]_\alpha$ by similar row operations. In fact, we may combine the three row operations to get

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{e}_1\ \vec{e}_2\ \vec{e}_3) = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

If we restrict the row operation to the first four columns $[\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{e}_1]$, then we get the coordinate $[\vec{e}_1]_\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ to be the fourth column of the reduced row echelon form. If we restrict to $[\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{e}_2]$, then we get the coordinate $[\vec{e}_2]_\alpha = (-\frac{2}{3}, \frac{1}{3}, \frac{1}{3})$ to be the fifth column. If we restrict to $[\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{e}_3]$, then we get the coordinate $[\vec{e}_3]_\alpha = (\frac{1}{3}, -\frac{2}{3}, \frac{1}{3})$ to be the fifth column.

By Proposition 1.15, the α-coordinate of a general vector in $\mathbb{R}^3$ is

$$\begin{aligned}
[(x, y, z)]_\alpha &= [x\vec{e}_1 + y\vec{e}_2 + z\vec{e}_3]_\alpha = x[\vec{e}_1]_\alpha + y[\vec{e}_2]_\alpha + z[\vec{e}_1]_\alpha \\
&= x(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}) + y(-\tfrac{2}{3}, \tfrac{1}{3}, \tfrac{1}{3}) + z(\tfrac{1}{3}, -\tfrac{2}{3}, \tfrac{1}{3}) \\
&= \begin{pmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.
\end{aligned}$$

Exercise 1.50. Find the coordinates of a general vector in Euclidean space with respect to basis.

1. $(0, 1), (1, 0)$.
2. $(1, 2), (3, 4)$.
3. $(a, 0), (0, b)$, $a, b \neq 0$.
4. $(\cos\theta, \sin\theta), (-\sin\theta, \cos\theta)$.
5. $(1, 1, 0), (1, 0, 1), (0, 1, 1)$.
6. $(1, 2, 3), (0, 1, 2), (0, 0, 1)$.
7. $(0, 1, 2), (0, 0, 1), (1, 2, 3)$.
8. $(-1, 1, 2), (0, 1, 1), (0, 1, 0)$.

Exercise 1.51. Determine when the collection is a basis of $\mathbb{R}^n$. Moreover, find the coordinates with respect to the basis.

1. $\vec{e}_1 - \vec{e}_2,\ \vec{e}_2 - \vec{e}_3,\ \dots,\ \vec{e}_{n-1} - \vec{e}_n,\ \vec{e}_n - \vec{e}_1$.
2. $\vec{e}_1 - \vec{e}_2,\ \vec{e}_2 - \vec{e}_3,\ \dots,\ \vec{e}_{n-1} - \vec{e}_n,\ \vec{e}_1 + \vec{e}_2 + \cdots + \vec{e}_n$.
3. $\vec{e}_1 + \vec{e}_2,\ \vec{e}_2 + \vec{e}_3,\ \dots,\ \vec{e}_{n-1} + \vec{e}_n,\ \vec{e}_n + \vec{e}_1$.
4. $\vec{e}_1 + \vec{e}_2 + \vec{e}_3,\ \vec{e}_2 + \vec{e}_3 + \vec{e}_4,\ \dots,\ \vec{e}_{n-1} + \vec{e}_n + \vec{e}_1,\ \vec{e}_n + \vec{e}_1 + \vec{e}_2$.
5. $\vec{e}_1, \vec{e}_1 + 2\vec{e}_2, \vec{e}_1 + 2\vec{e}_2 + 3\vec{e}_3,\ \dots,\ \vec{e}_1 + 2\vec{e}_2 + \cdots + n\vec{e}_n$.

Exercise 1.52. Determine when the collection is a basis of P_n. Moreover, find the coordinates with respect to the basis.

1. $1 - t,\ t - t^2,\ \dots,\ t^{n-1} - t^n,\ t^n - 1$.
2. $1 + t,\ t + t^2,\ \dots,\ t^{n-1} + t^n,\ t^n + 1$.

3. $1, 1+t, 1+t^2, \dots, 1+t^n$.
4. $1, t-1, (t-1)^2, \dots, (t-1)^n$.

Exercise 1.53. Find the coordinates of a general vector with respect to the basis in Exercise 1.42.

Exercise 1.54. Suppose $ad \neq bc$. Find the coordinate of a vector in $\mathbb{R}^2$ with respect to the basis $(a, b), (c, d)$ (see Exercise 1.44).

Construct Basis from Spanning Set

Basis means span plus linear independence. If we only have the span property, then we can achieve linear independence (and therefore basis) by deleting "unnecessary" vectors.

Definition 1.16. A vector space is *finite dimensional* if its is spanned by finitely many vectors.

Theorem 1.17. *In a finite dimensional vector space, a set of vectors is a basis if and only if it is a minimal spanning set. Moreover, any finite spanning set contains a minimal spanning set and therefore a basis.*

By a minimal spanning set, we mean that the set spans V, and any strictly smaller subset does not span V.

The theorem implies that any finite dimensional vector space has a basis.

Proof. Suppose finitely many vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ span V. The theorem is the consequence of the following two statements.

1. If α is not a minimal spanning set, then α is linearly dependent, and α contains a minimal spanning set.
2. If α is a minimal spanning set, then α is linearly independent, and is therefore a basis.

If α is not minimal, then a strictly smaller subset $\alpha' \subset \alpha$ also spans V. This implies that any vector in $\alpha - \alpha'$ is a linear combination of vectors in α'. By Proposition 1.7, we find that α is linearly dependent. Moreover, we may ask again whether α' is a minimal spanning set. If not, we may further get a strictly smaller spanning subset of α'. Since α is finite, we eventually obtain a minimal spanning set inside α. This proves that α contains a minimal spanning set.

If α is minimal and linearly dependent, then by Proposition 1.7 and without loss of generality, we may assume $\vec{v}_n = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_{n-1}\vec{v}_{n-1}$. Since α spans V, any vector $\vec{x} \in V$ is a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$, and we have

$$\begin{aligned}
\vec{x} &= x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \\
&= x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_{n-1}\vec{v}_{n-1} + x_n(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_{n-1}\vec{v}_{n-1}) \\
&= (x_1 + x_nc_1)\vec{v}_1 + (x_2 + x_nc_2)\vec{v}_2 + \cdots + (x_{n-1} + x_nc_{n-1})\vec{v}_{n-1}.
\end{aligned}$$

This shows that $\alpha' = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{n-1}\}$ is a spanning set strictly smaller than α, a contradiction to the minimality of α. $\square$

The intuition behind Theorem 1.17 is the following. Imagine that α is all the people in a company, and V is all the things the company wants to do. Then α spanning V means that the company can do all the things it wants to do. However, the company may not be efficient in the sense that if somebody's duty can be fulfilled by the others (the person is a linear combination of the others), then the company can fire the person and still do all the things. By firing unnecessary persons one after another, eventually everybody is indispensable. The result is that the company can do everything, and is also the most efficient.

Example 1.32. In Examples 1.19 and 1.21, we use the row operation

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -2 & -7 \\ 0 & 0 & -4 & 4 \end{pmatrix}$$

to show that $\vec{v}_1 = (1,2,3), \vec{v}_2 = (2,3,4), \vec{v}_3 = (3,4,1), \vec{v}_4 = (4,1,2)$ span $\mathbb{R}^3$ and are linearly dependent. If we view the 3×4 matrix as the augmented matrix of the system of linear equations $x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{v}_4$, then the row operation implies that the system has solution. This means that $\vec{v}_4$ is a linear combination of the other three vectors, and is therefore a "waste". By deleting $\vec{v}_4$, the same row operation applied to $\vec{v}_1, \vec{v}_2, \vec{v}_3$ shows that the three vectors are linearly independent (see Example 1.21). Therefore the three vectors form a basis.

Construct Basis from Linearly Independent Set

If we only have the linear independence, then we can achieve span property (and therefore basis) by adding "independent" vectors. Therefore we have two ways of constructing basis

$$\text{span vector space} \xrightarrow{\text{delete vectors}} \text{basis} \xleftarrow{\text{add vectors}} \text{linearly independent}$$

Using the analogy of company, linear independence means there is no waste. What we need to achieve is to do all the things the company wants to do. If there is a job that the existing employees cannot do, then we find somebody who can do the job. The new hire is linearly independent of the existing employees because the person can do something the others cannot do. We keep adding new necessary people (necessary means independent) until the company can do all the things, and therefore achieve the span.

Theorem 1.18. *In a finite dimensional vector space, a set of vectors is a basis if and only if it is a maximal linearly independent set. Moreover, any finite linearly independent set can be extended to a maximal linearly independent set and therefore a basis.*

By a maximal linearly independent set, we mean that the set is linearly independent, and any strictly bigger subset is linearly dependent.

Proof. Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is a linearly independent set in V.

If α is not maximal, then we will prove that α does not span V, and is therefore not a basis. Since α is not maximal, a strictly bigger subset $\alpha' \supset \alpha$ of V is also linearly independent.

By Proposition 1.7, any vector in $\alpha' - \alpha$ is not a linear combination of vectors in α. This shows that α does not span V. Moreover, we may ask again whether α' is a maximal linearly independent set. If not, we may further get a linearly independent strictly that is bigger than α'. We can keep going. The problem is why the process stops after finitely many steps, so that we eventually get a maximal linearly independent set. This is a consequence of finite dimension and is done after Proposition 1.20.

If α is maximal, then we will prove that α spans V, and therefore is a basis. For any $\vec{x} \in V$, since α is maximal linearly independent, adding $\vec{x}$ to α forms a linearly dependent set. This means that we have

$$a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n + c\vec{x} = \vec{0},$$

where some coefficients are nonzero. If $c = 0$, then $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n = \vec{0}$. By the linear independence of α, we get $a_1 = a_2 = \cdots = a_n = 0$. Since some coefficients are nonzero, this shows that $c \neq 0$, and we get

$$\vec{x} = -\frac{a_1}{c}\vec{v}_1 - \frac{a_2}{c}\vec{v}_2 - \cdots - -\frac{a_n}{c}\vec{v}_n.$$

This proves that every vector in V is a linear combination of vectors in α. Therefore α spans V. $\square$

Example 1.33. We take the transpose of the matrix in Examples 1.19 and 1.21 and carry out the row operation (this is the *column operation* on the matrix in the earlier examples)

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \\ 4 & 1 & 2 \end{pmatrix} \xrightarrow[R_4-4R_1]{\substack{R_2-2R_1 \\ R_3-3R_1}} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -2 & -8 \\ 0 & -7 & -10 \end{pmatrix} \xrightarrow[R_4-4R_2]{R_3-2R_2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \\ 0 & 0 & 4 \end{pmatrix} \xrightarrow{R_4+R_3} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -4 \\ 0 & 0 & 0 \end{pmatrix}.$$

This shows that $\vec{w}_1 = (1, 2, 3, 4), \vec{w}_2 = (2, 3, 4, 1), \vec{w}_3 = (3, 4, 1, 2)$ are linearly independent. However, since the last row is not pivot, the three vectors do not span $\mathbb{R}^4$.

We need to find a new vector $\vec{w}_4$ that is not a linear combination of $\vec{w}_1, \vec{w}_2, \vec{w}_3$. One idea is to postulate that the same row operation on $(\vec{w}_1\ \vec{w}_2\ \vec{w}_3\ \vec{w}_4)$ gives $(0, 0, 0, 1)$ as the last column. This means that $\vec{w}_4$ is obtained by reversing the operations on $(0, 0, 0, 1)$

$$\vec{w}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \xleftarrow[R_4+4R_1]{\substack{R_2+2R_1 \\ R_3+3R_1}} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \xleftarrow[R_4+4R_2]{R_3+2R_2} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \xleftarrow{R_4-R_3} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

This implies the row operation

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

We conclude that $\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4$ form a basis of $\mathbb{R}^4$.

We may try a more interesting vector $(4, 3, 2, 1)$

$$\begin{pmatrix} 4 \\ 11 \\ 20 \\ 27 \end{pmatrix} \xleftarrow[R_4+4R_1]{\substack{R_2+2R_1 \\ R_3+3R_1}} \begin{pmatrix} 4 \\ 3 \\ 8 \\ 11 \end{pmatrix} \xleftarrow[R_4+4R_2]{R_3+2R_2} \begin{pmatrix} 4 \\ 3 \\ 2 \\ -1 \end{pmatrix} \xleftarrow{R_4-R_3} \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}.$$

This shows that $\vec{w}_1, \vec{w}_2, \vec{w}_3$ and $(4, 11, 20, 27)$ form a basis of $\mathbb{R}^4$.

Dimension

Let V be a finite dimensional vector space. By Theorem 1.17, V has a basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$. Then the coordinate with respect to the basis translates the linear algebra of V to the Euclidean space $[\cdot]_\alpha \colon V \leftrightarrow \mathbb{R}^n$.

Let $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$ be another basis of V. Then the α-coordinate translates β into a basis $[\beta]_\alpha = \{[\vec{w}_1]_\alpha, [\vec{w}_2]_\alpha, \dots, [\vec{w}_m]_\alpha\}$ of $\mathbb{R}^n$. Therefore $[\beta]_\alpha$ spans $\mathbb{R}^n$ and is also linearly independent. Since $[\beta]_\alpha$ spans $\mathbb{R}^n$, by Proposition 1.11 (we have m vectors in $\mathbb{R}^n$, and the proposition is about n vectors in $\mathbb{R}^m$), we get $m \geq n$. Since $[\beta]_\alpha$ is linearly independent, by Proposition 1.13 (again m and n are switched), we get $m \leq n$. Therefore $m = n$.

We have argued that the following concept is well defined.

Definition 1.19. The *dimension* of a (finite dimensional) vector space is the number of vectors in a basis.

The dimension of V is denoted $\dim V$. By Examples 1.29, 1.30, 1.26, we have $\dim \mathbb{R}^n = n$, $\dim P_n = n + 1$, $\dim M_{m\times n} = mn$.

If $\dim V = n$, then V can be identified with the Euclidean space $\mathbb{R}^n$, as far as linear algebra problems are concerned. For example, we may change $\mathbb{R}^m$ in Propositions 1.11 and 1.13 to any vector space of dimension m, and get the following generalisations.

Proposition 1.20. *Suppose V is a finite dimensional vector space.*

1. *If n vectors span V, then $n \geq \dim V$.*
2. *If n vectors in V are linearly independent, then $n \leq \dim V$.*

Continuation of the proof of Theorem 1.18. The proof of the theorem creates bigger and bigger linearly independent set of vectors. However, by Proposition 1.20, the set is no longer linearly independent when the number of vectors is $> \dim V$. This means that, if the set α we start with has n vectors, then the construction in the proof stops after at most $\dim V - n$ steps.

We note that the argument uses Theorem 1.17, for the existence of basis and then the concept of dimension. The theorem we want to prove is not used in the argument. $\square$

If we combine the two statements, then the proposition says that the number of vectors in a basis of V is exactly $\dim V$. The following is sort of the converse of this statement.

Theorem 1.21. *Suppose α is a collection of vectors in a finite dimensional vector space V. If the number of vectors in α is $\dim V$, then the following are equivalent*

1. *α spans V.*
2. *α is linearly independent.*
3. *α is a basis.*

Proof. By using a basis of V to translate the properties to Euclidean space, we may assume that $V = \mathbb{R}^n$ is the Euclidean space. Then we use vectors in α as columns of a matrix A. The assumption on the number of vectors in α means that A is an $n \times n$ matrix (a square matrix).

By Proposition 1.10, α spanning V means that all rows of A are pivot. By Proposition 1.12, α linearly independent means that all columns of A are pivot. Since A is an $n \times n$ matrix, we have the following equivalences

$$\begin{aligned} \alpha \text{ spans } V &\iff \text{all } n \text{ rows pivot} \\ &\iff \text{number of pivots is } n \\ &\iff \text{all } n \text{ columns pivot} \\ &\iff \alpha \text{ is linearly independent.} \end{aligned}$$

Since basis means spanning V and linearly independent, the equivalence of the first two statements implies the equivalence of all three statements. $\square$

We may use Propositions 1.10 and 1.12 to rephrase Proposition 1.21 in terms of the existence and uniqueness of solutions of systems of linear equations.

Theorem 1.22. *Suppose in a system of linear equations $A\vec{x} = \vec{b}$, the number of equations is the same as the number of variables. Then $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$ if and only if the solution of $A\vec{x} = \vec{b}$ is unique.*

Example 1.34. By Proposition 1.20, the matrices $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}, \begin{pmatrix} 9 & 10 \\ 11 & 12 \end{pmatrix}$ cannot span $M_{2\times 2}$. Although this can be verified by row operations in Example 1.22, there is no need to calculate. In fact, we also know that $\begin{pmatrix} 1 & 0 \\ \sqrt{2} & \pi \end{pmatrix}, \begin{pmatrix} \log 2 & e \\ 100 & -0.5 \end{pmatrix}, \begin{pmatrix} \sqrt{3} & e^{-1} \\ \sin 1 & 2.3 \end{pmatrix}$ cannot span $M_{2\times 2}$. THe examples can be compared with Example 1.20.

Similar to Example 1.23, Proposition 1.20 also implies that the polynomials $1+2t+3t^2, 4+5t+6t^2, 7+8t+at^2, 10+11t+bt^2$ are linearly dependent, no matter what a, b are.

Example 1.35. By the equality

$$a+bt+ct^2 = (3a+5b-8c)(-2+3t+t^2)+(a+2b-4c)(3-5t-2t^2)+(a+b-c)(4-4t-t^2),$$

we know $-2+3t+t^2, 3-5t-2t^2, 4-4t-t^2$ span P_2. By Theorem 1.21, the three polynomials form a basis of P_2.

Exercise 1.55. Use Theorem 1.17 to give another proof of the first part of Proposition 1.20. Use Theorem 1.18 to give another proof of the second part of Proposition 1.20.

2 Linear Transformation

Chapter 1 discusses what happens within one vector space. This chapter discusses the relation between different vector spaces. The key concept here is the linear transformation, which are the maps that preserve the addition and scalar multiplication. We also discuss the general concepts about maps such as onto, one-to-one, and invertibility, and specialise these concepts to linear transformations.

For the calculation, we use matrix as the formula for linear transformations. The calculation of properties such as onto, one-to-one, and inverse are translated into the row operations on the corresponding matrices.

2.1 Definition

Definition 2.1. A map $L\colon V \to W$ between vector spaces is a *linear transformation* if it preserves two operations in vector spaces

$$L(\vec{u}+\vec{v}) = L(\vec{u})+L(\vec{v}), \quad L(a\vec{u}) = aL(\vec{u}).$$

If $V = W$, then we also call L a *linear operator*.

The geometrical meaning of linear transformation is preserving parallelogram and scaling. For example, the rotation of the plane by certain angle is a linear transformation, and the projection of $\mathbb{R}^3$ to a plane passing through the origin is also a linear transformation.

Example 2.1. The identity map $I(\vec{v}) = \vec{v}\colon V \to V$ is a linear operator. We also denote by I_V to emphasise the vector space V.

The zero map $O(\vec{v}) = \vec{0}\colon V \to W$ is a linear transformation.

Example 2.2. Proposition 1.15 means that the α-coordinate map is a linear transformation.

Example 2.3. The rotation R_θ of the plane by angle θ and the reflection (flipping) F_θ by angle ρ are linear because they clearly preserve parallelogram and scaling.

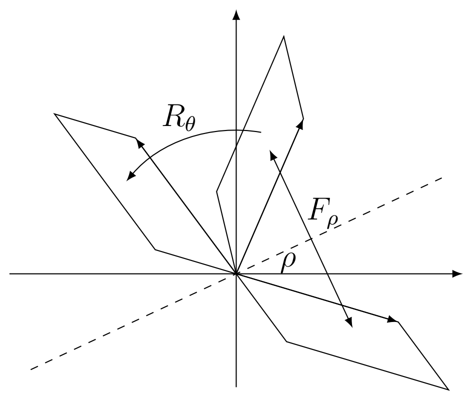

Figure 2.1: Rotation by angle θ and flipping by angle ρ.

Example 2.4. The projection of $\mathbb{R}^3$ to a plane $\mathbb{R}^2$ passing through the origin is a linear transformation.

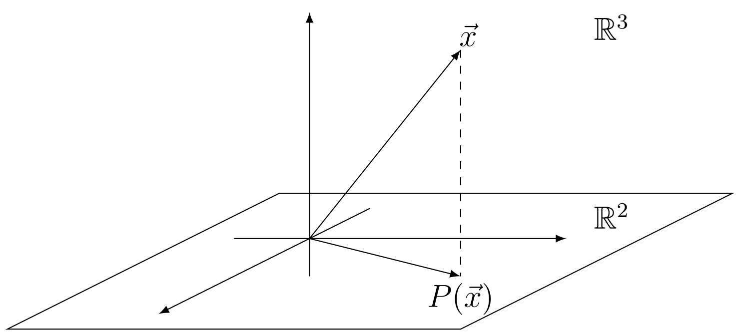

Figure 2.2: Projection from $\mathbb{R}^3$ to $\mathbb{R}^2$.

Example 2.5. The evaluation of functions at several places is a linear transformation

$$L(f) = (f(0), f(1), f(2)) \colon C^\infty \to \mathbb{R}^3.$$

In the reverse direction, the linear combination of some functions is a linear transformation

$$L(x_1, x_2, x_3) = x_1 \cos t + x_2 \sin t + x_3 e^t \colon \mathbb{R}^3 \to C^\infty.$$

The idea is extended in Exercise 2.14.

Example 2.6. In C^∞, taking the derivative is a linear operator

$$f \mapsto f' \colon C^\infty \to C^\infty.$$

The integration is a linear operator

$$f(t) \mapsto \int_0^t f(\tau) d\tau \colon C^\infty \to C^\infty.$$

Multiplying a fixed function $a(t)$ is also a linear operator

$$f(t) \mapsto a(t)f(t) \colon C^\infty \to C^\infty.$$

Exercise 2.1. Which are linear transformations?

1. $(x_1, x_2, x_3) \mapsto (x_1, x_2 + x_3) \colon \mathbb{R}^3 \to \mathbb{R}^2$.
2. $(x_1, x_2, x_3) \mapsto (x_1, x_2 x_3) \colon \mathbb{R}^3 \to \mathbb{R}^2$.
3. $(x_1, x_2, x_3) \mapsto (x_3, x_1, x_2) \colon \mathbb{R}^3 \to \mathbb{R}^3$.
4. $(x_1, x_2, x_3) \mapsto x_1 + 2x_2 + 3x_3 \colon \mathbb{R}^3 \to \mathbb{R}$.

Exercise 2.2. Which are linear transformations?

1. $f \mapsto f^2 \colon C^\infty \to C^\infty$.
2. $f(t) \mapsto f(t^2) \colon C^\infty \to C^\infty$.
3. $f \mapsto f'' \colon C^\infty \to C^\infty$.
4. $f(t) \mapsto f(t-2) \colon C^\infty \to C^\infty$.
5. $f(t) \mapsto f(2t) \colon C^\infty \to C^\infty$.
6. $f \mapsto f' + 2f \colon C^\infty \to C^\infty$.
7. $f \mapsto (f(0), f(1), f(2)) \colon C^\infty \to \mathbb{R}^3$.
8. $f \mapsto f(0)f(1) \colon C^\infty \to \mathbb{R}$.
9. $f \mapsto \int_0^1 f(t)dt \colon C^\infty \to \mathbb{R}$.
10. $f \mapsto \int_0^t \tau f(\tau)d\tau \colon C^\infty \to C^\infty$.

Combination of Linear Transformation

We can add two linear transformations $L, K \colon V \to W$ by

$$(L + K)(\vec{v}) = L(\vec{v}) + K(\vec{v}) \colon V \to W.$$

The following shows that $L + K$ is preserves the addition

$$\begin{aligned}
(L+K)(\vec{u}+\vec{v}) &= L(\vec{u}+\vec{v}) + K(\vec{u}+\vec{v}) && \text{(definition of } L+K\text{)} \\
&= (L(\vec{u}) + L(\vec{v})) + (K(\vec{u}) + K(\vec{v})) && \text{(}L \text{ and } K \text{ preserve addition)} \\
&= L(\vec{u}) + K(\vec{u}) + L(\vec{v}) + K(\vec{v}) && \text{(Axioms 1 and 2 of vector space)} \\
&= (L+K)(\vec{u}) + (L+K)(\vec{v}). && \text{(definition of } L+K\text{)}
\end{aligned}$$

We can similarly verify $(L + K)(a\vec{u}) = a(L + K)(\vec{u})$.

We may also multiply a number to a linear transformation by

$$(aL)(\vec{v}) = aL(\vec{v}) \colon V \to W.$$

The proof that aL is a linear transformation is similar.

We have introduced addition $L + K$ and scalar multiplication aL in the set $\mathrm{Hom}(V, W)$ of all linear transformations from V to W.

Proposition 2.2. $\mathrm{Hom}(V, W)$ *is a vector space.*

Proof. The following proves $L+K=K+L$

$$(L+K)(\vec{u})=L(\vec{u})+K(\vec{u})=K(\vec{u})+L(\vec{u})=(K+L)(\vec{u}).$$

The first and third equalities are due to the definition of addition in $\mathrm{Hom}(V,W)$. The second equality is due to Axiom 1 of vector space.

The associativity $(L+K)+K'=L+(K+K')$ can be proved similarly. The zero vector in $\mathrm{Hom}(V,W)$ is the zero transformation $O(\vec{v})=\vec{0}$ in Example 2.1. The negative of $L\in\mathrm{Hom}(V,W)$ is $K(\vec{v})=-L(\vec{v})$. The other axioms can also be verified, and are left as exercises. □

We can compose two linear transformations $K\colon U\to V$ and $L\colon V\to W$ if the domain and range match

$$(L\circ K)(\vec{v})=L(K(\vec{v}))\colon U\to W.$$

The composition preserves the addition

$$\begin{aligned}(L\circ K))(\vec{u}+\vec{v})&=L(K(\vec{u}+\vec{v}))=L(K(\vec{u})+K(\vec{v}))\\&=L(K(\vec{u}))+L(K(\vec{v}))=(L\circ K))(\vec{u})+(L\circ K))(\vec{v}).\end{aligned}$$

The first and fourth equalities are the definition of composition. The second and third equalities are the linearity of L and K. We can similarly prove that the composition also preserves the scalar multiplication. Therefore the composition is a linear transformation.

Example 2.7. The composition of two rotations is still a rotation: $R_{\theta_1}\circ R_{\theta_2}=R_{\theta_1+\theta_2}$.

Example 2.8. Consider the differential equation

$$f''+(1+t^2)f'+tf=3t+1.$$

The left is the addition of three transformations $f\mapsto f''$, $f\mapsto(1+t^2)f'$, $f\mapsto tf$.

Let $D(f)=f'$ be the derivative linear transformation in Example 2.6. Let $M_a(f)=af$ be the linear transformation of multiplying a function $a(t)$. Then $f\mapsto f''$ is the composition $D^2=D\circ D$, $(1+t^2)f'$ is the composition $M_{1+t^2}\circ D$, and $f\mapsto tf$ is the linear transformation M_t. Therefore the left side of the differential equation is the linear transformation $L=D^2+M_{1+t^2}\circ D+M_t$. The differential equation can be expressed as $L(f(t))=b(t)$ with $b(t)=3t+1\in C^\infty$.

In general, a *linear differential equation* of order n is

$$\frac{d^nf}{dt^n}+a_1(t)\frac{d^{n-1}f}{dt^{n-1}}+a_2(t)\frac{d^{n-2}f}{dt^{n-2}}+\cdots+a_{n-1}(t)\frac{df}{dt}+a_n(t)f=b(t),$$

because if the coefficient functions $a_1(t),a_2(t),\dots,a_n(t)$ are smooth, then the left side is a linear transformation $C^\infty\to C^\infty$.

Example 2.9. For the special case $W=\mathbb{R}^1=\mathbb{R}$. The vector space in Proposition 2.2 is the *dual space* $V^*=\mathrm{Hom}(V,\mathbb{R})$. The dual space consists of linear transformations $l\colon V\to\mathbb{R}$, which we call *linear functionals*.

Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be a basis of V. Then we may express any vector in V as unique linear combination of vectors in α

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n$$

and get

$$l(\vec{x}) = x_1 l(\vec{v}_1) + x_2 l(\vec{v}_2) + \cdots + x_n l(\vec{v}_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n, \quad a_i = l(\vec{v}_i).$$

This shows that linear functionals are linear combinations of the α-coordinates.

Using the basis, we can find the explicit vector space structure in the dual space. Let

$$l(\vec{x}) = a_1x_1 + a_2x_2 + \cdots + a_nx_n, \quad k(\vec{x}) = b_1x_1 + b_2x_2 + \cdots + b_nx_n.$$

Then

$$(l+k)(\vec{x}) = (a_1+b_1)x_1 + (a_2+b_2)x_2 + \cdots + (a_n+b_n)x_n, \quad (cl)(\vec{x}) = ca_1x_1 + ca_2x_2 + \cdots + ca_nx_n.$$

Exercise 2.3. Interpret the Newton-Leibniz formula $f(t) = f(0) + \int_0^t f(\tau)d\tau$ as an equality of linear transformations.

Exercise 2.4. Fix a vector $\vec{v} \in V$. Prove that the evaluation map $L \mapsto L(\vec{v})\colon \mathrm{Hom}(V,W) \to W$ is a linear transformation.

Exercise 2.5. The *trace* of a square matrix $A = (a_{ij})$ is the sum of its diagonal entries

$$\mathrm{tr}A = a_{11} + a_{22} + \cdots + a_{nn}.$$

Explain that the trace is a linear functional on the vector space of $M_{n\times n}$ of $n \times n$ matrices, and $\mathrm{tr}A^T = \mathrm{tr}A$.

Exercise 2.6. Let $L\colon V \to W$ be a linear transformation. Prove that $L \circ (K_1 + K_2) = L \circ K_1 + L \circ K_2$ and $L \circ (aK) = a(L \circ K)$. Explain that this means the map $L_* = L \circ \cdot\colon \mathrm{Hom}(U,V) \to \mathrm{Hom}(U,W)$ is a linear transformation.

Exercise 2.7. Let $L\colon U \to V$ be a linear transformation. Prove that $L^* = \cdot \circ L\colon \mathrm{Hom}(V,W) \to \mathrm{Hom}(U,W)$ is a linear transformation.

Exercise 2.8. For the induced linear transformation in Exercise 2.6, prove that

$$(L+K)_* = L_* + K_*, \quad (aL)_* = aL_*, \quad (L \circ K)_* = L_* \circ K_*.$$

This implies that $L \to L_*\colon \mathrm{Hom}(V,W) \to \mathrm{Hom}(\mathrm{Hom}(U,V),\ \mathrm{Hom}(U,W))$ is a linear transformation. Find the similar properties for induced linear transformation in Exercise 2.7.

Linear Transformation of Linear Combination

By combining addition and scalar multiplication, a linear transformation $L\colon V \to W$ preserves linear combination

$$\begin{aligned} L(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n) &= x_1L(\vec{v}_1) + x_2L(\vec{v}_2) + \cdots + x_nL(\vec{v}_n) \\ &= x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n, \qquad \vec{w}_i = L(\vec{v}_i). \end{aligned} \tag{2.1}$$

If $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span V, then any vector $\vec{x}$ in V as a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$, and the formula implies that a linear transformation is determined by its values on the spanning set.

Proposition 2.3. *If $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span V, then two linear transformations $L, K\colon V \to W$ are equal if and only if $L(\vec{v}_i) = K(\vec{v}_i)$ for each i.*

Conversely, given assigned values on a spanning set, the following says when the formula (2.1) gives a well defined linear transformation.

Proposition 2.4. *If $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span a vector space V, and $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n$ are vectors in W. Then* (2.1) *gives a well defined linear transformation $L\colon V \to W$ if and only if*

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{0} \implies x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n = \vec{0}.$$

A special case of Proposition 2.4 is that $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ is a basis of V. In this case, we have $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{0}$ implying $x_1 = x_2 = \cdots = x_n = 0$, which further implies $x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n = \vec{0}$.

For the case $\vec{v}_1 = \vec{e}_1, \vec{v}_2 = \vec{e}_2, \dots, \vec{v}_n = \vec{e}_n$ is a standard basis of $\mathbb{R}^n$, this means that linear transformations $L\colon \mathbb{R}^n \to V$ are in one-to-one correspondence with collections of n vectors $L(\vec{e}_1), L(\vec{e}_2), \dots, L(\vec{e}_n)$ in V.

Proof. The formula gives well defined map if and only if

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n$$

implies

$$x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n = y_1\vec{w}_1 + y_2\vec{w}_2 + \cdots + y_n\vec{w}_n.$$

Let $z_i = x_i - y_i$. Then the condition becomes

$$z_1\vec{v}_1 + z_2\vec{v}_2 + \cdots + z_n\vec{v}_n = \vec{0}$$

implying

$$z_1\vec{w}_1 + z_2\vec{w}_2 + \cdots + z_n\vec{w}_n = \vec{0}.$$

This is the condition in the proposition.

We still need to show that, if L is well defined, then L is a linear transformation. For

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n, \quad \vec{y} = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n,$$

by (2.1), we have

$$\begin{aligned}
L(\vec{x}+\vec{y}) &= L((x_1+y_1)\vec{v}_1 + (x_2+y_2)\vec{v}_2 + \cdots + (x_n+y_n)\vec{v}_n) \\
&= (x_1+y_1)\vec{w}_1 + (x_2+y_2)\vec{w}_2 + \cdots + (x_n+y_n)\vec{w}_n \\
&= (x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n) + (y_1\vec{w}_1 + y_2\vec{w}_2 + \cdots + y_n\vec{w}_n) \\
&= L(\vec{x}) + L(\vec{y}).
\end{aligned}$$

We can similarly show $L(c\vec{x}) = cL(\vec{x})$. □

Example 2.10. The rotation R_θ in Example 2.3 is determined by the values $R_\theta(\vec{e}_1) = (\cos\theta, \sin\theta)$ and $R_\theta(\vec{e}_2) = (-\sin\theta, \cos\theta)$ at the standard basis.

Example 2.11. The derivative linear transformation $P_3 \to P_2$ is determined by the derivatives of the monomials $1' = 0$, $t' = 1$, $(t^2)' = 2t$, $(t^3)' = 3t^2$. It is also determined by the derivatives of another basis (see Example 1.25) of P_3: $1' = 0$, $(t-1)' = 1$, $((t-1)^2)' = 2(t-1)$, $((t-1)^3)' = 3(t-1)^2$.

Exercise 2.9. Suppose $L\colon V \to W$ is a linear transformation, and $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are vectors in V. Prove the following.

1. If $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are linearly dependent, then $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ are linearly dependent.
2. If $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ are linearly independent, then $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are linearly independent.

2.2 Linear Transformation between Euclidean Spaces

Now we further specialise to the case $V = \mathbb{R}^n$ is a Euclidean space and $\vec{v}_i = \vec{e}_i$ is the standard basis. Then (2.1) and Proposition 2.4 show that linear transformations $L\colon \mathbb{R}^n \to W$ are in one-to-one correspondence with the collection of vectors $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n$ in W

$$L(x_1, x_2, \dots, x_n) = x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n, \qquad \vec{w}_i = L(\vec{e}_i). \tag{2.2}$$

In case $W = \mathbb{R}^m$ is also a Euclidean space, the vectors in W form the columns of a matrix

$$A = (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n) = (L(\vec{e}_1)\ L(\vec{e}_2)\ \cdots\ L(\vec{e}_n)), \tag{2.3}$$

and linear transformations $L\colon \mathbb{R}^n \to \mathbb{R}^m$ are in one-to-one correspondence with $m \times n$ matrices

$$L(\vec{x}) = x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n = A\vec{x}.$$

Here $A\vec{x}$ is the left of system of linear equations, and the second equality is from (1.5) and (1.6). We call A the *matrix of linear transformation* L.

Example 2.12. The matrix of the identity operator $I\colon \mathbb{R}^n \to \mathbb{R}^n$ is the *identity matrix*, still denoted by I

$$(I(\vec{e}_1)\ I(\vec{e}_2)\ \cdots\ I(\vec{e}_n)) = (\vec{e}_1\ \vec{e}_2\ \cdots\ \vec{e}_n) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I_{n\times n}.$$

The zero transformation $O = \mathbb{R}^n \to \mathbb{R}^m$ has the matrix $O_{m\times n}$ in which all entries are 0.

Example 2.13. The rotation R_θ in Example 2.3 takes $\vec{e}_1 = (1,0)$ to the vector of radius 1 and angle θ. Therefore $R_\theta(\vec{e}_1) = (\cos\theta, \sin\theta)$. It also takes $\vec{e}_2 = (0,2)$ to the vector of radius 1 and angle $\theta + \frac{\pi}{2}$. Therefore $R_\theta(\vec{e}_2) = (\cos\left(\theta+\frac{\pi}{2}\right), \sin\left(\theta+\frac{\pi}{2}\right)) = (-\sin\theta, \cos\theta)$. This gives the matrix of R_θ

$$(R_\theta(\vec{e}_1)\ R_\theta(\vec{e}_2)) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The reflection F_ρ of $\mathbb{R}^2$ with respect to the line of angle ρ takes $\vec{e}_1$ to the vector of radius 1 and angle 2ρ. Therefore $R_\theta(\vec{e}_1) = (\cos\theta, \sin\theta)$. It also takes $\vec{e}_2$ to the vector of radius 1 and angle $2\rho - \frac{\pi}{2}$. Therefore the matrix of F_ρ is

$$\begin{pmatrix} \cos 2\rho & \cos\left(2\rho - \frac{\pi}{2}\right) \\ \sin 2\rho & \sin\left(2\rho - \frac{\pi}{2}\right) \end{pmatrix} = \begin{pmatrix} \cos 2\rho & \sin 2\rho \\ \sin 2\rho & -\cos 2\rho \end{pmatrix}.$$

Example 2.14. The projection in Example 2.4 takes the standard basis $\vec{e}_1 = (1,0,0)$, $\vec{e}_2 = (0,1,0)$, $\vec{e}_3 = (0,0,1)$ of $\mathbb{R}^3$ to $(1,0)$, $(0,1)$, $(0,0)$ in $\mathbb{R}^2$. The matrix of the projection is

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

Example 2.15. The linear transformation corresponding to the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}$$

is

$$L\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 3x_3 + 4x_4 \\ 5x_1 + 6x_2 + 7x_3 + 8x_4 \\ 9x_1 + 10x_2 + 11x_3 + 12x_4 \end{pmatrix} : \mathbb{R}^4 \to \mathbb{R}^3.$$

We note that

$$L(\vec{e}_1) = \begin{pmatrix} 1 + 2\cdot 0 + 3\cdot 0 + 4\cdot 0 \\ 5\cdot 1 + 6\cdot 0 + 7\cdot 0 + 8\cdot 0 \\ 9\cdot 1 + 10\cdot 0 + 11\cdot 0 + 12\cdot 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ 9 \end{pmatrix}$$

is the first column of A. Similarly, $L(\vec{e}_2), L(\vec{e}_3), L(\vec{e}_4)$ are the second, third and fourth columns of A.

Example 2.16. The orthogonal projection P of the $\mathbb{R}^3$ to the plane $x+y+z=0$ is a linear transformation, similar to Example 2.4. The columns of the matrix of projection are the projections of the standard basis to the plane. These projections are not easy to see directly. On the other hand, we can easily find the projections of some other matrices.

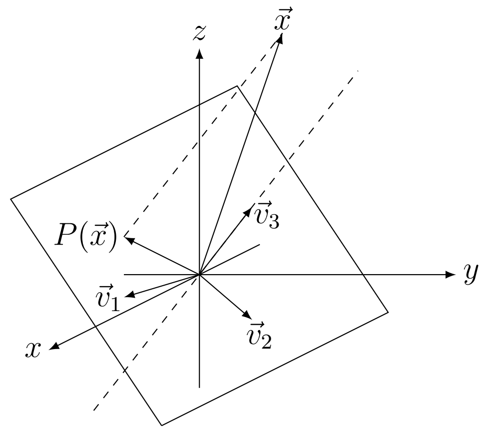

Figure 2.3: Projection to the plane $x + y + z = 0$.

First, the vectors $\vec{v}_1 = (1, -1, 0)$ and $\vec{v}_2 = (1, 0, -1)$ lie in the plane because they satisfy $x+y+z = 0$. Since the projection clearly preserves the vectors on the plane, we get $P(\vec{v}_1) = \vec{v}_1$ and $P(\vec{v}_2) = \vec{v}_2$.

Second, the vector $\vec{v}_3 = (1, 1, 1)$ is the coefficients of $x+y+z = 0$, and is therefore orthogonal to the plane. Since the projection kills the vectors orthogonal to the plane, we get $P(\vec{v}_3) = \vec{0}$.

Our idea of finding $P(\vec{e}_1)$ (the first column of the matrix of P) is to write $\vec{e}_1 = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3$ and then get

$$P(\vec{e}_1) = x_1P(\vec{v}_1) + x_2P(\vec{v}_2) + x_3P(\vec{v}_3) = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{0} = x_1\vec{v}_1 + x_2\vec{v}_2.$$

So the key here is to find the coordinates (x_1, x_2, x_3) of $\vec{e}_1$ with respect to the basis (see Example 1.27) $\alpha = \{\vec{v}_1, \vec{v}_2, , \vec{v}_3\}$. In Example 1.31, we found $[\vec{e}_1]_\alpha = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and therefore

$$P(\vec{e}_1) = \tfrac{1}{3}\vec{v}_1 + \tfrac{1}{3}\vec{v}_2 = (\tfrac{2}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}).$$

We can similarly use $[\vec{e}_2]_\alpha$ and $[\vec{e}_3]_\alpha$ in Example 1.31 to get

$$P(\vec{e}_2) = -\tfrac{2}{3}\vec{v}_1 + \tfrac{1}{3}\vec{v}_2 = (-\tfrac{1}{3}, \tfrac{2}{3}, -\tfrac{1}{3}), \quad P(\vec{e}_3) = \tfrac{1}{3}\vec{v}_1 - \tfrac{2}{3}\vec{v}_2 = (-\tfrac{1}{3}, -\tfrac{1}{3}, \tfrac{2}{3}).$$

We conclude the matrix of P is

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

Matrix Operation

We have the equivalence between linear transformations between Euclidean spaces and matrices

$$L \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \longleftrightarrow A \in M_{m\times n}, \quad L(\vec{x}) = A\vec{x}.$$

We also have addition, scalar multiplication and composition of linear transformations. These operations have corresponding operations on matrices.

Let $L, K\colon \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations, with respective matrices

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}.$$

Then the i-th column of the matrix of $L + K\colon \mathbb{R}^n \to \mathbb{R}^m$ is

$$(L+K)(\vec{e}_i) = L(\vec{e}_i) + K(\vec{e}_i) = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix} + \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{mi} \end{pmatrix} = \begin{pmatrix} a_{1i} + b_{1i} \\ a_{2i} + b_{2i} \\ \vdots \\ a_{mi} + b_{mi} \end{pmatrix}.$$

We define the addition of two matrices (of the same size) to be the matrix of $L + K$

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{pmatrix}.$$

Similarly, we define the scalar multiplication cA of a matrix to be the matrix of the linear transformation cL

$$cA = \begin{pmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{pmatrix}.$$

We emphasise that we have the formulae for $A + B$ and cA not because they are the obvious thing to do, but because they reflect the concepts of $L + K$ and cL for linear transformations.

Let $L\colon \mathbb{R}^n \to \mathbb{R}^m$ and $K\colon \mathbb{R}^k \to \mathbb{R}^n$ be linear transformations, with respective matrices

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nk} \end{pmatrix}.$$

To get the matrix of the composition linear transformation $L \circ K\colon \mathbb{R}^k \to \mathbb{R}^m$, we note that

$$B = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_k), \quad K(\vec{e}_i) = \vec{v}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{ni} \end{pmatrix}.$$

Then the i-th column of the matrix of $L \circ K$ is (see (1.5) and (1.6))

$$\begin{aligned}(L \circ K)(\vec{e}_i) &= L(K(\vec{e}_i)) = L(\vec{v}_i) = A\vec{v}_i \\ &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{ni} \end{pmatrix} = \begin{pmatrix} a_{11}b_{1i} + a_{12}b_{2i} + \dots + a_{1n}b_{ni} \\ a_{21}b_{1i} + a_{22}b_{2i} + \dots + a_{2n}b_{ni} \\ \vdots \\ a_{m1}b_{1i} + a_{m2}b_{2i} + \dots + a_{mn}b_{ni} \end{pmatrix}.\end{aligned}$$

We define the multiplication of two matrices (of matching size) to be the matrix of $L \circ K$

$$AB = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1k} \\ c_{21} & c_{22} & \dots & c_{2k} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mk} \end{pmatrix}, \quad c_{ij} = a_{i1}b_{1j} + \dots + a_{in}b_{nj}.$$

The (i, j)-entry of AB is obtained by multiplying the i-th row of A and the j-th column of B.

Example 2.17. The zero map O in Example 2.1 corresponds to the zero matrix O in Example 2.12. Since $O + L = L = L + O$, we get $O + A = A = A + O$.

The identity map I in Example 2.1 corresponds to the identity matrix I in Example 2.12. Since $I \circ L = L = L \circ I$, we get $IA = A = AI$.

Example 2.18. The composition of maps satisfies $(L \circ K) \circ T = L \circ (K \circ T)$. The equality is also satisfied by linear transformations. Correspondingly, we get the associativity $(AB)C = A(BC)$ of the matrix multiplication.

It is very complicated to verify $(AB)C = A(BC)$ by multiplying rows and columns. The conceptual explanation makes such computation unnecessary.

Example 2.19. In Example 2.13, we obtained the matrix of rotation R_θ in Example 2.3. Then the equality $R_{\theta_1} \circ R_{\theta_2} = R_{\theta_1+\theta_2}$ in Example 2.7 corresponds to the multiplication of the corresponding matrices

$$\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix}.$$

On the other hand, we calculate the left side by multiplying the rows of the first matrix with the columns of the second matrix. We get

$$\begin{pmatrix} \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_2 - \sin\theta_1\cos\theta_2 \\ \cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2 & \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 \end{pmatrix}.$$

By comparing the two sides, we get the familiar trigonometric identity

$$\cos(\theta_1+\theta_2) = \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2, \quad \sin(\theta_1+\theta_2) = \cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2.$$

Example 2.20. Let
$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$
We try to find matrices X satisfying $AX = I$. Let $X = \begin{pmatrix} x & z \\ y & w \end{pmatrix}$. Then the equality becomes
$$AX = \begin{pmatrix} x+2y & z+2w \\ 3x+4y & 3z+4w \end{pmatrix} = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$
This means two systems of linear equations
$$\begin{aligned} x + 2y &= 1, & z + 2w &= 0, \\ 3x + 4y &= 0; & 3z + 4w &= 1. \end{aligned}$$
We can solve two systems simultaneously by carrying out the row operation
$$(A\ I) = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$
By the first three columns, we get $x = -2$ and $y = \frac{3}{2}$. By the first, second and fourth columns, we get $z = 1$ and $w = -\frac{1}{2}$. Therefore
$$X = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$
In general, to solve $AX = B$, we may carry out the row operation on the matrix $(A\ B)$.

Exercise 2.10. Composing the reflection F_ρ in Examples 2.3 and 2.13 with itself is the identity. Explain that this means the trigonometric identity $\cos^2\theta + \sin^2\theta = 1$.

Exercise 2.11. Geometrically, one can see the following compositions of rotations and refelctions.

1. $R_\theta \circ F_\rho$ is a reflection. What is the angle of reflection?
2. $F_\rho \circ R_\theta$ is a reflection. What is the angle of reflection?
3. $F_{\rho_1} \circ F_{\rho_2}$ is a rotation. What is the angle of rotation?

Interpret the geometrical observations as trigonometric identities.

Exercise 2.12. Use some examples to show that for two $n \times n$ matrices, AB may not be equal to BA.

Exercise 2.13. Use the formula for matrix addition to show the commutativity $A+B = B+C$ and the associativity $(A+B)+C = A+(B+C)$. Then give a concept explanation to the properties without using calculation.

Exercise 2.14. Explain that the addition and scalar multiplication of matrices make the set $M_{m\times n}$ of $m\times n$ matrices into a vector space. Moreover, the matrix of linear transformation gives a (invertible) linear transformation $\mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m)\to M_{m\times n}$.

Exercise 2.15. Explain that Exercise 2.6 means that the matrix multiplication satisfies $A(B+C)=AB+AC$, $A(aB)=a(AB)$, and the left multiplication $X\mapsto AX$ is a linear transformation.

Exercise 2.16. Explain that Exercise 2.7 means that the matrix multiplication satisfies $(A+B)C=AC+BC$, $(aA)B=a(AB)$, and the right multiplication $X\mapsto XA$ is a linear transformation.

Exercise 2.17. Let A be an $m\times n$ matrix and let B be a $k\times m$ matrix. For the trace defined in Example 2.5, explain that $\mathrm{tr}AXB$ is a linear functional for $X\in M_{n\times k}$.

Exercise 2.18. Let A be an $m\times n$ matrix and let B be an $n\times m$ matrix. Show that the trace defined in Example 2.5 satisfies $\mathrm{tr}AB=\mathrm{tr}BA$.

Exercise 2.19. Add or multiply matrices, whenever you can.

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

Exercise 2.20. Find the n-th power matrix A^n (i.e., multiply the matrix to itself n times).

1. $\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}$.
2. $\begin{pmatrix}\cos\theta&\sin\theta\\\sin\theta&-\cos\theta\end{pmatrix}$.
3. $\begin{pmatrix}a_1&0&0&0\\0&a_2&0&0\\0&0&a_3&0\\0&0&0&a_4\end{pmatrix}$.
4. $\begin{pmatrix}0&a&0&0\\0&0&a&0\\0&0&0&a\\0&0&0&0\end{pmatrix}$.
5. $\begin{pmatrix}a&b&0&0\\0&a&b&0\\0&0&a&b\\0&0&0&a\end{pmatrix}$.

Exercise 2.21. Solve the matrix equations.

1. $\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}X=\begin{pmatrix}7&8\\9&10\end{pmatrix}$.
2. $\begin{pmatrix}1&4\\2&5\\3&6\end{pmatrix}X=\begin{pmatrix}7&10\\8&11\\a&b\end{pmatrix}$.
3. $\begin{pmatrix}1&2\\5&6\\9&10\end{pmatrix}X=\begin{pmatrix}3&4\\7&8\\11&b\end{pmatrix}$.

Exercise 2.22. For the transpose of 2×2 matrices, verify that $(AB)^T=B^TA^T$. (The general conceptual argument will be given by Proposition 4.6.) Then use this to solve the matrix equations.

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

Exercise 2.23. Let $A=\begin{pmatrix}-1&0\\0&1\end{pmatrix}$. Find all the matrices X satisfying $AX=XA$. Generalise your result to *diagonal matrix*

$$A = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}.$$

Exercise 2.24 (Elementary Matrix). There are three types of *elementary matrices*.

1. T_{ij} $(i \neq j)$ is the matrix obtained by exchanging the i-th and j-th rows of the identity matrix.
2. $D_i(a)$ is the diagonal matrix with the i-th diagonal entry a and all the other diagonal entries a.
3. $E_{ij}(a)$ $(i \neq j)$ is the identity matrix except the (i, j)-entry is a.

The following are some 4×4 examples

$$T_{23} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad D_2(-2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E_{13}(-2) = \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Prove the left multiplication by elementary matrices is the same as row operations.

1. $T_{ij}A$ exchanges i-th and j-th rows of A.
2. $D_i(a)A$ multiplies a to the i-th row of A.
3. $E_{ij}(a)A$ adds the a multiple of the j-th row to the i-th row.

Show that the right multiplication has the similar effect on the columns.

2.3 Onto and One-to-one

A map $f\colon X \to Y$ is *onto* (or *surjective*) if for any $y \in Y$, there is $x \in X$, such that $f(x) = y$. The map is *one-to-one* (or *injective*) if $f(x) = f(x')$ implies $x = x'$. This is equivalent to that $x \neq x'$ implies $f(x) \neq f(x')$. The map is a *one-to-one correspondence* (or *bijective*) if it is onto and one-to-one.

Proposition 2.5. *A map f is onto if and only if there is g, such that $f \circ g = id$. A map f is one-to-one if and only if there is g, such that $g \circ f = id$.*

Proof. We will prove the necessity. The sufficiency is left as Exercise 2.27.

Suppose $f\colon X \to Y$ is onto. We construct a map $g\colon Y \to X$ as follows. For any $y \in Y$, by f onto, we can find some $x \in X$ satisfying $f(x) = y$. We choose one such x (strictly speaking, this uses the Axiom of Choice) and define $g(y) = x$. Then the map g satisfies $(f \circ g)(y) = f(x) = y$.

Suppose $f\colon X \to Y$ is one-to-one. We fix a element $x_0 \in X$ and construct a map $g\colon Y \to X$ as follows. For any $y \in Y$, if $y = f(x)$ for some $x \in X$ (i.e., y lies in the image of f), then we define $g(y) = x$. If we cannot find such x, then we define $g(y) = x_0$. Note that in the first case, if $y = f(x')$ for another $x' \in X$, then by f one-to-one, we have $x' = x$. This shows that g is well defined. For the case $y = f(x)$, our construction of g implies $(g \circ f)(x) = g(f(x)) = x$. $\square$

Example 2.21. Consider the map ($f =$) Instructor: Courses $\to$ Professors.

The map is onto means every professor teaches some course. The map g in Proposition 2.5 can take a professor (say me) to any one course (say linear algebra) he or she teaches.

The map is one-to-one means any professor either teaches one course, or does not teach any course. This also means that no professor teaches two or more courses. If a professor (say me) teaches one course, then the map g in Proposition 2.5 takes the professor to the unique course (say linear algebra) he or she teaches. If a professor does not teach any course, then g takes the professor to any one existing course.

Exercise 2.25. Prove that the composition of onto maps is onto.

Exercise 2.26. Prove that the composition of one-to-one maps is one-to-one.

Exercise 2.27. Prove that if $f \circ g$ is onto, then f is onto. Prove that if $g \circ f$ is one-to-one, then f is one-to-one.

Exercise 2.28. Prove that f is onto if and only if there is g, such that $f \circ g = id$. Prove that g is one-to-one if and only if there is f, such that $f \circ g = id$.

In this section, we specialise the properties of maps to linear transformations. We also calculate these properties for the special case of Euclidean spaces.

Onto Linear Transformation

We verify the onto property for the linear transformations in Section 2.1. The identity map in Example 2.1 is always onto. The zero map is onto if and only if W is the zero vector space in Example 1.1.

The rotation R_θ in Example 2.3 is onto because for any vector $\vec{v} \in \mathbb{R}^2$, we can rotate $\vec{v}$ back by θ (i.e., rotate by $-\theta$) to get $\vec{w}$. Then $\vec{v} = R_\theta(\vec{w})$. Another way is to observe that $R_\theta \circ R_{-\theta}$ is the identity, and use Proposition 2.5. By the similar reason, the flipping of the plane is also onto.

The projection in Example 2.4 is also onto. The reason is that $(x_1, x_2) = P(x_1, x_2, 0)$.

The derivative map in Example 2.6 is onto because by the Newton-Leibniz formula, we have $f(t) = F'(t)$, for $F(t) = \int_0^t f(\tau)d\tau$. On the other hand, the integration map is not onto, because any function of the form $F(t) = \int_0^t f(\tau)d\tau$ satisfies $F(0) = 0$, which shows that the constant function 1 is not of the form $\int_0^t f(\tau)d\tau$.

Exercise 2.29. Show that the evaluation map $L(f) = (f(0), f(1), f(2)) \colon C^\infty \to \mathbb{R}^3$ in Example 2.5 is onto. However, the linear combination map $L(x_1, x_2, x_3) = x_1 \cos t + x_2 \sin t + x_3 e^t \colon \mathbb{R}^3 \to C^\infty$ is not onto.

Exercise 2.30. Show that the multiplication map $f(t) \mapsto a(t)f(t) \colon C^\infty \to C^\infty$ in Example 2.6 is onto if and only if $a(t) \neq 0$ everywhere.

Exercise 2.31. Suppose $L \circ K$ and K are linear transformations. Prove that if K is onto, then L is also a linear transformation.

Exercise 2.32. Suppose L is an onto linear transformation. Prove that two linear transformations K and K' are equal if and only if $K \circ L = K' \circ L$. What does this tell you about the linear transformation $L^* \colon \mathrm{Hom}(V, W) \to \mathrm{Hom}(U, W)$ in Exercise 2.7?

The following is the enhanced version of the onto part of Proposition 2.5.

Proposition 2.6. *Suppose W is a finite dimensional vector space. Then a linear transformation $L \colon V \to W$ is onto, if and only if there is a linear transformation $K \colon W \to V$, such that $L \circ K = I_W$.*

The proof is similar to the proof of Proposition 2.5. The extra care we need to take is to make sure K is a linear transformation. We achieve this by defining K on a basis and then extend to a linear transformation.

Proof. The sufficiency follows from Proposition 2.5.

For the necessity, we consider an onto linear transformation L. Let $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m$ be a basis of W. Since L is onto, we can find $\vec{v}_i \in V$ satisfying $L(\vec{v}_i) = \vec{w}_i$. By Proposition 2.4 (and the subsequent remark), the formula

$$K(x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_m\vec{w}_m) = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m$$

gives a well defined linear transformation $K \colon W \to V$. By $(L\circ K)(\vec{w}_i) = L(K(\vec{w}_i)) = L(\vec{v}_i) = \vec{w}_i$ and Proposition 2.3, we get $L \circ K = I_W$. □

The following shows the relation between the onto property and span.

Proposition 2.7. *A linear transformation $L \colon V \to W$ is onto if and only if it takes a spanning set of V to a spanning set of W.*

Proof. Suppose $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ span V. Then any vector $\vec{x} \in V$ is a linear combination of these vectors, and we get

$$\vec{x} \in V \iff \vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \implies L(\vec{x}) = x_1L(\vec{v}_1) + x_2L(\vec{v}_2) + \cdots + x_nL(\vec{v}_n).$$

Since L is onto if and only if any vector in W can be expressed as $L(\vec{x})$, we find that the property is the same as $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ spanning V. In other words, L takes a spanning set of V to a spanning set of W. □

We say that a (not necessarily finite) set α of vectors in V is a spanning set if any vector in V is a linear combination of finitely many vectors from α. With this extension, Proposition 2.7 remains true for infinite dimensional vector spaces.

Proposition 2.7 implies the following linear transformation version of Propositions 1.11 and 1.20. It reflects the intuition that, if every professor teaches course (see Example 2.21), then the number of courses is more than the number of professors.

Proposition 2.8. *If a linear transformation $L\colon V \to W$ is onto, then $\dim V \geq \dim W$.*

Proof. Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ be a basis of V, $n = \dim V$. By Proposition 2.7, $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ spans W. Then by the first part of Proposition 1.20, we have $n \geq \dim W$. $\square$

Consider a linear transformation $L(\vec{x}) = A\vec{x}\colon \mathbb{R}^n \to \mathbb{R}^m$ between Euclidean spaces. By the definition, L is onto means that $A\vec{x} = \vec{b}$ has solution for all $\vec{b} \in \mathbb{R}^m$. Alternatively, by Proposition 2.7, L is onto means that $L(\vec{e}_1), L(\vec{e}_2), \dots, L(\vec{e}_n)$ span $\mathbb{R}^m$. In other words, the columns of A span $\mathbb{R}^m$. In fact, Proposition 1.10 already shows that the two interpretations are equivalent, and can be calculated by showing that all rows of A being pivot.

Example 2.22. Consider the linear transformation given by the matrix in Example 1.14

$$L(\vec{x}) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & a & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 + 4x_2 + 7x_3 + 10x_4 \\ 2x_1 + 5x_2 + 8x_3 + 11x_4 \\ 3x_1 + 6x_2 + ax_3 + bx_4 \end{pmatrix} : \mathbb{R}^4 \to \mathbb{R}^3.$$

By the row echelon form in Example 1.18, L is onto if and only if $a \neq 9$ or $b \neq 9$.

Example 2.23. By Proposition 2.8, the linear transformation

$$L(\vec{x}) = \begin{pmatrix} x_1 + (\log 2)x_2 + \sqrt{3}x_3 \\ ex_2 + e^{-1}x_3 \\ \sqrt{2}x_1 + 100x_2 + (\sin 1)x_3 \\ \pi x_1 - 0.5x_2 + 2.3x_3 \end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^4$$

cannot be onto.

Example 2.24. The differential equation in Example 2.8

$$f'' + (1+t^2)f' + tf = b(t)$$

can be interpreted as $L(f(t)) = b(t)$ for a linear transformation $L\colon C^\infty \to C^\infty$. If we regard L as a linear transformation $L\colon P_n \to P_{n+1}$ (restricting L to polynomials), then by Proposition 2.8, the restriction linear transformation is not onto. For example, we can find a polynomial $b(t)$ of degree 5, such that $f'' + (1+t^2)f' + tf = b(t)$ cannot be solved for a polynomial $f(t)$ of degree 4.

Exercise 2.33. Strictly speaking, the property in Proposition 2.7 can be stated for one spanning set or all spanning sets. Show that the two versions are equivalent.

Exercise 2.34. Prove that the linear transformation $L\colon \mathbb{R}^n \to V$ is onto if and only if $\vec{v}_1 = L(\vec{e}_1), \vec{v}_2 = L(\vec{e}_2), \dots, \vec{v}_n = L(\vec{e}_n)$ span V.

Exercise 2.35. Suppose $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ span V. Suppose the columns of an $m \times n$ matrix $A = (a_{ij})$ span $\mathbb{R}^m$. Prove that $a_{11}\vec{v}_1 + a_{21}\vec{v}_2 + \cdots + a_{m1}\vec{v}_m, a_{12}\vec{v}_1 + a_{22}\vec{v}_2 + \cdots + a_{m2}\vec{v}_m, \dots, a_{1n}\vec{v}_1 + a_{2n}\vec{v}_2 + \cdots + a_{mn}\vec{v}_m$ span V. This generalizes Exercise 1.39.

Exercise 2.36. Let A be an $m \times n$ matrix. Explain that a system of linear equations $A\vec{x} = \vec{b}$ has solution for all $\vec{b} \in \mathbb{R}^m$ if and only if there is an $n \times m$ matrix B, such that $AB = I_{m \times m}$.

One-to-one Linear Transformation

We verify the one-to-one property for the linear transformations in Section 2.1. The identity map in Example 2.1 is always one-to-one. The zero map is one-to-one if and only if V is the zero vector space in Example 1.1.

The rotation R_θ in Example 2.3 is one-to-one because if the same (i.e., same rotation angle) rotations of two vectors are equal, then the two vectors are equal. Another way is to observe that $R_{-\theta} \circ R_\theta$ is the identity, and use Proposition 2.5. By the similar reason, the flipping of the plane is also one-to-one.

The projection in Example 2.4 is not one-to-one, because all the points on the line perpendicular to the plane are mapped to the same point.

The derivative map in Example 2.6 is not one-to-one because two functions differing by a constant have the same derivative. On the other hand, the integration map is one-to-one, because by the Newton-Leibniz formula, taking the derivative of $\int_0^t f(\tau)d\tau = \int_0^t g(\tau)d\tau$ gives $f(t) = g(t)$.

Exercise 2.37. Show that the evaluation map $L(f) = (f(0), f(1), f(2))\colon C^\infty \to \mathbb{R}^3$ in Example 2.5 is not one-to-one. However, the linear combination map $L(x_1, x_2, x_3) = x_1 \cos t + x_2 \sin t + x_3 e^t\colon \mathbb{R}^3 \to C^\infty$ is one-to-one.

Exercise 2.38. Show that the multiplication map $f(t) \mapsto a(t)f(t)\colon C^\infty \to C^\infty$ in Example 2.6 is one-to-one if $a(t) = 0$ at only finitely many places.

Exercise 2.39. Suppose $L \circ K$ and L are linear transformations. Prove that if L is one-to-one, then K is also a linear transformation.

Exercise 2.40. Suppose L is a one-to-one linear transformation. Prove that two linear transformations K and K' are equal if and only if $L \circ K = L \circ K'$. What does this tell you about the linear transformation $L_*\colon \mathrm{Hom}(U, V) \to \mathrm{Hom}(U, W)$ in Exercise 2.6?

The following is the enhanced version of the one-to-one part of Proposition 2.5.

Proposition 2.9. *Suppose W is a finite dimensional vector space. Then a linear transformation $L\colon V \to W$ is one-to-one if and only if there is a linear transformation $K\colon W \to V$, such that $K \circ L = I_V$.*

Proof. The sufficiency follows from Proposition 2.5.

For the necessity, we consider a one-to-one linear transformation L. By Proposition 2.10, L takes linearly independent set in V to a linearly independent set in W. Since W is finite dimensional, by Proposition 1.20, linearly independent sets in W contains at most $\dim W$ vectors. Therefore linearly independent sets in V contains at most $\dim W$ vectors. By Theorem 1.18, this implies that V is finite dimensional.

Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ be a basis of V. By Proposition 2.10, $\vec{w}_1 = L(\vec{v}_1), \vec{w}_2 = L(\vec{v}_2), \dots, \vec{w}_n = L(\vec{v}_n)$ are also linearly independent. By Theorem 1.18, this can be extended to a basis $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n, \vec{w}_{n+1}, \dots, \vec{w}_m$. By Proposition 2.4 (and the subsequent remark), the formula

$$K(x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_m\vec{w}_m) = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n$$

gives a well defined linear transformation $K\colon W \to V$. By $(K \circ L)(\vec{v}_i) = K(L(\vec{v}_i)) = K(\vec{w}_i) = \vec{v}_i$ and Proposition 2.3, we get $K \circ L = I_V$. $\square$

The following shows the relation between the one-to-one property and linear independence.

Proposition 2.10. *A linear transformation $L\colon V \to W$ is one-to-one if and only if it takes a linearly independent set in V to a linearly independent set in W. This is also equivalent to that $L(\vec{v}) = \vec{0} \implies \vec{v} = \vec{0}$.*

Proof. Suppose L is one-to-one and vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ in V are linearly independent. Then the following shows that $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ are linearly independent

$$\begin{aligned}
& x_1L(\vec{v}_1) + x_2L(\vec{v}_2) + \cdots + x_nL(\vec{v}_n) = y_1L(\vec{v}_1 + y_2L(\vec{v}_2) + \cdots + y_nL(\vec{v}_n) \\
& \implies L(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n) = L(y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n) \quad (L \text{ is linear}) \\
& \implies x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n \quad (L \text{ is one-to-one}) \\
& \implies x_1 = y_1,\ x_2 = y_2,\ \dots,\ x_n = y_n. \quad (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n \text{ linearly independent})
\end{aligned}$$

Now suppose L takes linearly independent set to linearly independent set. This means that, if $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ are linearly dependent, then $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are also linearly dependent. Consider the special case $L(\vec{v}) = \vec{0}$. By Proposition 1.8, the single vector $\vec{0}$ is linearly dependent. By the general assumption, therefore, the single vector $\vec{v}$ is also linearly dependent. By Proposition 1.8, this means $\vec{v} = \vec{0}$.

Finally, suppose $L(\vec{v}) = \vec{0} \implies \vec{v} = \vec{0}$. Then

$$L(\vec{x}) = L(\vec{y}) \implies L(\vec{x} - \vec{y}) = L(\vec{x}) - L(\vec{y}) = \vec{0} \implies \vec{x} - \vec{y} = \vec{0} \implies \vec{x} = \vec{y}.$$

This means that L is one-to-one. $\square$

We note that the proof does not require the vector spaces to be finite dimensional.

Proposition 2.10 implies the following linear transformation version of Propositions 1.13 and 1.20. It reflects the intuition that, if any professor teaches at most one course (see Example 2.21), then the number of courses is less than the number of professors.

Proposition 2.11. *If a linear transformation* $L\colon V \to W$ *is one-to-one, then* $\dim V \le \dim W$.

Proof. Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ be a basis of V, $n = \dim V$. By Proposition 2.10, $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ are linearly independent in W. Then by the second part of Proposition 1.20, we have $n \le \dim W$. $\square$

Consider a linear transformation $L(\vec{x}) = A\vec{x}\colon \mathbb{R}^n \to \mathbb{R}^m$ between Euclidean spaces. By the definition, L is one-to-one means that the solution of $A\vec{x} = \vec{b}$ is unique. Alternatively, by Proposition 2.10, L is one-to-one means that $L(\vec{e}_1), L(\vec{e}_2), \dots, L(\vec{e}_n)$ are linearly independent. In other words, the columns of A are linearly independent. In fact, Proposition 1.12 already shows that the two interpretations are equivalent, and can be calculated by showing that all columns of A being pivot.

Example 2.25. Consider the linear transformation given by the matrix in Example 1.22

$$L(\vec{x}) = \begin{pmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 5x_2 + 9x_3 \\ 2x_1 + 6x_2 + 10x_3 \\ 3x_1 + 7x_2 + 11x_3 \\ 4x_1 + 8x_2 + ax_3 \end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^4.$$

By the row echelon form in Example 1.22, L is one-to-one if and only if $a \ne 12$.

Example 2.26. By Proposition 2.11, the orthogonal projection in Example 2.16 cannot be one-to-one.

Example 2.27. By Proposition 2.11, the linear transformation

$$L(\vec{x}) = \begin{pmatrix} x_1 + \sqrt{2}x_3 + \pi x_4 \\ (\log 2)x_1 + ex_2 + 100x_3 - 0.5x_4 \\ \sqrt{3}x_1 + e^{-1}x_2 + (\sin 1)x_3 + 2.3x_4 \end{pmatrix} : \mathbb{R}^4 \to \mathbb{R}^3$$

cannot be one-to-one.

Exercise 2.41. Prove that for any vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$, the vectors $\vec{v}_1 + 2\vec{v}_2 + 3\vec{v}_3, 4\vec{v}_1 + 5\vec{v}_2 + 6\vec{v}_3, 7\vec{v}_1 + 8\vec{v}_2 + 9\vec{v}_3$ are always linearly dependent. Find the condition on a, such that the linear independence of $\vec{v}_1, \vec{v}_2, \vec{v}_3$ implies the linear independence of $\vec{v}_1 + 2\vec{v}_2 + 3\vec{v}_3, 4\vec{v}_1 + 5\vec{v}_2 + 6\vec{v}_3, 7\vec{v}_1 + 8\vec{v}_2 + a\vec{v}_3$.

Exercise 2.42. Strictly speaking, the property in Proposition 2.10 can be stated for one linearly independent set or all linearly independent sets. Show that the two versions are equivalent.

Exercise 2.43. Prove that the linear transformation $L\colon \mathbb{R}^n \to V$ is one-to-one if and only if $\vec{v}_1 = L(\vec{e}_1), \vec{v}_2 = L(\vec{e}_2), \dots, \vec{v}_n = L(\vec{e}_n)$ are linearly independent.

Exercise 2.44. Suppose $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ are linearly independent. Suppose the columns of an $m \times n$ matrix $A = (a_{ij})$ are also linearly independent. Prove that $a_{11}\vec{v}_1 + a_{21}\vec{v}_2 + \cdots + a_{m1}\vec{v}_m, a_{12}\vec{v}_1 + a_{22}\vec{v}_2 + \cdots + a_{m2}\vec{v}_m, \dots, a_{1n}\vec{v}_1 + a_{2n}\vec{v}_2 + \cdots + a_{mn}\vec{v}_m$ are linearly independent. This generalizes Exercise 1.41.

Exercise 2.45. Let A be an $m \times n$ matrix. Explain that the solution of a system of linear equations $A\vec{x} = \vec{b}$ is unique if and only if there is an $n \times m$ matrix B, such that $BA = I_{n\times n}$.

2.4 Isomorphism

A map $f\colon X \to Y$ is *invertible* if there is another map $g\colon Y \to X$, such that $g(f(x)) = x$ for all $x \in X$ and $f(g(y)) = y$ for all $y \in Y$. In other words, $g \circ f = id_X$ and $f \circ g = id_Y$. The map g is the *inverse* of f and is denoted $g = f^{-1}$.

The following is a basic result in set theory.

Proposition 2.12. *A map is invertible if and only if it is onto and one-to-one.*

Proof. Proposition 2.5 tells us that a map f is onto and one-to-one if and only if there are maps g and h, such that $f \circ g = id$ and $h \circ f = id$. Compared with the definition of invertibility, we only need to show $g = h$. This follows from $g = id \circ g = h \circ f \circ g = h \circ id = h$. □

Exercise 2.46. Prove that the composition of invertible maps is invertible. Moreover, we have $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.

Invertible Linear Transformation

Definition 2.13. An invertible linear transformation is called an *isomorphism*. If there is an isomorphism between two vector spaces V and W, then we say V and W are *isomorphic*, and denote $V \cong W$.

Example 2.28. Given a basis α of V, we explained in Section 1.5 that the α-coordinate map $[\cdot]_\alpha\colon V \to \mathbb{R}^n$ has an inverse. Therefore the map is an isomorphism.

Example 2.29. A linear transformation $L\colon \mathbb{R} \to V$ gives a vector $L(1) \in V$. This is a linear map (see Exercise 2.4)

$$L \in \mathrm{Hom}(\mathbb{R}, V) \mapsto L(1) \in V.$$

Conversely, for any $\vec{v} \in V$, we may construct a linear transformation $L(x) = x\vec{v}\colon \mathbb{R} \to V$. The construction gives a map

$$\vec{v} \in V \mapsto (L(x) = x\vec{v}) \in \mathrm{Hom}(\mathbb{R}, V).$$

We can verify that the two maps are inverse to each other. Therefore we get an isomorphism $\mathrm{Hom}(\mathbb{R}, V) \cong V$. In particular, the second map is also a linear transformation.

Example 2.30. If we switch $\mathbb{R}$ and V in Example 2.29, then we get the dual space $V^* = \mathrm{Hom}(\mathbb{R}, V)$ in Example 2.9. The vectors in V^* are the linear functionals. According to Example 2.9, a basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of V induces explicit formula for linear functionals

$$l(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n, \quad a_i = l(\vec{v}_i).$$

This shows that linear functionals are linear combinations of the α-coordinates. This gives an isomorphism (dependent on the choice of α) between the dual space and the Euclidean space

$$l \in \mathrm{Hom}(V, \mathbb{R}) \longleftrightarrow (a_1, a_2, \dots, a_n) \in \mathbb{R}^n, \quad a_i = l(\vec{v}_i).$$

In particular, we get $\dim V^* = \dim V$.

Example 2.31. The matrix of linear transformation between Euclidean spaces gives an invertible map

$$L \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \longleftrightarrow A \in M_{m\times n}, \quad A = (L(\vec{e}_1)\ L(\vec{e}_2)\ \cdots\ L(\vec{e}_n)), \quad L(\vec{x}) = A\vec{x}.$$

The vector space structure on $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ was established by Proposition 2.2. Then the addition and scalar multiplication in $M_{m\times n}$ are defined for the purpose of making the map an isomorphism.

Example 2.32. The transpose of matrices is an isomorphism

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \in M_{m\times n} \mapsto A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix} \in M_{n\times m}.$$

In fact, we have $(A^T)^T = A$, which means that the inverse of the transpose map is the transpose map.

Property of Isomorphism

The isomorphism can be used to translate the linear algebra in one vector space to the linear algebra in another vector space.

Theorem 2.14. *If a linear transformation $L\colon V \to W$ is an isomorphism, then the inverse map $L^{-1}\colon W \to V$ is also a linear transformation. Moreover, suppose $\vec{v}_1, \dots, \vec{v}_n$ are vectors in V.*

1. *$\vec{v}_1, \dots, \vec{v}_n$ span V if and only if $L(\vec{v}_1), \dots, L(\vec{v}_n)$ span W.*

2. $\vec{v}_1, \dots, \vec{v}_n$ *are linearly independent if and only if* $L(\vec{v}_1), \dots, L(\vec{v}_n)$ *are linearly independent.*

3. $\vec{v}_1, \dots, \vec{v}_n$ *form a basis of* V *if and only if* $L(\vec{v}_1), \dots, L(\vec{v}_n)$ *form a basis of* W*.*

Proof. To prove that the inverse map is also linear, we apply L to $L^{-1}(\vec{x}) + L^{-1}(\vec{y})$

$$L(L^{-1}(\vec{x}) + L^{-1}(\vec{y})) = L(L^{-1}(\vec{x})) + L(L^{-1}(\vec{y})) = \vec{x} + \vec{y} = L(L^{-1}(\vec{x} + \vec{y})).$$

Here the first equality uses the linearity of L, and the other equalities use $L \circ L^{-1} = id$. Then by L one-to-one, the equality implies $L^{-1}(\vec{x}) + L^{-1}(\vec{y}) = L^{-1}(\vec{x} + \vec{y})$. By the similar method, we can prove $L^{-1}(a\vec{x}) = aL^{-1}(\vec{x})$.

The rest of the proposition follows from Propositions 2.7 and 2.10. □

The theorem implies the following, which is also the combination of Propositions 2.8 and 2.11.

Proposition 2.15. *Isomorphic vector spaces have the same dimension.*

The following is the linear transformation version of Theorems 1.21 and 1.22.

Theorem 2.16. *Suppose* $\dim V = \dim W$*. Then for a linear transformation* $L \colon V \to W$*, the following are equivalent.*

1. *L is onto.*

2. *L is one-to-one.*

3. *L is an isomorphism.*

Proof. Let $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ be a basis of V, $n = \dim V = \dim W$. If L is onto, then by Proposition 2.8, $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ span W. By Theorem 1.21, $L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)$ are also linearly independent. Then by Proposition 2.11, L is one-to-one. By the similar method, we can show that one-to-one implies onto. □

Example 2.33 (Lagrange interpolation)**.** An important fact (to be ????) about polynomials is that, if two polynomials of degree n have equal value at $n+1$ distinct locations, then the two polynomials are equal. Let $t_0, t_1, \dots, t_n$ be n distinct numbers. Then the fact means that the evaluation linear transformation (see Example 2.5)

$$L(f(t)) = (f(t_0), f(t_1), \dots, f(t_n)) \colon P_n \to \mathbb{R}^{n+1}$$

is one-to-one. Since $\dim P_n = n + 1 = \dim \mathbb{R}^{n+1}$. The linear transformation is an isomorphism.

For each $0 \le k \le n$, the polynomial

$$p_k(t) = (t - t_0)(t - t_1) \dots (t - t_{k-1})(t - t_{k+1}) \dots (t - t_n) = \prod_{0 \le j \le n,\ j \ne k} (t - t_j).$$

satisfies

$$L(p_k(t)) = (0, 0, \dots, 0, p_k(t_k), 0, \dots, 0) = p_k(t_k)\vec{e}_k,$$

where $\vec{e}_0, \vec{e}_1, \dots, \vec{e}_n$ is the standard basis of $\mathbb{R}^{n+1}$ (vectors in $\mathbb{R}^{n+1}$ are $(x_0, x_1, \dots, x_n)$), and

$$p_k(t_k) = (t_k - t_0)(t_k - t_1)\dots(t_k - t_{k-1})(t_k - t_{k+1})\dots(t_k - t_n).$$

For any $f \in P_n$, we have

$$\begin{aligned}
L(f(t)) &= f(t_0)\vec{e}_0 + f(t_1)\vec{e}_1 + \dots + f(t_n)\vec{e}_n \\
&= \frac{f(t_0)}{p_0(t_0)}L(p_0(t)) + \frac{f(t_1)}{p_1(t_1)}L(p_1(t)) + \dots + \frac{f(t_n)}{p_n(t_n)}L(p_n(t)) \\
&= L\left(\frac{f(t_0)}{p_0(t_0)}p_0(t) + \frac{f(t_1)}{p_1(t_1)}p_1(t) + \dots + \frac{f(t_n)}{p_n(t_n)}p_n(t)\right).
\end{aligned}$$

Since L is an isomorphism, we get

$$f(t) = \frac{f(t_0)}{p_0(t_0)}p_0(t) + \frac{f(t_1)}{p_1(t_1)}p_1(t) + \dots + \frac{f(t_n)}{p_n(t_n)}p_n(t) = \sum_{k=0}^{n} f(t_k) \prod_{0 \le j \le n,\, j \ne k} \frac{t - t_j}{t_k - t_j}.$$

The polynomials $\prod_{0 \le j \le n,\, j \ne k} \frac{t-t_j}{t_k-t_j} \in P_n$ are the *Lagrange polynomials*, and the equality above is the *Lagrange interpolation*. It gives an explicit formula of a polynomial in terms of its values. For example, a quadratic polynomial $f(t)$ satisfying $f(-1) = 1, f(0) = 2, f(1) = 3$ is uniquely given by

$$f(t) = 1\frac{t(t-1)}{(-1)\cdot(-2)} + 2\frac{(t+1)(t-1)}{1\cdot(-1)} + 3\frac{(t+1)t}{4\cdot 3} = 2 - \frac{1}{4}t - \frac{5}{4}t^2.$$

Exercise 2.47. Suppose $\dim V = \dim W$ and $L\colon V \to W$ is a linear transformation. Prove that the following are equivalent.

1. L is invertible.
2. L has left inverse: There is $K\colon W \to V$, such that $K \circ L = I$.
3. L has right inverse: There is $K\colon W \to V$, such that $L \circ K = I$.

Moreover, show that the two K in the second and third parts must be the same.

Exercise 2.48. Explain $\dim \mathrm{Hom}(V, W) = \dim V \dim W$.

Invertible Matrix

The addition, scalar multiplication, and multiplication of matrices are defined as the calculations of the addition, scalar multiplication, and composition of linear transformations. The inverse matrix corresponds to the inverse of linear transformation.

A matrix A is *invertible* if $L(\vec{x}) = A\vec{x}\colon \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation. By Proposition 2.15, the matrix must be a square matrix. The *inverse matrix* A^{-1} is the matrix of the inverse linear transformation L^{-1}.

A linear transformation K is the inverse of L means $L \circ K = I$ and $K \circ L = I$. Correspondingly, a matrix B is the inverse of A means $AB = BA = I$ is the identity matrix.

Exercise 2.47 shows that, in case of equal dimension, $L \circ K = I$ is equivalent to $K \circ L = I$. Correspondingly, for square matrices, $AB = I$ is equivalent to $BA = I$. See Exercise 2.49.

Example 2.34. The inverse of the identity linear transformation is the identity. Therefore the inverse of the identity matrix is the identity matrix: $I_{n\times n}^{-1} = I_{n\times n}$.

Example 2.35. The rotation R_θ of the plane by angle θ in Example 2.3 is invertible, with the inverse $R_\theta^{-1} = R_{-\theta}$ being the rotation by angle $-\theta$. Therefore the matrix of $R_{-\theta}$ is the inverse of the matrix of R_θ

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

One can directly verify that the multiplication of the two matrices is the identity.

The flipping F_ρ in Example 2.3 is also invertible, with the inverse $F_\rho^{-1} = F_\rho$ being the flipping itself. Therefore the matrix of F_ρ is the inverse of itself ($\theta = 2\rho$)

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

Exercise 2.49. Suppose A is a square matrix. Prove that the following are equivalent.

1. A is invertible.

2. A has left inverse: There is a matrix B, such that $BA = I$.

3. A has right inverse: There is a matrix B, such that $AB = I$.

Moreover, show that the two B in the second and third parts must be the same.

Exercise 2.50. Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ and

$$\beta = \{a_{11}\vec{v}_1 + a_{21}\vec{v}_2 + \cdots + a_{n1}\vec{v}_n, a_{12}\vec{v}_1 + a_{22}\vec{v}_2 + \cdots + a_{n2}\vec{v}_n, \dots, a_{1n}\vec{v}_1 + a_{2n}\vec{v}_2 + \cdots + a_{nn}\vec{v}_n\}.$$

Prove that any two of following statements imply the third.

1. α is a basis.

2. β is a basis.

3. The coefficient matrix (a_{ij}) is invertible.

The result is a combination of Exercises 2.35, 2.44, and generalises Exercises 1.45.

Exercise 2.51. Prove that the trace defined in Example 2.5 satisfies $\mathrm{tr}AXA^{-1} = \mathrm{tr}X$.

By Theorem 2.12, a linear transformation is invertible if and only if it is onto and one-to-one. Correspondingly, a matrix A is invertible if and only if the system of linear equation $A\vec{x} = \vec{b}$ has unique solution for all $\vec{b}$. See the discussion before Examples 2.22 and 2.25. By Propositions 1.10 and 1.12, we get the many criteria for the invertibility of matrix (also see Exercise 2.49).

Proposition 2.17. *The following are equivalent for an $n \times n$ matrix A.*

1. *A is invertible.*
2. *$A\vec{x} = \vec{b}$ has solution for all $\vec{b} \in \mathbb{R}^n$.*
3. *The solution of $A\vec{x} = \vec{b}$ is unique.*
4. *The homogeneous system $A\vec{x} = \vec{0}$ has only trivial solution $\vec{x} = \vec{0}$.*
5. *$A\vec{x} = \vec{b}$ has unique solution for all $\vec{b} \in \mathbb{R}^n$.*
6. *The columns of A span $\mathbb{R}^n$.*
7. *The columns of A are linearly independent.*
8. *The columns of A form a basis of $\mathbb{R}^n$.*
9. *All rows of A are pivot.*
10. *All columns of A are pivot.*
11. *The reduced row echelon form of A is the identity matrix I.*

Let A be the matrix of an invertible linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$. Let $A^{-1} = (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n)$ be the matrix of the inverse linear transformation L^{-1}. Then $\vec{w}_i = L^{-1}(\vec{e}_i) = A^{-1}\vec{e}_i$. This implies $A\vec{w}_i = L(\vec{w}_i) = \vec{e}_i$. Therefore the i-th column of A^{-1} is the solution of the system of linear equation $A\vec{x} = \vec{e}_i$. The solution can be calculated by the reduced row echelon form of the augmented matrix

$$(A\ \vec{e}_i) \to (I\ \vec{w}_i).$$

Here the row operations can reduce A to I by Proposition 2.17. Then the solution of $A\vec{x} = \vec{e}_i$ is exactly the last column of the reduced row echelon form $(I\ \vec{w}_i)$.

Since the systems of liner equations $A\vec{x} = \vec{e}_1$, $A\vec{x} = \vec{e}_2$, $\dots$, $A\vec{x} = \vec{e}_n$ have the same coefficient matrix A, we may solve these equations simultaneously by combining the row operations

$$(A\ I) = (A\ \vec{e}_1\ \vec{e}_2\ \cdots\ \vec{e}_n) \to (I\ \vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n) = (I\ A^{-1}).$$

Example 2.36. The row operation in Example 2.20 shows

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{pmatrix}.$$

In general, one can directly verify that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad ad \neq bc.$$

Example 2.37. The row operation in Example 1.14 suggests that the matrix

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 8 \end{pmatrix}$$

is invertible. Then we carry out the row operations

$$\begin{aligned}(A\ I) = \begin{pmatrix} 1 & 4 & 7 & 1 & 0 & 0 \\ 2 & 5 & 8 & 0 & 1 & 0 \\ 3 & 6 & 8 & 0 & 0 & 1 \end{pmatrix} &\to \begin{pmatrix} 1 & 4 & 7 & 1 & 0 & 0 \\ 0 & -3 & -6 & -2 & 1 & 0 \\ 0 & -6 & -13 & -3 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{pmatrix} \\ &\to \begin{pmatrix} 1 & 0 & -1 & -\frac{5}{3} & \frac{4}{3} & 0 \\ 0 & 1 & 2 & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & -\frac{8}{3} & \frac{10}{3} & -1 \\ 0 & 1 & 0 & \frac{8}{3} & -\frac{13}{3} & 2 \\ 0 & 0 & 1 & -1 & 2 & -1 \end{pmatrix}.\end{aligned}$$

Therefore

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 8 \end{pmatrix}^{-1} = \begin{pmatrix} -\frac{8}{3} & \frac{10}{3} & -1 \\ \frac{8}{3} & -\frac{13}{3} & 2 \\ -1 & 2 & -1 \end{pmatrix}.$$

Example 2.38. The row operation in Example 1.31 shows that

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}^{-1} = \frac{1}{3}\begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix}.$$

We note that Example 1.27 already shows that the three column vectors on the left form a basis. By Proposition 2.17, this means that the matrix is invertible.

In terms of linear transformation, the result means that the linear transformation

$$L\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 + x_3 \\ -x_1 + x_3 \\ -x_2 + x_3 \end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^3$$

is invertible, and the inverse is

$$L^{-1}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} x_1 - 2x_2 + x_3 \\ x_1 + x_2 - 2x_3 \\ x_1 + x_2 + x_3 \end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^3.$$

Exercise 2.52. What is the inverse of 1×1 matrix (a)?

Exercise 2.53. Verify the formula for the inverse of 2×2 matrix in Example 2.36 by multiplying the two matrices together. Moreover, show that the 2×2 matrix is not invertible when $ad = bc$.

Exercise 2.54. Find inverse matrix.

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

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

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

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

5. $\begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}$.

6. $\begin{pmatrix} 1 & a & b & c \\ 0 & 1 & a & b \\ 0 & 0 & 1 & a \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

7. $\begin{pmatrix} b & c & 1 \\ 1 & 0 & 0 \\ a & 1 & 0 \end{pmatrix}$.

8. $\begin{pmatrix} c & a & b & 1 \\ b & 1 & a & 0 \\ a & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$.

Exercise 2.55. Find inverse matrix.

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

2. $\begin{pmatrix} 1 & a_1 & a_2 & \cdots & a_{n-1} \\ 0 & 1 & a_1 & \cdots & a_{n-2} \\ 0 & 0 & 1 & \cdots & a_{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$.

3. $\begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & a_1 \\ 0 & 0 & \cdots & a_1 & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & a_1 & \cdots & a_{n-2} & a_{n-1} \end{pmatrix}$.

2.5 Matrix of Linear Transformation

Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$ be (ordered) bases of (finite dimensional) vector spaces V and W. Then a linear transformation $L\colon V \to W$ can be translated into a linear transformation $L_{\beta\alpha}$ between Euclidean spaces.

$$\begin{tikzcd} V \arrow[r, "L"] \arrow[d, "{[\cdot]_\alpha}"', "\cong"] & W \arrow[d, "\cong"', "{[\cdot]_\beta}"] \\ \mathbb{R}^n \arrow[r, "L_{\beta\alpha}"] & \mathbb{R}^n \end{tikzcd} \qquad L_{\beta\alpha}([\vec{v}]_\alpha) = [L(\vec{v})]_\beta. \tag{2.4}$$

Please pay attention to the notation, that $L_{\beta\alpha}$ is the corresponding linear transformation from α to β. The matrix $[L]_{\beta\alpha}$ of the linear transformation $L_{\beta\alpha}$ is the *matrix of L with respect to the bases α and β*. This means

$$[L]_{\beta\alpha}[\vec{v}]_\alpha = [L(\vec{v})]_\beta. \tag{2.5}$$

Proposition 2.18. *The matrix of linear transformation has the following properties*

$$[I_V]_{\alpha\alpha} = I, \quad [L+K]_{\beta\alpha} = [L]_{\beta\alpha} + [K]_{\beta\alpha}, \quad [aL]_{\beta\alpha} = a[L]_{\beta\alpha}, \quad [L\circ K]_{\gamma\alpha} = [L]_{\gamma\beta}[K]_{\beta\alpha}.$$

Proof. The equality $[I_V]_{\alpha\alpha} = I$ for matrices is equivalent to that $(I_V)_{\alpha\alpha}$ is the identity linear transformation. This follows from

$$(I_V)_{\alpha\alpha}([\vec{v}]_\alpha) = [I_V(\vec{v})]_\alpha = [\vec{v}]_\alpha.$$

The equality $[L+K]_{\beta\alpha} = [L]_{\beta\alpha}+[K]_{\beta\alpha}$ for matrices is equivalent to the equality $(L+K)_{\beta\alpha} = L_{\beta\alpha}+K_{\beta\alpha}$ for linear transformations. We verify the equality for linear transformations by using Proposition 1.15 and (2.4)

$$\begin{aligned}(L+K)_{\beta\alpha}([\vec{v}]_\alpha) &= [(L+K)(\vec{v})]_\beta = [L(\vec{v})+K(\vec{v})]_\beta \\ &= [L(\vec{v})]_\beta + [K(\vec{v})]_\beta = L_{\beta\alpha}([\vec{v}]_\alpha) + K_{\beta\alpha}([\vec{v}]_\alpha).\end{aligned}$$

The verification of the other equalities are left as exercise. □

Exercise 2.56. Prove that $(aL)_{\beta\alpha} = aL_{\beta\alpha}$ and $(L\circ K)_{\gamma\alpha} = L_{\gamma\beta}K_{\beta\alpha}$. This implies the equalities $[aL]_{\beta\alpha} = a[L]_{\beta\alpha}$ and $[L\circ K]_{\gamma\alpha} = [L]_{\gamma\beta}[K]_{\beta\alpha}$ in Proposition 2.18.

Exercise 2.57. Prove that L is invertible if and only if $[L]_{\beta\alpha}$ is invertible.

The i-th column of the matrix $[L]_{\beta\alpha}$ is

$$L_{\beta\alpha}(\vec{e}_i) = L_{\beta\alpha}([\vec{v}_i]_\alpha) = [L(\vec{v}_i)]_\beta.$$

Therefore we get the formula

$$[L]_{\beta\alpha} = ([L(\vec{v}_1)]_\beta\ [L(\vec{v}_2)]_\beta\ \cdots\ [L(\vec{v}_n)]_\beta) = [L(\alpha)]_\beta, \quad L(\alpha) = \{L(\vec{v}_1), L(\vec{v}_2), \dots, L(\vec{v}_n)\}. \quad (2.6)$$

For the special case that $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ are Euclidean spaces, and α and β are the standard bases ϵ, we get back the standard matrix (2.3) of linear transformation

$$[L]_{\epsilon\epsilon} = (L(\vec{e}_1)\ L(\vec{e}_2)\ \cdots\ L(\vec{e}_n)).$$

Example 2.39. With respect to the standard monomial bases $\alpha = \{1, t, t^2, t^3\}$ and $\beta = \{1, t, t^2\}$ of P_3 and P_2, the derivative linear transformation $D\colon P_3 \to P_2$ has matrix

$$[D]_{\beta\alpha} = [(1)', (t)', (t^2)', (t^3)']_{\{1,t,t^2\}} = [0, 1, 2t, 3t^2]_{\{1,t,t^2\}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

For example, the derivative $(1+2t+3t^2+4t^3)' = 2+6t+12t^2$ fits into

$$\begin{pmatrix} 2 \\ 6 \\ 12 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

If we modify the basis β to $\gamma = \{1, t-1, (t-1)^2\}$ in Example 1.25, then

$$\begin{aligned}[D]_{\gamma\alpha} &= [0, 1, 2t, 3t^2]_{\{1,t-1,(t-1)^2\}} = [0, 1, 2[1+(t-1)], 3[1+(t-1)]^2]_{\{1,t-1,(t-1)^2\}} \\ &= [0, 1, 2+2(t-1), 3+6(t-1)+3(t-1)^2]_{\{1,t-1,(t-1)^2\}} \\ &= \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 3 \end{pmatrix}.\end{aligned}$$

Example 2.40. Consider the linear transformation $L(f) = f'' + (1+t^2)f' + tf\colon P_3 \to P_4$ inspired by Example 2.8. By

$$L(1) = t, \quad L(t) = 1 + 2t^2, \quad L(t^2) = 2 + 2t + 3t^3, \quad L(t^3) = 6t + 3t^2 + 4t^4,$$

we get

$$[L]_{\{1,t,t^2,t^3,t^4\}\{1,t,t^2,t^3\}} = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & 2 & 6 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

The row operation

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

shows that L is one-to-one and not onto.

Example 2.41 (Vandermonde Matrix)**.** Applying the evaluation linear transformation in Example 2.33 to the monomials, we get the matrix of the linear transformation

$$[L]_{\epsilon\{1,t,\dots,t^n\}} = \begin{pmatrix} 1 & t_0 & t_0^2 & \dots & t_0^n \\ 1 & t_1 & t_1^2 & \dots & t_1^n \\ 1 & t_2 & t_2^2 & \dots & t_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \dots & t_n^n \end{pmatrix}.$$

This is called the *Vandermonde Matrix*. Example 2.33 tells us that the matrix is invertible if and only if all t_i are distinct. Moreover, the interpolation formula in Example 2.33 can be regarded as the formula for the inverse linear transformation L^{-1}. This means that the k-th column of the inverse of the Vandermonde Matrix is the coefficients in

$$L^{-1}(\vec{e}_k) = \frac{p_k(t)}{p_k(t_k)} = \prod_{0\le j\le n,\, j\ne k} \frac{t-t_j}{t_k-t_j}.$$

For example, for $n=2$, the first column (i.e., $k=0$) of the inverse of the Vandermonde Matrix is

$$[L^{-1}(\vec{e}_0)]_{\{1,t,t^2\}} = \left[\frac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)}\right]_{\{1,t,t^2\}} = \frac{(t_1t_2, -t_1-t_2, 1)}{(t_0-t_1)(t_0-t_2)}.$$

Then

$$\begin{pmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{t_1t_2}{(t_0-t_1)(t_0-t_2)} & \frac{t_0t_2}{(t_1-t_0)(t_1-t_2)} & \frac{t_0t_1}{(t_2-t_0)(t_2-t_1)} \\ \frac{-t_1-t_2}{(t_0-t_1)(t_0-t_2)} & \frac{-t_0-t_2}{(t_1-t_0)(t_1-t_2)} & \frac{-t_0-t_1}{(t_2-t_0)(t_2-t_1)} \\ \frac{1}{(t_0-t_1)(t_0-t_2)} & \frac{1}{(t_1-t_0)(t_1-t_2)} & \frac{1}{(t_2-t_0)(t_2-t_1)} \end{pmatrix}.$$

Example 2.42. With respect to the standard basis of $M_{2\times 2}$

$$\sigma = \{S_1, S_2, S_3, S_4\} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

the transpose linear transformation $?^T \colon M_{2\times 2} \to M_{2\times 2}$ has matrix

$$[?^T]_{\sigma\sigma} = [\sigma^T]_\sigma = [S_1, S_3, S_2, S_4]_{\{S_1,S_2,S_3,S_4\}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Moreover, the linear transformation $A\cdot \colon M_{2\times 2} \to M_{2\times 2}$ of left multiplying by $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ has matrix

$$\begin{aligned} [A\cdot]_{\sigma\sigma} &= [A\sigma]_\sigma = [AS_1, AS_2, AS_3, AS_4]_{\{S_1,S_2,S_3,S_4\}} \\ &= [S_1 + 3S_3, S_2 + 3S_4, 2S_1 + 4S_3, 2S_2 + 4S_4]_{\{S_1,S_2,S_3,S_4\}} = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{pmatrix}. \end{aligned}$$

Exercise 2.58. Suppose two sets of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$ in V are related by a matrix

$$\vec{w}_i = a_{1i}\vec{v}_i + a_{2i}\vec{v}_i + \cdots + a_{ni}\vec{v}_n, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

1. Show that when α, β are bases, A is the matrix linear transformation given by (2.1).
2. Prove that if β spans V, then α spans V.
3. Prove that if α spans V, and $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$, then β spans V.
4. Prove that if β is linearly independent, and the solution of $A\vec{x} = \vec{b}$ is unique, then α is linearly independent.
5. If α is linearly independent, can you always say that β is linearly independent?
6. Prove that if A is invertible, then α is a basis of V if and only if β is a basis of V.

Change of Basis

The matrix $[L]_{\beta\alpha}$ of linear transformation $L \colon V \to W$ depends on the choice of bases α and β. If α' and β' are also (ordered) bases of V and W, then by Proposition 2.18, the matrix of L with respect to the new choice is

$$[L]_{\beta'\alpha'} = [I_W \circ L \circ I_V]_{\beta'\alpha'} = [I_W]_{\beta'\beta}[L]_{\beta\alpha}[I_V]_{\alpha\alpha'}. \tag{2.7}$$

This shows that the matrix of linear transformation is modified by multiplying matrices of the identity operator with respect to various bases.

Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$ are bases of V. By (2.5), the matrix $[I_V]_{\beta\alpha} = [I]_{\beta\alpha}$ is characterised by

$$[I]_{\beta\alpha}[\vec{v}]_\alpha = [\vec{v}]_\beta.$$

This shows that the matrix $[I]_{\beta\alpha}$ converts the α-coordinate to the β-coordinate. For this reason, we call $[I]_{\beta\alpha}$ the *matrix for the change of basis* from α to β. By (2.6), the matrix

$$[I]_{\beta\alpha} = ([\vec{v}_1]_\beta \ [\vec{v}_2]_\beta \ \cdots \ [\vec{v}_n]_\beta) = [\alpha]_\beta$$

is given by the β-coordinates of vectors in α.

Proposition 2.18 implies the following properties of the matrix for the change of basis.

Proposition 2.19. *The matrix for the change of basis has the following properties*

$$[I]_{\alpha\alpha} = I, \quad [I]_{\beta\alpha} = [I]_{\alpha\beta}^{-1}, \quad [I]_{\gamma\alpha} = [I]_{\gamma\beta}[I]_{\beta\alpha}.$$

Example 2.43. Let ϵ be the standard basis of $\mathbb{R}^n$. Then for any other basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of $\mathbb{R}^n$, the matrix for changing from α to ϵ is

$$[I]_{\epsilon\alpha} = [\alpha]_\epsilon = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_n).$$

The matrix has vectors in α as columns.

In general, the matrix for changing from $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ to $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$ is

$$[I]_{\beta\alpha} = [I]_{\beta\epsilon}[I]_{\epsilon\alpha} = [I]_{\epsilon\beta}^{-1}[I]_{\epsilon\alpha} = [\beta]_\epsilon^{-1}[\alpha]_\epsilon.$$

For example, the matrix for changing from the basis in Examples 1.19, 1.21 and 1.32

$$\alpha = \{(1,2,3),(2,3,4),(3,4,1)\}$$

to the basis in Examples 1.27, 1.31 and 2.16

$$\beta = \{(1,-1,0),(1,0,-1),(1,1,1)\}$$

is (see Example 2.38)

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 0 & 0 & -4 \\ -3 & -5 & 1 \\ 6 & 9 & 8 \end{pmatrix}.$$

Example 2.44. Consider the basis $\alpha_\theta = \{(\cos\theta, \sin\theta), (-\sin\theta, \cos\theta)\}$ of unit length vectors on the plane at angles θ and $\theta + \frac{\pi}{2}$. The matrix for the change of basis from α_{θ_1} to α_{θ_2} is obtained from the α_{θ_2}-coordinates of vectors in α_{θ_1}. Since α_{θ_1} is obtained from α_{θ_2} by rotating $\theta = \theta_1 - \theta_2$, the coordinates are the same as the ϵ-coordinates of vectors in α_θ. This means

$$[I]_{\alpha_{\theta_2}\alpha_{\theta_1}} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta_1 - \theta_2) & -\sin(\theta_1 - \theta_2) \\ \sin(\theta_1 - \theta_2) & \cos(\theta_1 - \theta_2) \end{pmatrix}.$$

This is consistent with the formula in Example 2.43

$$[I]_{\alpha_{\theta_2}\alpha_{\theta_1}} = [\alpha_{\theta_2}]_\epsilon^{-1}[\alpha_{\theta_1}]_\epsilon = \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix}^{-1} \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix}.$$

Example 2.45. The matrix for the change from the basis $\alpha = \{1, t, t^2, t^3\}$ of P_3 to another basis $\beta = \{1, t-1, (t-1)^2, (t-1)^3\}$ is

$$\begin{aligned}
[I]_{\beta\alpha} &= [1, t, t^2, t^3]_{\{1,t-1,(t-1)^2,(t-1)^3\}} \\
&= [1, 1+(t-1), 1+2(t-1)+(t-1)^2, 1+3(t-1)+3(t-1)^2+(t-1)^3]_{\{1,t-1,(t-1)^2,(t-1)^3\}} \\
&= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

The inverse of this matrix is

$$\begin{aligned}
[I]_{\alpha\beta} &= [1, t-1, (t-1)^2, (t-1)^3]_{\{1,t,t^2,t^3\}} \\
&= [1, -1+t, 1-2t+t^2, -1+3t-3t^2+t^3]_{\{1,t,t^2,t^3\}} \\
&= \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned}$$

We can also use the method outlined before Example 2.37 to calculate the inverse. But the method above is simpler.

The equality

$$(t+1)^3 = 1 + 3t + 3t^2 + t^3 = ((t-1)+2)^3 = 8 + 12(t-1) + 6(t-1)^2 + (t-1)^3$$

gives coordinates

$$[(t+1)^3]_\alpha = (1, 3, 3, 1), \quad [(t+1)^3]_\beta = (8, 12, 6, 1).$$

The two coordinates are related by the matrices for the change of basis

$$\begin{pmatrix} 8 \\ 12 \\ 6 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 3 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 3 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 8 \\ 12 \\ 6 \\ 1 \end{pmatrix}.$$

Exercise 2.59. Directly verify that the matrices in Example 2.39 satisfy $[D]_{\gamma\alpha} = [I]_{\gamma\beta}[D]_{\beta\alpha}$, where $[I]_{\gamma\beta}$ is given by Example 2.45.

Exercise 2.60. For the linear transformation L in Example 2.40, find $[L]_{\{1,t,t^2,t^3,t^4\}\{1,t-1,(t-1)^2,(t-1)^3\}}$ by using matrix for the change of basis in Example 2.45.

Similar Matrix

For linear operators $L\colon V \to V$, we usually choose the same basis α for the domain V and the range V. The matrix of the linear operator with respect to the basis α is $[L]_{\alpha\alpha}$. The matrices with respect to two bases of V are related by

$$[L]_{\beta\beta} = [I]_{\beta\alpha}[L]_{\alpha\alpha}[I]_{\alpha\beta} = [I]_{\alpha\beta}^{-1}[L]_{\alpha\alpha}[I]_{\alpha\beta} = [I]_{\beta\alpha}[L]_{\alpha\alpha}[I]_{\beta\alpha}^{-1}.$$

We say the two matrices $A = [L]_{\alpha\alpha}$ and $B = [L]_{\beta\beta}$ are *similar* in the sense that they are related by

$$B = P^{-1}AP = QAQ^{-1},$$

where P (matrix for changing from β to α) is an invertible matrix with $P^{-1} = Q$ (matrix for changing from α to β).

Example 2.46. The orthogonal projection P in Example 2.16 is a linear operator on $\mathbb{R}^3$. From its geometrical meaning, we know $P(\vec{v}_1) = \vec{v}_1$, $P(\vec{v}_2) = \vec{v}_2$, $P(\vec{v}_3) = \vec{0}$ for the vectors in the basis

$$\alpha = \{(1,-1,0),(1,0,-1),(1,1,1)\}.$$

This can be interpreted as

$$[P]_{\alpha\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The usual matrix of P is the matrix $[P]_{\epsilon\epsilon}$ with respect the the standard basis. By Example 2.43, we have

$$[I]_{\epsilon\alpha} = [\alpha]_\epsilon = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}.$$

Therefore

$$[P]_{\epsilon\epsilon} = [I]_{\epsilon\alpha}[P]_{\alpha\alpha}[I]_{\epsilon\alpha}^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}^{-1} = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

The matrix is the same as the one we obtained in Example 2.16 by another method.

Example 2.47. Consider the linear operator $L(f(t)) = tf'(t) + f(t)\colon P_3 \to P_3$. Applying the operator to the bases $\alpha = \{1, t, t^2, t^3\}$, we get

$$L(1) = 1, \quad L(t) = 2t, \quad L(t^2) = 3t^2, \quad L(t^3) = 4t^3.$$

Therefore

$$[L]_{\alpha\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

Consider another basis $\beta = \{1, t-1, (t-1)^2, (t-1)^3\}$ of P_2. By Example 2.45, we have

$$[I]_{\beta\alpha} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad [I]_{\alpha\beta} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore

$$\begin{aligned} [L]_{\beta\beta} &= [I]_{\beta\alpha}[L]_{\alpha\alpha}[I]_{\alpha\beta} \\ &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 4 \end{pmatrix}. \end{aligned}$$

We can verify the result by directly applying $L(f) = (tf(t))'$ to vectors in β

$$\begin{aligned} L(1) &= 1, \\ L(t-1) &= [(t-1) + (t-1)^2]' = 1 + 2(t-1), \\ L((t-1)^2) &= [(t-1)^2 + (t-1)^3]' = 2(t-1) + 3(t-1)^2, \\ L((t-1)^3) &= [(t-1)^3 + (t-1)^4]' = 3(t-1)^2 + 4(t-1)^3. \end{aligned}$$

Exercise 2.61. Explain that if A is similar to B, then B is similar to A.

Exercise 2.62. Explain that if A is similar to B, and B is similar to C, then A is similar to C.

Trace

For a finite dimensional vector space V, we choose a basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of V and define the *trace* of an operator $L\colon V \to V$ to be the trace of the matrix $[L]_{\alpha\alpha}$ introduced in Example 2.5. Specifically, this means that, if

$$\begin{aligned} L(\vec{v}_1) &= a_{11}\vec{v}_1 + a_{21}\vec{v}_2 + \cdots + a_{n1}\vec{v}_n, \\ L(\vec{v}_2) &= a_{12}\vec{v}_1 + a_{22}\vec{v}_2 + \cdots + a_{n2}\vec{v}_n, \\ &\vdots \\ L(\vec{v}_n) &= a_{1n}\vec{v}_1 + a_{2n}\vec{v}_2 + \cdots + a_{nn}\vec{v}_n, \end{aligned}$$

then

$$\mathrm{tr}L = a_{11} + a_{22} + \cdots + a_{nn}.$$

Since choosing another basis will change the matrix $[L]_{\alpha\alpha}$ to a similar matrix, by Exercise 2.51, the trace is independent of the choice of basis. Moreover, by Exercise 2.5, the trace is a linear functional on the vector space $\mathrm{Hom}(V, V)$ of all linear operators on V.

The trace of linear operators can be characterised as the "universal" linear functional on the vector space of linear operators.

Proposition 2.20. *Suppose there is a linear functional t_V on Hom(V, V) for each finite dimensional vector space V. If $t_W(PLP^{-1}) = t_V(L)$ for any linear operator $L\colon V \to V$ and isomorphism $P\colon V \to W$, then there is a constant c, such that $t_V(L) = c\,\mathrm{tr}L$.*

????????????????

3 Subspace

As human civilisation got more sophisticated, they found it necessary to extend their number systems. The ancient Greeks found that rational numbers $\mathbb{Q}$ was not sufficient for describing lengths in geometry, and the problem was solved later by the Arabs who extended the rational numbers to the real numbers $\mathbb{R}$. Then the Italians found it useful to take the square root of -1 in their search for the roots of cubic equations, and the idea led to the extension of real numbers to complex numbers $\mathbb{C}$.

In extending the number system, we still wish to preserve the key features of the old system. This means that the inclusions $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ is compatible with the four arithmetic operations. In other words, $2 + 3 = 5$ is an equality of rational numbers, and is also an equality of real (or complex) numbers. In this sense, we may call $\mathbb{Q}$ a sub-number system of $\mathbb{R}$ and $\mathbb{C}$.

When the idea is applied to vector spaces, we are interested in inclusions that are compatible with addition and scalar multiplication.

Definition 3.1. A subset H of a vector space V is a *subspace* if it satisfies

$$\vec{u}, \vec{v} \in H,\ a, b \in \mathbb{R} \implies a\vec{u} + b\vec{v} \in H.$$

Using the addition and scalar multiplication of V, the subset H is also a vector space. One should imagine that a subspace is a flat and infinite (with the only exception of the trivial subspace) subset passing through the origin.

The smallest subspace is the *trivial subspace* $\{\vec{0}\}$. The biggest subspace is the whole space V itself. Polynomials of degree ≤ 3 is a subspace of polynomials of degree ≤ 5. All polynomials is a subspace of all functions. Although $\mathbb{R}^3$ can be identified with a subspace of $\mathbb{R}^5$ (in many different ways), $\mathbb{R}^3$ is *not* a subspace of $\mathbb{R}^5$.

Proposition 3.2. *If H is a subspace of a finite dimensional vector space V, then* $\dim H \le \dim V$. *Moreover, $H = V$ if and only if* $\dim H = \dim V$.

Proof. Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be a basis of H. Then α is a linearly independent set in V. By Theorem 1.18, we can get a basis of V by adding more vectors to α. This shows that a basis of V contains more vectors than α and therefore proves $\dim V \ge \dim H$. Moreover, the case $\dim V = \dim H$ means the number of vectors in α is $\dim V$. By Proposition 1.21, this means that α is also a basis of V (in addition to being a basis of H). Therefore $H = V$. □

Exercise 3.1. Determine whether the subset is a subspace of $\mathbb{R}^2$.

1. $\{(x, 0) \colon x \in \mathbb{R}\}$.
2. $\{(x, y) \colon x + y = 0\}$.
3. $\{(x, y) \colon 2x - 3y = 0\}$.
4. $\{(x, y) \colon 2x - 3y = 1\}$.
5. $\{(x, y) \colon xy = 0\}$.
6. $\{(x, y) \colon x, y \in \mathbb{Q}\}$.

Exercise 3.2. Determine whether the subset is a subspace of $\mathbb{R}^3$.

1. $\{(x, 0, z) \colon x, z \in \mathbb{R}\}$.
2. $\{(x, y, z) \colon x + y + z = 0\}$.
3. $\{(x, y, z) \colon x + y + z = 1\}$.
4. $\{(x, y, z) \colon x + y + z = 0, x + 2y + 3z = 0\}$.

Exercise 3.3. Determine whether the subset is a subspace of P_n.

1. even polynomials.
2. polynomials satisfying $f(1) = 0$.
3. polynomials satisfying $f(0) = 1$.
4. polynomials satisfying $f'(0) = f(1)$.

Exercise 3.4. Determine whether the subset is a subspace of C^∞.

1. odd functions.
2. functions satisfying $f'' + f = 0$.
3. functions satisfying $f'' + f = 1$.
4. functions satisfying $f'(0) = f(1)$.
5. functions satisfying $\lim_{t\to\infty} f(t) = 0$.
6. functions satisfying $\lim_{t\to\infty} f(t) = 1$.
7. functions such that $\lim_{t\to\infty} f(t)$ diverges.
8. functions satisfying $\int_0^1 f(t)dt = 0$.

Exercise 3.5. Determine whether the subset is a subspace of the space of all sequences (x_n).

1. x_n converges.
2. x_n diverges.
3. The series $\sum x_n$ converges.
4. The series $\sum x_n$ absolutely converges.

Exercise 3.6. Prove that H is a subspace if and only if $\vec{0} \in H$ and $a\vec{v} + \vec{w} \in H$ for any $a \in \mathbb{R}$ and $\vec{v}, \vec{w} \in H$.

Exercise 3.7. Suppose H is a subspace of V, and $\vec{v} \in V$. Prove that $\vec{v} + H = \{\vec{v} + \vec{h} \colon \vec{h} \in H\}$ is still a subspace if and only if $\vec{v} \in H$.

Exercise 3.8. Suppose H is a subspace of V. Prove that the *inclusion* $i(\vec{h}) = \vec{h} \colon H \to V$ is a one-to-one linear transformation.

For any linear transformation $L \colon V \to W$, the *restriction* $L|_H = L \circ i \colon H \to W$ is still a linear transformation.

Exercise 3.9. If H is a subspace of V and V is a subspace of H, what can you conclude?

Exercise 3.10. Suppose H and H' are subspaces of V.

1. Prove that the sum $H + H' = \{\vec{h} + \vec{h}' \colon \vec{h} \in H,\ \vec{h}' \in H'\}$ is a subspace.

2. Prove that the intersection $H \cap H'$ is a subspace.

When is the union $H \cup H'$ a subspace?

3.1 Span, Range and Rank

The *span* of a set of vectors

$$\text{Span}\alpha = \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} = \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n : x_i \in \mathbb{R}\},$$

is a subspace because the linear combination of linear combinations is still a linear combination (see (1.1) and (1.2)).

The span of a nonzero vector

$$\text{Span}\{\vec{v}\} = \{x\vec{v} : x \in \mathbb{R}\} = \mathbb{R}\vec{v}$$

is the straight line in the direction of the vector. The span of two non-parallel vectors is the 2-dimensional plane containing the origin and the two vectors (or containing the parallelogram formed by the two vectors, see Figure 1.2). If two vectors are parallel, then the span is reduced to a line in the direction of the two vectors.

Exercise 3.11. Prove that $\text{Span}\alpha$ is the smallest subspace containing α.

Exercise 3.12. Prove that $\alpha \subset \beta$ implies $\text{Span}\alpha \subset \text{Span}\beta$.

Exercise 3.13. Prove that $\vec{v}$ is a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ if and only if $\text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n, \vec{v}\} = \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$.

Exercise 3.14. Prove that if one vector is a linear combination of the other vectors, then deleting the vector does not change the span.

Exercise 3.15. Exercise 1.12 shows that a linear combination of linear combinations is still a linear combination. Use the exercise to prove that $\text{Span}\alpha \subset \text{Span}\beta$ if and only if vectors in α are linear combinations of vectors in β. In particular, $\text{Span}\alpha = \text{Span}\beta$ if and only if vectors in α are linear combinations of vectors in β, and vectors in β are linear combinations of vectors in α.

First Calculation of Span

By the definition, α is already a spanning set of the vector space $\text{Span}\alpha$. By Theorem 1.17, we can obtain a basis of $\text{Span}\alpha$ by finding a maximal linearly independent set in α.

Example 3.1. To find a basis for the span of

$$\vec{v}_1 = (1, 2, 3), \quad \vec{v}_2 = (4, 5, 6), \quad \vec{v}_3 = (7, 8, 9), \quad \vec{v}_4 = (10, 11, 12),$$

Consider the row operation

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

If we restrict the row operations to the first two columns, then we find that $\{\vec{v}_1, \vec{v}_2\}$ is linearly independent. If we restrict the row operations to the first three columns, then we see that adding $\vec{v}_3$ gives linearly dependent set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, because the third column is not pivot. By the same reason, $\{\vec{v}_1, \vec{v}_2, \vec{v}_4\}$ is also linearly dependent. Therefore $\{\vec{v}_1, \vec{v}_2\}$ is a maximal linearly independent set inside $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$. By Theorem 1.17, $\{\vec{v}_1, \vec{v}_2\}$ is a basis of $\mathrm{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$.

The example shows the following way of finding a basis of a span inside $\mathbb{R}^n$. Given $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$, we carry out row operation on the matrix $(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_k)$. Then the pivot columns in $(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_k)$ form a basis of $\mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$. In particular, we have

$$\dim \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} = \text{number of pivots (after row operation) in } (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_k). \tag{3.1}$$

Exercise 3.16. Show that $\{\vec{v}_1, \vec{v}_3\}$ and $\{\vec{v}_1, \vec{v}_4\}$ are also bases in Example 3.1.

Exercise 3.17. Find a basis of the span.

1. $(1,2,3,4), (2,3,4,5), (3,4,5,6)$.
2. $(1,5,9), (2,6,10), (3,7,11), (4,8,12)$.
3. $(1,1,0,0), (1,0,1,0), (1,0,0,1), (0,1,1,0), (0,1,0,1), (0,0,1,1)$.
4. $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$.
5. $1-t, t-t^3, 1-t^3, t^3-t^5, t-t^5$.

Second Calculation of Span

The other way of finding a basis of $\mathrm{Span}\alpha$ is based on the following result. The proof is given in Exercises 1.16, 1.17, 1.18, 1.46.

Proposition 3.3. *The following operations do not change the span.*

1. $\{\vec{v}_1, \dots, \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n\} \to \{\vec{v}_1, \dots, \vec{v}_j, \dots, \vec{v}_i, \dots, \vec{v}_n\}$.
2. $\{\vec{v}_1, \dots, \vec{v}_i, \dots, \vec{v}_n\} \to \{\vec{v}_1, \dots, c\vec{v}_i, \dots, \vec{v}_n\}$, $c \neq 0$.
3. $\{\vec{v}_1, \dots, \vec{v}_i, \dots, \vec{v}_j, \dots, \vec{v}_n\} \to \{\vec{v}_1, \dots, \vec{v}_i + c\vec{v}_j, \dots, \vec{v}_j, \dots, \vec{v}_n\}$.

For vectors in Euclidean space, the operations in Proposition 3.3 are the *column operations* on the matrix $(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_k)$.

Example 3.2. For the four vectors in Example 3.1, we carry out the column on the matrix $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4)$

$$\begin{pmatrix}1&4&7&10\\2&5&8&11\\3&6&9&12\end{pmatrix}\xrightarrow[C_2-C_1]{\substack{C_4-C_3\\C_3-C_2}}\begin{pmatrix}1&3&3&3\\2&3&3&3\\3&3&3&3\end{pmatrix}\xrightarrow{\substack{C_4-C_3\\C_3-C_2\\\frac{1}{3}C_2}}\begin{pmatrix}1&1&0&0\\2&1&0&0\\3&1&0&0\end{pmatrix}\xrightarrow[C_1\leftrightarrow C_2]{C_1-C_2}\begin{pmatrix}1&0&0&0\\1&1&0&0\\1&2&0&0\end{pmatrix}.$$

The result is a *column echelon form*. By Proposition 3.3, we get

$$\mathrm{Span}\{\vec{v}_1,\vec{v}_2,\vec{v}_3,\vec{v}_4\}=\mathrm{Span}\{(1,1,1),(0,1,2),(0,0,0),(0,0,0)\}=\mathrm{Span}\{(1,1,1),(0,1,2)\}.$$

The two pivot columns $(1,1,1),(0,1,2)$ of the column echelon form are always linearly independent, and therefore form a basis of $\mathrm{Span}\{\vec{v}_1,\vec{v}_2,\vec{v}_3,\vec{v}_4\}$.

Example 3.3. By taking the transpose of the row operation in Example 3.1, we get the column operation

$$\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\\10&11&12\end{pmatrix}\to\begin{pmatrix}1&0&0\\4&-3&-6\\7&-6&-12\\10&-9&-18\end{pmatrix}\to\begin{pmatrix}1&0&0\\4&1&0\\7&2&0\\10&3&0\end{pmatrix}.$$

This implies that $(1,4,7,10),(0,1,2,3)$ form a basis of $\mathrm{Span}\{(1,4,7,10),(2,5,8,11),(3,6,9,12)\}$.

The examples show the following way of finding a basis of a span inside $\mathbb{R}^n$. Given $\vec{v}_1,\dots,\vec{v}_k\in\mathbb{R}^n$, we carry out column operation on the matrix $(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_k)$ and get a column echelon form. Then the pivot (i.e., nonzero) columns in the column echelon form is a basis of $\mathrm{Span}\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_k\}$. In particular, we have

$$\begin{aligned}\dim\mathrm{Span}\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n\}&=\text{number of pivots (after column operation) in }(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)\\&=\text{number of pivots (after row operation) in }(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)^T.\end{aligned}\tag{3.2}$$

Comparing (3.1) and (3.2), we conclude that applying row operation to A and A^T gives the same number of pivots.

We emphasise that, in Example 3.1, we carry out row operation on a matrix A to determine the pivot columns. Then the pivot columns of the *original matrix* A form a basis of the spanning subspace. In Examples 3.2 and 3.3, however, we carry out column operation on a matrix A. Then the pivot columns of the column echelon form (*not* the original A) form a basis of the spanning subspace.

Exercise 3.18. Explain how Proposition 3.3 follows from Exercises 1.16, 1.17, 1.18, 1.46.

Exercise 3.19. List all the 2×3 column echelon forms.

Exercise 3.20. Explain that the nonzero columns in a column echelon form are linearly independent.

Exercise 3.21. Explain that, if the columns of an $n \times n$ matrix is a basis of $\mathbb{R}^n$, then the rows of the matrix is also a basis of $\mathbb{R}^n$.

Exercise 3.22. Use another way find a basis of the span in Exercise 3.17.

Calculation of Extension to Basis

The column echelon form also gives a way to extend linearly independent vectors to a basis. By the proof of Theorem 1.18, this can be achieved by finding vectors not in the span.

Example 3.4. In Example 1.33, we have linearly independent vectors $\vec{w}_1 = (1,2,3,4), \vec{w}_2 = (2,3,4,1), \vec{w}_3 = (3,4,1,2)$ in $\mathbb{R}^4$. To extend to a basis of $\mathbb{R}^4$, we need to add one more vector $\vec{v}$ that is not in $\text{Span}\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$. To find the vector $\vec{v}$, we carry out column operation

$$(\vec{w}_1\ \vec{w}_2\ \vec{w}_3) = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 1 \\ 4 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & -3 \\ 4 & -3 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & -4 \\ -2 & 7 & 4 \end{pmatrix}.$$

We add the fourth column $\vec{v} = (0,0,0,1)$ apply and the same column operation to the first three columns, followed by the reduced column echelon form

$$(\vec{w}_1\ \vec{w}_2\ \vec{w}_3\ \vec{v}) = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 2 & 3 & 4 & 0 \\ 3 & 4 & 1 & 0 \\ 4 & 1 & 2 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & 1 & 1 & 0 \\ 3 & 1 & -3 & 0 \\ 4 & -3 & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & -4 & 0 \\ -2 & 7 & 4 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

By Proposition 3.3, we have

$$\text{Span}\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{v}\} = \text{Span}\{(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)\} = \mathbb{R}^4.$$

Then by Theorem 1.21, we conclude that $\vec{w}_1 = (1,2,3,4), \vec{w}_2 = (2,3,4,1), \vec{w}_3 = (3,4,1,2), \vec{v} = (0,0,0,1)$ form a basis of $\mathbb{R}^4$.

The reader should compare the method here with the method in the ealrier example.

Example 3.4 suggests the following practical way of extending a linearly independent set in $\mathbb{R}^n$ to a basis of $\mathbb{R}^n$. Suppose column operations on three linearly independent vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^5$ gives

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3) \xrightarrow{\text{col op}} \begin{pmatrix} \bullet & 0 & 0 \\ * & 0 & 0 \\ * & \bullet & 0 \\ * & * & \bullet \\ * & * & * \end{pmatrix}.$$

Then we may add $\vec{u}_1 = (0, \bullet, *, *, *)$ and $\vec{u}_2 = (0, 0, 0, 0, \bullet)$ to create pivots in the second and the fifth rows

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{u}_1\ \vec{u}_2) \xrightarrow[\text{first 3 col}]{\text{col op on}} \begin{pmatrix} \bullet & 0 & 0 & 0 & 0 \\ * & 0 & 0 & \bullet & 0 \\ * & \bullet & 0 & * & 0 \\ * & * & \bullet & * & 0 \\ * & * & * & * & \bullet \end{pmatrix} \xrightarrow[\text{col}]{\text{exchange}} \begin{pmatrix} \bullet & 0 & 0 & 0 & 0 \\ * & \bullet & 0 & 0 & 0 \\ * & * & \bullet & 0 & 0 \\ * & * & * & \bullet & 0 \\ * & * & * & * & \bullet \end{pmatrix} \xrightarrow[\text{form}]{\text{col echelon}} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

By Proposition 3.3, this implies that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{u}_1, \vec{u}_2\}$ has the same span as the standard basis. Then by Theorem 1.21, the five vectors form a basis of $\mathbb{R}^5$.

Range

The *range* (or *image*) of a map $f\colon X \to Y$ is

$$\text{Ran} f = f(X) = \{f(x)\colon \text{all } x \in X\} \subset Y.$$

For the map Instructor: Courses $\to$ Professors, the range is all the professors who teaches some courses.

The map is onto if and only if $f(X) = Y$. This suggests that we may consider the same map with smaller target

$$\tilde{f}\colon X \to f(X), \quad \tilde{f}(x) = f(x).$$

For the Instructor map, this means $\widetilde{\text{Instructor}}$: Courses $\to$ Teaching Professors. The advantage of the modification is the following.

Proposition 3.4. *For any map $f\colon X \to Y$, the corresponding map $\tilde{f}\colon X \to f(X)$ has the following properties.*

1. *$\tilde{f}$ is onto.*
2. *$\tilde{f}$ is one-to-one if and only if f is one-to-one.*

For a linear transformation $L\colon V \to W$, the range $L(V)$ is a subspace of W

$$\begin{aligned} \vec{w}, \vec{w}\,' \in L(V) &\implies \vec{w} = L(\vec{v}), \vec{w}\,' = L(\vec{v}\,'),\ \vec{v}, \vec{v}\,' \in V \\ &\implies \vec{w} + \vec{w}\,' = L(\vec{v}) + L(\vec{v}\,') = L(\vec{v} + \vec{v}\,') \in L(W). \end{aligned}$$

The argument for the scalar multiplication is similar.

Let $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be a basis of V. Then the linear transform is determined by

$$L(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n) = x_1\vec{w}_1 + x_2\vec{w}_2 + \cdots + x_n\vec{w}_n, \quad \vec{w}_i = L(\vec{v}_i).$$

Therefore the range subspace is a span

$$\text{Ran} L = L(V) = \text{Span}\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\},$$

and a basis can be calculated by methods in Examples 3.1, 3.2, 3.3.

A linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is given by an $m \times n$ matrix A. The range of L is the span of the column vectors of A, and is called the *column space* of A

$$\text{Ran}L = \text{Col}A \subset \mathbb{R}^m.$$

By $L(\vec{x}) = A\vec{x}$, we have

$$\text{Col}A = \{L(\vec{x})\colon \text{all } \vec{x} \in \mathbb{R}^n\} = \{A\vec{x}\colon \text{all } \vec{x} \in \mathbb{R}^n\}.$$

This shows that

$$A\vec{x} = \vec{b} \text{ has solution } \iff \vec{b} \in \text{Col}A.$$

Again a basis of the column space can be calculated by methods in Examples 3.1, 3.2, 3.3.

Of course we can also consider the span of the rows of A and get the *row space*. The row space and column space are clearly related by the transpose of the matrix

$$\text{Row}A = \text{Col}A^T \subset \mathbb{R}^n.$$

The row space is identified as the range $\text{Ran}(L^*)$ of the *adjoint transformation* $L^*\colon \mathbb{R}^m \to \mathbb{R}^n$ to be introduced later (i.e., A^T is the matrix of L^*).

Example 3.5. The derivative of a polynomial of degree n is a polynomial of degree $n-1$. Therefore we have linear transform $D(f) = f'\colon P_n \to P_m$ for $m \ge n-1$, and $\text{Ran}D = P_{n-1} \subset P_m$. The linear transformation is onto if and only if $m = n-1$.

Example 3.6. Consider the linear transformation $L(f) = f'' + (1+t^2)f' + tf\colon P_3 \to P_4$ in Example 2.40. The row operation of the earlier example shows that

$$L(1) = t, \quad L(t) = 1 + 2t^2, \quad L(t^2) = 2 + 2t + 3t^3, \quad L(t^3) = 6t + 3t^2 + 4t^4$$

form a basis of $\text{Ran}L$. Alternatively, we can carry out the column operation

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & 2 & 6 \\ 0 & 2 & 0 & 3 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & -4 & 3 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & -4 & 0 \\ 0 & 0 & 3 & \frac{9}{4} \\ 0 & 0 & 0 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & -4 & 0 \\ 0 & 0 & 3 & 9 \\ 0 & 0 & 0 & 16 \end{pmatrix}.$$

This shows that $1+2t^2, t, -4t^2+3t^3, 9t^3+16t^4$ form a basis of $\text{Ran}L$. It is also easy to see that adding t^4 gives a basis of P_4.

Example 3.7. Consider the linear transformation $L(A) = A + A^T\colon M_{n\times n} \to M_{n\times n}$. Since $(A+A^T)^T = A^T + (A^T)^T = A^T + A = A + A^T$, and $X = A + A^T \in \text{Ran}L$ satisfies $X^T = X$. Such matrices are called *symmetric* because they are of the form (for $n = 3$, for example)

$$X = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} = \begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix}.$$

Conversely, if $X = X^T$, then for $A = \frac{1}{2}X$, we have

$$L(A) = A + A^T = \frac{1}{2}X + \frac{1}{2}X^T = \frac{1}{2}X + \frac{1}{2}X = X.$$

This shows that any symmetric matrix lies in $\mathrm{Ran}L$. Therefore the range consists of all the symmetric matrices. A basis of 3×3 symmetric matrices is given by

$$\begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix} = a\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + d\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + e\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} + f\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Exercise 3.23. Prove Proposition 3.4.

Exercise 3.24. Prove that $\mathrm{Ran}(f \circ g) \subset \mathrm{Ran}f$. Moreover, if g is onto, then $\mathrm{Ran}(f \circ g) = \mathrm{Ran}f$.

Exercise 3.25. Suppose $L \colon V \to W$ is a linear transformation.

1. Prove the modified linear transformation (see Proposition 3.4) $\tilde{L} \colon V \to L(V)$ is an onto linear transformation.

2. Let $i \colon L(V) \to W$ be the inclusion linear transformation in Exercise 3.8. Show that $L = i \circ \tilde{L}$.

Exercise 3.26. Suppose $L \colon V \to W$ is a linear transformation. Prove that L is one-to-one if and only if $\tilde{L} \colon V \to L(V)$ is an isomorphism.

Exercise 3.27. Suppose $L \colon V \to W$ is a linear transformation, and $H \subset V$ is a subspace. Prove that $L(H) = \{L(\vec{v}) \colon \text{all } \vec{v} \in H\}$ is a subspace.

Exercise 3.28. Show that the range of the linear transformation $L(A) = A - A^T \colon M_{n\times n} \to M_{n\times n}$ consists of matrices X satisfying $X^T = -X$. These are called *skew-symmetric* matrices.

Exercise 3.29. Find the dimensions of the subspaces of symmetric and skew-symmetric matrices.

Rank

The span of a set of vectors, the range of a linear transformation, and the column space of a matrix are different presentations of the same concept. Their size, which is their dimension, is the *rank*

$$\begin{aligned} \mathrm{rank}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} &= \dim \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}, \\ \mathrm{rank}L &= \dim \mathrm{Ran}L, \\ \mathrm{rank}A &= \dim \mathrm{Col}A. \end{aligned}$$

By the calculations illustrated by Examples 3.1, 3.2, 3.3, the rank of a matrix A is the number of pivots in the row echelon form, and is also the number of pivots in the column echelon form. This implies

$$\mathrm{rank}A^T = \mathrm{rank}A.$$

In terms of the adjoint transformation (to be introduced later), this means $\text{rank}L^* = \text{rank}L$.

Since the number of pivots of an $m \times n$ matrix is always no more than m and n, we have

$$\text{rank}A_{m\times n} \le \min\{m, n\}.$$

If the equality holds, then the matrix has *full rank*. This means either all rows are pivot, or all columns are pivot. The following rephrases Propositions 1.10, 1.12, 2.17.

Proposition 3.5. *Let A be an $m \times n$ matrix. Then* $\text{rank}A \le \min\{m, n\}$. *Moreover,*

1. *$A\vec{x} = \vec{b}$ has solution for all $\vec{b}$ if and only if* $\text{rank}A = m$.
2. *The solution of $A\vec{x} = \vec{b}$ is unique if and only if* $\text{rank}A = n$.
3. *A is invertible if and only if* $\text{rank}A = m = n$.

The following are the corresponding characterisations of rank for a set of vectors. Although the result can be proved by translating Proposition 3.5, an independent proof is given.

Proposition 3.6. *Let $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in V$. Then* $\text{rank}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \le \min\{n, \dim V\}$, *Moreover,*

1. *The vectors span V if and only if* $\text{rank}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} = \dim V$.
2. *The vectors are linearly independent if* $\text{rank}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} = n$.
3. *The vectors form a basis of V if and only if* $\text{rank}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} = \dim V = n$.

Proof. Let $H = \text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. Then $\dim H = \text{rank}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. By Proposition 1.20, we have $n \ge \dim H$. By Proposition 3.2, we also have $\dim V \ge \dim H$. Moreover, $\dim V = \dim H$ if and only if $H = V$. This is exactly the first statement.

By the definition, the vectors already span H. Then by taking V in Proposition 1.20 and Theorem 1.21 to be H, the vectors are linearly independent if and only if $n = \dim H$. This is exactly the second statement. $\square$

The following are the corresponding characterisations of rank for a linear transformation.

Proposition 3.7. *Let $L\colon V \to W$ be a linear transformation. Then* $\text{rank}L \le \min\{\dim V, \dim W\}$. *Moreover,*

1. *L is onto if and only if* $\text{rank}L = \dim W$.
2. *L is one-to-one if and only if* $\text{rank}L = \dim V$.
3. *L is invertible if and only if* $\text{rank}L = \dim V = \dim W$.

Proof. Let $H = L(V)$. We consider the modified linear transformation (see Proposition 3.4) $\tilde{L}\colon V \to H = L(V)$ that replaces W by H. Since $\tilde{L}(V) = L(V) = H$, we find that $\tilde{L}$ is onto and rank$\tilde{L}$ = rank$L = \dim H$.

Applying Proposition 2.8 to the onto linear transformation $\tilde{L}$, we get $\dim V \geq \dim H$. By applying Proposition 3.2 to $H \subset W$, we also have $\dim W \geq \dim H$. Moreover, $\dim W = \dim H$ if and only if $H = W$. This is exactly the first statement.

By Proposition 3.4, L is one-to-one if and only if $\tilde{L}$ is one-to-one. If $\tilde{L}$ is one-to-one, then by $\tilde{L}$ always onto, we know $\tilde{L}$ is an isomorphism. By Propositions 2.15, we have $\dim V = \dim H$. Conversely, if $\dim V = \dim H$, then by Theorem 2.16, $\tilde{L}$ being onto implies $\tilde{L}$ being one-to-one. This proves the second statement. □

3.2 Kernel

The *kernel* of a linear transformation $L\colon V \to W$ is the *preimage* of the zero vector

$$\text{Ker}L = L^{-1}(\vec{0}) = \{\vec{v}\colon \vec{v} \in V \text{ and } L(\vec{v}) = \vec{0}\} \subset V.$$

The following shows that the kernel is a subspace of V

$$\vec{v}, \vec{v}\,' \in \text{Ker}(L) \implies L(\vec{v}) = \vec{0},\ L(\vec{v}\,') = \vec{0} \implies L(\vec{v} + \vec{v}\,') = L(\vec{v}) + L(\vec{v}\,') = \vec{0} + \vec{0} = \vec{0}.$$

The argument for the scalar multiplication is similar.

For a linear transformation $L(\vec{x}) = A\vec{x}\colon \mathbb{R}^n \to \mathbb{R}^m$ between vector spaces, the kernel is the *null space* of the matrix

$$\text{Nul}A = \text{Ker}L = \{\vec{v}\colon \text{all } \vec{v} \in \mathbb{R}^n \text{ satisfying } A\vec{v} = \vec{0}\}.$$

Therefore the null space is all the solutions of the *homogeneous system of linear equations* $A\vec{x} = \vec{0}$.

Exercise 3.30. Prove that $\text{Ker}(L \circ K) \supset \text{Ker}K$. Moreover, if L is one-to-one, then $\text{Ker}(L \circ K) = \text{Ker}K$.

Exercise 3.31. Suppose $L\colon V \to W$ is a linear transformation, and $H \subset W$ is a subspace. Prove that $L^{-1}(H) = \{\vec{v} \in V\colon L(\vec{v}) \in H\}$ is a subspace.

Exercise 3.32. Prove that $L(H) \cap \text{Ker}K = L(H \cap \text{Ker}(L \circ K))$.

Basis of Kernel

Example 3.8. The null space of

$$A = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}$$

is given by the general solution of the homogeneous system $A\vec{x} = \vec{0}$. The general solution obtained in Example 1.16 can be rephrased in vector form

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_3 + 2x_4 \\ -2x_3 - 3x_4 \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix} = x_3\vec{v}_1 + x_4\vec{v}_2.$$

This shows that Nul$A$ is spanned by $\vec{v}_1 = (1, -2, 1, 0), \vec{v}_2 = (2, -3, 0, 1)$. Since the projections of the two vectors to the free variable coordinates (x_3, x_4) is the standard basis $\vec{e}_1 = (1, 0), \vec{e}_2 = (0, 1)$ of $\mathbb{R}^2$, the two vectors are linearly independent (see Exercise 2.9). The following is another way to show that $\vec{v}_1, \vec{v}_2$ are linearly independent

$$x_3\vec{v}_1 + x_4\vec{v}_2 = \vec{0} \iff \text{the solution } \vec{x} = \vec{0} \implies \text{coordinates of the solution } x_3 = x_4 = 0.$$

The example shows that the general solution of a homogeneous equation $A\vec{x} = \vec{0}$ is $\vec{x} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k$, where $c_1, c_2, \dots, c_k$ are all the free variables. Since we can arbitrarily and independently choose the values of free variables, the null space Nul$A$ is spanned by $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$. Moreover, the last argument in the example shows that the vectors are always linearly independent. Therefore $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$ form a basis of Nul$A$. In particular, the *nullity* $\dim \text{Nul}A$ is the number of non-pivot columns (corresponding to free variables) of A. Since the rank $\dim \text{Col}A$ is the number of pivot columns (corresponding to non-free variables) of A, we find that $\dim \text{Nul}A + \dim \text{Col}A$ is the number of columns (corresponding to all variables) of A.

Theorem 3.8. *If A is an $m \times n$ matrix, then* $\dim \text{Nul}A + \text{rank}A = n$.

Next we extend the discussion of kernel to general vector space.

Example 3.9. For the linear transformation $L(f) = f''\colon P_5 \to P_3$, we have

$$\text{Ker}L = \{f \in P_5\colon f'' = 0\} = \{a + bt\colon a, b \in \mathbb{R}\}.$$

The monomials $1, t$ form a basis of the kernel, and $\dim \text{Ker}L = 2$. Since $L(P_5) = P_3$ is onto, we have $\text{rank}L = \dim L(P_5) = \dim P_3 = 4$. Then

$$\dim \text{Ker}L + \text{rank}L = 2 + 4 = 6 = \dim P_5.$$

This is the linear transformation version of the equality in Theorem 3.8.

Example 3.10. By Exercise 2.15, the left multiplication by an $m \times n$ matrix A is a linear transformation

$$L_A(X) = AX\colon M_{k\times n} \to M_{k\times m}.$$

Let $X = (\vec{x}_1\ \vec{x}_2\ \cdots\ \vec{x}_k)$. Then $AX = (A\vec{x}_1\ A\vec{x}_2\ \cdots\ A\vec{x}_k)$. Therefore $Y \in \text{Ran}L_A$ if and only if all columns of Y lie in Col$A$, and $X \in \text{Ker}L_A$ if and only if all columns of X lie in Nul$A$.

Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r\}$ be a basis of Col$A$ ($r = \text{rank}A$). Then for the special case $k = 2$, the following is a basis of RanL_A

$$(\vec{v}_1\ \vec{0}),\ (\vec{0}\ \vec{v}_1),\ (\vec{v}_2\ \vec{0}),\ (\vec{0}\ \vec{v}_2),\ \dots,\ (\vec{v}_r\ \vec{0}),\ (\vec{0}\ \vec{v}_r).$$

Therefore $\dim \text{Ran}L_A = 2r = 2\dim \text{Col}A$. Similarly, we have $\dim \text{Ker}L_A = 2r = 2\dim \text{Nul}A$.

In general, we have

$$\dim \text{Ran}L_A = k\dim \text{Col}A = k\text{rank}A, \quad \dim \text{Ker}L_A = k\dim \text{Nul}A = k(n - \text{rank}A).$$

Example 3.11. In Example 3.7, we saw the range of linear transformation $L(A) = A + A^T\colon M_{n\times n} \to M_{n\times n}$ is exactly all symmetric matrices. The kernel of the linear transformation consists those A satisfying $A = A^T$, or $A^T = -A$. These are the skew-symmetric matrices. See Exercises 3.28 and 3.29.

Exercise 3.33. An $m \times n$ matrix A induces four subspaces Col$A$, Row$A$, Nul$A$, NulA^T.

1. Which are subspaces of $\mathbb{R}^m$? Which are subspaces of $\mathbb{R}^n$?
2. Which basis can you calculate by using the row operation on A?
3. Which basis can you calculate by using the column operation on A?

Exercise 3.34. Find the dimensions of the range and the kernel of right multiplication by an $m \times n$ matrix A

$$R_A(X) = XA\colon M_{m\times k} \to M_{n\times k}.$$

Combination of Range and Kernel

Let $L\colon V \to W$ be a linear transformation. For $\vec{b} \in W$, we try to find $\vec{x} \in V$ satisfying $L(\vec{x}) = \vec{b}$. In case V, W are Euclidean spaces and A is the matrix of L, the problem is to solve the system of linear equations $A\vec{x} = \vec{b}$.

For any problem, the first question we should ask is always the existence of solution. Only after we know the existence, then we can further ask the second question of how many solutions.

First, the equation $L(\vec{x}) = \vec{b}$ has solution if and only if $\vec{b} \in L(V)$ lies in the range of L. The "size" of all such $\vec{b}$ is $\dim L(V) = \text{rank}L$.

Second, suppose $L(\vec{x}) = \vec{b}$ has a solution $\vec{x}_0$. This means that we already know $L(\vec{x}_0) = \vec{b}$. Then $\vec{x}$ is a solution if and only if $L(\vec{x}) = \vec{b} = L(\vec{x}_0)$. Since L is linear, this means $L(\vec{x} - \vec{x}_0) = \vec{b} - \vec{b} = \vec{0}$, or $\vec{x} - \vec{x}_0 \in \text{Ker}L$. The difference $\vec{v} = \vec{x} - \vec{x}_0$ can be considered as the change from the initial solution $\vec{x}_0$. The "size" of all such changes is $\dim \text{Ker}L$. Theorem 3.8 suggests $\dim \text{Ker}L = \dim V - \text{rank}L$.

Proposition 3.9. *Suppose $L\colon V \to W$ is a linear transformation, and $\vec{b} \in \text{Ran}L$. Then $L(\vec{x}) = \vec{b}$ has solution, and all the solutions form the subset*

$$\vec{x}_0 + \text{Ker}L = \{\vec{x}_0 + \vec{v}\colon L(\vec{v}) = \vec{0}\}.$$

In particular, L is one-to-one (i.e., the solution is unique) if and only if $\mathrm{Ker}L = \{\vec{0}\}$ *(i.e., $L(\vec{v}) = \vec{0}$ implies $\vec{v} = \vec{0}$).*

The range (and $\vec{x}_0$) manifests the existence, while (the triviality of) the kernel manifests the uniqueness.

Example 3.12. In Example 1.7, we solve the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= 11, \\ 3x_1 + 6x_2 + 9x_3 &= 12. \end{aligned}$$

The general solution

$$x_1 = -2 + x_3, \quad x_2 = 3 - 2x_3, \quad x_3 \text{ arbitrary}$$

can be rephrased as

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 + x_3 \\ 3 - 2x_3 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \vec{x}_0 + \vec{v}, \quad \vec{x}_0 = \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix}, \ \vec{v} \in \mathbb{R} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

Geometrically, the solutions is the straight line (subspace) $\mathrm{Span}\{(1, -2, 1)\} = \mathbb{R}(1, -2, 1)$ in the direction of $(1, -2, 1)$ shifted by $\vec{x}_0 = (-2, 3, 0)$.

Example 3.13. The system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 + 10x_4 &= 1, \\ 2x_1 + 5x_2 + 8x_3 + 11x_4 &= 1, \\ 3x_1 + 6x_2 + 9x_3 + 12x_4 &= 1, \end{aligned}$$

has an obvious solution $\vec{x}_0 = \frac{1}{3}(-1, 1, 0, 0)$. In Example 3.8, we found that $\vec{v}_1 = (1, -2, 1, 0)$ and $\vec{v}_2 = (2, -3, 0, 1)$ form a basis of the kernel. Therefore the general solution is

$$\vec{x} = \vec{x}_0 + c_1\vec{v}_1 + c_2\vec{v}_2 = \frac{1}{3} \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} + c_1 + 2c_2 \\ \frac{1}{3} - 2c_1 - 3c_2 \\ c_1 \\ c_2 \end{pmatrix}, \quad c_1, c_2 \in \mathbb{R}.$$

Geometrically, the solutions is the plane (subspace) $\mathrm{Span}\{(1, -2, 1, 0), (1, -2, 1, 0)\}$ shifted by $\vec{x}_0 = \frac{1}{3}(-1, 1, 0, 0)$.

We may also use another obvious solution $\frac{1}{3}(0, -1, 1, 0)$ and get an alternative formula for the general solution

$$\vec{x} = \frac{1}{3} \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c_1 + 2c_2 \\ -\frac{1}{3} - 2c_1 - 3c_2 \\ \frac{1}{3} + c_1 \\ c_2 \end{pmatrix}, \quad c_1, c_2 \in \mathbb{R}.$$

Example 3.14. The general solution of the linear differential equation $f' = \sin t$ is

$$f = -\cos t + C.$$

Here $f_0 = -\cos t$ is one special solution, and the arbitrary constants C form the kernel of the derivative linear transform

$$\mathrm{Ker}(f \mapsto f') = \{C : C \in \mathbb{R}\} = \mathbb{R}1.$$

Similarly, $f'' = \sin t$ has a special solution $f_0 = -\sin t$. Moreover, we have (see Example 3.9)

$$\mathrm{Ker}(f \mapsto f'') = \{C + Dt : C, D \in \mathbb{R}\}.$$

Therefore the general solution of $f'' = \sin t$ is $f = -\sin t + C + Dt$.

Example 3.15. The left side of a *linear differential equation* of order n (see Example 2.7)

$$L(f) = \frac{d^n f}{dt^n} + a_1(t)\frac{d^{n-1} f}{dt^{n-1}} + a_2(t)\frac{d^{n-2} f}{dt^{n-2}} + \cdots + a_{n-1}(t)\frac{df}{dt} + a_n(t)f = b(t)$$

is a linear transformation $C^\infty \to C^\infty$. A fundamental theorem in the theory of differential equations says that $\dim \mathrm{Ker}L = n$. Therefore to solve the differential equation, we need to find one special function f_0 satisfying $L(f_0) = b(t)$ and n linearly independent functions $f_1, f_2, \dots, f_n$ satisfying $L(f_i) = 0$. Then the general solution is

$$f = f_0 + c_1 f_1 + c_2 f_2 + \cdots + c_n f_n.$$

Take the second order differential equation $f'' + f = e^t$ as an example. We try the special solution $f_0 = ae^t$ and find $(ae^t)'' + ae^t = 2ae^t = e^t$ implying $a = \frac{1}{2}$. Therefore $f_0 = \frac{1}{2}e^t$ is a solution. Moreover, we know that both $f_1 = \cos t$ and $f_2 = \sin t$ satisfy the homogeneous equation $f'' + f = 0$. By Example 1.11, f_1 and f_2 are linearly independent. This leads to the general solution of the differential equation

$$f = \frac{1}{2}e^t + c_1 \cos t + c_2 \sin t, \quad c_1, c_2 \in \mathbb{R}.$$

The following is the linear transformation version of Theorem 3.8.

Theorem 3.10. *If* $L\colon V \to W$ *is a linear transformation, then* $\dim \mathrm{Ker}L + \mathrm{rank}L = \dim V$.

Proof. Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ be a basis of $\mathrm{Ker}L \subset V$. Let $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_r\}$ be a basis of $\mathrm{Ran}L = L(V)$. By the definition of range, we have $\vec{w}_i = L(\vec{v}_{k+i})$ for some $\vec{v}_{k+i} \in V$. Then L maps the last r vectors in $\alpha' = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k, \vec{v}_{k+1}, \vec{v}_{k+2}, \dots, \vec{v}_{k+r}\}$ to β. We will prove that α' is a basis of V. This implies $\dim \mathrm{Ker}L + \mathrm{rank}L = k + r = \dim V$.

For any $\vec{x} \in V$, we have $L(\vec{x}) \in L(V)$. Since β is a basis of $L(V)$, we can write

$$\begin{aligned}
L(\vec{x}) &= y_1\vec{w}_1 + y_2\vec{w}_2 + \cdots + y_r\vec{w}_r \\
&= y_1L(\vec{v}_{k+1}) + y_2L(\vec{v}_{k+2}) + \cdots + y_rL(\vec{v}_{k+r}) \\
&= L(y_1\vec{v}_{k+1} + y_2\vec{v}_{k+2} + \cdots + y_r\vec{v}_{k+r}).
\end{aligned}$$

This implies $L(\vec{x} - (y_1\vec{v}_{k+1} + y_2\vec{v}_{k+2} + \cdots + y_r\vec{v}_{k+r})) = \vec{0}$. Since α is a basis of $\mathrm{Ker}L$, we can write

$$\vec{x} - (y_1\vec{v}_{k+1} + y_2\vec{v}_{k+2} + \cdots + y_r\vec{v}_{k+r}) = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k.$$

This implies the linear combination

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k + y_1\vec{v}_{k+1} + y_2\vec{v}_{k+2} + \cdots + y_r\vec{v}_{k+r},$$

and proves that α' spans V.

To show that α' is linearly independent, we start with

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k + x_{k+1}\vec{v}_{k+1} + x_{k+2}\vec{v}_{k+2} + \cdots + x_{k+r}\vec{v}_{k+r} = \vec{0}.$$

Applying the linear transformation L, we get

$$\begin{aligned}
\vec{0} &= x_1L(\vec{v}_1) + x_2L(\vec{v}_2) + \cdots + x_kL(\vec{v}_k) + x_{k+1}L(\vec{v}_{k+1}) + x_{k+2}L(\vec{v}_{k+2}) + \cdots + x_{k+r}L(\vec{v}_{k+r}) \\
&= x_1\vec{0} + x_2\vec{0} + \cdots + x_k\vec{0} + x_{k+1}\vec{w}_1 + x_{k+2}\vec{w}_2 + \cdots + x_{k+r}\vec{w}_r \\
&= x_{k+1}\vec{w}_1 + x_{k+2}\vec{w}_2 + \cdots + x_{k+r}\vec{w}_r.
\end{aligned}$$

Since β is linearly independent, we get $x_{k+1} = x_{k+2} = \cdots = x_{k+r} = 0$. Substituting into the linear combination we started with, we get

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{0}.$$

Since α is linearly independent, this implies $x_1 = x_2 = \cdots = x_k = 0$. This completes the proof that α' is linearly independent. □

Exercise 3.35. Use Theorem 3.10 to show that $L\colon V \to W$ is one-to-one if and only if $\mathrm{rank}L = \dim V$ (the second statement of Proposition 3.7).

3.3 Sum and Direct Sum

The sum of subspaces generalises the span. The direct sum of subspaces generalises the linear independence. Subspace, sum, and direct sum are the deeper linear algebra concepts that replace vector, span, and linear independence.

Sum of Subspace

The *sum* of subspaces $H_1, H_2, \dots, H_k \subset V$ is

$$H_1 + H_2 + \cdots + H_k = \{\vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_k \colon \vec{h}_i \in H_i\}.$$

For the special case that $H_i = \text{Span}\vec{v}_i = \mathbb{R}\vec{v}_i$ are straight line subspaces, the sum is the span of vectors

$$\mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_k = \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k \colon x_i \in \mathbb{R}\} = \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}.$$

Therefore the sum of subspaces generalises the concept of span.

Exercise 3.36. Prove that the sum is indeed a subspace. Moreover, prove the following properties

$$H_1 + H_2 = H_2 + H_1, \quad (H_1 + H_2) + H_3 = H_1 + H_2 + H_3 = H_1 + (H_2 + H_3).$$

Exercise 3.37. Prove that the intersection $H_1 \cap \cdots \cap H_k$ is a subspace.

Exercise 3.38. Prove that $H_1 + H_2 + \cdots + H_k$ is the smallest subspace containing all H_i.

Exercise 3.39. Prove that $\text{Span}\alpha + \text{Span}\beta = \text{Span}(\alpha \cup \beta)$.

Exercise 3.40. Prove that $\text{Span}(\alpha \cap \beta) \subset (\text{Span}\alpha) \cap (\text{Span}\beta)$. Show that the two sides may or may not equal.

Exercise 3.41. Extend the sum of subspaces to the sum of possibly infinitely many subspaces.

Proposition 3.11. *We have* $\dim(H + H') = \dim H + \dim H' - \dim(H \cap H')$. *Moreover, the following are equivalent.*

1. $H \cap H' = \{\vec{0}\}$.
2. $\dim(H + H') = \dim H + \dim H'$.
3. *Any vector in* $H + H'$ *can be expressed as* $\vec{h} + \vec{h}'$ *for unique* $\vec{h} \in H$ *and* $\vec{h}' \in H'$.

The proposition implies

$$\dim(H_1 + H_2 + \cdots + H_k) \leq \dim H_1 + \dim H_2 + \cdots + \dim H_k.$$

For the special case $H_i = \mathbb{R}\vec{v}_i$, this means (see the discussion before Proposition 3.5)

$$\text{rank}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} = \dim \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} \leq k.$$

Proof. Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ be a basis of $H \cap H'$. Then α is a linearly independent set in H. By Theorem 1.18, α extends to a basis $\alpha \cup \beta$ of H and a basis $\alpha \cup \beta'$ of H', by adding

$$\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_l\}, \quad \beta' = \{\vec{w}'_1, \vec{w}'_2, \dots, \vec{w}'_{l'}\}.$$

We claim that $\alpha \cup \beta \cup \beta'$ is a basis of $H + H'$. This then implies

$$\dim(H + H') = k + l + l' = (k + l) + (k + l') - k = \dim H + \dim H' - \dim(H \cap H').$$

Since vectors in H and H' are linear combinations of vectors in $\alpha \cup \beta \subset \alpha \cup \beta \cup \beta'$ and vectors in $\alpha \cup \beta' \subset \alpha \cup \beta \cup \beta'$. The vectors in $H + H'$ are also linear combinations of vectors $\alpha \cup \beta \cup \beta'$. This shows that $\alpha \cup \beta \cup \beta'$ spans $H + H'$.

To show the linear independence of $\alpha \cup \beta \cup \beta'$, we consider the equality $\vec{v} + \vec{h} + \vec{h}' = \vec{0}$, where

$$\begin{aligned}
\vec{v} &= x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k \in H \cap H', \\
\vec{h} &= y_1\vec{w}_1 + y_2\vec{w}_2 + \cdots + y_l\vec{w}_l \in H, \\
\vec{h}' &= z_1\vec{w}'_1 + z_2\vec{w}'_2 + \cdots + z_{l'}\vec{w}'_{l'} \in H'.
\end{aligned}$$

Since $\vec{v} + \vec{h} = -\vec{h}' \in H \cap H'$, $\vec{v} + \vec{h}$ can be expressed as a linear combination of vectors in α. We view this as a linear combination of vectors in $\alpha \cup \beta$, with 0 coefficient for vectors in β. On the other hand, we have another linear combination expression

$$\vec{v} + \vec{h} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k + y_1\vec{w}_1 + y_2\vec{w}_2 + \cdots + y_l\vec{w}_l.$$

Since $\alpha \cup \beta$ is linearly independent, the two linear combinations have the same coefficients. This implies that $y_1 = y_2 = \cdots = y_l = 0$ (0 coefficient for vectors in β). By the same reason, we get $z_1 = z_2 = \cdots = z_{l'} = 0$. Therefore $\vec{h} = \vec{h}' = \vec{0}$, and we get $\vec{v} = -\vec{h} - \vec{h}' = \vec{0}$. Since α is linearly independent, we get $x_1 = x_2 = \cdots = x_k = 0$. This proves that $\alpha \cup \beta \cup \beta'$ is linearly independent.

For the three equivalent statement, we note that the equivalence of the first two follows from $\dim(H + H') = \dim H + \dim H' - \dim(H \cap H')$.

If $H \cap H' = \{\vec{0}\}$, then for $\vec{h}_1, \vec{h}_2 \in H$ and $\vec{h}'_1, \vec{h}'_2 \in H'$, we have

$$\vec{h}_1 + \vec{h}'_1 = \vec{h}_2 + \vec{h}'_2 \iff H \ni \vec{h}_1 - \vec{h}_2 = \vec{h}'_2 - \vec{h}'_1 \in H' \overset{H\cap H'=\{\vec{0}\}}{\Longrightarrow} \vec{h}_1 - \vec{h}_2 = \vec{0} = \vec{h}'_2 - \vec{h}'_1.$$

This proves that the expression $\vec{h} + \vec{h}'$ is unique. Conversely, suppose the expression $\vec{h} + \vec{h}'$ is unique. Then for $\vec{v} \in H \cap H'$, we have $\vec{v} = \vec{0} + \vec{v} = \vec{v} + \vec{0}$, which are two ways of expressing $\vec{v}$ as sum of vectors in H and H'. Applying the uniqueness to the two ways, we get $\vec{v} = \vec{0}$. This proves $H \cap H' = \{\vec{0}\}$. □

Direct Sum of Subspace

Definition 3.12. A sum $H_1 + H_2 + \cdots + H_k$ of subspaces is a *direct sum* if the expression $\vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_k$ for vectors in the sum is unique. We indicate the direct sum by writing $H_1 \oplus H_2 \oplus \cdots \oplus H_k$.

The definition means

$$\vec{h}_1+\vec{h}_2+\cdots+\vec{h}_k=\vec{h}'_1+\vec{h}'_2+\cdots+\vec{h}'_k,\ \vec{h}_i,\vec{h}'_i\in H_i \implies \vec{h}_i=\vec{h}'_i \text{ for all } i.$$

For the special case of $H_i=\mathbb{R}\vec{v}_i$, $\vec{v}_i\neq\vec{0}$ (i.e., H_i are 1-dimensional lines), we have $\vec{h}_i=a_i\vec{v}_i$, and the definition is exactly Definition 1.5 of linear independence of vectors. Therefore direct sum generalises and concept of linear independence.

Example 3.16. Let P_n^{even} be all the even polynomials in P_n and P_n^{odd} be all the odd polynomials in P_n. Then $P_n=P_n^{\text{even}}\oplus P_n^{\text{odd}}$.

Example 3.17 (Abstract Direct Sum)**.** Let V and W be vector spaces. Construct a vector space $V\oplus W$ to be the set $V\times W=\{(\vec{v},\vec{w})\colon \vec{v}\in V,\vec{w}\in W\}$, together with addition and scalar multiplication

$$(\vec{v}_1,\vec{w}_1)+(\vec{v}_2,\vec{w}_2)=(\vec{v}_1+\vec{v}_2,\vec{w}_1+\vec{w}_2),\quad a(\vec{v},\vec{w})=(a\vec{v},a\vec{w}).$$

It is easy to verify that $V\oplus W$ is a vector space. Moreover, V and W are isomorphic to subspaces

$$V\cong V\oplus\vec{0}=\{(\vec{v},\vec{0}_W)\colon \vec{v}\in V\},\quad W\cong\vec{0}\oplus W=\{(\vec{0}_V,\vec{w})\colon \vec{w}\in W\}.$$

Since any vector in $V\oplus W$ can be uniquely expressed as $(\vec{v},\vec{w})=(\vec{v},\vec{0}_W)+(\vec{0}_V,\vec{w})$, we find that $V\oplus W$ is the direct sum of two subspaces

$$V\oplus W=\{(\vec{v},\vec{0}_W)\colon \vec{v}\in V\}\oplus\{(\vec{0}_V,\vec{w})\colon \vec{w}\in W\}.$$

This is the reason why $V\oplus W$ is called *abstract direct sum*.

The construction allows us to write $\mathbb{R}^m\oplus\mathbb{R}^n=\mathbb{R}^{m+n}$. Note that strictly speaking $\mathbb{R}^m$ and $\mathbb{R}^n$ are note subspaces of $\mathbb{R}^{m+n}$. The equality means

1. $\mathbb{R}^m$ is isomorphic to the subspace of vectors in $\mathbb{R}^{m+n}$ with the last n coordinates vanishing.

2. $\mathbb{R}^n$ is isomorphic to the subspace of vectors in $\mathbb{R}^{m+n}$ with the first m coordinates vanishing.

3. $\mathbb{R}^{m+n}$ is the direct sum of two subspaces.

Exercise 3.42. Show that $M_{n\times n}$ is the direct of subspace of symmetric matrices (see Example 3.7) and skew-symmetric matrices (see Exercise 3.28). In other words, any square matrix is the sum of a unique symmetric matrix and a unique skew-symmetric matrix.

Exercise 3.43. Let H,H' be subspaces of V. We have the sum $H+H'\subset V$ and we also have the abstract direct sum $H\oplus H'$ from Example 3.17. Prove that $L(\vec{h},\vec{h}')=\vec{h}+\vec{h}'\colon H\oplus H'\to H+H'$ is a linear transformation. Moreover, show that $\text{Ker}L$ is isomorphic to $H\cap H'$.

Exercise 3.44. Extend the direct sum of subspaces to the direct sum of possibly infinitely many subspaces.

The following generalises Proposition 1.6.

Proposition 3.13. *A sum $H_1 + H_2 + \cdots + H_k$ is direct if and only if the sum expression for $\vec{0}$ is unique*

$$\vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_k = \vec{0},\ \vec{h}_i \in H_i \implies \vec{h}_1 = \vec{h}_2 = \cdots = \vec{h}_k = \vec{0}.$$

In the three equivalent properties in Proposition 3.11, the third is exactly the definition that $H + H'$ is a direct sum. Therefore

$$H + H' \text{ is a direct sum} \iff H \cap H' = \{\vec{0}\} \iff \dim(H + H') = \dim H + \dim H'.$$

For the special case $H = \mathbb{R}\vec{u}$, $H' = \mathbb{R}\vec{v}$, $\vec{u}, \vec{v} \neq \vec{0}$, this means that $\vec{u}$ and $\vec{v}$ are linearly independent if and only if the two vectors are not scalar multiples of each other.

The dimension criterion for the direct sum has the following generalisation.

Proposition 3.14. *We always have* $\dim(H_1 + H_2 + \cdots + H_k) \leq \dim H_1 + \dim H_2 + \cdots + \dim H_k$, *and the sum is direct if and only if the equality holds.*

Proof. We already see that the proposition for $k = 2$ is Proposition 3.11. Suppose the proposition holds for $k - 1$. Then by Proposition 3.11, we have

$$\begin{aligned} \dim(H_1 + H_2 + \cdots + H_k) &\leq \dim(H_1 + H_2 + \cdots + H_{k-1}) + \dim H_k \\ &\leq \dim H_1 + \dim H_2 + \cdots + \dim H_{k-1} + \dim H_k. \end{aligned}$$

Therefore the equality $\dim(H_1 + H_2 + \cdots + H_k) = \dim H_1 + \dim H_2 + \cdots + \dim H_k$ means

$$\begin{aligned} \dim(H_1 + H_2 + \cdots + H_k) &= \dim(H_1 + H_2 + \cdots + H_{k-1}) + \dim H_k, \\ \dim(H_1 + H_2 + \cdots + H_{k-1}) &= \dim H_1 + \dim H_2 + \cdots + \dim H_{k-1}. \end{aligned}$$

By the case $k = 2$, the first equality means that $(H_1 + H_2 + \cdots + H_{k-1}) + H_k$ is a direct sum. By induction, the second equality means that $H_1 + H_2 + \cdots + H_{k-1}$ is a direct sum. The two direct sums combine to mean that $H_1 + H_2 + \cdots + H_k$ is a direct sum. See the simplest case of Proposition 3.15. □

Next we consider when a sum of a sum such as $(H_1 + H_2) + H_3 + (H_4 + H_5) = H_1 + H_2 + H_3 + H_4 + H_5$. We will show that $H_1 + H_2 + H_3 + H_4 + H_5$ is a direct sum if and only if

1. $H_1 + H_2$, H_3, $H_4 + H_5$ are direct sums.

2. $(H_1 + H_2) + H_3 + (H_4 + H_5)$ is a direct sum.

In particular, we have

$$H_1 + H_2 + H_3 + H_4 + H_5 \text{ is direct} \implies H_1 + H_2 \text{ is direct.}$$

This means that any partial sum of a direct sum is direct. Moreover, for the special case $H_i = \mathbb{R}\vec{v}_i$, the second property means that $\alpha_1 = \{\vec{v}_1, \vec{v}_2\}$, $\alpha_2 = \{\vec{v}_3\}$, $\alpha_3 = \{\vec{v}_4, \vec{v}_5\}$ are linearly independent. Under this assumption, we have

$$\text{Span}\alpha_1 + \text{Span}\alpha_2 + \text{Span}\alpha_3 \text{ is direct} \iff \alpha_1 \cup \alpha_2 \cup \alpha_3 \text{ is linearly independent.}$$

To state the general result, we consider n sums

$$H_i = +_j H_{ij} = +_{j=1}^{k_i} H_{ij} = H_{i1} + H_{i2} + \cdots + H_{ik_i}, \quad i = 1, 2, \dots, n.$$

Then we consider the sum

$$H = +_i(+_j H_{ij}) = +_{i=1}^{n} H_i = H_1 + H_2 + \cdots + H_n,$$

and consider the further splitting of the sum

$$\begin{aligned} H = +_{ij} H_{ij} = H_{11} + H_{12} + \cdots + H_{1k_1} + H_{21} + H_{22} + \cdots + H_{2k_2} \\ + \cdots\cdots + H_{n1} + H_{n2} + \cdots + H_{nk_n}. \end{aligned}$$

Proposition 3.15. *The sum $+_{ij} H_{ij}$ is direct if and only if the sum $+_i(+_j H_{ij})$ is direct and the sum $+_j H_{ij}$ is direct for each i.*

Proof. Suppose $H = +_{ij} H_{ij}$ is a direct sum. To prove that $H = +_i H_i = +_i(+_j H_{ij})$ is a direct sum, we consider a vector $\vec{h} = \sum_i \vec{h}_i = \vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_n$, $\vec{h}_i \in H_i$, in the sum. By $H_i = +_j H_{ij}$, we have $\vec{h}_i = \sum_j \vec{h}_{ij} = \vec{h}_{i1} + \vec{h}_{i2} + \cdots + \vec{h}_{ik_i}$, $\vec{h}_{ij} \in H_{ij}$. Then $\vec{h} = \sum_{ij} \vec{h}_{ij}$. Since $H = +_{ij} H_{ij}$ is a direct sum, we find that $\vec{h}_{ij}$ are uniquely determined by $\vec{h}$. This implies that $\vec{h}_i$ are also uniquely determined by $\vec{h}$. This proves that $H = +_i H_i$ is a direct sum.

Next we further prove that $H_i = +_j H_{ij}$ is also a direct sum. We consider a vector $\vec{h} = \sum_j \vec{h}_{ij} = \vec{h}_{i1} + \vec{h}_{i2} + \cdots + \vec{h}_{ik_i}$, $\vec{h}_{ij} \in H_{ij}$, in the sum. By taking $\vec{h}_{i'j} = \vec{0}$ for all $i' \neq i$, we form the double sum $\vec{h} = \sum_{ij} \vec{h}_{ij}$. Since $H = +_{ij} H_{ij}$ is a direct sum, all $\vec{h}_{pj}$, $p = i$ or $p = i'$, are uniquely determined by $\vec{h}$. In particular, $\vec{h}_{i1}, \vec{h}_{i2}, \dots, \vec{h}_{ik_i}$ are uniquely determined by $\vec{h}$. This proves that $H_i = +_j H_{ij}$ is a direct sum.

Conversely, suppose the sum $+_i(+_j H_{ij})$ is direct and the sum $+_j H_{ij}$ is direct for each i. To prove that $H = +_{ij} H_{ij}$ is a direct sum, we consider a vector $\vec{h} = \sum_{ij} \vec{h}_{ij}$, $\vec{h}_{ij} \in H_{ij}$, in the sum. We have $\vec{h} = \sum_i \vec{h}_i$ for $\vec{h}_i = \sum_j \vec{h}_{ij} \in +_j H_{ij}$. Since $H = +_i(+_j H_{ij})$ is a direct sum, we find that $\vec{h}_i$ are uniquely determined by $\vec{h}$. Since $H_i = +_j H_{ij}$ is a direct sum, we also find that $\vec{h}_{ij}$ are uniquely determined by $\vec{h}_i$. Therefore all $\vec{h}_{ij}$ are uniquely determined by $\vec{h}$. This proves that $+_{ij} H_{ij}$ is a direct sum. $\square$

Exercise 3.45. We may regard a subspace H as a sum of single subspace. Explain that the single sum is always direct.

Exercise 3.46. Prove Proposition 3.13.

Exercise 3.47. Prove that a sum of subspaces is not direct if and only if a nonzero vector in one subspace is sum of vectors from other subspaces. This generalises Proposition 1.7.

Exercise 3.48. If a sum is direct, prove that the sum of a selection of subspaces is also direct.

Exercise 3.49. Suppose α_i is a basis of H_i. Prove that $\alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_n$ is a basis of $H_1 + H_2 + \cdots + H_n$ if and only if the sum $H_1 + H_2 + \cdots + H_n$ is direct.

Projection

A direct sum $V = H \oplus H'$ induces a map (by picking the first term in the unique expression)

$$P(\vec{v}) = \vec{h}, \quad \text{if } \vec{v} = \vec{h} + \vec{h}',\ \vec{h} \in H,\ \vec{h}' \in H'.$$

The direct sum implies that P is a well defined linear transformation satisfying $P^2 = P$. See Exercise 3.50.

Definition 3.16. A linear operator $P\colon V \to V$ is a *projection* if $P^2 = P$.

Conversely, given any projection P, we have $\vec{v} = P(\vec{v}) + (I - P)(\vec{v})$. By $P(I-P)(\vec{v}) = (P - P^2)(\vec{v}) = \vec{0}$, we have $P(\vec{v}) \in \text{Ran}P$ and $(I - P)(\vec{v}) \in \text{Ker}P$. Therefore $V = \text{Ran}P + \text{Ker}P$. On the other hand, if $\vec{v} = \vec{h} + \vec{h}'$ with $\vec{h} \in \text{Ran}P$ and $\vec{h}' \in \text{Ker}P$, then $\vec{h} = P(\vec{w})$ for some $\vec{w} \in V$, and

$$\begin{aligned}
P(\vec{v}) &= P(\vec{h}) + P(\vec{h}') \\
&= P(\vec{h}) && (\vec{h}' \in \text{Ker}P) \\
&= P^2(\vec{w}) && (\vec{h} = P(\vec{w})) \\
&= P(\vec{w}) && (P^2 = P) \\
&= \vec{h}.
\end{aligned}$$

This shows that $\vec{h}$ is unique. Therefore the decomposition $\vec{v} = \vec{h} + \vec{h}'$ is also unique, and we have direct sum

$$V = \text{Ran}P \oplus \text{Ker}P.$$

We conclude that there is a one-to-one correspondence between projections of V and decomposition of V into direct sum of two subspaces.

Exercise 3.50. Given a direct sum $V = H \oplus H'$, verify that $P(\vec{v}) = \vec{h}$ is well defined, is a linear operator, and satisfies $P^2 = P$.

Exercise 3.51. If P is a projection, prove that $\text{Ran}P = \text{Ker}(I - P)$ and $\text{Ran}(I - P) = \text{Ker}P$.

Exercise 3.52. If P is a projection, prove that $Q = I - P$ is also a projection satisfying

$$P + Q = I, \quad PQ = QP = O.$$

Moreover, P and Q induce the same direct sum decomposition $V = H \oplus H'$. The only difference is the order of two subspaces (corresponding to basis versus ordered basis).

Exercise 3.53. Given a direct sum $V = H_1 \oplus H_2 \oplus H_3$, we have three maps

$$P_1(\vec{v}) = \vec{h}_1,\ P_2(\vec{v}) = \vec{h}_2,\ P_3(\vec{v}) = \vec{h}_3, \quad \text{if } \vec{v} = \vec{h}_1 + \vec{h}_2 + \vec{h}_3, \quad \vec{h}_i \in H_i.$$

1. Prove that P_1 is the projection to H_1, with respect to the direct sum $V = H_1 \oplus (H_2 \oplus H_3)$.
2. Prove that $P_i P_j = O$ for $i \neq j$ and $P_1 + P_2 + P_3 = I$.
3. Prove that conversely, three operators satisfying the second part induce a direct sum of three subspaces.

Blocks of Linear Transformation

The matrix of linear transformation can be interpreted in terms of direct sum. Consider a linear transformation together with bases $\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and $\beta = \{\vec{w}_1, \vec{w}_2\}$ (i.e., direct sum decomposition into 1-dimensional subspaces)

$$L\colon V = \mathbb{R}\vec{v}_1 \oplus \mathbb{R}\vec{v}_2 \oplus \mathbb{R}\vec{v}_3 \to W = \mathbb{R}\vec{w}_1 \oplus \mathbb{R}\vec{w}_2.$$

The matrix means

$$[L]_{\beta\alpha} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, \quad \begin{aligned} L(\vec{v}_1) &= a_{11}\vec{w}_1 + a_{21}\vec{w}_2, \\ L(\vec{v}_2) &= a_{12}\vec{w}_1 + a_{22}\vec{w}_2, \\ L(\vec{v}_3) &= a_{13}\vec{w}_1 + a_{23}\vec{w}_2. \end{aligned}$$

Let $P_1\colon W \to W$ be the projection to $\mathbb{R}\vec{w}_1$. Then $P_1 L|_{\mathbb{R}\vec{v}_2}(x\vec{v}_2) = a_{12}x\vec{w}_1$. This means that a_{12} is the 1×1 matrix of the linear transformation $L_{21} = P_1 L|_{\mathbb{R}\vec{v}_2}\colon \mathbb{R}\vec{v}_2 \to \mathbb{R}\vec{w}_1$. Similarly, a_{ji} is the 1×1 matrix of the linear transformation $L_{ji}\colon \mathbb{R}\vec{v}_i \to \mathbb{R}\vec{w}_j$ obtained by restricting L to the direct sum components, and we may write

$$L = \begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \end{pmatrix}.$$

In general, a linear transformation $L\colon V_1 \oplus V_2 \oplus \cdots \oplus V_n \to W_1 \oplus W_2 \oplus \cdots \oplus W_m$ has the *block matrix*

$$L = \begin{pmatrix} L_{11} & L_{12} & \dots & L_{1n} \\ L_{21} & L_{22} & \dots & L_{2n} \\ \vdots & \vdots & & \vdots \\ L_{m1} & L_{m2} & \dots & L_{mn} \end{pmatrix}, \quad L_{ji} = P_j L|_{V_i}\colon V_i \to W_j \subset W.$$

In line with expressing vectors in Euclidean spaces by vertical vectors, we should write

$$\begin{pmatrix} \vec{w}_1 \\ \vec{w}_2 \\ \vdots \\ \vec{w}_m \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} & \dots & L_{1n} \\ L_{21} & L_{22} & \dots & L_{2n} \\ \vdots & \vdots & & \vdots \\ L_{m1} & L_{m2} & \dots & L_{mn} \end{pmatrix} \begin{pmatrix} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \\ \vec{v}_n \end{pmatrix},$$

which means

$$L(\vec{v}_1 + \vec{v}_2 + \cdots + \vec{v}_n) = \vec{w}_1 + \vec{w}_2 + \cdots + \vec{w}_m, \quad \vec{w}_i = L_{i1}(\vec{v}_1) + L_{i2}(\vec{v}_2) + \cdots + L_{in}(\vec{v}_n).$$

Example 3.18. Suppose a projection $P\colon V \to V$ corresponds to a direct sum $V = H \oplus H'$. Then

$$P = \begin{pmatrix} I & O \\ O & O \end{pmatrix}$$

with respect to the direct sum.

Example 3.19. The *direct sum* $L = L_1 \oplus \cdots \oplus L_n$ of the linear transformations $L_i\colon V_i \to W_i$ is the *diagonal block matrix*

$$L = \begin{pmatrix} L_1 & O & \dots & O \\ O & L_2 & \dots & O \\ \vdots & \vdots & & \vdots \\ O & O & \dots & L_n \end{pmatrix} : V_1 \oplus V_2 \oplus \cdots \oplus V_n \to W_1 \oplus W_2 \oplus \cdots \oplus W_n,$$

given by

$$L(\vec{v}_1 + \vec{v}_2 + \cdots + \vec{v}_n) = L_1(\vec{v}_1) + L_2(\vec{v}_2) + \cdots + L_n(\vec{v}_n), \quad \vec{v}_i \in V_i, \quad L_i(\vec{v}_i) \in W_i.$$

For example, the identity on $V_1 \oplus \cdots \oplus V_n$ is the direct sum of identities

$$I = \begin{pmatrix} I_{V_1} & O & \dots & O \\ O & I_{V_2} & \dots & O \\ \vdots & \vdots & & \vdots \\ O & O & \dots & I_{V_n} \end{pmatrix}.$$

Exercise 3.54. What is the block matrix for switching the factors in a direct sum $V \oplus W \to W \oplus V$?

The operations of block matrices are similar to the usual matrices, as long as the direct sums match. For example, for linear transformations $V_1 \oplus V_2 \oplus V_3 \xrightarrow{L,K} W_1 \oplus W_2$, we have

$$\begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \end{pmatrix} + \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \end{pmatrix} = \begin{pmatrix} L_{11}+K_{11} & L_{12}+K_{12} & L_{13}+K_{13} \\ L_{21}+K_{21} & L_{22}+K_{22} & L_{23}+K_{23} \end{pmatrix},$$

$$a\begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \end{pmatrix} = \begin{pmatrix} aL_{11} & aL_{12} & aL_{13} \\ aL_{21} & aL_{22} & aL_{23} \end{pmatrix}.$$

For the composition of linear tranformations $U_1 \oplus U_2 \xrightarrow{K} V_1 \oplus V_2 \oplus V_3 \xrightarrow{L} W_1 \oplus W_2$, we have

$$\begin{pmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \end{pmatrix} \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \\ K_{31} & K_{32} \end{pmatrix} = \begin{pmatrix} L_{11}K_{11} + L_{12}K_{21} + L_{13}K_{31} & L_{11}K_{12} + L_{12}K_{22} + L_{13}K_{32} \\ L_{21}K_{11} + L_{22}K_{21} + L_{23}K_{31} & L_{21}K_{12} + L_{22}K_{22} + L_{23}K_{32} \end{pmatrix}.$$

Note that $L_{ij}K_{jk}$ is the abbreviation of the composition $L_{ij} \circ K_{jk}$ of linear transformations.

Example 3.20. We have

$$\begin{pmatrix} I & L \\ O & I \end{pmatrix} \begin{pmatrix} I & K \\ O & I \end{pmatrix} = \begin{pmatrix} I & L+K \\ O & I \end{pmatrix}.$$

In particular, this implies

$$\begin{pmatrix} I & L \\ O & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -L \\ O & I \end{pmatrix}.$$

Exercise 3.55. Let L and K be invertible. Find the inverses of $\begin{pmatrix} L & M \\ O & K \end{pmatrix}$, $\begin{pmatrix} L & O \\ M & K \end{pmatrix}$, $\begin{pmatrix} O & L \\ K & M \end{pmatrix}$.

Exercise 3.56. Find the n-th power of

$$J = \begin{pmatrix} \lambda I & L & O & \dots & O \\ O & \lambda I & L & \dots & O \\ O & O & \lambda I & \dots & O \\ \vdots & \vdots & \vdots & & \vdots \\ O & O & O & \dots & L \\ O & O & O & \dots & \lambda I \end{pmatrix}.$$

Exercise 3.57. Use block matrix to explain that

$$\mathrm{Hom}(V_1 \oplus V_2 \oplus \cdots \oplus V_n, W) = \mathrm{Hom}(V_1, W) \oplus \mathrm{Hom}(V_2, W) \oplus \cdots \oplus \mathrm{Hom}(V_n, W),$$
$$\mathrm{Hom}(V, W_1 \oplus W_2 \oplus \cdots \oplus W_n) = \mathrm{Hom}(V, W_1) \oplus \mathrm{Hom}(V, W_2) \oplus \cdots \oplus \mathrm{Hom}(V, W_n).$$

3.4 Quotient Space

Let H be a subspace of V. The *quotient space* V/H measures the "difference" between H and V. This is achieved by ignoring the differences in H.

Specifically, we regard two vectors to be *equivalent* if they differ by a vector in H

$$\vec{v} \sim \vec{w} \iff \vec{v} - \vec{w} \in H.$$

The equivalence relation has the following three properties.

1. Reflexivity: $\vec{v} \sim \vec{v}$.

2. Symmetry: $\vec{v} \sim \vec{w} \implies \vec{w} \sim \vec{v}$.

3. Transitivity: $\vec{u} \sim \vec{v}$ and $\vec{v} \sim \vec{w} \implies \vec{u} \sim \vec{w}$.

The reflexivity follows from $\vec{v} - \vec{v} = \vec{0} \in H$. The symmetry follows from $\vec{w} - \vec{v} = -(\vec{v} - \vec{w}) \in H$. The transitivity follows from $\vec{u} - \vec{w} = (\vec{u} - \vec{v}) + (\vec{v} - \vec{w}) \in H$.

The *equivalence class* of a vector $\vec{v}$ is all the vectors equivalent to $\vec{v}$

$$\bar{v} = \{\vec{w} \colon \vec{w} - \bar{v} \in H\} = \{\vec{v} + \bar{h} \colon \bar{h} \in H\} = \vec{v} + H.$$

Definition 3.17. Let H be a subspace of V. The *quotient space* is the collection of all equivalence classes (by difference lying in H)

$$\bar{V} = V/H = \{\vec{v} + H \colon \vec{v} \in V\},$$

together with the addition and scalar multiplication

$$(\vec{v} + H) + (\vec{w} + H) = (\vec{v} + \vec{w}) + H, \quad a(\vec{v} + H) = a\vec{v} + H.$$

Moreover, the following natural map is the *quotient map*

$$\pi(\vec{v}) = \bar{v} = \vec{v} + H \colon V \to \bar{V}.$$

The addition in the quotient space means $\bar{v} + \bar{w} = \overline{v + w}$. Therefore to show the addition is well defined means $\vec{v} \sim \vec{v}\,'$ and $\vec{w} \sim \vec{w}\,'$ implying $\vec{v} + \vec{w} \sim \vec{v}\,' + \vec{w}\,'$. The following is the argument

$$\begin{aligned} \vec{v} \sim \vec{v}\,',\ \vec{w} \sim \vec{w}\,' &\iff \vec{v} - \vec{v}\,', \vec{w} - \vec{w}\,' \in H \\ &\implies (\vec{v} + \vec{w}) - (\vec{v}\,' + \vec{w}\,') = (\vec{v} - \vec{v}\,') + (\vec{w} - \vec{w}\,') \in H \\ &\iff \vec{v} + \vec{w} \sim \vec{v}\,' + \vec{w}\,'. \end{aligned}$$

The argument for the scalar multiplication is similar.

We still need to verify the axioms for vector spaces. The commutativity and associativity of the addition in $\bar{V}$ follows from the commutativity and associativity of the addition in V. The zero vector $\bar{0} = \vec{0} + H = H$. The negative vector $-(\vec{v} + H) = -\vec{v} + H$. The axioms for the scalar multiplications can be similarly verified.

Proposition 3.18. *The quotient map $\pi \colon V \to \bar{V}$ is an onto linear transformation with kernel H.*

The proposition and Theorem 3.10 gives the dimension of the quotient space

$$\dim V/H = \mathrm{rank}\pi = \dim V - \dim \mathrm{Ker}\pi = \dim V - \dim H.$$

Proof. The onto property of π is tautology. The linearity of π follows from the definition of the vector space operations in $\bar{V}$. In fact, we can say that the operations in $\bar{V}$ are defined for the purpose of making π a linear transformation. Moreover, the kernel of π consists of $\vec{v}$ satisfying $\vec{v} + H = \vec{0} + H = H$. This is the same as $\vec{v} \sim \vec{0}$, or $\vec{v} = \vec{v} - \vec{0} \in H$. $\square$

Example 3.21. Let $V = \mathbb{R}^2$ and $H = \mathbb{R}\vec{e}_1 = \mathbb{R} \times 0 = \{(x, 0) \colon x \in \mathbb{R}$. Then $(a, b) + H = \{(a+x, b) \colon x \in \mathbb{R} = \{(x, b) \colon x \in \mathbb{R}$ are all the horizontal lines. See Figure 3.1. These horizontal lines are in one-to-one correspondence with the y-coordinate

$$(a, b) + H \in \mathbb{R}^2/H \longleftrightarrow b \in \mathbb{R}.$$

This identifies the quotient space $\mathbb{R}^2/H$ with $\mathbb{R}$. The identification is a linear transformation because it simply picks the second coordinate. Therefore we have isomorphism $\mathbb{R}^2/H \cong \mathbb{R}$ of vector spaces.

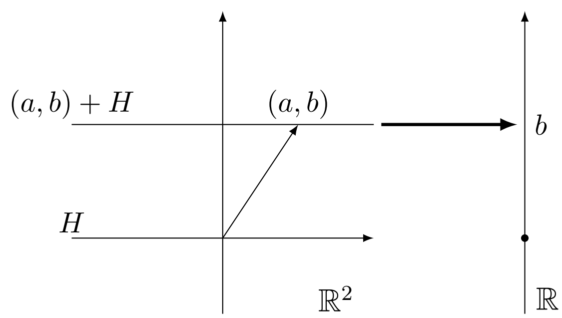

Figure 3.1: Quotient space $\mathbb{R}^2/\mathbb{R} \times 0$.

Exercise 3.58. Define the addition and scalar multiplication of subsets X, Y of vector space V

$$X + Y = \{\vec{v} + \vec{w} \colon \vec{v} \in X, \vec{w} \in Y\}, \quad aX = \{a\vec{v} \colon \vec{v} \in X\}.$$

Verify the following properties similar to some axioms of vector space.

1. $X + Y = Y + X$.
2. $(X + Y) + Z = X + (Y + Z)$.
3. $\{\vec{0}\} + X = X = X + \{\vec{0}\}$.
4. $1X = X$.
5. $(ab)X = a(bX)$.
6. $a(X + Y) = aX + aY$.

Exercise 3.59. Prove that a subset H is a vector space is a subspace if and only if $H + H = H$ and $aH = H$ for $a \neq 0$.

Exercise 3.60. A subset A of a vector space is an *affine subspace* if $aA + (1 - a)A = A$ for any $a \in \mathbb{R}$. Prove the following.

1. Prove that sum of two affine subspaces is an affine subspace.

2. Show that a finite subset is an affine subspace if and only if it is a single vector.

3. Prove that an affine subspace A is a vector subspace if and only if $\vec{0} \in A$.

4. Prove that A is an affine subspace if and only if $A = \vec{v} + H$ for a vector $\vec{v}$ and a subspace H.

Exercise 3.61. An *equivalence relation* on a set X is a collection of ordered pairs $x \sim y$ (regarded as elements in $X \times X$) satisfying the reflexivity, symmetry, and transitivity. The equivalence class of $x \in X$ is

$$\bar{x} = \{y \in X : y \sim x\} \subset X.$$

Prove the following.

1. For any $x, y \in X$, either $\bar{x} = \bar{y}$ or $\bar{x} \cap \bar{y} = \emptyset$.

2. $X = \cup_{x \in X} \bar{x}$.

If we choose one element from each equivalence class, and let I be the set of all such elements, then the two properties imply $X = \sqcup_{x \in I} \bar{x}$ is a decomposition of X into a *disjoint union* of non-empty subsets.

Exercise 3.62. Suppose $X = \sqcup_{i \in I} X_i$ is a *partition* (i.e., disjoint union of non-empty subsets). Define $x \sim y \iff x$ and y are in the same subset X_i. Prove that $x \sim y$ is an equivalence relation, and the equivalence classes are exactly X_i.

Exercise 3.63. Let $f : X \to Y$ be a map. Define $x \sim x' \iff f(x) = f(x')$. Prove that $x \sim x'$ is an equivalence relation, and the equivalence classes are exactly the preimages $f^{-1}(y) = \{x \in X : f(x) = y\}$ for $y \in f(X)$ (otherwise the preimage is empty).

Universal Property

The quotient map $\pi : V \to \bar{V}$ is a linear transformation constructed for the purpose of ignoring (i.e., vanishing on) H. The map is *universal* because it can be used to construct all linear transformations on V that vanish on H.

Theorem 3.19. *Suppose H is a subspace of V. Then a linear transformation $L : V \to W$ satisfies $L(\vec{h}) = \vec{0}$ for all $\vec{h} \in H$ (i.e., $H \subset \mathrm{Ker}L$) if and only if it is the composition of a linear transformation $\bar{L} : \bar{V} \to W$ with the quotient map.*

The linear transformation $\bar{L}$ can be described by the following *commutative diagram*.

$$\begin{tikzcd} V \arrow[rr, "L"] \arrow[dr, "\pi"'] & & W \\ & V/H \arrow[ur, "\bar{L}"'] & \end{tikzcd}$$

Proof. If $L = \bar{L} \circ \pi$, then $\mathrm{Ker}L \supset \mathrm{Ker}\pi = H$. Conversely, suppose $\mathrm{Ker}L \supset H$. Then define $\bar{L}(\bar{v}) = L(\vec{v})$. The following shows that $\bar{L}$ is well defined

$$\bar{v} = \bar{w} \implies \vec{v} - \vec{w} \in H \implies L(\vec{v}) - L(\vec{w}) = L\vec{v} - \vec{w}) = \vec{0} \implies L(\vec{v}) = L(\vec{w}).$$

Then it is easy to verify that $\bar{L}$ is linear, and $L = \bar{L} \circ \pi$. □

Theorem 3.19 implies that any linear transformation $L\colon V \to W$ is the composition of onto π and one-to-one $\bar{L}$

$$V \xrightarrow{\pi} V/\mathrm{Ker}L \xrightarrow{\bar{L}} W.$$

In particular, if L is onto, then $\bar{L}$ is an isomorphism.

Proposition 3.20. *If $L\colon V \to W$ is an onto linear transformation, then $V/\mathrm{Ker}L \cong W$.*

Example 3.22. The picking of the second coordinate $(x, y) \in \mathbb{R}^2 \to y \in \mathbb{R}$ is an onto linear transformation with kernel $H = \mathbb{R}\vec{e}_1 = \mathbb{R} \times 0$. By Proposition 3.20, we get $\mathbb{R}^2/H \cong \mathbb{R}$. This is the isomorphism in Example 3.26.

In general, if $H = \mathbb{R}^k \times \vec{0}$ is the subspace of $\mathbb{R}^n$ such that the last k coordinates vanish, then $\mathbb{R}^n/H \cong \mathbb{R}^{n-k}$ by picking the first $n - k$ coordinates.

Example 3.23. The orthogonal projection P of the $\mathbb{R}^3$ in Example 2.16 is onto the range $H = \{(x, y, z)\colon x + y + z = 0\}$. The geometrical meaning of P shows that $\mathrm{Ker}P = \mathbb{R}(1, 1, 1)$ is the line in direction $(1, 1, 1)$ and passing through the origin. Then by Proposition 3.22, we have $\mathbb{R}^3/\mathbb{R}(1, 1, 1) \cong H$. Note that the vectors in the quotient space $\mathbb{R}^3/\mathbb{R}(1, 1, 1)$ are defined as all the lines in direction $(1, 1, 1)$ (but not necessarily passing through the origin). The isomorphism identifies the collection of such lines with the plane H.

Example 3.24. The linear functional $l(x, y, z) = x + y + z\colon \mathbb{R}^3 \to \mathbb{R}$ is onto and has kernel $H = \{(x, y, z)\colon x + y + z = 0\}$ being the plane in Examples 2.16 and 3.24. Therefore $\bar{l}\colon \mathbb{R}^3/H \to \mathbb{R}$ is an isomorphism. The equivalence classes are planes $(a, 0, 0) + H = \{(a + x, y, z)\colon x + y + z = 0\} = \{(x, y, z)\colon x + y + z = a\} = l^{-1}(a)$ parallel to H.

Example 3.25. The derivative map $D(f) = f'\colon C^\infty \to C^\infty$ is onto, and the kernel is all the constant functions $\mathrm{Ker}D = \{C\colon C \in \mathbb{R}\} = \mathbb{R}$. This induces an isomorphism $C^\infty/\mathbb{R} \cong C^\infty$.

The second order derivative map $D_2(f) = f''\colon C^\infty \to C^\infty$ vanishes on constant functions. By Theorem 3.19, we have $D_2 = \bar{D}_2 \circ D$. Of course we know $\bar{D}_2 = D$ and $D_2 = D^2$.

Exercise 3.64. Prove that the map $\bar{L}$ in Theorem 3.19 is one-to-one if and only if $H = \mathrm{Ker}L$.

Exercise 3.65. Suppose H and H' are subspaces of V. Prove that

$$(\vec{h}, \vec{h}') \in H \oplus H' \mapsto \vec{h} + \vec{h}' \in H + H'$$

is an onto linear transformation, with kernel isomorphic to $H \cap H'$. Then give an alternative proof of the equality in Proposition 3.11

$$\dim(H + H') + \dim(H \cap H') = \dim H + \dim H'.$$

Exercise 3.66. Show that the linear transformation by the matrix is onto. Then explain the implication in terms of quotient space.

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

2. $\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \end{pmatrix}$.

3. $\begin{pmatrix} a_1 & -1 & 0 & \cdots & 0 \\ a_2 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_n & 0 & 0 & \cdots & -1 \end{pmatrix}$.

Exercise 3.67. Let $k \le n$ and $t_1, t_2, \dots, t_k$ be distinct. Let

$$H = (t - t_1)(t - t_2) \cdots (t - t_k)P_n = \{(t - t_1)(t - t_2) \cdots (t - t_k)f(t) \colon f \in P_{n-k}\}.$$

Show that H is a subspace of P_n and the evaluations at $t_1, t_2, \dots, t_k$ gives an isomorphism between P_n/H and $\mathbb{R}^k$.

Exercise 3.68. For fixed t_0, the map

$$f \in C^\infty \mapsto (f(t_0), f'(t_0), \dots, f^{(k)}(t_0)) \in \mathbb{R}^{n+1}$$

can be regarded as the n-th order Taylor expansion at t_0. Prove that the Taylor expansion is an onto linear transformation. Find the kernel of the linear transformation and interpret your result in terms of quotient space.

Exercise 3.69. Suppose $\sim$ is an equivalence relation on a set X. Define the *quotient set* $\bar{X} = X/\sim$ to be the collection of equivalence classes.

1. Prove that the *quotient map* $\pi(x) = \bar{x} \colon X \to \bar{X}$ is onto.

2. Prove that a map $f \colon X \to Y$ satisfies $x \sim x'$ implying $f(x) = f(x')$ if and only if it is the composition of a map $\bar{f} \colon \bar{X} \to Y$ with the quotient map.

Direct Summand

A *direct summand* of H in V is a subspace H' satisfying $V = H \oplus H'$. The following shows that the direct summand fills the gap between H and V, similar to that 3 fills the gap between 2 and 5 by $2 + 3 = 5$. Therefore the direct summand is a more precise measure of the difference between H and V.

Proposition 3.21. *A subspace H' is a direct summand of H in V if and only if the composition $H' \subset V \to V/H$ is an isomorphism.*

Proof. The proposition is the consequence of the following two claims and Proposition 3.11.

1. $V = H + H'$ if and only if the composition $H' \subset V \to V/H$ is onto.

2. The kernel of the composition $H' \subset V \to V/H$ is $H \cap H'$.

For the first claim, we note that the composition is onto means that for any $\vec{v} \in V$, there is $\vec{h}' \in H'$, such that $\vec{v} + H = \vec{h}' + H$. Since $\vec{v} + H = \vec{h}' + H$ means $\vec{v} - \vec{h}' \in H$, we find that onto means any $\vec{v} \in V$ can be expressed as $\vec{h} + \vec{h}'$ for some $\vec{h} \in H$ and $\vec{h}' \in H'$. This is exactly $V = H + H'$.

For the second claim, we note that the kernel of the composition is

$$\{\vec{h}' \in H' \colon \pi(\vec{h}') = \vec{0}\} = \{\vec{h}' \in H' \colon \vec{h}' \in H\} = H \cap H'. \qquad \square$$

Example 3.26. A direct summand of $H = \mathbb{R}\vec{e}_1 = \mathbb{R} \times 0$ in $V = \mathbb{R}^2$ is a 1-dimensional subspace $H' = \mathbb{R}(a, b)$, such that $\vec{e}_1$ and (a, b) form a basis of $\mathbb{R}^2$. The direct summand is also characterised by that it is isomorphic to $\mathbb{R}$ via the map $\mathbb{R}^2 \to \mathbb{R}$ of picking the second coordinate. Either way, we see that H' is a direct summand if and only if $b \neq 0$. We may get the same H' by modifying b to 1 ((a, b) changed to $(\frac{a}{b}, 1)$). Then the direct summands of H are exactly $H' = \mathbb{R}(a, 1)$, in one-to-one correspondence with $a \in \mathbb{R}$.

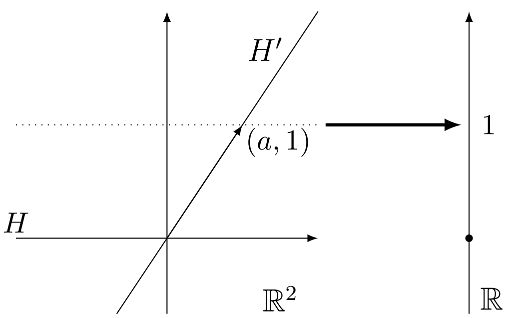

Figure 3.2: Direct summands of $\mathbb{R} \times 0$ in $\mathbb{R}^2$.

Example 3.27. By Example 3.25, the derivative induces an isomorphism $C^\infty/\mathbb{R} \cong C^\infty$, where $\mathbb{R}$ is the subspace of all constant functions. By Proposition 3.21, a direct summand of constant functions $\mathbb{R}$ in C^∞ is then a subspace $H \subset C^\infty$, such that $D|_H \colon f \in H \to f' \in C^\infty$ is an isomorphism. For any fixed t_0, we may choose

$$H(t_0) = \{f \in C^\infty \colon f(t_0) = 0\}.$$

For any $g \in C^\infty$, we have $f(t) = \int_{t_0}^{t} g(\tau)d\tau$ satisfying $f' = g$ and $f(t_0) = 0$. This shows that $D|_H$ is onto. Since $f' = 0$ and $f(t_0) = 0$ implies $f = 0$, we also know that the kernel of $D|_H$ is trivial. Therefore $D|_H$ is an isomorphism, and $H(t_0)$ is a direct summand.

Exercise 3.70. What is the dimension of a direct summand?

Exercise 3.71. Describe all the direct summands of $\mathbb{R}^k \times \vec{0}$ in $\mathbb{R}^n$.

Exercise 3.72. Is it true that any direct summand of $\mathbb{R}$ in C^∞ is $H(t_0)$ for some t_0?

Exercise 3.73. Suppose α is a basis of H and $\alpha \cup \beta$ is a basis of V. Prove that β spans a direct summand of H in V. Moreover, all the direct summands are obtained in this way.

Exercise 3.74. Prove that $P \mapsto H' = \text{Ker}P$ is a one-to-one correspondence between projection operators P on V satisfying $P(V) = H$ and direct summands of H in V.

Exercise 3.75. A *splitting* of a linear transformation $L \colon V \to W$ is a linear transformation $K \colon W \to V$ satisfying $L \circ K = I$. Let $H = \text{Ker}L$.

1. Prove that L has a splitting if and only if L is onto. By Proposition 3.20, L induces an isomorphism $\bar{L} \colon V/H \cong W$.
2. Prove that K is a splitting of L if and only if $K(W)$ is a direct summand of H in V.
3. Prove that splittings of L are in one-to-one correspondence with direct summands of H in V.

Exercise 3.76. Suppose K is a splitting of L. Prove that $K \circ L$ is a projection. Then discuss the relation between two interpretations of direct sum in Exercises 3.74 and 3.75.

Exercise 3.77. Suppose H' and H'' are two direct summands of H in V. Prove that there is a self isomorphism $L \colon V \to V$, such that $L(H) = H$ and $L(H') = H''$. Moreover, prove that it is possible to further require that L satisfies the following, and such L is unique.

1. L fixes H: $L(\vec{h}) = \vec{h}$ for all $\vec{h} \in H$.
2. L is natural: $\vec{h}' + H = L(\vec{h}') + H$ for all $\vec{h}' \in H'$.

Exercise 3.78. Suppose $V = H \oplus H'$. Prove that

$$A \in \text{Hom}(H', H) \mapsto H'' = \{(A(\vec{h}'), \vec{h}') \colon \vec{h}' \in H'\}$$

is a one-to-one correspondence to all direct sums H'' of H in V. This extends Example 3.26.

Suppose H' and H'' are direct summands of H in V. Then by Proposition 3.21, both natural linear transformations $H' \subset V \to V/H$ and $H'' \subset V \to V/H$ are isomorphisms. Combining the two isomorphisms, we find that H' and H'' are naturally isomorphic. Since the direct summand is unique up to natural isomorphism, we denote the direct summand by $V \ominus H$.

Exercise 3.79. Prove that $H + H' = (H \ominus (H \cap H')) \oplus (H \cap H') \oplus (H' \ominus (H \cap H'))$. Then prove that

$$\dim(H + H') + \dim(H \cap H') = \dim H + \dim H'.$$

4 Inner Product

The inner product introduces geometry into a vector space. By geometry, we mean length, angle, area, volume, etc.

4.1 Definition

Definition 4.1. An *inner product* on a real vector space V is a map

$$\langle \vec{u}, \vec{v} \rangle \colon V \times V \to \mathbb{R},$$

satisfying the following properties.

1. Bilinearity: $\langle a\vec{u} + b\vec{u}', \vec{v} \rangle = a\langle \vec{u}, \vec{v} \rangle + b\langle \vec{u}', \vec{v} \rangle$, $\langle \vec{u}, a\vec{v} + b\vec{v} \rangle = a\langle \vec{u}, \vec{v} \rangle + b\langle \vec{u}, \vec{v}' \rangle$.
2. Symmetry: $\langle \vec{v}, \vec{u} \rangle = \langle \vec{u}, \vec{v} \rangle$.
3. Positivity: $\langle \vec{u}, \vec{u} \rangle \geq 0$ and $\langle \vec{u}, \vec{u} \rangle = 0$ if and only if $\vec{u} = \vec{0}$.

An *inner product space* is a vector space equipped with an inner product.

Example 4.1. The *dot product* on the Euclidean space is

$$(x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) = x_1 y_1 + \cdots + x_n y_n.$$

If we use the convention of expressing Euclidean vectors as vertical $n \times 1$ matrices, then we have

$$\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y}.$$

This is especially convenient when we combine the dot product with matrices. For example, for matrices $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_m)$ and $B = (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n)$, where all column vectors are in the same Euclidean space, we have

$$A^T B = \begin{pmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_m^T \end{pmatrix} (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n) = \begin{pmatrix} \vec{v}_1 \cdot \vec{w}_1 & \vec{v}_1 \cdot \vec{w}_2 & \ldots & \vec{v}_1 \cdot \vec{w}_n \\ \vec{v}_2 \cdot \vec{w}_1 & \vec{v}_2 \cdot \vec{w}_2 & \ldots & \vec{v}_2 \cdot \vec{w}_n \\ \vdots & \vdots & & \vdots \\ \vec{v}_m \cdot \vec{w}_1 & \vec{v}_m \cdot \vec{w}_2 & \ldots & \vec{v}_m \cdot \vec{w}_n \end{pmatrix}.$$

In particular, we have

$$A^T \vec{x} = \begin{pmatrix} \vec{v}_1 \cdot \vec{x} \\ \vec{v}_2 \cdot \vec{x} \\ \vdots \\ \vec{v}_m \cdot \vec{x} \end{pmatrix}.$$

Example 4.2. The dot product is not the only inner product on the Euclidean space. For example, if all $a_i > 0$, then the following is also an inner product

$$\langle \vec{x}, \vec{y} \rangle = a_1x_1y_1 + a_2x_2y_2 + \cdots + a_nx_ny_n = \vec{x}^T \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & a_n \end{pmatrix} \vec{y}.$$

In general, if $A = (a_{ij})$ is a symmetric matrix (i.e., $a_{ij} = a_{ji}$), then

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^T A\vec{y} = \sum_{i,j} x_iy_j$$

satisfies the symmetric and bilinear conditions. The positivity condition is more complicated. For example, for $A = \begin{pmatrix} 1 & 2 \\ 2 & a \end{pmatrix}$, we have

$$\vec{x}^T A\vec{x} = x_1^2 + 4x_1x_2 + ax_2^2 = (x_1 + 2x_2)^2 + (a-4)x_2^2.$$

Then it is easy to see that $\vec{x}^T A\vec{y}$ has the positivity (and is therefore an inner product) if and only if $a > 4$.

A symmetric matrix A satisfying $\vec{x}^T A\vec{x} > 0$ for all $\vec{x} \neq \vec{0}$ is called *positive definite*. Propositions 9.6 and 9.7 will give a criterion for a matrix to be positive definite.

Example 4.3. On the vector space P_n of polynomials of degree $\leq n$, we may introduce the inner product

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

This is also an inner product on the vector space $C[0,1]$ of all continuous functions on $[0,1]$, or the vector space of continuous periodic functions on $\mathbb{R}$ of period 1.

More generally, if $K(t) > 0$, then $\langle f, g \rangle = \int_0^1 f(t)g(t)K(t)dt$ is an inner product. In fact, we may even consider $\langle f, g \rangle = \int_0^1 \int_0^1 f(t)K(t,s)g(s)dtds$ for a function K satisfying $K(s,t) = K(t,s)$ and certain (not so easy to verify) positivity condition.

Example 4.4. On the vector space $M_{m\times n}$ of $m \times n$ matrices, we use the trace introduced in Exercise 2.5 to introduce

$$\langle A, B \rangle = \mathrm{tr} A^T B = \sum_{i,j} a_{ij}b_{ij}, \quad A = (a_{ij}), B = (b_{ij}).$$

By Exercises 2.5 and 2.17, the symmetry and bilinear conditions are satisfied. By $\langle A, B \rangle = \sum_{i,j} a_{ij}^2 \geq 0$, the positivity condition is satisfied. Therefore $\mathrm{tr} A^T B$ is an inner product on $M_{m\times n}$.

In fact, if we use the usual isomorphism between $M_{m\times n}$ and $\mathbb{R}^{mn}$, the inner product is translated into the dot product on the Euclidean space.

Exercise 4.1. Suppose $\langle\ ,\ \rangle_1$ and $\langle\ ,\ \rangle_2$ are two inner products on V. Prove that for any $a, b > 0$, $a\langle\ ,\ \rangle_1 + b\langle\ ,\ \rangle_2$ is also an inner product.

Exercise 4.2. Prove that $\vec{u}$ satisfies $\langle \vec{u}, \vec{v} \rangle = 0$ for all $\vec{v}$ (i.e., $\vec{u}$ is orthogonal to all vectors) if and only if $\vec{u} = \vec{0}$.

Exercise 4.3. Prove that $\vec{v}_1 = \vec{v}_2$ if and only if $\langle \vec{u}, \vec{v}_1 \rangle = \langle \vec{u}, \vec{v}_2 \rangle$ for all $\vec{u}$. In other words, two vectors are equal if and only if their inner products with all other vectors are equal.

Exercise 4.4. Prove that two linear transformations $L, K \colon V \to W$ are equal if and only if $\langle \vec{u}, L(\vec{v}) \rangle = \langle \vec{u}, K(\vec{v}) \rangle$ for all $\vec{u}$ and $\vec{v}$.

Exercise 4.5. Prove that two matrices A and B are equal if and only if $\vec{x} \cdot A\vec{y} = \vec{x} \cdot B\vec{y}$ (i.e., $\vec{x}^T A\vec{y} = \vec{x}^T B\vec{y}$) for all $\vec{x}$ and $\vec{y}$.

Exercise 4.6. Show that $\langle f, g \rangle = f(0)g(0) + f(1)g(1) + \cdots + f(n)g(n)$ is an inner product on P_n.

Exercise 4.7. By Example 4.1, the symmetric property of the dot product means $\vec{x}^T\vec{y} = \vec{y}^T\vec{x}$. Use the formula for $A^T B$ in the example to verify that $(AB)^T = B^T A^T$.

Exercise 4.8. For vectors in Euclidean spaces, show that the symmetry and bilinearity imply $\langle \vec{x}, \vec{y} \rangle = \vec{x}^T A\vec{y}$ for a symmetric matrix A. Therefore inner products on Euclidean spaces are in one-to-one correspondence with positive definite matrices.

Exercise 4.9. Prove that $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is positive definite if and only if $a > 0$ and $ac > b^2$.

Geometry

The usual Euclidean length is given by the Pythagorian theorem

$$\|\vec{x}\| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\vec{x} \cdot \vec{x}}.$$

In general, the *length* (or *norm*) with respect to an inner product is

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

The positivity allows us to take the square root.

Inspired by the geometry in $\mathbb{R}^2$, we define the *angle* θ between two nonzero vectors $\vec{u}, \vec{v}$ by

$$\cos\theta = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\|\,\|\vec{v}\|}.$$

Then we may compute the area of the parallelogram spanned by the two vectors

$$\text{Area}(\vec{u}, \vec{v}) = \|\vec{u}\|\,\|\vec{v}\| \sin\theta = \|\vec{u}\|\,\|\vec{v}\| \sqrt{1 - \left(\frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\|\,\|\vec{v}\|}\right)^2} = \sqrt{\|\vec{u}\|^2 \|\vec{v}\|^2 - (\langle \vec{u}, \vec{v} \rangle)^2}.$$

For the definition of angle to make sense, however, we need the following result.

Proposition 4.2 (Cauchy-Schwarz Inequality). $|\langle\vec{u},\vec{v}\rangle| \le \|\vec{u}\|\,\|\vec{v}\|$.

Proof. For any real number t, we have

$$0 \le \langle\vec{u}+t\vec{v},\vec{u}+t\vec{v}\rangle = \langle\vec{u},\vec{u}\rangle + 2t\langle\vec{u},\vec{v}\rangle + t^2\langle\vec{v},\vec{v}\rangle.$$

For the quadratic function of t to be always non-negative, the coefficients must satisfy

$$(\langle\vec{u},\vec{v}\rangle)^2 \le \langle\vec{u},\vec{u}\rangle\langle\vec{v},\vec{v}\rangle.$$

This is the same as $|\langle\vec{u},\vec{v}\rangle| \le \|\vec{u}\|\,\|\vec{v}\|$. □

Proposition 4.3. *The length of vectors has the following properties.*

1. *Positivity:* $\|\vec{u}\| \ge 0$, *and* $\|\vec{u}\| = 0$ *if and only if* $\vec{u} = \vec{0}$.
2. *Scaling:* $\|a\vec{u}\| = |a|\,\|\vec{u}\|$.
3. *Triangle inequality:* $\|\vec{u}+\vec{v}\| \le \|\vec{u}\| + \|\vec{v}\|$.

The first two properties are easy to verify, and the triangle inequality is a consequence of the Cauchy-Schwarz inequality

$$\begin{aligned}\|\vec{u}+\vec{v}\|^2 &= \langle\vec{u}+\vec{v},\vec{u}+\vec{v}\rangle \\ &= \langle\vec{u},\vec{u}\rangle + \langle\vec{u},\vec{v}\rangle + \langle\vec{v},\vec{u}\rangle + \langle\vec{v},\vec{v}\rangle \\ &\le \|\vec{u}\|^2 + \|\vec{u}\|\,\|\vec{v}\| + \|\vec{v}\|\,\|\vec{u}\| + \|\vec{v}\|^2 \\ &= (\|\vec{u}\| + \|\vec{v}\|)^2.\end{aligned}$$

By the scaling property in Proposition 4.3, if $\vec{v} \ne \vec{0}$, then by dividing the length, we get a unit length vector (i.e., length 1)

$$\vec{u} = \frac{\vec{v}}{\|\vec{v}\|}.$$

Note that $\vec{u}$ indicates the *direction* of the vector $\vec{v}$ by "forgetting" its length. In fact, all the directions in the inner product space form the *unit sphere*

$$S_1 = \{\vec{u} \in V : \|\vec{u}\| = 1\} = \{\vec{u} \in V : \vec{u}\cdot\vec{u} = 1\}.$$

Any nonzero vector has unique *polar decomposition*

$$\vec{v} = r\vec{u}, \quad r = \|\vec{v}\| > 0,\ \|\vec{u}\| = 1.$$

Example 4.5. With respect to the dot product, the lengths of $(1,1,1)$ and $(1,2,3)$ are

$$\|(1,1,1)\| = \sqrt{1^2+1^2+1^2} = \sqrt{3}, \quad \|(1,2,3)\| = \sqrt{1^2+2^2+3^2} = \sqrt{14}.$$

Their polar decompositions are

$$(1,1,1) = \sqrt{3}(\tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}), \quad (1,2,3) = \sqrt{14}(\tfrac{1}{\sqrt{14}}, \tfrac{2}{\sqrt{14}}, \tfrac{3}{\sqrt{14}}).$$

The angle between the two vectors is given by

$$\cos\theta = \frac{1\cdot 1 + 1\cdot 2 + 1\cdot 3}{\|(1,1,1)\|\,\|(1,2,3)\|} = \frac{6}{\sqrt{42}}.$$

Therefore the angle is $\arccos\frac{6}{\sqrt{42}} = 0.1234\pi = 22.2077^\circ$.

Example 4.6. The area of the triangle with vertices $\vec{a} = (1,-1,0), \vec{b} = (2,0,1), \vec{c} = (2,1,3)$ is half of the parallelogram spanned by $\vec{u} = \vec{b} - \vec{a} = (1,1,1)$ and $\vec{v} = \vec{c} - \vec{a} = (1,2,3)$

$$\frac{1}{2}\sqrt{\|(1,1,1)\|^2\|(1,2,3)\|^2 - ((1,1,1)\cdot(1,2,3))^2} = \frac{1}{2}\sqrt{3\cdot 14 - 6^2} = \sqrt{\frac{3}{2}}.$$

Example 4.7. By the inner product in Example 4.3, the lengths of 1 and t are

$$\|1\| = \sqrt{\int_0^1 dt} = 1, \quad \|t\| = \sqrt{\int_0^1 t^2 dt} = \frac{1}{\sqrt{3}}.$$

Therefore 1 has unit length, and t has polar decomposition $t = \dfrac{1}{\sqrt{3}}(\sqrt{3}t)$. The angle between 1 and t is given by

$$\cos\theta = \frac{\int_0^1 t dt}{\|1\|\,\|t\|} = \frac{\sqrt{3}}{2}.$$

Therefore the angle is $\frac{1}{6}\pi$. Moreover, the area of the parallelogram spanned by 1 and t is

$$\sqrt{\int_0^1 dt \int_0^1 t^2 dt - \left(\int_0^1 t dt\right)^2} = \frac{1}{2\sqrt{3}}.$$

Exercise 4.10. Show that the area of the triangle with vertices at $(0,0), (a,b), (c,d)$ is $\frac{1}{2}\sqrt{|ad-bc|}$.

Exercise 4.11. In Example 4.6, we calculated the area of the triangle by subtracting $\vec{a}$. By the obvious symmetry, we can also calculate the ares by by subtracting $\vec{b}$ or $\vec{c}$. Please verify that the alternative calculations give the same results. Can you provie an argument for the general case.

Exercise 4.12. Prove that the *distance* $d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|$ in an inner product space has the following properties.

1. Positivity: $d(\vec{u}, \vec{v}) \geq 0$, and $d(\vec{u}, \vec{v}) = 0$ if and only if $\vec{u} = \vec{v}$.
2. Symmetry: $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$.
3. Triangle inequality: $d(\vec{u}, \vec{w}) \leq d(\vec{u}, \vec{v}) + d(\vec{v}, \vec{w})$.

Exercise 4.13. Show that the area of the parallelogram spanned by two vectors is zero if and only if the two vectors are parallel.

Exercise 4.14. Prove the *polarisation* identities

$$\langle \vec{u}, \vec{v} \rangle = \frac{1}{4}(\|\vec{u}+\vec{v}\|^2 - \|\vec{u}-\vec{v}\|^2) = \frac{1}{2}(\|\vec{u}+\vec{v}\|^2 - \|\vec{u}\|^2 - \|\vec{v}\|^2).$$

Exercise 4.15. Prove the parallelogram identity

$$\|\vec{u}+\vec{v}\|^2 + \|\vec{u}-\vec{v}\|^2 = 2(\|\vec{u}\|^2 + \|\vec{v}\|^2).$$

Exercise 4.16. Find the length of vectors and the angle between vectors.

1. $(1,0)$, $(1,1)$.
2. $(1,0,1)$, $(1,1,0)$.
3. $(1,2,3)$, $(2,3,4)$.
4. $(1,0,1,0)$, $(0,1,0,1)$.
5. $(0,1,2,3)$, $(4,5,6,7)$.
6. $(1,1,1,1)$, $(1,-1,1,-1)$.

Exercise 4.17. Find the area of the triangle with the given vertices.

1. $(1,0)$, $(0,1)$, $(1,1)$.
2. $(1,0,0)$, $(0,1,0)$, $(0,0,1)$.
3. $(1,2,3)$, $(2,3,4)$, $(3,4,5)$.
4. $(1,1,0)$, $(1,0,1)$, $(0,1,1)$.
5. $(1,0,1,0)$, $(0,1,0,1)$, $(1,0,0,1)$.
6. $(0,1,2,3)$, $(4,5,6,7)$, $(8,9,10,11)$.

Exercise 4.18. Find the length of vectors and the angle between vectors.

1. $(1,0,1,0,\dots)$, $(0,1,0,1,\dots)$.
2. $(1,2,3,\dots,n)$, $(n,n-1,n-2,\dots,1)$.
3. $(1,1,1,1,\dots)$, $(1,-1,1,-1,\dots)$.
4. $(1,1,1,1,\dots)$, $(1,2,3,4,\dots)$.

Exercise 4.19. Redo Exercises 4.16, 4.17, 4.18 with respect to the inner product

$$\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + \cdots + nx_ny_n.$$

Exercise 4.20. Find the area of the triangle with the given vertices, with respect to the inner product in Example 4.3.

1. 1, t, t^2.
2. 0, $\sin t$, $\cos t$.
3. 1, a^t, b^t.
4. $1-t$, $t-t^2$, t^2-1.

Exercise 4.21. Redo Exercise 4.20 with respect to the inner product

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

Exercise 4.22. Redo Exercise 4.20 with respect to the inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)dt.$$

Adjoint

The dot product gives the following interpretation of the transpose of an $m \times n$ matrix A. For $\vec{x} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$, we have

$$A\vec{x} \cdot \vec{y} = \sum_i (a_{i1}x_1 + \cdots + a_{in}x_n)y_i = \sum_{ij} a_{ij}x_iy_j = \sum_j x_j(a_{1j}y_1 + \cdots + a_{mj}y_m) = \vec{x} \cdot A^T\vec{y}.$$

Alternatively, we have $(A\vec{x})^T = \vec{x}^TA^T$ by Exercise 4.7, and

$$A\vec{x} \cdot \vec{y} = (A\vec{x})^T\vec{y} = \vec{x}^TA^T\vec{y} = \vec{x} \cdot A^T\vec{y}$$

by Example 4.1. We note that the dot product on the left happens in $\mathbb{R}^m$, and the dot product on the right happens in $\mathbb{R}^n$. We also remark that, by Exercise 4.5 for the given A, the transpose A^T is the only matrix B satisfying $A\vec{x} \cdot \vec{y} = \vec{x} \cdot B\vec{y}$ for all $\vec{x}, \vec{y}$. Therefore the equality $A\vec{x} \cdot \vec{y} = \vec{x} \cdot A^T\vec{y}$ can be used as an alternative definition of the transpose matrix.

The interpretation of transpose above suggests the following generalisation.

Definition 4.4. Suppose $L\colon V \to W$ is a linear transformation between two inner product spaces. Then the *adjoint* of L is the linear transformation $L^*\colon W \to V$ satisfying

$$\langle L(\vec{v}), \vec{w}\rangle = \langle \vec{v}, L^*(\vec{w})\rangle \quad \text{for all } \vec{v} \in V,\ \vec{w} \in W.$$

For Euclidean spaces with dot products, the definition means that the transpose of a matrix A is the matrix of the adjoint of the linear transformation $L(\vec{x}) = A\vec{x}\colon \mathbb{R}^n \to \mathbb{R}^m$. If we denote by ϵ_n is the standard basis of $\mathbb{R}^n$, then this also means

$$[L^*]_{\epsilon_n\epsilon_m} = [L]_{\epsilon_m\epsilon_n}.$$

Strictly speaking, we need to justify the existence and uniqueness of the adjoint L^*. We have done the justification for the Euclidean spaces with dot products. If we use the forthcoming Theorem 4.12 that identifies any finite dimensional inner product space with Euclidean space with dot product, then the justification extends to general finite dimensional inner product spaces. By the way, the adjoint may not exist if the vector spaces are infinite dimensional.

Alternatively, we may give a more conceptual justification for the existence of the adjoint. This is based on the following fact.

Proposition 4.5. *In a finite dimensional inner product space V, the map*

$$\vec{v} \in V \mapsto \vec{v}^* = \langle ?, \vec{v}\rangle \in V^*$$

is an isomorphism of vector spaces.

Proof. The symbol $\langle ?, \vec{v}\rangle$ means the function $\vec{v}^*(\vec{x}) = \langle \vec{x}, \vec{v}\rangle$ on V. The linear property of the inner product with respect to the second vector means that the function $\vec{v}^*$ is linear and therefore belongs to the dual vector space V^* in Example 2.9. Then the linear property of the inner product with respect to the first vector means exactly that $\vec{v} \mapsto \vec{v}^*$ is a linear transformation. To show that this is an isomorphism, we note that $\dim V^* = \dim V$ by Example 2.30. Therefore it is sufficient to argue that the linear transformation in the proposition is one-to-one, or it has trivial kernel. A vector $\vec{v}$ is in this kernel means that $\langle \vec{x}, \vec{v}\rangle = 0$ for all $\vec{x}$. By taking $\vec{x} = \vec{v}$ and applying the positivity property of the inner product, we get $\vec{v} = \vec{0}$. $\square$

We may now regard $\vec{v}$ to be the variable and interpret the definition of adjoint as the following composition (i.e., $\vec{w}^* \circ L = (L^*(\vec{w}))^*$)

$$\vec{w} \in W \underset{\cong}{\xrightarrow{*}} \vec{w}^* \in W^* \xrightarrow{\circ L} \vec{w}^* \circ L \in V^* \underset{\cong}{\xleftarrow{*}} L^*(\vec{w}) \in V.$$

The last arrow can be reversed because it is an isomorphism. Since the first arrow is also an isomorphism, we find that L and L^* determine each other (this justifies the uniqueness).

Proposition 4.6. *The adjoint has the following properties.*

$$(L+K)^* = L^* + K^*, \quad (aL)^* = aL^*, \quad (L \circ K)^* = K^* \circ L^*, \quad (L^*)^* = L.$$

The properties give conceptual explanation of the following properties of transpose (see Exercises 1.7 and 2.22)

$$(A+B)^T = A^T + B^T, \quad (aA)^T = aA^T, \quad (AB)^T = B^T A^T, \quad (A^T)^T = A.$$

Proof. We have

$$\begin{aligned}\langle \vec{v}, (L \circ K)^*(\vec{w})\rangle &= \langle (L \circ K)(\vec{v}), \vec{w}\rangle = \langle L(K(\vec{v})), \vec{w}\rangle = \langle K(\vec{v}), L^*(\vec{w})\rangle \\ &= \langle \vec{v}, K^*(L^*(\vec{w}))\rangle = \langle \vec{v}, (K^* \circ L^*)(\vec{w})\rangle.\end{aligned}$$

By Exercise 4.4, we get $(L \circ K)^* = K^* \circ L^*$. The proof of the other equalities are similar. $\square$

Example 4.8. Consider the vector space of polynomials with inner product in Example 4.3. The adjoint $D^*\colon P_{n-1} \to P_n$ of the derivative linear transformation $D(f) = f'\colon P_n \to P_{n-1}$ is characterised by

$$\int_0^1 t^p D^*(t^q)dt = \int_0^1 pt^{p-1}t^q dt = \frac{p}{p+q}, \quad 0 \le p \le n,\ 0 \le q \le n-1.$$

For fixed q, let $D^*(t^q) = x_0 + x_1 t + \cdots + x_n t^n$. Then we get a system of linear equations

$$\frac{1}{p+1}x_0 + \frac{1}{p+2}x_1 + \cdots + \frac{1}{p+n+1}x_n = \frac{p}{p+q}, \quad 0 \le p \le n.$$

The solution is quite non-trivial. For $n = 2$, we have

$$D^*(t^q) = \frac{2}{(1+q)(2+q)}(11 - 2q + 6(17+q)t - 90t^2), \quad q = 0, 1.$$

Example 4.9. Let V be the vector space of all smooth periodic functions on $\mathbb{R}$ of period 2π, with inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)g(t)dt.$$

The derivative operator $D(f) = f' \colon V \to V$ takes periodic function to periodic function. By the integration by parts and period 2π, we have

$$\begin{aligned}\langle D(f), g\rangle &= \frac{1}{2\pi}\int_0^{2\pi} f'(t)g(t)dt = f(2\pi)g(2\pi) - f(0)g(0) - \frac{1}{2\pi}\int_0^{2\pi} f(t)g'(t)dt \\ &= -\frac{1}{2\pi}\int_0^{2\pi} f(t)g'(t)dt = -\langle f, D(g)\rangle.\end{aligned}$$

This implies $D^* = -D$.

The same argument can be applied to the vector space of all smooth functions f on $\mathbb{R}$ satisfying $\lim_{t\to\infty} f^{(n)}(t) = 0$ for all $n \ge 0$, with inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(t)g(t)dt.$$

We still get $D^* = -D$.

Exercise 4.23. Prove the other properties in Proposition 4.6.

Exercise 4.24. Prove that a linear transformation $L\colon V \to W$ is an isometry if and only if $L^*L = I$. Moreover, if L is an isometric isomorphism, then L^* is an isometric isomorphism.

Exercise 4.25. Prove that $\text{rank}L = \text{rank}L^*$.

Exercise 4.26. Prove that $\text{Ker}L = \text{Ker}L^*L$ and $\text{rank}L = \text{rank}L^*L$.

Exercise 4.27. Prove that the following are equivalent.

1. L is one-to-one.
2. L^* is onto.
3. L^*L is invertible.

Exercise 4.28. Prove that a linear operator $L\colon V \to V$ satisfies

$$\langle L(\vec{u}), \vec{v}\rangle + \langle L(\vec{v}), \vec{u}\rangle = \langle L(\vec{u}+\vec{v}), \vec{u}+\vec{v}\rangle - \langle L(\vec{u}), \vec{u}\rangle - \langle L(\vec{v}), \vec{v}\rangle.$$

Then prove that $\langle L(\vec{v}), \vec{v}\rangle = 0$ for all $\vec{v}$ if and only if $L + L^* = 0$.

Exercise 4.29. Calculate the adjoint of the derivative linear transformation $D(f) = f' \colon P_n \to P_{n-1}$ with respect to the inner products in Exercises 4.6, 4.21, 4.22.

4.2 Orthogonal Vectors

By using coordinates with respect to a basis, any finite dimensional vector space is isomorphic to a Euclidean space. For an inner product space, if the basis has the special orthonormal property, then the isomorphism identifies the inner product space with Euclidean space with dot product. This makes it possible to extend discussions about the dot product to general inner product.

Orthogonal Set

Two vectors are orthogonal if the angle between them is $\frac{\pi}{2}$. Since $\cos\frac{\pi}{2} = 0$, we have the following definition.

Definition 4.7. Two vectors $\vec{u}$ and $\vec{v}$ in an inner product space are *orthogonal*, and denoted $\vec{u} \perp \vec{v}$, if

$$\langle \vec{u}, \vec{v} \rangle = 0.$$

A set of vectors is an *orthogonal set* if the vectors are pairwise orthogonal. The set is *orthonormal* if additionally all vectors in the set have length 1.

If all vectors in an orthogonal set $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ are nonzero, then we get an orthonormal set $\{\frac{\vec{v}_1}{\|\vec{v}_1\|}, \frac{\vec{v}_2}{\|\vec{v}_2\|}, \dots, \frac{\vec{v}_n}{\|\vec{v}_n\|}\}$ by dividing the length.

Orthogonality makes it easy to calculate the coefficients in a linear combination.

Proposition 4.8. *If $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are nonzero orthogonal vectors, and $\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n$, then*

$$x_i = \frac{\langle \vec{x}, \vec{v}_i \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle}.$$

In particular, if the vectors are orthonormal, then $x_i = \langle \vec{x}, \vec{v}_i \rangle$.

Proof. By taking the inner product of $\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n$ with $\vec{v}_i$, we get

$$\langle \vec{x}, \vec{v}_i \rangle = x_1\langle \vec{v}_1, \vec{v}_i \rangle + x_2\langle \vec{v}_2, \vec{v}_i \rangle + \cdots + x_n\langle \vec{v}_n, \vec{v}_i \rangle = x_i\langle \vec{v}_i, \vec{v}_i \rangle.$$

Here the first equality is due to the bilinearity, and the second equality is due to orthogonality. Since $\vec{v}_i \neq \vec{0}$, we have $\langle \vec{v}_i, \vec{v}_i \rangle > 0$ by the positivity. Therefore we can divide $\langle \vec{v}_i, \vec{v}_i \rangle$ to get the formula for x_i. □

The proposition has the following consequence. In fact, since $\frac{\pi}{2}$ is the biggest possible angle between two vectors, the geometric meaning is intuitively clear.

Proposition 4.9. *Nonzero orthogonal vectors are linearly independent.*

Example 4.10. All vectors (x, y, z) orthogonal to $(1, 1, 1)$ is the solutions of $x + y + z = 0$. This is the plane in Example 2.16. In general, all vectors orthogonal to a nonzero vector $\vec{a} = (a_1, a_2, \dots, a_n)$ is the solutions of the homogeneous equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

Geometrically, this is a *hyperplane* of $\mathbb{R}^n$.

If there are two (non-trivial) homogeneous equations, then we have two hyperplanes. The solutions of the system of two homogeneous equations is the intersection of two hyperplanes. For example, the solutions of the system

$$x + y + z = 0, \quad x + 2y + 3z = 0$$

is the intersection of two planes, respectively orthogonal to $(1, 1, 1)$ and $(1, 2, 3)$. By solving the system, we find that the intersection of the two planes in $\mathbb{R}^3$ is actually the line $\mathbb{R}(1, 1, -2)$. The line is also all the vectors orthogonal to both $(1, 1, 1)$ and $(1, 2, 3)$.

In general, the solutions of a homogeneous system of linear equations $A\vec{x} = \vec{0}$ is the intersection of hyperplanes orthogonal to the nonzero *row vectors* of A.

Example 4.11. By the product of two matrices in Example 4.1, the column vectors of a matrix A are orthogonal with respect to the dot product if and only if A^TA is a diagonal matrix, and the column vectors are orthonormal if and only if $A^TA = I$.

Example 4.12. By the inner product in Example 4.3, we have

$$\langle t, t-a \rangle = \int_0^1 t(t-a)dt = \frac{1}{3} - \frac{1}{2}a.$$

Therefore t is orthogonal to $t - a$ if and only if $a = \frac{2}{3}$.

Exercise 4.30. Find all vectors orthogonal to the given vectors.

1. $(1, 4, 7)$, $(2, 5, 8)$, $(3, 6, 9)$.
2. $(1, 2, 3)$, $(4, 5, 6)$, $(7, 8, 9)$.
3. $(1, 2, 3, 4)$, $(2, 3, 4, 1)$, $(3, 4, 1, 2)$.
4. $(1, 0, 1, 0)$, $(0, 1, 0, 1)$, $(1, 0, 0, 1)$.

Exercise 4.31. Redo Exercise 4.30 with respect to the inner product in Exercise 4.19.

Exercise 4.32. Find all polynomials of degree 2 orthogonal to the given functions, with respect to the inner product in Example 4.3.

1. 1, t.
2. 1, t, $1 + t$.
3. $\sin t$, $\cos t$.
4. 1, t, t^2.

Exercise 4.33. Redo Exercise 4.32 with respect to the inner product in Exercises 4.21, 4.22.

Exercise 4.34. What are vectors that are orthogonal to itself?

Exercise 4.35. For an orthogonal set $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$, prove the *Pythagorian identity*

$$\|\vec{v}_1 + \vec{v}_2 + \cdots + \vec{v}_n\|^2 = \|\vec{v}_1\|^2 + \|\vec{v}_2\|^2 + \cdots + \|\vec{v}_n\|^2.$$

Exercise 4.36. Suppose $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is an orthonormal basis of V. Prove the *Parsival's identity*

$$\langle \vec{x}, \vec{y} \rangle = \langle \vec{x}, \vec{v}_1 \rangle \langle \vec{v}_1, \vec{y} \rangle + \langle \vec{x}, \vec{v}_2 \rangle \langle \vec{v}_2, \vec{y} \rangle + \cdots + \langle \vec{x}, \vec{v}_n \rangle \langle \vec{v}_n, \vec{y} \rangle.$$

Exercise 4.37. Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_m\}$ are orthonormal bases of inner product spaces V and W. By Proposition 4.8, a linear transformation $L\colon V \to W$ is given by the formula

$$L(\vec{v}_i) = \langle L(\vec{v}_i), \vec{w}_1 \rangle \vec{w}_1 + \langle L(\vec{v}_i), \vec{w}_2 \rangle \vec{w}_2 + \cdots + \langle L(\vec{v}_i), \vec{w}_m \rangle \vec{w}_m.$$

Write down the similar formula for $L^*\colon W \to V$ and prove that $[L^*]_{\alpha\beta} = [L]_{\beta\alpha}^T$.

Isometry

Linear transformations between vector spaces are maps preserving the two key structures of addition and scalar multiplication. Similarly, we should consider maps of inner product spaces that, in addition to being linear, also preserve the third key structure of the inner product.

Definition 4.10. A *isometry* between inner product spaces is a linear transformation preserving the inner product. If the isometry is also invertible, then it is an *isometric isomorphism.*

By the definition, a linear transformation $L\colon V \to W$ of inner product spaces is isometric if

$$\langle L(\vec{u}), L(\vec{v}) \rangle_W = \langle \vec{u}, \vec{v} \rangle_V \quad \text{for all } \vec{u}, \vec{v} \in V.$$

This implies that all the concepts defined by the inner product, such as the length, the angle, the orthogonality, and the area are preserves. For example, an isometry preserves the *distance*

$$\|L(\vec{u}) - L(\vec{v})\| = \|L(\vec{u} - \vec{v})\| = \|\vec{u} - \vec{v}\|.$$

Conversely, by the polarisation formula in Exercise 4.14, a linear transformation preserving the length (or distance) is an isometry.

The following shows that an isometry is always one-to-one

$$\vec{u} \neq \vec{v} \implies \|\vec{u} - \vec{v}\| \neq 0 \implies \|L(\vec{u}) - L(\vec{v})\| \neq 0 \implies L(\vec{u}) \neq L(\vec{v}).$$

Therefore an isometry is an isomorphism if and only if it is onto. By Proposition 2.15 and Theorem 2.16, this is also equivalent to that the spaces on both sides have the same dimension.

A finite dimensional vector space V is isomorphic to a Euclidean space $\mathbb{R}^n$ by the coordinates with respect to an ordered basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$

$$[\vec{x}]_\alpha = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \longleftrightarrow \vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \in V.$$

By the bilinear property of the inner product, we have

$$\langle x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n, y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n\rangle = \sum \langle \vec{v}_i, \vec{v}_j\rangle x_i y_j.$$

This is equal to the dot product in $\mathbb{R}^n$ if and only if

$$\langle \vec{v}_i, \vec{v}_i\rangle = 1, \quad \langle \vec{v}_i, \vec{v}_j\rangle = 0 \text{ for } i \neq j.$$

The condition means exactly that α is orthonormal.

Proposition 4.11. *Suppose α is a basis of an inner product space. Then the α-coordinate is an isometric isomorphism between V and the Euclidean space (with dot product) if and only if α is an orthonormal basis.*

Exercise 4.40 gives a similar characterisation for isometric linear transformations. We will also prove Proposition 4.14, which shows that any linearly independent set can be modified to an orthonormal set, such that the span remain the same. Combined with Proposition 4.11, we get the following result.

Theorem 4.12. *Any finite dimensional inner product space is isometrically isomorphic to the Euclidean space with dot product.*

Exercise 4.38. Prove that the composition of isometries is an isometry.

Exercise 4.39. Prove that the inverse of an isometric isomorphism is also an isometric isomorphism.

Orthogonal Matrix

Consider the matrix A of an isometric linear transformation $L(\vec{x}) = A\vec{x}\colon \mathbb{R}^n \to \mathbb{R}^m$ between Euclidean spaces with dot product. By Example 4.1, we have $\vec{x}\cdot\vec{y} = \vec{x}^T\vec{y}$, and

$$A\vec{x}\cdot A\vec{y} = (A\vec{x})^T(A\vec{y}) = \vec{x}^T A^T A\vec{y}.$$

Therefore L is an isometry if and only if $\vec{x}^T A^T A\vec{y} = \vec{x}^T\vec{y}$ for all $\vec{x}$ and $\vec{y}$. By Exercise 4.5, this means exactly $A^TA = I$ is the identity matrix. By Example 4.11, this means that the columns of A form an orthonormal set. This can also be explained by applying Exercise 4.40 to the standard basis of $\mathbb{R}^n$ (which are the columns of A).

Since isometries are always one-to-one, the isometric linear transformation L becomes an isometric isomorphism when $m = n$. Then A is a square matrix.

Definition 4.13. An *orthogonal* matrix is a square matrix U satisfying $U^TU = I$.

It is more appropriate to call U an orthonormal matrix instead of orthogonal, because the columns of U form an orthonormal basis. We note that U must be invertible. Then $U^TU = I$ implies $U^{-1} = U^T$, and $UU^T = I$. In particular, the rows of U is also an orthonormal basis.

Suppose $L\colon V \to W$ is an isometric linear transformation. Suppose α and β are orthonormal bases of V and W. Then (by Exercises 4.38, 4.39) the induced linear transformation $L_{\beta\alpha}\colon \mathbb{R}^n \to \mathbb{R}^m$ is isometric. This means that the matrix of L with respect to the orthonormal bases satisfies $[L]_{\beta\alpha}^T[L]_{\beta\alpha} = I$. If we further know that $\dim V = \dim W$, then L is invertible and $[L]_{\beta\alpha}$ is an orthogonal matrix.

Example 4.13. The vectors $\vec{v}_1 = (2, 2, -1)$ and $\vec{v}_2 = (2, -1, 2)$ are orthogonal. To get an orthogonal basis, we need to add one vector $\vec{v}_3 = (x, y, z)$ satisfying

$$\vec{v}_3 \cdot \vec{v}_1 = 2x + 2y - z = 0, \quad \vec{v}_3 \cdot \vec{v}_2 = 2x - y + 2z = 0.$$

The solution is $y = z = -2x$. Taking $x = -1$, or $\vec{v}_3 = (-1, 2, 2)$, we get an orthogonal basis $\{(2, 2, -1), (2, -1, 2), (-1, 2, 2)\}$. By dividing the length $\|\vec{v}_1\| = \|\vec{v}_2\| = \|\vec{v}_3\| = 3$, we get an orthonormal basis $\{\frac{1}{3}(2, 2, -1), \frac{1}{3}(2, -1, 2), \frac{1}{3}(-1, 2, 2)\}$. Using three vectors as columns, we get an orthogonal matrix

$$U = \frac{1}{3}\begin{pmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{pmatrix}, \quad U^T U = I.$$

The inverse of the matrix is simply the transpose

$$U^{-1} = U^T = \frac{1}{3}\begin{pmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{pmatrix}.$$

Exercise 4.40. Prove that a linear transformation $L\colon V \to W$ between inner product spaces is isometric if and only if L takes an orthonormal bases of V to an orthonormal set of W.

Exercise 4.41. Prove that the transpose, inverse and multiplication of orthogonal matrices are orthogonal matrices.

Exercise 4.42. Suppose α is an orthonormal basis. Prove that another basis β is orthonormal if and only if $[I]_{\beta\alpha}$ is an orthogonal matrix.

Gram-Schmidt Process

Theorem 4.12 says that the linear algebra of inner product spaces is the same as the linear algebra of Euclidean spaces with dot product. The theorem is based on the existence of orthonormal basis, which is a consequence of the following result.

Proposition 4.14. *Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is a linearly independent set in an inner product space. Then there is an orthonormal set $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$, such that*

$$\text{Span}\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_i\} = \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_i\}, \quad i = 1, 2, \dots, n.$$

Proof. The set β is constructed by the following *Gram-Schmidt process*. First we fix $\vec{w}_1 = \vec{v}_1$ and get

$$\mathrm{Span}\{\vec{w}_1\} = \mathrm{Span}\{\vec{v}_1\}, \quad \vec{w}_1 \neq \vec{0}.$$

Next we try to modify $\vec{v}_2$ to

$$\vec{w}_2 = \vec{v}_2 + c\vec{w}_1 = \vec{v}_2 + c\vec{v}_1,$$

such that $\vec{w}_2$ is orthogonal to $\vec{w}_1$

$$0 = \langle \vec{w}_2, \vec{w}_1 \rangle = \langle \vec{v}_2, \vec{w}_1 \rangle + c\langle \vec{w}_1, \vec{w}_1 \rangle.$$

This shows that, if we take $c = -\frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}$, then

$$\vec{w}_2 = \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}\vec{w}_1$$

satisfies $\vec{w}_2 \perp \vec{w}_1$, and

$$\mathrm{Span}\{\vec{w}_1, \vec{w}_2\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2 + c\vec{v}_1\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2\}.$$

Here the second equality follows from Proposition 3.3. We also note that $\vec{w}_2 \neq \vec{0}$ because $\vec{v}_1, \vec{v}_2$ are linearly independent.

Inductively, suppose we already find a nonzero orthogonal set $\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_i\}$ satisfying

$$\mathrm{Span}\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_i\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_i\}.$$

Then we try to modify $\vec{v}_{i+1}$ to

$$\vec{w}_{i+1} = \vec{v}_{i+1} + c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_i\vec{w}_i,$$

such that $\vec{w}_{i+1}$ is orthogonal to each $\vec{w}_j$, $1 \leq j \leq i$,

$$\begin{aligned} 0 = \langle \vec{w}_{i+1}, \vec{w}_j \rangle &= \langle \vec{v}_{i+1}, \vec{w}_j \rangle + c_1\langle \vec{w}_1, \vec{w}_j \rangle + c_2\langle \vec{w}_2, \vec{w}_j \rangle + \cdots + c_i\langle \vec{w}_i, \vec{w}_j \rangle \\ &= \langle \vec{v}_{i+1}, \vec{w}_j \rangle + c_j\langle \vec{w}_j, \vec{w}_j \rangle. \end{aligned}$$

This shows that, if we take $c_j = -\frac{\langle \vec{v}_{i+1}, \vec{w}_j \rangle}{\langle \vec{w}_j, \vec{w}_j \rangle}$, then

$$\vec{w}_{i+1} = \vec{v}_{i+1} - \frac{\langle \vec{v}_{i+1}, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}\vec{w}_1 - \frac{\langle \vec{v}_{i+1}, \vec{w}_2 \rangle}{\langle \vec{w}_2, \vec{w}_2 \rangle}\vec{w}_2 - \cdots - \frac{\langle \vec{v}_{i+1}, \vec{w}_i \rangle}{\langle \vec{w}_i, \vec{w}_i \rangle}\vec{w}_i$$

satisfies

$$\vec{w}_{i+1} \perp \vec{w}_1,\ \vec{w}_{i+1} \perp \vec{w}_2,\ \dots,\ \vec{w}_{i+1} \perp \vec{w}_i.$$

Moreover, we have

$$\vec{v}_{i+1} = \vec{w}_{i+1} - c_1\vec{w}_1 - c_2\vec{w}_2 - \cdots - c_i\vec{w}_i \in \mathrm{Span}\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_{i+1}\}.$$

On the other hand, by Proposition 3.3, we have

$$\vec{w}_{i+1} = \vec{v}_{i+1} + c'_1\vec{v}_1 + c'_2\vec{v}_2 + \cdots + c'_i\vec{v}_i \in \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{i+1}\}.$$

Then we conclude

$$\mathrm{Span}\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_{i+1}\} = \mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{i+1}\}.$$

By the linear independence of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{i+1}$, we also know that $\vec{w}_{i+1} \neq \vec{0}$.

The proposition is almost proved after we get $\vec{w}_n$, except the lengths of the vectors are not necessarily 1. This can be acheived by dividing the lengths $\|\vec{w}_i\|$. □

The proof means α and β are related in "triangular" way, with $a_{ii} \neq 0$

$$\begin{aligned}
\vec{v}_1 &= a_{11}\vec{w}_1, \\
\vec{v}_2 &= a_{12}\vec{w}_1 + a_{22}\vec{w}_2, \\
&\vdots \\
\vec{v}_n &= a_{1n}\vec{w}_1 + a_{2n}\vec{w}_2 + \cdots + a_{nn}\vec{w}_n.
\end{aligned}$$

The vectors $\vec{w}_i$ can also be expressed in $\vec{v}_i$ in similar triangular way. The relation can be rephrased in the matrix form, called *QR-decomposition*

$$A = QR, \quad A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n), \quad Q = (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n), \quad R = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

Proposition 4.15. *If the columns of a matrix A are linearly independent, then $A = QR$ for a unique matrix Q with orthonormal column and a unique upper triangular matrix R.*

Example 4.14. The subspace H of $\mathbb{R}^3$ given by the equation $x + y + z = 0$ has basis given by $\vec{v}_1 = (1, -1, 0)$ and $\vec{v}_2 = (1, 0, -1)$ (see Example 2.16). Then we derive an orthogonal basis of H

$$\begin{aligned}
\vec{w}_1 &= \vec{v}_1 = (1, -1, 0), \\
\vec{w}'_2 &= \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}\vec{w}_1 = (1, 0, -1) - \frac{1+0+0}{1+1+0}(1, -1, 0) = \frac{1}{2}(1, 1, -2), \\
\vec{w}_2 &= 2\vec{w}'_2 = (1, 1, -2).
\end{aligned}$$

By dividing the lengths, we further get an orthonormal basis

$$\frac{\vec{w}_1}{\|\vec{w}_1\|} = \frac{1}{\sqrt{2}}(1, -1, 0), \quad \frac{\vec{w}_2}{\|\vec{w}_2\|} = \frac{1}{\sqrt{6}}(1, 1, -2).$$

To get the corresponding QR-decomposition, we note that

$$\vec{v}_1 = \vec{w}_1 = \sqrt{2}\frac{\vec{w}_1}{\|\vec{w}_1\|},$$
$$\vec{v}_2 = \frac{\langle \vec{v}_2, \vec{w}_1\rangle}{\langle \vec{w}_1, \vec{w}_1\rangle}\vec{w}_1 + \vec{w}_2' = \frac{1}{2}\vec{w}_1 + \frac{1}{2}\vec{w}_2 = \frac{1}{\sqrt{2}}\frac{\vec{w}_1}{\|\vec{w}_1\|} + \frac{\sqrt{3}}{\sqrt{2}}\frac{\vec{w}_2}{\|\vec{w}_2\|}.$$

Alternatively, we can use Proposition 4.20 to get

$$\vec{v}_1 = \frac{\langle \vec{v}_1, \vec{w}_1\rangle}{\langle \vec{w}_1, \vec{w}_1\rangle}\vec{w}_1 = \frac{1+1+0}{1+1+0}\vec{w}_1 = \vec{w}_1,$$
$$\vec{v}_2 = \frac{\langle \vec{v}_2, \vec{w}_1\rangle}{\langle \vec{w}_1, \vec{w}_1\rangle}\vec{w}_1 + \frac{\langle \vec{v}_2, \vec{w}_2\rangle}{\langle \vec{w}_2, \vec{w}_2\rangle}\vec{w}_2 = \frac{1+0+0}{1+1+0}\vec{w}_1 + \frac{1+0+2}{1+1+4}\vec{w}_2 = \frac{1}{2}\vec{w}_1 + \frac{1}{2}\vec{w}_2.$$

Then we get the QR-decomposition

$$\begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}\begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} \end{pmatrix}.$$

Example 4.15. By Example 1.28, we know $\vec{v}_1 = (1,2,3)$, $\vec{v}_2 = (4,5,6)$, $\vec{v}_3 = (7,8,0)$ form a basis of $\mathbb{R}^3$. We apply the Gram-Schmidt process to the basis.

$$\begin{aligned}
\vec{w}_1 &= \vec{v}_1 = (1,2,3),\\
\vec{w}_2' &= \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1\rangle}{\langle \vec{w}_1, \vec{w}_1\rangle}\vec{w}_1 = (4,5,6) - \frac{4+10+18}{1+4+9}(1,2,3) = -\frac{3}{7}(4,1,-2),\\
\vec{w}_2 &= (4,1,-2),\\
\vec{w}_3' &= \vec{v}_3 - \frac{\langle \vec{v}_3, \vec{w}_1\rangle}{\langle \vec{w}_1, \vec{w}_1\rangle}\vec{w}_1 - \frac{\langle \vec{v}_3, \vec{w}_2\rangle}{\langle \vec{w}_2, \vec{w}_2\rangle}\vec{w}_2\\
&= (7,8,0) - \frac{7+16+0}{1+4+9}(1,2,3) - \frac{28+8+0}{16+1+4}(4,1,-2) = -\frac{1}{14}(21,-62,41),\\
\vec{w}_3 &= (21,-62,41).
\end{aligned}$$

The result can be expressed as a matrix product

$$\begin{pmatrix} 1 & 4 & 21 \\ 2 & 1 & -62 \\ 3 & -2 & 41 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 0 \end{pmatrix}\begin{pmatrix} 1 & \frac{16}{7} & \frac{23}{14} \\ 0 & -\frac{3}{7} & \frac{12}{7} \\ 0 & 0 & -\frac{1}{14} \end{pmatrix}.$$

Then we get the QR-decomposition

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 21 \\ 2 & 1 & -62 \\ 3 & -2 & 41 \end{pmatrix}\begin{pmatrix} 1 & \frac{16}{7} & \frac{23}{14} \\ 0 & -\frac{3}{7} & \frac{12}{7} \\ 0 & 0 & -\frac{1}{14} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 4 & 21 \\ 2 & 1 & -62 \\ 3 & -2 & 41 \end{pmatrix}\begin{pmatrix} 1 & \frac{16}{3} & 151 \\ 0 & -\frac{7}{3} & -56 \\ 0 & 0 & -14 \end{pmatrix}.$$

Example 4.16. The natural basis $\{1, t, t^2\}$ of P_2 is not orthogonal with respect to the inner product in Example 4.3. We improve the basis to become orthogonal

$$\begin{aligned}
f_1 &= 1, \\
f_2 &= t - \frac{\int_0^1 t \cdot 1dt}{\int_0^1 1^2 dt} 1 = t - \frac{1}{2}, \\
f_3 &= t^2 - \frac{\int_0^1 t^2 \cdot 1dt}{\int_0^1 1^2 dt} 1 - \frac{\int_0^1 t^2 \left(t - \frac{1}{2}\right) dt}{\int_0^1 \left(t - \frac{1}{2}\right)^2 dt} \left(t - \frac{1}{2}\right) = t^2 - t + \frac{1}{6}.
\end{aligned}$$

Exercise 4.43. Find an orthogonal basis of the subspace in Example 4.14 by starting with $\vec{v}_2$ and then use $\vec{v}_1$.

Exercise 4.44. Find an orthogonal basis of the subspace in Example 4.14 with respect to the inner product $\langle (x_1, x_2, x_3), (y_1, y_2, y_3) \rangle = x_1 y_1 + 2x_2 y_2 + 3x_3 y_3$.

Exercise 4.45. Apply the Gram-Schmidt process to $1, t, t^2$ with respect to the inner products in Exercises 4.6, 4.21, 4.22.

Exercise 4.46. Find QR-decomposition. ??????????

4.3 Orthogonal Subspace

In Section 3.3, we learned that the essence of linear algebra is not individual vectors, but subspaces. The essence of span is sum of subspace, and the essence of linear independence is that the sum is direct. Similarly, the essence of orthogonal vectors is orthogonal subspaces.

Definition 4.16. Two spaces H and H' are *orthogonal* and denoted $H \perp H'$, if $\langle \vec{v}, \vec{w} \rangle = 0$ for all $\vec{v} \in H$ and $\vec{w} \in H'$.

The following generalises Proposition 4.9.

Theorem 4.17. *If subspaces $H_1, H_2, \dots, H_n$ are pairwise orthogonal, then $H_1 + H_2 + \cdots + H_n$ is a direct sum.*

Proof. Suppose $\vec{h}_i \in H_i$ satisfies $\vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_n = \vec{0}$. Then by the pairwise orthogonality, we have

$$0 = \langle \vec{h}_i, \vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_n \rangle = \langle \vec{h}_i, \vec{h}_1 \rangle + \langle \vec{h}_i, \vec{h}_2 \rangle + \cdots + \langle \vec{h}_i, \vec{h}_n \rangle = \langle \vec{h}_i, \vec{h}_i \rangle.$$

By the positivity of the inner product, this implies $\vec{h}_i = \vec{0}$. □

Exercise 4.47. What is the subspace orthogonal to itself?

Exercise 4.48. Prove that $H_1 + H_2 + \cdots + H_m \perp H'_1 + H'_2 + \cdots + H'_n$ if and only if $H_i \perp H'_j$ for all i and j. What does this tell you about $\text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\} \perp \text{Span}\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$?

Orthogonal Complement

Definition 4.18. The *orthogonal complement* of a subspace H of an inner product space V is

$$H^\perp = \{\vec{v} \colon \langle \vec{v}, \vec{h} \rangle = 0 \text{ for all } \vec{h} \in H\}.$$

The orthogonal complement is easily seen to be a subspace. If $H = \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$, then

$$\vec{x} \in H^\perp \implies \langle \vec{x}, \vec{v}_1 \rangle = \langle \vec{x}, \vec{v}_2 \rangle = \cdots = \langle \vec{x}, \vec{v}_k \rangle = 0.$$

Conversely, if the right side holds, then any $\vec{h} \in H$ is a linear combination $\vec{h} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$, and we have

$$\langle \vec{x}, \vec{h} \rangle = c_1\langle \vec{x}, \vec{v}_1 \rangle + c_2\langle \vec{x}, \vec{v}_2 \rangle + \cdots + c_k\langle \vec{x}, \vec{v}_k \rangle = 0.$$

This gives the following practical way of calculating the orthogonal complement

Proposition 4.19. *The orthogonal complement of* $\text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ *is all the vector orthogonal to* $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$.

Consider an $m \times n$ matrix $A = (\vec{v}_1 \; \vec{v}_2 \; \cdots \; \vec{v}_n)$, $\vec{v}_i \in \mathbb{R}^m$. By Proposition 4.19, the orthogonal complement of the column space $\text{Col}A$ consists of vectors $\vec{x} \in \mathbb{R}^m$ satisfying $\vec{v}_1 \cdot \vec{x} = \vec{v}_2 \cdot \vec{x} = \cdots = \vec{v}_n \cdot \vec{x} = 0$. By the formula in Example 4.1, this means $A^T\vec{x} = \vec{0}$, or the null space of A^T. Therefore we get

$$(\text{Col}A)^\perp = \text{Nul}A^T.$$

Taking transpose, we also get

$$(\text{Row}A)^\perp = \text{Nul}A.$$

Translated into linear transformations $L \colon V \to W$ between general inner product spaces, we expect

$$(\text{Ran}L)^\perp = \text{Ker}L^*.$$

The following is a direct argument for the equality

$$\begin{aligned}
\vec{x} \in (\text{Ran}L)^\perp &\iff \langle \vec{w}, \vec{x} \rangle = 0 \text{ for all } \vec{w} \in \text{Ran}L \subset W \\
&\iff \langle L(\vec{v}), \vec{x} \rangle = 0 \text{ for all } \vec{v} \in V \\
&\iff \langle \vec{v}, L^*(\vec{x}) \rangle = 0 \text{ for all } \vec{v} \in V \\
&\iff L^*(\vec{x}) = \vec{0} \text{ for all } \vec{v} \in V \\
&\iff \vec{x} \in \text{Ker}L^*.
\end{aligned}$$

Substituting L^* in place of L, we get

$$(\text{Ran}L^*)^\perp = \text{Ker}L.$$

Example 4.17. The orthogonal complement of the line $H = \mathbb{R}(1,1,1)$ consists of vectors $(x,y,z) \in \mathbb{R}^3$ satisfying $(1,1,1)\cdot(x,y,z) = x+y+z = 0$. This is the plane in Example 2.16. In general, the solutions of $a_1x_1 + a_2x_2 + \cdots + a_mx_m = 0$ is the hyperplane orthogonal to the line $\mathbb{R}(a_1, a_2, \dots, a_m)$.

For another example, Example 3.8 shows that the orthogonal complement of the span of $(1,4,7,10), (2,5,8,11), (3,6,9,12)$ is spanned by $(1,-2,0,0), (2,-3,0,1)$.

Example 4.18. We try to calculate the orthogonal complement of P_1 (span of 1 and t) in P_3 with respect to the inner product in Example 4.3. A polynomial $f = a_0 + a_1t + a_2t^2 + a_3t^3$ is in the orthogonal complement if and only if

$$\langle 1, f\rangle = \int_0^1 (a_0 + a_1t + a_2t^2 + a_3t^3)dt = a_0 + \frac{1}{2}a_1 + \frac{1}{3}a_2 + \frac{1}{4}a_3 = 0,$$
$$\langle t, f\rangle = \int_0^1 t(a_0 + a_1t + a_2t^2 + a_3t^3)dt = \frac{1}{2}a_0 + \frac{1}{3}a_1 + \frac{1}{4}a_2 + \frac{1}{5}a_3 = 0.$$

We find two linearly independent solutions $2 - 9t + 10t^3$ and $1 - 18t^2 + 20t^3$, which form a basis of the orthogonal complement.

Exercise 4.49. Prove that $H \cap H^\perp$ consists of the zero vector only. In other words, $H + H^\perp$ is a direct sum.

Exercise 4.50. Prove that $H \subset H'$ implies $H^\perp \supset H'^\perp$.

Exercise 4.51. Prove that $H \subset (H^\perp)^\perp$.

Exercise 4.52. Prove that $(H_1 + H_2 + \cdots + H_n)^\perp = H_1^\perp \cap H_2^\perp \cap \cdots \cap H_n^\perp$.

Exercise 4.53. Find the orthogonal complement of the span of $(1,4,7,10), (2,5,8,11), (3,6,9,12)$ with respect to the inner product $\langle (x_1,x_2,x_3,x_4), (y_1,y_2,y_3,y_4)\rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3 + 4x_4y_4$.

Exercise 4.54. Find the orthogonal complement of P_1 in P_3 respect to the inner products in Exercises 4.6, 4.21, 4.22.

Proposition 4.20. *Suppose H is a subspace of an (finite dimensional) inner product space V. Then $(H^\perp)^\perp = H$ and $V = H \oplus H^\perp$.*

Proof. Let $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ be an orthogonal basis of H. By Theorem 1.18, we may extend the set to a basis $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of V. By applying the Gram-Schmidt process in Proposition 4.14, the basis can be improved to an orthogonal basis of V. Since $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ is already orthogonal, the improved orthogonal basis is of the form $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k, \vec{w}_{k+1}, \vec{w}_{k+2}, \dots, \vec{w}_n\}$.

By Proposition 4.8, any vector $\vec{x} \in V$ is of the form

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k + x_{k+1}\vec{w}_{k+1} + x_{k+2}\vec{w}_{k+2} + \cdots + x_n\vec{w}_n,$$

with

$$x_i = \frac{\langle \vec{x}, \vec{v}_i \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \text{ for } 1 \le i \le k, \quad x_i = \frac{\langle \vec{x}, \vec{w}_i \rangle}{\langle \vec{w}_i, \vec{w}_i \rangle} \text{ for } k+1 \le i \le n.$$

Therefore $\vec{x} \perp \vec{v}_i$ if and only if $x_i = 0$. By Proposition 4.19, we find that $H^\perp$ consists of those $\vec{x}$ satisfying $x_1 = x_2 = \cdots = x_k = 0$. This means exactly $\vec{x} = x_{k+1}\vec{w}_{k+1} + x_{k+2}\vec{w}_{k+2} + \cdots + x_n\vec{w}_n$, or

$$H^\perp = \text{Span}\{\vec{w}_{k+1}, \vec{w}_{k+2}, \ldots, \vec{w}_n\}.$$

The explicit formula of H and $H^\perp$ by span shows that $V = H \oplus H^\perp$, and

$$(H^\perp)^\perp = (\text{Span}\{\vec{w}_{k+1}, \vec{w}_{k+2}, \ldots, \vec{w}_n\})^\perp = \text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\} = H.$$

Here the second equality is obtained by switching $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ and $\{\vec{w}_{k+1}, \vec{w}_{k+2}, \ldots, \vec{w}_n\}$. □

Combining Proposition 4.20 with the orthogonal complements of the column space, row space and range space, we get the orthogonal complements of the null space and kernel space

$$(\text{Nul}A)^\perp = \text{Row}A, \quad (\text{Nul}A^T)^\perp = \text{Col}A, \quad (\text{Ker}L)^\perp = \text{Ran}L^*, \quad (\text{Ker}L^*)^\perp = \text{Ran}L.$$

These are called the *complementarity principles*. For example, the equality $(\text{Nul}A^T)^\perp = \text{Col}A$ means that $A\vec{x} = \vec{b}$ has solution if and only if $\vec{b}$ is orthogonal to all the solutions of the equation $A^T\vec{x} = \vec{0}$. Similarly, the equality $(\text{Row}A)^\perp = \text{Nul}A$ means that $A\vec{x} = \vec{0}$ if and only if $\vec{x}$ is orthogonal to all $\vec{b}$ such that $A^T\vec{x} = \vec{b}$ has solution.

Example 4.19. By Example 4.17 and $(H^\perp)^\perp = H$, the orthogonal complement of the span of $(1, -2, 0, 0), (2, -3, 0, 1)$ is spanned by $(1, 4, 7, 10), (2, 5, 8, 11), (3, 6, 9, 12)$.

Similarly, in Example 4.18, the orthogonal complement of the span of $2 - 9t + 10t^3$ and $1 - 18t^2 + 20t^3$ in P_3 is P_1.

Exercise 4.55. Prove that $H \subset H'$ if and only if $H^\perp \supset H'^\perp$.

Exercise 4.56. Prove that V/H is naturally isomorphic to $H^\perp$.

Exercise 4.57. Prove that the orthogonal complement of H in V is the unique subspace H' satisfying $H' \perp H$ and $H + H' = V$.

Orthogonal Projection

The direct sum in Proposition 4.20 induces the *orthogonal projection* from V to H

$$\text{proj}_H\vec{v} = \vec{h}, \quad \text{if } \vec{v} = \vec{h} + \vec{h}', \ \vec{h} \in H, \ \vec{h}' \in H^\perp.$$

By $(H^\perp)^\perp = H$, the role of H and H' can be replaced, and we get

$$\vec{v} = \text{proj}_H\vec{v} + \text{proj}_{H^\perp}\vec{v}.$$

The following is the orthogonal version of Exercise 3.53.

Proposition 4.21. *If $H_1, H_2, \dots, H_n$ are pairwise orthogonal subspaces, then*

$$\mathrm{proj}_{H_1+H_2+\cdots+H_n}\vec{v} = \mathrm{proj}_{H_1}\vec{v} + \mathrm{proj}_{H_2}\vec{v} + \cdots + \mathrm{proj}_{H_n}\vec{v}.$$

Proof. Let $H = H_1 + H_2 + \cdots + H_n$. Then

$$\vec{v} = \vec{h} + \vec{h}', \quad \vec{h} = \mathrm{proj}_{H_1+H_2+\cdots+H_n}\vec{v} \in H,\ \vec{h}' \in H^\perp.$$

Moreover, we have

$$\vec{h} = \vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_n, \quad \vec{h}_i \in H_i.$$

By Exercise 4.48, we have $H_2 + \cdots + H_n + H' \subset H_1^\perp$ and

$$\vec{v} = \vec{h}_1 + (\vec{h}_2 + \cdots + \vec{h}_n + \vec{h}'), \quad \vec{h}_1 \in H_1,\ \vec{h}_2 + \cdots + \vec{h}_n + \vec{h}' \in H_1^\perp.$$

Therefore $\vec{h}_1 = \mathrm{proj}_{H_1}\vec{v}$. Similarly, we get $\vec{h}_i = \mathrm{proj}_{H_i}\vec{v}$ for all $i = 1, 2, \dots, n$, and

$$\mathrm{proj}_{H_1+H_2+\cdots+H_n}\vec{v} = \vec{h} = \vec{h}_1 + \vec{h}_2 + \cdots + \vec{h}_n = \mathrm{proj}_{H_1}\vec{v} + \mathrm{proj}_{H_2}\vec{v} + \cdots + \mathrm{proj}_{H_n}\vec{v}. \qquad \square$$

Exercise 4.58. Prove the formula for the orthogonal projection to a line $\mathbb{R}\vec{u}$

$$\mathrm{proj}_{\mathbb{R}\vec{u}} = \frac{\langle \vec{v}, \vec{u} \rangle}{\langle \vec{u}, \vec{u} \rangle}\vec{u}.$$

Exercise 4.59. Prove that $I = \mathrm{proj}_H + \mathrm{proj}_{H'}$ if and only if H' is the orthogonal complement of H.

To calculate the orthogonal projection, we start with an orthogonal basis $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ of H. By the argument in the proof of Proposition 4.20, the basis extends to an orthogonal basis $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of V. By Proposition 4.8, any vector $\vec{x} \in V$ is of the form

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n, \quad x_i = \frac{\langle \vec{x}, \vec{v}_i \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \text{ for } 1 \le i \le n.$$

Then its is easy to verify that

$$\vec{h} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k, \quad \vec{h}' = x_{k+1}\vec{v}_{k+1} + x_{k+2}\vec{v}_{k+2} + \cdots + x_n\vec{v}_n,$$

in the definition of orthogonal projection.

Proposition 4.22. *If $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ is an orthogonal basis of a subspace $H \subset V$, then*

$$\mathrm{proj}_H\vec{x} = \frac{\langle \vec{x}, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}\vec{v}_1 + \frac{\langle \vec{x}, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle}\vec{v}_2 + \cdots + \frac{\langle \vec{x}, \vec{v}_k \rangle}{\langle \vec{v}_k, \vec{v}_k \rangle}\vec{v}_k.$$

Example 4.20. In Examples 2.16 and 2.46, we used two different ways to find the matrix of the orthogonal projection onto the subspace in $\mathbb{R}^3$ given by $x + y + z = 0$. Here we find the matrix by using Proposition 4.22.

By the orthogonal basis obtained in Example 4.14, we have

$$\mathrm{proj}_{\{x+y+z=0\}}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\frac{x_1-x_2}{1+1+0}\begin{pmatrix}1\\-1\\0\end{pmatrix}+\frac{x_1+x_2-2x_3}{1+1+4}\begin{pmatrix}1\\1\\-2\end{pmatrix}=\frac{1}{3}\begin{pmatrix}2x_1-x_2-x_3\\-x_1+2x_2-x_3\\-x_1-x_2+2x_3\end{pmatrix}.$$

The coefficient matrix is the same as the one we get in the earlier examples.

More generally we wish to calculate the matrix of the orthogonal projection to the subspace of $\mathbb{R}^n$ of dimension $n-1$

$$H=\{(x_1,x_2,\dots,x_n)\colon a_1x_1+a_2x_2+\cdots+a_nx_n=0\}=(\mathbb{R}\vec{a})^{\perp},\quad \vec{a}=(a_1,a_2,\dots,a_n).$$

It would be complicated to find an orthogonal basis of H. Instead, we take advantage of the equality $\mathrm{proj}_H\vec{x}=\vec{x}-\mathrm{proj}_{H^{\perp}}\vec{x}$ and Exercise 4.58. If $\|\vec{a}\|=1$, then

$$\begin{aligned}\mathrm{proj}_{\{a_1x_1+a_2x_2+\cdots+a_nx_n=0\}}\vec{x}&=\vec{x}-\mathrm{proj}_{\mathbb{R}\vec{a}}\vec{x}=\vec{x}-(a_1x_1+a_2x_2+\cdots+a_nx_n)\vec{a}\\&=\begin{pmatrix}x_1-a_1(a_1x_1+a_2x_2+\cdots+a_nx_n)\\x_2-a_2(a_1x_1+a_2x_2+\cdots+a_nx_n)\\\vdots\\x_n-a_n(a_1x_1+a_2x_2+\cdots+a_nx_n)\end{pmatrix}\\&=\begin{pmatrix}1-a_1^2&-a_1a_2&\cdots&-a_1a_n\\-a_2a_1&1-a_2^2&\cdots&-a_2a_n\\\vdots&\vdots&&\vdots\\-a_na_1&-a_na_2&\cdots&1-a_n^2\end{pmatrix}\vec{x}.\end{aligned}$$

For $\vec{a}=\frac{1}{\sqrt{3}}(1,1,1)$, we get the same matrix as earlier examples.

Example 4.21. Let

$$A=\begin{pmatrix}1&4&7&10\\2&5&8&11\\3&6&9&12\end{pmatrix}.$$

In Example 3.8, we get a basis $(1,-2,1,0),(2,-3,0,1)$ for the null space $\mathrm{Nul}A$. Before using the formula in Proposition 4.22, we note that the two vectors are not orthogonal. By

$$(2,-3,0,1)-\frac{(2,-3,0,1)\cdot(1,-2,1,0)}{(1,-2,1,0)\cdot(1,-2,1,0)}(1,-2,1,0)=\frac{1}{3}(2,-1,-4,3),$$

$(1,-2,1,0),(2,-1,-4,3)$ form an orthogonal basis. Then

$$\mathrm{proj}_{\mathrm{Nul}A}\vec{x}=\frac{x_1-2x_2+x_3}{1+4+1}\begin{pmatrix}1\\-2\\1\\0\end{pmatrix}+\frac{2x_1-x_2-4x_3+3x_4}{4+1+16+9}\begin{pmatrix}2\\-1\\-4\\3\end{pmatrix}=\frac{1}{10}\begin{pmatrix}3&4&-1&2\\-4&7&-2&-1\\-1&-2&7&-4\\2&-1&-4&3\end{pmatrix}\vec{x}.$$

Since Row$A$ is the orthogonal complement of Nul$A$, the matrix of the orthogonal projection to Row$A$ is

$$I-\frac{1}{10}\begin{pmatrix}3&4&-1&2\\-4&7&-2&-1\\-1&-2&7&-4\\2&-1&-4&3\end{pmatrix}=\frac{1}{10}\begin{pmatrix}7&-4&1&-2\\4&3&2&1\\1&2&3&4\\-2&1&4&7\end{pmatrix}.$$

Example 4.22. We calculate the orthogonal projection of P_3 to P_1 with respect to the inner product in Example 4.3. First we note that the basis $1, t$ of P_1 is not orthogonal. We can use the Gram-Schmidt process to get an orthogonal basis $1, 1-2t$. Then

$$\|1\|^2=\int_0^1 1^2dt=1,\quad \|1-2t\|^2=\int_0^1 t^2dt=\frac{1}{3},$$

and

$$\begin{aligned}\mathrm{proj}_{P_1}t^2&=\frac{\int_0^1 t^2dt}{\|1\|^2}1+\frac{\int_0^1(1-2t)t^2dt}{\|1-2t\|^2}(1-2t)=\frac{1}{3}1-\frac{1}{2}(1-2t)=-\frac{1}{6}+t,\\ \mathrm{proj}_{P_1}t^3&=\frac{\int_0^1 t^3dt}{\|1\|^2}1+\frac{\int_0^1(1-2t)t^3dt}{\|1-2t\|^2}(1-2t)=\frac{1}{4}1-\frac{9}{20}(1-2t)=-\frac{1}{5}+\frac{9}{10}t.\end{aligned}$$

Combined with $\mathrm{proj}_{P_1}1=1$ and $\mathrm{proj}_{P_1}t=t$, we get

$$\begin{aligned}\mathrm{proj}_{P_1}(a_0+a_1t+a_2t^2+a_3t^3)&=a_0+a_1t+a_2\left(-\frac{1}{6}+t\right)+a_3\left(-\frac{1}{5}+\frac{9}{10}t\right)\\&=\left(a_0-\frac{1}{6}a_2-\frac{1}{5}a_3\right)+\left(a_1+a_2+\frac{9}{10}a_3\right)t.\end{aligned}$$

Exercise 4.60. Among four subspaces Col$A$, Row$A$, Nul$A$, NulA^T, which two form complementary orthogonal projections

Exercise 4.61. Directly prove Proposition 4.22 by using the idea for the proof of Proposition 4.8.

Exercise 4.62. Directly prove Proposition 4.22 by using Proposition 4.21 and Exercise 4.58.

Exercise 4.63. Find the orthogonal projection the subspace $x+y+z=0$ in $\mathbb{R}^3$ with respect to the inner product $\langle(x_1,x_2,x_3),(y_1,y_2,y_3)\rangle=x_1y_1+2x_2y_2+3x_3y_3$.

Exercise 4.64. Directly find the orthogonal projection to the row space in Example 4.21.

Exercise 4.65. Use the orthogonal basis in Example 4.12 to calculate the orthogonal projection in Example 4.22.

Exercise 4.66. Find the orthogonal projection of general polynomial on P_1 with respect to the inner product in Example 4.3. What about the inner products in Exercises 4.6, 4.21, 4.22?

5 Determinant

Determinant first appeared as the numerical criterion that "determines" whether a system of linear equations has a unique solution. More properties of the determinant were discovered later, especially its relation to geometry. We will define determinant by axioms, derive the calculation technique from the axioms, and then discuss the geometric meaning of determinant.

5.1 Algebra

Definition 5.1. The *determinant* of $n \times n$ matrices A is the function $\det A$ satisfying the following properties.

1. Multilinear: The function is linear in each column vector
$$\det(\cdots\ a\vec{u}+b\vec{v}\ \cdots) = a\det(\cdots\ \vec{u}\ \cdots) + b\det(\cdots\ \vec{v}\ \cdots).$$
2. Alternating: Exchanging two columns introduces a negative sign
$$\det(\cdots\ \vec{v}\ \cdots\ \vec{u}\ \cdots) = -\det(\cdots\ \vec{u}\ \cdots\ \vec{v}\ \cdots).$$
3. Normal: The determinant of the identity matrix is 1
$$\det I = \det(\vec{e}_1\ \vec{e}_2\ \cdots\ \vec{e}_n) = 1.$$

For a multilinear function D, taking $\vec{u} = \vec{v}$ in the alternating property gives
$$D(\cdots\ \vec{u}\ \cdots\ \vec{u}\ \cdots) = 0.$$
Conversely, if D satisfies the equality above, then
$$\begin{aligned}
0 &= D(\cdots\ \vec{u}+\vec{v}\ \cdots\ \vec{u}+\vec{v}\ \cdots)\\
&= D(\cdots\ \vec{u}\ \cdots\ \vec{u}\ \cdots) + D(\cdots\ \vec{u}\ \cdots\ \vec{v}\ \cdots) + D(\cdots\ \vec{v}\ \cdots\ \vec{u}\ \cdots) + D(\cdots\ \vec{v}\ \cdots\ \vec{v}\ \cdots)\\
&= D(\cdots\ \vec{u}\ \cdots\ \vec{v}\ \cdots) + D(\cdots\ \vec{v}\ \cdots\ \vec{u}\ \cdots).
\end{aligned}$$
We recover the alternating property.

Explicit Formula for Determinant

A multilinear and alternating function D of 2×2 matrices is given by
$$\begin{aligned}
D\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} &= D(x_{11}\vec{e}_1 + x_{21}\vec{e}_2 \quad x_{12}\vec{e}_1 + x_{22}\vec{e}_2)\\
&= D(\vec{e}_1\ \vec{e}_1)x_{11}x_{12} + D(\vec{e}_1\ \vec{e}_2)x_{11}x_{22} + D(\vec{e}_2\ \vec{e}_1)x_{21}x_{12} + D(\vec{e}_2\ \vec{e}_2)x_{21}x_{22}\\
&= 0x_{11}x_{12} + D(\vec{e}_1\ \vec{e}_2)x_{11}x_{22} - D(\vec{e}_1\ \vec{e}_2)x_{21}x_{12} + 0x_{21}x_{22}\\
&= D(\vec{e}_1\ \vec{e}_2)(x_{11}x_{22} - x_{21}x_{12}).
\end{aligned}$$

A multilinear and alternating function D of 3×3 matrices is given by

$$\begin{aligned}
&D(x_{11}\vec{e}_1 + x_{21}\vec{e}_2 + x_{31}\vec{e}_3 \quad x_{12}\vec{e}_1 + x_{22}\vec{e}_2 + x_{32}\vec{e}_3 \quad x_{13}\vec{e}_1 + x_{23}\vec{e}_2 + x_{33}\vec{e}_3) \\
&= D(\vec{e}_1\ \vec{e}_2\ \vec{e}_3)x_{11}x_{22}x_{33} + D(\vec{e}_2\ \vec{e}_3\ \vec{e}_1)x_{21}x_{32}x_{13} + D(\vec{e}_3\ \vec{e}_1\ \vec{e}_2)x_{31}x_{12}x_{23} \\
&\quad + D(\vec{e}_1\ \vec{e}_3\ \vec{e}_2)x_{11}x_{32}x_{23} + D(\vec{e}_3\ \vec{e}_2\ \vec{e}_1)x_{31}x_{22}x_{13} + D(\vec{e}_2\ \vec{e}_1\ \vec{e}_3)x_{21}x_{12}x_{33} \\
&= D(\vec{e}_1\ \vec{e}_2\ \vec{e}_3)(x_{11}x_{22}x_{33} + x_{21}x_{32}x_{13} + x_{31}x_{12}x_{23} - x_{11}x_{32}x_{23} - x_{31}x_{22}x_{13} - x_{21}x_{12}x_{33}).
\end{aligned}$$

Here in the first equality, we used the alternating property that $D(\vec{e}_i\ \vec{e}_j\ \vec{e}_k) = 0$ whenever two of i, j, k are equal. In the second equality, we exchange distinct i, j, k (which must be a rearrangement of $1, 2, 3$) to the usual order. For example, we have

$$D(\vec{e}_3\ \vec{e}_1\ \vec{e}_2) = -D(\vec{e}_1\ \vec{e}_3\ \vec{e}_2) = D(\vec{e}_1\ \vec{e}_2\ \vec{e}_3).$$

In general, a multilinear and alternating function D of $n \times n$ matrices $A = (x_{ij})$ is

$$D(A) = D(\vec{e}_1\ \vec{e}_2\ \cdots\ \vec{e}_n) \sum \pm x_{i_1 1} x_{i_2 2} \cdots x_{i_n n},$$

where the sum runs over all the rearrangements (or *permutations*) $(i_1, i_2, \dots, i_n)$ of $(1, 2, \dots, n)$. Moreover, the $\pm$ sign, usually denoted by $\operatorname{sign}(i_1, i_2, \dots, i_n)$, is the parity of the number of exchanges (i.e., switching two indices) it takes to change from $(i_1, i_2, \dots, i_n)$ to $(1, 2, \dots, n)$.

The argument above can be carried out on any vector space, by replacing the standard basis of $\mathbb{R}^n$ by an (ordered) basis of V.

Theorem 5.2. *Multilinear and alternating functions of n vectors in an n-dimensional vector space V are unique up to multiplying constants. Specifically, in terms of the coordinates with respect to a basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of V, such a function D is given by*

$$D(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n) = c \sum_{(i_1, i_2, \dots, i_n)} \operatorname{sign}(i_1, i_2, \dots, i_n) x_{i_1 1} x_{i_2 2} \cdots x_{i_n n}, \quad [\vec{x}_j]_\alpha = (x_{1j}, x_{2j}, \dots, x_{nj}).$$

In case constant $c = D(\vec{v}_1, \dots, \vec{v}_n) = 1$, the formula in the theorem is the determinant.

Exercise 5.1. Write down the explicit formula for the determinant of 2×2 matrices.

Exercise 5.2. How many terms are in the explicit formula for the determinant of $n \times n$ matrices?

Exercise 5.3. Show that any alternating and bilinear function on 2×3 matrices is zero. Can you generalise this observation?

Exercise 5.4. Find explicit formula for an alternating and bilinear function on 3×2 matrices.

Theorem 5.2 is a very useful tool for deriving properties of the determinant. The following is a typical example.

Proposition 5.3. $\det AB = \det A \det B$.

Proof. For fixed A, we consider the function $D(B) = \det AB$. Since AB is obtained by multiplying A to the columns of B, $D(B)$ is multilinear and alternating in the columns of B. By Theorem 5.2, we get $D(B) = c\det B$. To determine the constant c, we let B to be the identity matrix and get $\det A = D(I) = c\det I = c$. Therefore $\det AB = D(B) = c\det B = \det A\det B$. $\square$

Exercise 5.5. Prove that $\det A^{-1} = \frac{1}{\det A}$. More generally, we have $\det A^n = (\det A)^n$ for any integer n.

Exercise 5.6. Use the explicit formula for the determinant to verify $\det AB = \det A \det B$ for 2×2 matrices.

Column Operation

The explicit formula in Theorem 5.2 is too complicated to be a practical way of calculating the determinant. Since matrices can be simplified by row and column operations, it is useful to know how the determinant is changed by the operations.

The alternating property means that the column operation $C_i \leftrightarrow C_j$ introduces a negative sign

$$\det(\cdots\ \vec{v}\ \cdots\ \vec{u}\ \cdots) = -\det(\cdots\ \vec{u}\ \cdots\ \vec{v}\ \cdots).$$

The multilinear property implies that the column operation $c\,C_i$ multiplies the determinant by the scalar

$$\det(\cdots\ c\vec{u}\ \cdots) = c\det(\cdots\ \vec{u}\ \cdots).$$

Combining the multilinear and alternating properties, the column operation $C_i + c\,C_j$ preserves the determinant

$$\begin{aligned}\det(\cdots\ \vec{u}+c\vec{v}\ \cdots\ \vec{v}\ \cdots) &= \det(\cdots\ \vec{u}\ \cdots\ \vec{v}\ \cdots) + c\det(\cdots\ \vec{v}\ \cdots\ \vec{v}\ \cdots)\\ &= \det(\cdots\ \vec{u}\ \cdots\ \vec{v}\ \cdots) + c\cdot 0\\ &= \det(\cdots\ \vec{u}\ \cdots\ \vec{v}\ \cdots).\end{aligned}$$

Example 5.1. By column operations, we have

$$\begin{aligned}\det\begin{pmatrix}1&4&7\\2&5&8\\3&6&a\end{pmatrix} &= \det\begin{pmatrix}1&0&0\\2&-3&-6\\3&-6&a-21\end{pmatrix} = -3\det\begin{pmatrix}1&0&0\\2&1&0\\3&2&a-9\end{pmatrix}\\ &= -3(a-9)\det\begin{pmatrix}1&0&0\\2&1&0\\3&2&1\end{pmatrix} = -3(a-9)\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} = -3(a-9).\end{aligned}$$

The column operation can simplify a matrix to a column echelon form, which is a *lower triangular matrix*. By column operation, the determinant of a lower triangular matrix is the product of diagonal entries

$$\det\begin{pmatrix} x_1 & 0 & \cdots & 0 \\ * & x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & x_n \end{pmatrix} = x_1x_2\cdots x_n \det\begin{pmatrix} 1 & 0 & \cdots & 0 \\ * & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & 1 \end{pmatrix}$$
$$= x_1x_2\cdots x_n \det\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = x_1x_2\cdots x_n.$$

The argument above assumes all $x_i \neq 0$. If some $x_i = 0$, then the last column in the column echelon form is the zero vector $\vec{0}$. By the linearity of the determinant in the last column, the determinant is 0. This shows that the equality always holds.

We note that the same argument also shows the same formula for the determinant of *upper triangular matrices*

$$\det\begin{pmatrix} x_1 & * & \cdots & * \\ 0 & x_2 & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_n \end{pmatrix} = x_1x_2\cdots x_n = \det\begin{pmatrix} x_1 & 0 & \cdots & 0 \\ * & x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & x_n \end{pmatrix}.$$

A square matrix is invertible if and only if all the diagonal entries in the column echelon form are pivots, i.e., all $x_i \neq 0$. This proves the following.

Theorem 5.4. *A square matrix is invertible if and only if its determinant is nonzero.*

In terms of a system of linear equations with equal number of equations and variables, this means that whether $A\vec{x} = \vec{b}$ has unique solution is "determined" by $\det A \neq 0$.

Row Operation

To find out the effect of row operation on the determinant, we use the idea for the proof of Proposition 5.3. The key is that a row operation preserves the multilinear and alternating property. In other words, suppose $A \mapsto \tilde{A}$ is a row operation, then $\det \tilde{A}$ is still a multilinear and alternating function of A.

By Theorem 5.2, therefore, we have $\det \tilde{A} = a \det A$ for a constant a. The constant a can be determined from the special case that $A = I$ is the identity matrix. For the operation $R_1 \leftrightarrow R_2$, we have

$$\tilde{I} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \quad a = \det \tilde{I} = -1.$$

For the operation $R_1 \leftrightarrow cR_1$, we have

$$\tilde{I} = \begin{pmatrix} c & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad a = \det \tilde{I} = c.$$

For the operation $R_1 + cR_2$, we have

$$\tilde{I} = \begin{pmatrix} 1 & c & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \quad a = \det \tilde{I} = 1.$$

The effect of the operation on other rows is similar.

Example 5.2. calculate by mixing row and column operations.....

Cofactor Expansion

The first column of an $n \times n$ matrix $A = (x_{ij}) = (\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_n)$ is

$$\vec{x}_1 = (x_{11} \ \cdots \ x_{n1})^T = x_{11}\vec{e}_1 \cdots + x_{n1}\vec{e}_n.$$

Using the linearity of $\det A$ in the first column, we get

$$\det A = x_{11}D_1(A) + \cdots + x_{n1}D_n(A), \quad D_i(A) = \det(\vec{e}_i \ \vec{x}_2 \ \cdots \ \vec{x}_n). \tag{5.1}$$

Then D_i is multilinear and alternating in the columns $\vec{x}_2, \dots, \vec{x}_n$ of A. Moreover, by the alternating property $\det(\vec{e}_i \ \cdots \ \vec{e}_i \ \cdots) = 0$, $D_i(A)$ depends only on the "non-i-th" coordinates of $\vec{x}_2, \dots, \vec{x}_n$. These coordinates form the $(n-1) \times (n-1)$ matrix A_{1i} obtained by deleting the first column and the i-th row of A. Then $D_i(A)$ is a multilinear and alternating function of the columns of the square matrix A_{1i}. By Theorem 5.2, we have

$$D_i(A) = c_i \det A_{1i} \text{ for a constant } c_i.$$

For the special case $A_{1i} = I$ is the identity matrix, we have

$$c_i = \det \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1_{(i)} & * & * & \cdots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = \det \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1_{(i)} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = (-1)^{i-1} \det I = (-1)^{i-1}.$$

Here the sign $(-1)^{i-1}$ is obtained by moving the first column to the i-th column after $i-1$ exchanges. We conclude the *cofactor expansion* formula

$$\det A = x_{11}\det A_{11} - x_{21}\det A_{21} + \cdots + (-1)^{n-1}x_{n1}\det A_{n1}.$$

We may carry out the same argument with respect to the i-th column of instead of the first one. Let A_{ij} be the matrix obtained by deleting the i-th row and j-th column from A. Then we get the cofactor expansions along the i-th column

$$\det A = (-1)^{1-i}x_{1i}\det A_{1i} + (-1)^{2-i}x_{2i}\det A_{2i} + \cdots + (-1)^{n-i}x_{ni}\det A_{ni}.$$

By $\det A^T = \det A$, we also have the cofactor expansion along the i-th row

$$\det A = (-1)^{i-1}x_{i1}\det A_{i1} + (-1)^{i-2}x_{i2}\det A_{i2} + \cdots + (-1)^{i-n}x_{in}\det A_{in}.$$

The cofactor expansions suggest the *adjugate matrix* of a square matrix

$$\mathrm{adj}(A) = \frac{1}{\det A}\begin{pmatrix} \det A_{11} & -\det A_{21} & \cdots & (-1)^{n-1}\det A_{n1} \\ -\det A_{12} & \det A_{22} & \cdots & (-1)^{n-2}\det A_{n2} \\ \vdots & \vdots & & \vdots \\ (-1)^{1-n}\det A_{1n} & (-1)^{2-n}\det A_{2n} & \cdots & \det A_{nn} \end{pmatrix}.$$

The cofactor expansion by rows is the same as [explain the off diagonal entries are 0 ??????]

$$A\,\mathrm{adj}(A) = (\det A)I.$$

For the case A is invertible, this gives an explicit formula for the inverse

$$A^{-1} = \frac{1}{\det A}\mathrm{adj}(A).$$

A consequence of the formula is the following explicit formula for the solution of $A\vec{x} = \vec{b}$.

Proposition 5.5 (Cramer's Rule). *If $A = (\vec{v}_1 \ \cdots \ \vec{v}_n)$ is an invertible matrix, then the solution of $A\vec{x} = \vec{b}$ is given by*

$$x_i = \frac{\det(\vec{v}_1 \ \cdots \ \overset{(i)}{\vec{b}} \ \cdots \ \vec{v}_n)}{\det A}.$$

Cramer's rule is not a practical way of calculating the solution for two reasons. The first is that it only applies to the case A is invertible. The second is that the row operation method in Example **??** is much more efficient.

Exercise 5.7. Use cofactor expansion to explain the determiant of upper and lower triangular matrices.

Exercise 5.8. The *Vandermonde matrix* is

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix}.$$

1. Find the determinant of the Vandermonde matrix.
2. Find the condition for the Vandermonde matrix to be invertible.
3. Interpret the invertible condition as the fact that a polynomial of degree $\leq n$ is determined by its values at $n+1$ distinct places.

Property of Determinant

We conclude that the effect of the row operation on the determinant is the same as the column operation. Moreover, the row operation can simplify a square matrix to an upper triangular matrix, and (**??**) shows that the determinant of lower and upper triangular matrices are calculated in the same way. Therefore we get the following result form the computational viewpoint.

Proposition 5.6. $\det A^T = \det A$.

Exercise 5.9. Prove that the determinant is the function that is multilinear and alternating on the row vectors, and satisfies $\det I = 1$.

Exercise 5.10. The matrices $\tilde{I}$ obtained by the row operations are the *elementary matrices*.

1. Prove that applying a row operation to A gives $\tilde{A} = \tilde{I}A$. Then prove the effect of the row operation by using Proposition 5.3 and direct computation of $\det \tilde{I}$.
2. Prove that applying a column operation to A gives $A\tilde{I}$. Then prove the effect of the column operation by using Proposition 5.3 and direct computation of $\det \tilde{I}$.

Exercise 5.11. Prove that any orthogonal matrix has determinant ± 1.

Exercise 5.12. What is $\det cA$?

Exercise 5.13. What is $\det \bar{A}$?

Exercise 5.14. Suppose A and B are square matrices. Suppose O is the zero matrix. Prove that

$$\det \begin{pmatrix} A & * \\ O & B \end{pmatrix} = \det A \det B = \det \begin{pmatrix} A & 0 \\ * & B \end{pmatrix}$$

Hint: Consider the multilinear and alternating property on columns of A and rows of B.

Exercise 5.15. Let r be the rank of a matrix A.

1. Prove that there is an $r \times r$ submatrix B of A, such that $\det B \neq 0$.
2. Prove that for any $s > r$ and any $s \times s$ submatrix B of A, we have $\det B = 0$.

Some textbooks use the two properties to define the rank of a matrix.

Determinant of Linear Operator

For a linear transformation $L \colon V \to W$ and bases α, β of V, W, we have the matrix $[L]_{\beta\alpha}$ of the linear transformation with respect to the two bases. For the matrix to be square, we need $\dim V = \dim W$. Then we can certainly define $\det[L]_{\beta\alpha}$ to be the "determinant with respect to the bases". The problem is how much is the determinant dependent on the basis.

Suppose $L \colon V \to V$ is a linear transformation from a vector space to itself. Then we can choose $\alpha = \beta$ and consider $\det[L]_{\alpha\alpha}$. If we choose another basis of V, then by (**??**), the matrix of L with respect to the other basis is $P[L]_{\alpha\alpha}P^{-1}$, where P is the matrix between the two bases. Further by Proposition 5.3, we have

$$\det(P[L]_{\alpha\alpha}P^{-1}) = (\det P)(\det[L]_{\alpha\alpha})(\det P)^{-1} = \det[L]_{\alpha\alpha}.$$

This shows that the determinant of a linear transformation from a vector space to itself is well defined (i.e., independent of the choice of basis).

Exercise 5.16. Prove that $\det(L \circ L') = \det L \det L'$, $\det L^{-1} = \frac{1}{\det L}$.

Suppose $L \colon V \to W$ is a linear transformation between oriented inner product spaces. Then we may consider $[L]_{\beta\alpha}$ for oriented orthonormal bases α and β. Since the matrices between oriented orthonormal bases of the same oriented inner product space has determinant 1, the change of basis formula and Proposition 5.3 imply that $\det[L]_{\beta\alpha}$ is independent of the choice of the bases.

Exercise 5.17. Prove that $\det L^* = \det L$.

The two cases of well defined determinant of linear transformations suggests a common and deeper concept of determinant of linear transformation between vector spaces of the same dimension. The concept will be clarified in the theory of exterior algebra.

5.2 Geometry

Orientation

Let $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $\beta = \{\vec{w}_1, \dots, \vec{w}_n\}$ be ordered bases of a vector space V. We say α and β are *compatibly oriented* if there is $\alpha(t) = \{\vec{v}_1(t), \dots, \vec{v}_n(t)\}$, $t \in [0, 1]$, such that

1. $\alpha(t)$ is continuous, i.e., each $\vec{v}_i(t)$ is a continuous function of t.

2. $\alpha(t)$ is a basis for each t.

3. $\alpha(0) = \alpha$ and $\alpha(1) = \beta$.

In other words, α and β are connected by a continuous family of ordered bases.

We can imagine $\alpha(t)$ to be a "movie" that starts from α and ends at β. If $\alpha(t)$ connects α to β, then the reverse movie $\alpha(1-t)$ connects β to α. If $\alpha(t)$ connects α to β, and $\beta(t)$ connects β to γ, then we may stitch the two movies together and get a movie

$$\begin{cases} \alpha(2t) & \text{for } t \in [0, \frac{1}{2}], \\ \beta(2t-1) & \text{for } t \in [\frac{1}{2}, 1], \end{cases}$$

that connects α to γ. This shows that the compatible orientation is an equivalence relation.

In $\mathbb{R}^1$, a basis is a nonzero number. If two nonzero numbers a and b have the same sign, then $a(t) = (1-t)a + tb$, $t \in [0,1]$, is never zero. This shows that a and b are compatibly oriented. Conversely, if $a(t)$ is a continuous function satisfying $a = a(0), b = a(1)$ and $a(t) \neq 0$ for all $t \in [0,1]$, then a and b have the same sign. We conclude that two bases of $\mathbb{R}^1$ are compatibly oriented if and only if they have the same sign.

In $\mathbb{R}^2$, a basis is a pair of non-parallel vectors $\vec{u}$ and $\vec{v}$. We may first rotate the whole plane so that $\vec{u}$ becomes $\vec{u}' = (a\ 0)^T$ with $a > 0$, and $\vec{v}$ becomes $\vec{v}' = (c\ d)^T$ with $d \neq 0$. Then $\vec{u}'(t) = \vec{u}' = (a\ 0)^T$ and $\vec{v}'(t) = ((1-t)c\ d)^T$ (fixing $\vec{u}'$ and sliding $\vec{v}'$) move to $\vec{u}'' = (a\ 0)^T$ and $\vec{v}'' = (0\ d)^T$. Next like in $\mathbb{R}^1$, $\vec{u}'' = ((1-t)a + t\ 0)^T$ and $\vec{v}'' = (0\ (1-t)d \pm t)^T$ ($\pm$ is the aign of d) move to $\vec{e}_1 = (1\ 0)^T$ and $\pm\vec{e}_2 = \pm(0\ 1)^T$. Stitching the three moves together, we see that any ordered basis in $\mathbb{R}^2$ is either compatibly oriented as $\{\vec{e}_1, \vec{e}_2\}$, or compatibly oriented as $\{\vec{e}_1, -\vec{e}_2\}$. The first case happens when moving from $\vec{u}$ to $\vec{v}$ is "counterclockwise", and the second happens when moving from $\vec{u}$ to $\vec{v}$ is "clockwise".

[picture of movie in $\mathbb{R}^2$]

Proposition 5.7. *Two ordered bases α, β are compatibly oriented if and only if* $\det[I]_{\beta\alpha} > 0$.

Proof. Suppose α, β are compatibly oriented. Then there is a continuous family $\alpha(t)$ of ordered bases connecting the two. The function $f(t) = \det[I]_{\alpha(t)\alpha}$ is a continuous function satisfying $f(0) = \det[I]_{\alpha\alpha} = 1$ and $f(t) \neq 0$ for all $t \in [0,1]$. By the intermediate value theorem, f never changes sign, so that $\det[I]_{\beta\alpha} = \det[I]_{\alpha(1)\alpha} = f(1) > 0$. This proves the necessity.

For the sufficiency, suppose $\det[I]_{\beta\alpha} > 0$. Let us study the operations on sets of vectors in Proposition 3.3 that preserve the orientation. We will describe the operations on $\vec{u}$ and $\vec{v}$, which are regarded as $\vec{v}_i$ and $\vec{v}_j$. The other vectors in the set are understood as being fixed.

1. The sets $\{\vec{u}, \vec{v}\}$ and $\{\vec{u} + c\vec{v}, \vec{v}\}$ are connected by the sliding $\{\vec{u} + tc\vec{v}, \vec{v}\}$. Therefore the third operation preserves the orientation.

2. The sets $\{\vec{u}, \vec{v}\}$ and $\{\vec{v}, -\vec{u}\}$ are connected by the "90° rotation" $\{\cos\frac{t\pi}{2}\vec{u} + \sin\frac{t\pi}{2}\vec{v}, -\sin\frac{t\pi}{2}\vec{u} + \cos\frac{t\pi}{2}\vec{v}\}$ (the quotation mark indicates that we only pretend $\{\vec{u}, \vec{v}\}$ to be orthonormal). Therefore the first operation plus a sign change preserves the orientation. Combining two such rotations also shows that $\{\vec{u}, \vec{v}\}$ and $\{-\vec{u}, -\vec{v}\}$ are connected by "180° rotation" and are therefore compatibly oriented.

3. For any $a > 0$, $\vec{v}$ and $a\vec{v}$ are connected by $(1 - t + ta)\vec{v}$. Therefore the second operation with positive scalar preserves the orientation.

By these operations, any ordered basis can be modified to become certain "reduced column echelon set". If $V = \mathbb{R}^n$, then the "reduced column echelon set" is either $\{\vec{e}_1, \dots, \vec{e}_{n-1}, \vec{e}_n\}$ or $\{\vec{e}_1, \dots, \vec{e}_{n-1}, -\vec{e}_n\}$. In general, using the β-coordinate isomorphism $V \cong \mathbb{R}^n$ to translate into Euclidean space, we can use a sequence of the operations above to modify α to either $\beta = \{\vec{w}_1, \dots, \vec{w}_{n-1}, \vec{w}_n\}$ or $\beta' = \{\vec{w}_1, \dots, \vec{w}_{n-1}, -\vec{w}_n\}$. Correspondingly, we have a continuous family $\alpha(t)$ of ordered bases connecting α to β or β'.

By the just proved necessity, we have $\det[I]_{\alpha(1)\alpha} > 0$. On the other hand, the assumption $\det[I]_{\beta\alpha} > 0$ implies $\det[I]_{\beta'\alpha} = \det[I]_{\beta'\beta}\det[I]_{\beta\alpha} = -\det[I]_{\beta\alpha} < 0$. Therefore $\alpha(1) \neq \beta'$, and we must have $\alpha(1) = \beta$. This proves that α and β are compatibly oriented. □

Proposition 5.7 shows that the orientation compatibility gives exactly two equivalence classes. We denote the equivalence class represented by α by

$$o_\alpha = \{\beta\colon \alpha, \beta \text{ compatibly oriented}\} = \{\beta\colon\ \det[I]_{\beta\alpha} > 0\}.$$

We also denote the other equivalence class by

$$-o_\alpha = \{\beta\colon \alpha, \beta \text{ incompatibly oriented}\} = \{\beta\colon\ \det[I]_{\beta\alpha} < 0\}.$$

In general, the set of all ordered bases is the disjoint union $o \cup o'$ of two equivalence classes. The choice of one from o or o' specifies an *orientation* of the vector space. In other words, an *oriented* vector space is a vector space equipped with a preferred choice of the equivalence class. An ordered basis α of an oriented vector space is *positively oriented* if α belongs to the preferred equivalence class, and is otherwise *negatively oriented*.

The *standard (positive) orientation* of $\mathbb{R}^n$ is the equivalence class represented by the standard basis $\{\vec{e}_1, \dots, \vec{e}_{n-1}, \vec{e}_n\}$. The standard negative orientation is then represented by $\{\vec{e}_1, \dots, \vec{e}_{n-1}, -\vec{e}_n\}$.

Volume

Proposition 5.8. *The absolute value of the determinant of $A = (\vec{x}_1 \ \cdots \ \vec{x}_n)$ is the volume of the parallelotope spanned by column vectors*

$$P(A) = \{c_1\vec{x}_1 + \cdots + c_n\vec{x}_n\colon 0 \le c_i \le 1\}.$$

Suppose A is invertible. Let $\epsilon = \{\vec{e}_1, \dots, \vec{e}_n\}$ be the standard basis and let $\alpha = \{\vec{x}_1 \ \cdots \ \vec{x}_n\}$ be the column basis of A. Then $A = [I]_{\alpha\epsilon}$, and by Proposition 5.7, we have $\det A > 0$ or < 0 according to whether the columns of A is positively or negatively oriented. If A is not invertible, then by Propostion 5.4, we have $\det A = 0$. Combined with Proposition 5.8, we find that $\det A$ is the "signed volume" of the parallelotope $P(A)$.

Proof. We show that the effects of a column operation on the volume of $P(A)$ and on $|\det A|$ are the same. We illustrate the idea by only looking at the operations on the first two columns, similar to the proof of Proposition 5.6.

The operation $C_1 \leftrightarrow C_2$ is

$$A = (\vec{x}_1 \; \vec{x}_2 \; \cdots) \mapsto \tilde{A} = (\vec{x}_2 \; \vec{x}_1 \; \cdots).$$

Since $P(\tilde{A}) = P(A)$, we get $|V(\tilde{A})| = |V(A)|$. This is the same as $|\det \tilde{A}| = |-\det A| = |\det A|$.

The operation $C_1 \to aC_2$ is

$$A = (\vec{x}_1 \; \cdots) \mapsto \tilde{A} = (a\vec{x}_1 \; \cdots).$$

Since $P(\tilde{A})$ is obtained by stretching $P(A)$ by factor of $|a|$ in the $\vec{x}_1$ direction, the volume of $P(\tilde{A})$ is $|a|$ times the volume of $P(A)$. This is the same as $|\det \tilde{A}| = |a \det A| = |a||\det A|$.

The operation $C_1 \leftrightarrow C_1 + aC_2$ is

$$A = (\vec{x}_1 \; \vec{x}_2 \; \cdots) \mapsto \tilde{A} = (\vec{x}_1 + a\vec{x}_2 \; \vec{x}_2 \; \cdots).$$

Since $P(\tilde{A})$ is obtained by "shearing" the paralleogram spanned by $\vec{x}_1$ and $\vec{x}_2$ in the direction of $\vec{x}_2$, $P(\tilde{A})$ and $P(A)$ have the same volume. This is the same as $|\det \tilde{A}| = |\det A|$.

If A is invertible, then the column operation reduces A to the identity matrix I. The volume of $P(I)$ is 1 and we also have $\det I = 1$. If A is not invertible, then $P(A)$ is degenerate and has volume 0. We also have $\det A = 0$ by Proposition 5.4. This completes the proof. □

.... The geometry of determinant suggests the following alternative definition of the determinant. The determinant is the function defined on all square matrices, satisfying

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

Another definition motivated by geometry is the function defined on all square matrices, satisfying

$$\det(AB) = \det A \det B, \quad \det I = 1.$$

6 Advanced Vector Space

A vector space is characterised by addition $V \times V \to V$ and scalar multiplication $\mathbb{R} \times V \to V$. The addition is and *internal* operation, and the scalar multiplication uses $\mathbb{R}$ which is *external* to V. Therefore the concept of vector space can be understood in two steps. For we single out the internal structure of addition.

Definition 6.1. A *abelian group* is a set V, together with the operations of addition

$$u + v \colon V \times V \to V,$$

such that the following are satisfied.

1. Commutativity: $u + v = v + u$.

2. Associativity: $(u + v) + w = u + (v + w)$.

3. Zero: There is an element $0 \in V$ satisfying $u + 0 = u = 0 + u$.

4. Negative: For any u, there is v (to be denoted $-u$), such that $u + v = 0 = v + u$.

On top of the concept of abelian group, we add scalar multiplication. The scalar does not have to be real numbers. In general, the scalars are the elements of a *ring* R.

Definition 6.2. A *module* over a ring R (called R-module) is an abelian group V, together with the operations of scalar multiplication

$$av \colon R \times V \to V,$$

such that the following are satisfied.

1. One: $1u = u$.

2. Associativity: $(ab)u = a(bu)$.

3. Distributivity in R: $(a + b)u = au + bu$.

4. Distributivity in V: $a(u + v) = au + au$.

From the axioms, we see that R must have 1, addition, and scalar multiplication. In order to prove some basis results such as Propositions 1.2, 1.3, 1.4, we find it necessary to also require R to have 0 and subtraction. Therefor a ring is a system with $+, -, \times$ operations, satisfying the usual properties.

Definition 6.3. A *ring* is a set R together with the operations of addition and multiplication

$$a + b \colon R \times R \to R, \quad ab \colon R \times R \to R,$$

such that the following are satisfied.

1. Commutativity in $+$: $a + b = b + a$.

2. Associativity in $+$: $(a + b) + c = a + (b + c)$.

3. Zero: There is an element $0 \in R$ satisfying $a + 0 = a = 0 + a$.

4. Negative: For any a, there is b (to be denoted $-a$), such that $a + b = 0 = b + a$.

5. Associativity in $\times$: $(ab)c = a(bc)$ satisfying $a1 = a = 1a$.

6. Distributivity: $(a + b)c = ac + bc$, $a(b + c) = ab + ac$.

The first four axions basically says that $(R, +)$ (i.e., ignoring $\times$) is an abelian group. We note that the multiplication is not necessarily commutative, and there may not be division $\div$.

6.1 Module over Ring

6.2 Abelian Group

6.3 Polynomial

6.4 Field and Complex Number

The key structures in a vector space are addition and scalar multiplication. In the theory of (real) vector spaces we developed so far, only the four arithmetic operations of real numbers are used until the introduction of inner product. If we regard any system with four arithmetic operations as "generalised numbers", then the similar linear algebra theory without inner product can be developed over such generalised numbers. The relation between $\mathbb{R}$ and generalised numbers is analogous to the relation between the Euclidean spaces and general vector spaces.

Definition 6.4. A *field* is a set with arithmetic operations $+, -, \times, \div$ satisfying the usual properties.

In fact, only two operations $+, \times$ are required in a field, and $-, \div$ are regarded as the "opposite" or "inverse" of the two operations. Besides real numbers, the other examples of fields are the rational numbers $\mathbb{Q}$, the complex numbers $\mathbb{C}$.

Example 6.1. The field of $\sqrt{2}$-rational numbers is $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \colon a, b \in \mathbb{Q}\}$. The four arithmetic operations are

$$\begin{aligned}
(a+b\sqrt{2}) + (c+d\sqrt{2}) &= (a+c) + (b+d)\sqrt{2},\\
(a+b\sqrt{2}) - (c+d\sqrt{2}) &= (a-c) + (b-d)\sqrt{2},\\
(a+b\sqrt{2})(c+d\sqrt{2}) &= (ac+2bd) + (ad+bc)\sqrt{2},\\
\frac{a+b\sqrt{2}}{c+d\sqrt{2}} &= \frac{(a+b\sqrt{2})(c-d\sqrt{2})}{(c+d\sqrt{2})(c-d\sqrt{2})} = \frac{ac-2bd}{c^2-2d^2} + \frac{-ad+bc}{c^2-2d^2}\sqrt{2}.
\end{aligned}$$

The field is a *subfield* of $\mathbb{R}$, just like a subsapce.

Example 6.2. The field of integers modulo a prime number p is

$$\mathbb{Z}_p = \{\bar{0}, \bar{1}, \dots, \overline{p-1}\}.$$

The addition and multiplication are the obvious operations

$$\bar{m} + \bar{n} = \overline{m+n}, \quad \bar{m}\bar{n} = \overline{mn}.$$

The two operations satisfy the usual properties, and the addition has the opposite operation of subtraction $\bar{m} - \bar{n} = \overline{m-n}$.

The only place that requires p to be a prime number is the division. The quotient $\frac{\bar{m}}{\bar{n}} = \bar{q}$ means finding integer q satisfying $\overline{nq} = \bar{m}$. The existence of such q means that the map of multiplying $\bar{n}$ is onto

$$\mathbb{Z}_p \to \mathbb{Z}_p, \quad \bar{x} \mapsto \bar{n}\bar{x} = \overline{nx}.$$

Since $\mathbb{Z}_p$ is finite, the onto property is the same as one-to-one. Since the map is "linear", the one-to-one property is equivalent to the trivial kernel property

$$\overline{nx} = \bar{0} \implies \bar{x} = \bar{0}.$$

The following is the proof of trivial kernel implying one-to-one

$$\begin{aligned}
\overline{nx} = \overline{ny} &\implies \overline{n(x-y)} = \overline{nx - ny} = \overline{nx} - \overline{ny} = \bar{0} \\
&\implies \overline{x-y} = \bar{0} \quad \text{(trivial kernel)} \\
&\implies \bar{x} - \bar{y} = \overline{x-y} = \bar{0}.
\end{aligned}$$

It remains to prove the trivial kernel property under the assumption that p is prime and $\bar{n} \neq \bar{0}$. The assumption $\bar{n} \neq \bar{0}$ means that n is not divisible by p. If $\bar{x} \neq \bar{0}$, then x is also not divisible by p. Since p is prime, we conclude that nx is not divisible by p. This proves

$$\bar{x} \neq \bar{0} \implies \overline{nx} \neq \bar{0}.$$

The statement is equivalent to the trivial kernel property.

In practice, the division is calculated by the *Euclidean algorithm*. The algorithm gives an alternative calculational proof of the existence of division.

Exercise 6.1. Prove that the conjugation $a+b\sqrt{2} \mapsto a-b\sqrt{2}$ is an operator on $\mathbb{Q}[\sqrt{2}]$ preserving the four arithmetic operations.

Exercise 6.2. Suppose n is a non-square natural number. Prove that $\mathbb{Q}[\sqrt{n}] = \{a+b\sqrt{n} \colon a, b \in \mathbb{Q}\}$ is a field.

Exercise 6.3. Suppose $r^3 + c_2t^2 + c_1t + c_0 = 0$ for some rational numbers c_0, c_1, c_2. Let

$$\mathbb{Q}[r] = \{a_nr^n + a_{n-1}r^{n-1} + \cdots + a_1r + a_0 \colon a_i \in \mathbb{Q}\}$$

be the set of all rational polynomials of r.

1. Prove that $\mathbb{Q}[r]$ is a $\mathbb{Q}$-vector space spanned by $r^2, r, 1$.
2. Prove that for any $x \in \mathbb{Q}[r]$, the "vectors" $x^3, x^2, x, 1$ are linearly dependent over $\mathbb{Q}$.
3. Prove that if $x \in \mathbb{Q}[r]$ is nonzero, then there is $y \in \mathbb{Q}[r]$ satisfying $xy = 1$.
4. Prove that $\mathbb{Q}[r]$ is a field.

6.5 Complex Number

A complex number is of the form $a+ib$, with $a, b \in \mathbb{R}$ and $i=\sqrt{-1}$. The square root of -1 means that $i^2=-1$. A complex number has the *real* and *imaginary* parts

$$\operatorname{Re}(a+ib)=a, \quad \operatorname{Im}(a+ib)=b.$$

The addition and multiplications of complex numbers are

$$(a+ib)+(c+id)=(a+c)+i(b+d), \quad (a+ib)(c+id)=(ac-bd)+i(ad+bc).$$

It can be easily verified that the operations satisfy the usual properties of arithmetic operations. In particular, the subtraction is

$$(a+ib)-(c+id)=(a-c)+i(b-d),$$

and the division is

$$\frac{a+ib}{c+id}=\frac{(a+ib)(c-id)}{(c+id)(c-id)}=\frac{(ac+bd)+i(-ad+bc)}{c^2+d^2}.$$

A complex number can be regarded as a vector in the 2-dimensional Euclidean space $\mathbb{R}^2$, with the real part as the x-coordinate and the imaginary part as the y-coordinate. The vector has length r and angle θ (i.e., polar coordinates), and we have

$$a+ib=r\cos\theta+ir\sin\theta=re^{i\theta}.$$

The first equality is simple trigonometry, and the second equality uses the expansion (the theoretical explanation is the complex analytic continuation of the exponential function of real numbers)

$$\begin{aligned}
e^{i\theta} &= 1+\frac{1}{1!}i\theta+\frac{1}{2!}(i\theta)^2+\cdots+\frac{1}{n!}(i\theta)^n+\cdots \\
&= \left(1-\frac{1}{2!}\theta^2+\frac{1}{4!}\theta^4+\cdots\right)+i\left(\theta-\frac{1}{3!}\theta^3+\frac{1}{5!}\theta^5+\cdots\right) \\
&= \cos\theta+i\sin\theta.
\end{aligned}$$

The complex exponential has the usual properties of the real exponential (because of complex analytic continuation), and we can easily get the multiplication and division of complex numbers

$$(re^{i\theta})(r'e^{i\theta'})=rr'e^{i(\theta+\theta')}, \quad \frac{re^{i\theta}}{r'e^{i\theta'}}=\frac{r}{r'}e^{i(\theta-\theta')}.$$

Therefore multiplying $re^{i\theta}$ means stretching by r and rotating by θ.

The *complex conjugation* $\overline{a+bi}=a-bi$ is an *automorphism* (self-isomorphism) of $\mathbb{C}$. This means that the conjugation preserves the four arithmetic operations

$$\overline{z_1+z_2}=\bar{z}_1+\bar{z}_2, \quad \overline{z_1-z_2}=\bar{z}_1-\bar{z}_2, \quad \overline{z_1z_2}=\bar{z}_1\bar{z}_2, \quad \overline{\left(\frac{z_1}{z_2}\right)}=\frac{\bar{z}_1}{\bar{z}_2}.$$

Geometrically, the conjugation means flipping with respect to the x-axis. This gives the conjugation in polar coordinates

$$\overline{re^{i\theta}} = re^{-i\theta}.$$

The length can also be expressed in terms of the conjugation

$$z\bar{z} = a^2 + b^2 = r^2, \quad |z| = r = \sqrt{z\bar{z}}.$$

This suggests that the positivity of the inner product can be extended to complex vector spaces, as long as we modify the inner product by using the complex conjugation.

A major difference between $\mathbb{R}$ and $\mathbb{C}$ is that the polynomial $t^2 + 1$ has no root in $\mathbb{R}$ but has a pair of roots $\pm i$ in $\mathbb{C}$. In fact, complex numbers has the following so called *algebraically closed* property.

Theorem 6.5 (Fundamental Theorem of Algebra). *Any non-constant complex polynomial has roots.*

The real number is not algebraically closed.

Exercise 6.4. A real vector in $\mathbb{R}^n$ can also be regarded as a complex vector in $\mathbb{C}^n$. Prove that a set of real vectors spans $\mathbb{R}^n$ if and only if it spans $\mathbb{C}^n$ as complex vectors. Prove that the set is linearly independent over $\mathbb{R}$ if and only if it is linearly independent over $\mathbb{C}$.

Exercise 6.5. Define the conjugation of a complex vector in $\mathbb{C}^n$ by taking the complex conjugation of each coordinate

$$\vec{v} = (z_1 \ \cdots \ z_n)^T \implies \bar{\vec{v}} = (\bar{z}_1 \ \cdots \ \bar{z}_n)^T.$$

For a complex subspace $H \subset \mathbb{C}^n$, prove that the conjugation of all vectors in H

$$\bar{H} = \{\bar{\vec{v}} \colon \vec{v} \in H\}$$

is still a complex subspace.

Exercise 6.6. For a set of complex vectors α, let $\bar{\alpha}$ be the set of conjugations of vectors in α.

1. Prove that $\text{Span}\bar{\alpha} = \overline{\text{Span}\alpha}$.

2. Prove that α is linearly independent if and only if $\bar{\alpha}$ is linearly independent.

3. Prove that α is a basis of V if and only if $\bar{\alpha}$ is a basis of V.

Exercise 6.7. The conjugation $\bar{A}$ of a matrix A is obtained by taking the complex conjugation of all entries in A. Prove that

$$\text{Ran}\bar{A} = \overline{\text{Ran}A}, \quad \text{Ker}\bar{A} = \overline{\text{Ker}A}, \quad \text{rank}\bar{A} = \text{rank}A.$$

Exercise 6.8. A map $L\colon V \to W$ is *conjugate linear* if $L(\vec{u}+\vec{v}) = L(\vec{u})+L(\vec{v})$ and $L(a\vec{v}) = \bar{a}L(\vec{v})$.

1. Prove that $\vec{v} \mapsto \bar{\vec{v}}\colon \mathbb{C}^n \to \mathbb{C}^n$ conjugate linear isomorphism.

2. Prove that any finite dimensional complex vector space has a conjugate linear operator.

3. Prove that any conjugate linear transformation $V \to W$ is the composition of a linear transformation $V \to W$ and a conjugate linear operator $V \to V$.

4. Prove that any conjugate linear transformation $V \to W$ is the composition of a linear transformation $V \to W$ and a conjugate linear operator $W \to W$.

5. What is the composition of two conjugate linear transformations?

6.6 Complex Inner Product

For complex inner product spaces, we cannot directly use the formula for the dot product of real Euclidean spaces because of the failure of positivity. The *dot product* on the complex Euclidean space $\mathbb{C}^n$ should be

$$(x_1 \ \cdots \ x_n)^T \cdot (y_1 \ \cdots \ y_n)^T = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n = \vec{x}^T\bar{\vec{y}}.$$

Inspired by this, a *(Hermitian) inner product* on a complex vector space V is a map

$$\langle \vec{u}, \vec{v}\rangle \colon V \times V \to \mathbb{C},$$

satisfying the following properties.

1. Conjugate symmetric: $\langle \vec{v}, \vec{u}\rangle = \overline{\langle \vec{u}, \vec{v}\rangle}$.

2. Sesquilinear: $\langle a\vec{u} + b\vec{u}', \vec{v}\rangle = a\langle \vec{u}, \vec{v}\rangle + b\langle \vec{u}', \vec{v}\rangle$, $\langle \vec{u}, a\vec{v} + b\vec{v}\rangle = \bar{a}\langle \vec{u}, \vec{v}\rangle + \bar{b}\langle \vec{u}, \vec{v}'\rangle$.

3. Positive: $\langle \vec{u}, \vec{u}\rangle \geq 0$ and $\langle \vec{u}, \vec{u}\rangle = 0$ if and only if $\vec{u} = \vec{0}$.

The sesquilinear property is the linearity in the first vector and the *conjugate linearity* in the second vector. The conjugate symmetry and the linearity in the first vector forces the conjugate linearity in the second vector

$$\begin{aligned}
\langle \vec{u}, a\vec{v} + b\vec{v}\rangle &= \overline{\langle a\vec{v} + b\vec{v}, \vec{u}\rangle} && \text{(conjugate symmetry)} \\
&= \overline{a\langle \vec{v}, \vec{u}\rangle + b\langle \vec{v}', \vec{u}\rangle} && \text{(linearity in the first vector)} \\
&= \bar{a}\overline{\langle \vec{v}, \vec{u}\rangle} + \bar{b}\overline{\langle \vec{v}', \vec{u}\rangle} \\
&= \bar{a}\langle \vec{u}, \vec{v}\rangle + \bar{b}\langle \vec{u}, \vec{v}'\rangle. && \text{(conjugate symmetry)}
\end{aligned}$$

The length of a vector is still $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v}\rangle}$. Due to the complex value of the inner product, the angle between nonzero vectors is not defined, and the area is not defined. The Cauchy-Schwarz inequality still holds, so that the length still satisfies the three properties.

Exercise 6.9. Prove the polarisation identity in the complex inner product space (compare Exercise 4.14)

$$\langle \vec{u}, \vec{v} \rangle = \frac{1}{4}(\|\vec{u}+\vec{v}\|^2 - \|\vec{u}-\vec{v}\|^2 + i\|\vec{u}+i\vec{v}\|^2 - i\|\vec{u}-i\vec{v}\|^2).$$

Exercise 6.10. Prove the parallelogram identity in the complex inner product space (compare Exercise 4.15)

$$\|\vec{u}+\vec{v}\|^2 + \|\vec{u}-\vec{v}\|^2 = 2(\|\vec{u}\|^2 + \|\vec{v}\|^2).$$

Exercise 6.11. Prove the complex version of the Cauchy-Schwarz inequality.

The orthogonality $\vec{u} \perp \vec{v}$ is still defined by $\langle \vec{u}, \vec{v} \rangle = 0$, and hence the concepts of orthogonal set, orthonormal set, orthonormal basis, orthogonal complement, orthogonal projection, etc. Due to conjugate linearity in the second variable, one needs to be careful with the order of vectors in formula. For example, the formula for orthogonal projeciton is given by (**??**), with $\vec{v}$ appearing as the first vector in inner products.

The Gram-Schmidt process still works, with the formula for the real Gram-Schmidt process still valid (one needs to be careful with the order of vectors in the inner product due to conjugate symmetry). Consequently, any finite dimensional complex inner product space is isometrically isomorphic to the complex Euclidean space with dot product.

Example 6.3. Gram-Schmidt

A linear transformation $L\colon V \to W$ between complex inner product spaces has the *(Hermitian) adjoint* $L^*\colon W \to V$ defined by

$$\langle L(\vec{v}), \vec{w} \rangle = \langle \vec{v}, L^*(\vec{w}) \rangle \text{ for all } \vec{v} \in V, \vec{w} \in W.$$

The adjoint is still complex linear, and the complex version of the complementarity principle (**??**) is still valid.

The argument about the relation between the matrices of L and L^* with respect to orthonormal bases $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ and $\beta = \{\vec{w}_1, \dots, \vec{w}_m\}$ of V and W is mostly the same as before, except

$$\langle L(\vec{v}_i), \vec{w}_j \rangle = \langle \vec{v}_i, L^*(\vec{w}_j) \rangle = \overline{\langle L^*(\vec{w}_j), \vec{v}_i \rangle}.$$

This implies that the two matrices are related by *conjugate transpose*

$$[L^*]_{\alpha\beta} = [L]^*_{\beta\alpha}, \quad A^* = \bar{A}^T. \tag{6.1}$$

For the special case of dot products of complex Euclidean spaces and the standard bases, this can be verified directly

$$A\vec{x} \cdot \vec{y} = (A\vec{x})^T \overline{\vec{y}} = \vec{x}^T A^T \overline{\vec{y}} = \vec{x}^T \overline{A^* \vec{y}} = \vec{x} \cdot A^* \vec{y}.$$

Exercise 6.12. Prove that $(\text{Ran}A)^\perp = \text{Ker}A^*$. This is need to prove $(H^\perp)^\perp = H$ for complex subspace.

Exercise 6.13. Prove the the complex version of the complementarity principle (**??**)

$$(\mathrm{Ran}L)^\perp = \mathrm{Ker}L^*, \quad (\mathrm{Ran}L^*)^\perp = \mathrm{Ker}L, \quad (\mathrm{Ker}L)^\perp = \mathrm{Ran}L^*, \quad (\mathrm{Ker}L^*)^\perp = \mathrm{Ran}L.$$

Exercise 6.14. Establish the complex version of Exercise 4.28. Prove that for a linear operator L on a complex inner product space, $\langle L(\vec{v}), \vec{v}\rangle = 0$ for all $\vec{v}$ if and only if $L = 0$.

Exercise 6.15. Prove that the columns of a complex matrix A form an orthogonal set with respect tot he dot product if and only if A^*A is diagonal. Moreover, the columns form an orthonormal set if and only if $A^*A = I$.

Exercise 6.16. Prove that the following are equivalent for a linear transformation $L\colon V \to W$.

1. L is an isometry.

2. L preserves length.

3. L takes an orthonormal basis to an orthonormal set.

4. $L^*L = I$.

5. The matrix A of L with respect to orthonormal bases of V and W satisfies $A^*A = I$.

A square complex matrix U is called a *unitary matrix* if it satisfies $U^*U = I$. A real unitary matrix is an orthogonal matrix. A unitary matrix is always invertible with $U^{-1} = U^*$. Unitary matrices are precisely the matrices of isometric isomorphisms with respect to orthonormal bases.

6.7 Complex vs Real Structure

By restricting the scalar multiplication by complex numbers to real numbers only, a complex vector space becomes a real vector space. Conversely, for a real vector space V to be the restriction of a complex vector space, we need to add the scalar multiplication by i, which is a special real linear operator J of V satisfying $J^2 = -I$. Given such an operator, we define the complex multiplication by

$$(a + ib)\vec{v} = a\vec{v} + bJ(\vec{v}).$$

Then we can verify the axioms of the complex vector space. For example,

$$\begin{aligned}
(a+ib)((c+id)\vec{v}) &= (a+ib)(c\vec{v} + dJ(\vec{v})) = a(c\vec{v} + dJ(\vec{v})) + bJ(c\vec{v} + dJ(\vec{v})) \\
&= ac\vec{v} + adJ(\vec{v}) + bcJ(\vec{v}) + bdJ^2(\vec{v})) = (ac - bd)\vec{v} + (ad + bc)J(\vec{v}), \\
((a+ib)(c+id))\vec{v} &= ((ac - bd) + i(ad + bc))\vec{v} = (ac - bd)\vec{v} + (ad + bc)J(\vec{v}).
\end{aligned}$$

Therefore the operator J is a *complex structure* on the real space.

Proposition 6.6. *A complex vector space is a real vector space equipped with a linear operator J satisfying $J^2 = -I$.*

If $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ is a complex basis of V, then $\alpha \cup i\alpha = \{\vec{v}_1, \dots, \vec{v}_n, i\vec{v}_1, \dots, i\vec{v}_n\}$ is a real basis of V. In other words, we have

$$V = W \oplus iW, \quad W = \text{Span}_{\mathbb{R}}\alpha = \mathbb{R}\vec{v}_1 \oplus \cdots \oplus \mathbb{R}\vec{v}_n, \quad iW = \text{Span}_{\mathbb{R}} i\alpha = \mathbb{R}i\vec{v}_1 \oplus \cdots \oplus \mathbb{R}i\vec{v}_n.$$

In particular, we have $\dim_{\mathbb{R}} V = 2\dim_{\mathbb{C}} V$, which is consistent with our intuition that $\mathbb{C}$ is a real vector space of real dimension 2.

In general, for any real vector space W, we may construct its *complexification* $V = W \oplus iW$. Here iW is a copy of W in which a vector $\vec{w} \in W$ is denoted $i\vec{w}$. Then V becomes a complex vector space by

$$\begin{aligned}(\vec{w}_1 + i\vec{w}_2) + (\vec{w}_1' + i\vec{w}_2') &= (\vec{w}_1 + \vec{w}_1') + i(\vec{w}_2 + \vec{w}_2'),\\ (a + ib)(\vec{w}_1 + i\vec{w}_2) &= (a\vec{w}_1 - b\vec{w}_2) + i(b\vec{w}_1 + a\vec{w}_2).\end{aligned}$$

We see that a complex vector space is always a complexification of a real subspace. However, the real subspace is not unique, because it depends on the choice of a complex basis α.

To get unique real subspace for the complexification, we may to add the extra structure of *(complex) conjugation* of vectors

$$\overline{\vec{w}_1 + i\vec{w}_2} = \vec{w}_1 - i\vec{w}_2.$$

In general, a conjugation of a complex vector space V is an operator C satisfying

$$C(\vec{u} + \vec{v}) = C(\vec{u}) + C(\vec{v}), \quad C(a\vec{u}) = \bar{a}C(\vec{u}), \quad C^2 = I.$$

The first two properties mean that C is *conjugate linear*. The third property means that conjugation twice goes back to itself. Exercise 6.17 shows that there are many other conjugations on $\mathbb{C}^n$ besides the standard one. Therefore the conjugation is an extra structure imposed on a complex vector space.

The conjugation of vectors induces the real and imaginary parts of the complex vector space

$$\text{Re}V = \{\vec{v} \colon \vec{v} = C(\vec{v})\}, \quad \text{Im}V = \{\vec{v} \colon \vec{v} = -C(\vec{v})\}.$$

Each is a real (not complex) subspace of V. Moreover, we have (see Exercise 6.18)

$$\text{Im}V = i\text{Re}V, \quad V = \text{Re}V \oplus \text{Im}V.$$

This shows that V is the complexification of $\text{Re}V$.

Proposition 6.7. *A complex vector space with conjugation is the same as the complexification of a real vector space.*

Exercise 6.17. Describe all the conjugations on $\mathbb{C}^n$.

Exercise 6.18. Suppose C is a conjugation on complex vector space.

1. Prove that ReV and ImV are real subspaces, and multiplying i is an isomorphism between them.

2. If $\vec{x} = \vec{v} + \vec{w}$ with $\vec{v} \in \mathrm{Re}V$ and $\vec{w} \in \mathrm{Im}V$, prove that $C(\vec{x}) = \vec{v} - \vec{w}$ and derive the unique formula for $\vec{v}$ and $\vec{w}$ in terms of $\vec{x}$ and $C(\vec{x})$.

3. Prove that $V = \mathrm{Re}V \oplus \mathrm{Im}V$.

Exercise 6.19. Suppose H is a complex subspace of a complex vector space V with conjugation. Prove that the conjugation of H

$$\bar{H} = \{C(\vec{v}) \colon \vec{v} \in H\} = \{\vec{w}_1 - i\vec{w}_2 \colon \vec{w}_1, \vec{w}_2 \in \mathrm{Re}V,\ \vec{w}_1 + i\vec{w}_2 \in H\}$$

is still a complex subspace.

Exercise 6.20. Suppose $V = W \oplus iW$ and $V' = W' \oplus iW'$ are complexifications. Prove that a complex linear transformation $L \colon V \to V'$ is determined by

$$L(\vec{w}) = L_1(\vec{w}) + iL_2(\vec{w}), \quad L_1, L_2 \colon W \to W',$$

and has the block form (using $iW \cong W$)

$$L = \begin{pmatrix} L_1 & -L_2 \\ L_2 & L_1 \end{pmatrix} : W \oplus (i)W \to W' \oplus (i)W'.$$

Moreover, prove that $L(W) \subset W$ if and only if it preserves the conjugation: $L(\bar{\vec{v}}) = \overline{L(\vec{v})}$.

Let H be a complex subspace of complixification $V = W \oplus iW$ satisfying $V = H \oplus \bar{H}$, where $\bar{H}$ is the conjugation subspace in Exercise 6.19. Let $\alpha = \{\vec{u}_1 - i\vec{w}_1, \dots, \vec{u}_m - i\vec{w}_m\}$, $\vec{u}_j, \vec{w}_j \in H$, be a (complex) basis of (complex subspace) H. Then $\bar{\alpha} = \{\vec{u}_1 + i\vec{w}_1, \dots, \vec{u}_m + i\vec{w}_m\}$ is a basis of $\bar{H}$, and $\alpha \cup \bar{\alpha}$ is a (complex) basis of $V = H \oplus \bar{H}$. Since vectors in $\alpha \cup \bar{\alpha}$ and vectors in $\beta = \{\vec{u}_1, \vec{w}_1, \dots, \vec{u}_m, \vec{w}_m\}$ are linear combinations of each other, and $\alpha \cup \bar{\alpha}$ and β have the same number of vectors, we find that β is also a complex basis of V. Since the vectors in β are real, the set β is actually a real basis of the real suspace W.

We introduce real subspaces

$$E = \mathrm{Span}\{\vec{u}_1, \dots, \vec{u}_m\}, \quad E^\dagger = \mathrm{Span}\{\vec{w}_1, \dots, \vec{w}_m\}, \quad W = E \oplus E^\dagger.$$

We also introduce a real isomorphism by corresponding real bases

$$\dagger \colon E \cong E^\dagger, \quad \vec{u}_j^\dagger = \vec{w}_j.$$

The isomorphism has the following property

$$\vec{u} \in E \implies \vec{u} - i\vec{u}^\dagger \in H.$$

Then for any $\vec{u}_1, \vec{u}_2 \in E$, we have

$$\vec{u}_1 + \vec{u}_2^\dagger + i\vec{u}_2 - i\vec{u}_1^\dagger = \vec{u}_1 - i\vec{u}_1^\dagger + i(\vec{u}_2 - i\vec{u}_2^\dagger) \in H.$$

Conversely, any vector in H lies in $V = W \oplus iW = E \oplus E^\dagger \oplus iE \oplus iE^\dagger$. So the vector in H can be written as $\vec{h} = \vec{u}_1 + \vec{u}_2^\dagger + i\vec{u}_3 - i\vec{u}_4^\dagger$ for unique $\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4 \in E$. On the other hand, we have $\vec{k} = \vec{u}_1 + \vec{u}_2^\dagger + i\vec{u}_2 - i\vec{u}_1^\dagger \in H$. Then

$$H \ni i(\vec{h} - \vec{k}) = \vec{u}_2 - \vec{u}_3 + \vec{u}_4^\dagger - \vec{u}_1^\dagger \in E \oplus E^\dagger = W.$$

We have

$$\vec{v} \in H \cap W \implies \vec{v} = \bar{\vec{v}} \in \bar{H} \implies \vec{v} \in H \cap \bar{H} = \{\vec{0}\}.$$

Therefore we conclude $i(\vec{h} - \vec{k}) = \vec{0}$. By the direct sum $E \oplus E^\dagger$, we further get $\vec{u}_2 = \vec{u}_3$ and $\vec{u}_4 = \vec{u}_1$. This proves that

$$H = \{\vec{u}_1 + \vec{u}_2^\dagger + i\vec{u}_2 - i\vec{u}_1^\dagger \colon \vec{u}_1, \vec{u}_2 \in E\}.$$

A special case of the equality above is when $\vec{u}_2 = \vec{0}$, we have

$$\vec{u} \in E,\ \vec{w} \in E^\dagger,\ \vec{u} + i\vec{w} \in H \implies \vec{w} = -\vec{u}^\dagger.$$

Therefore the isomorphism $\dagger \colon E \cong E^\dagger$ is uniquely charaterised by the property $\vec{u} - i\vec{u}^\dagger \in H$.

Proposition 6.8. *Suppose $V = W \oplus iW$ is a complexfication of a real vector space. Suppose $H \subset V$ is a complex subspace satisfying $V = H \oplus \bar{H}$. Then there are real subspaces $E, E^\dagger$ and an isomorphism $\vec{u} \leftrightarrow \vec{u}^\dagger$ between E and $E^\dagger$, such that*

$$W = E \oplus E^\dagger, \quad H = \{(\vec{u}_1 + \vec{u}_2^\dagger) + i(\vec{u}_2 - \vec{u}_1^\dagger) \colon \vec{u}_1, \vec{u}_2 \in E\}.$$

Conversely, any complex subspace H constructed in this way satisfies $V = H \oplus \bar{H}$.

Exercise 6.21. The construction E in Proposition 6.8 makes use of a choice of basis of H and is therefore not unique. However, prove that E uniquely determines $E^\dagger$ by

$$E^\dagger = \{\vec{w} \in W \colon \vec{u} - i\vec{w} \in H \text{ for some } \vec{u} \in E\}.$$

Under the situation in Proposition 6.8, we consider a linear operator $L \colon V \to V$ satisfying $L(W) \subset W$ and $L(H) \subset H$. The condition $L(W) \subset W$ means L is a *real operator* with respect to the complex conjugation structure on V. The property is equivalent to L commuting with the conjugation operation (see Exercise 6.20). The condition $L(H) \subset H$ means that H is an *invariant subspace* of L.

By $L(H) \subset H$ and the description of H in Proposition 6.8, we have real linear transformations $L_1, L_2 \colon E \to E$, such that for $\vec{u} \in E$, we have

$$L(\vec{u}) - iL(\vec{u}^\dagger) = L(\vec{u} - i\vec{u}^\dagger) = (L_1(\vec{u}) + L_2(\vec{u})^\dagger) + i(L_2(\vec{u}) - L_1(\vec{u})^\dagger).$$

By $V = W \oplus iW$ and $L(\vec{u}), L(\vec{u}^\dagger) \in L(W) \subset W$, we get

$$L(\vec{u}) = L_1(\vec{u}) + L_2(\vec{u})^\dagger, \quad L(\vec{u}^\dagger) = -L_2(\vec{u}) + L_1(\vec{u})^\dagger,$$

or

$$L\begin{pmatrix}\vec{u}_1 \\ \vec{u}_2^\dagger\end{pmatrix} = \begin{pmatrix}L_1(\vec{u}_1) \\ L_2(\vec{u}_1)^\dagger\end{pmatrix} + \begin{pmatrix}-L_2(\vec{u}_2) \\ L_1(\vec{u}_2)^\dagger\end{pmatrix} = \begin{pmatrix}L_1 & -L_2 \\ L_2 & L_1\end{pmatrix}\begin{pmatrix}\vec{u}_1 \\ \vec{u}_2^\dagger\end{pmatrix}.$$

In other words, with respect to the direct sum $W = E \oplus E^\dagger$ and using $E \cong E^\dagger$, the restriction $L|_W \colon W \to W$ has the block matrix form

$$L|_W = \begin{pmatrix}L_1 & -L_2 \\ L_2 & L_1\end{pmatrix}.$$

7 Eigenvalue and Eigenvector

The famous Fibonacci numbers $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \dots$ is defined through the recursive relation

$$F_0 = 0, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}.$$

Given a specific number, say 100, we can certainly calculate F_{100} by repeatedly applying the recursive relation. However, it is not clear what the general formula for F_n should be.

The essential difficulty for finding the general formula is due to the lack of understanding of the *structure* of the recursion process. The Fibonacci numbers is a linear system because it is governed by a linear equation $F_n = F_{n-1} + F_{n-2}$. Many differential equations such as Newton's second law $F = m\vec{x}''$ are also linear. Understanding the structure of the linear transformation inherent in linear systems helps us to solve problems about the system.

Example 7.1. Suppose a pair of numbers x_n, y_n is defined through the recursive relation

$$x_0 = 1, \quad y_0 = 0, \quad x_n = x_{n-1} - y_{n-1}, \quad y_n = x_{n-1} + y_{n-1}.$$

To find the general formula for x_n and y_n, we rewrite the recursive relation as a linear transformation

$$\vec{x}_n = \begin{pmatrix}x_n \\ y_n\end{pmatrix} = \begin{pmatrix}1 & -1 \\ 1 & 1\end{pmatrix}\vec{x}_{n-1} = A\vec{x}_{n-1}, \quad \vec{x}_0 = \begin{pmatrix}1 \\ 0\end{pmatrix} = \vec{e}_1.$$

By

$$A = \sqrt{2}\begin{pmatrix}\cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4}\end{pmatrix},$$

the linear transformation is the rotation by $\frac{\pi}{4}$ and scalar multiplication by $\sqrt{2}$. Therefore vector $\vec{x}_n$ is obtained by rotating $\vec{e}_1$ by $\frac{n\pi}{4}$ and has length $(\sqrt{2})^n$. We conclude that

$$x_n = 2^{\frac{n}{2}}\cos\frac{n\pi}{4}, \quad y_n = 2^{\frac{n}{2}}\sin\frac{n\pi}{4}.$$

Example 7.2. Suppose a linear system is obtained by repeatedly applying the matrix

$$A = \begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}.$$

To find A^n, we note that

$$A\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 5\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad A\begin{pmatrix} -2 \\ 1 \end{pmatrix} = 15\begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

This means that, with respect to the basis given by the two new vectors, the linear transformation simply multiplies 5 in the direction of the first vector and multiplies 15 in the direction of the second vector. The understanding immediately implies

$$A^n\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 5^n\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad A^n\begin{pmatrix} -2 \\ 1 \end{pmatrix} = 15^n\begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Then we may apply A^n to the other vectors by expressing the vectors in terms of the new basis

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{2}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{5}\begin{pmatrix} -2 \\ 1 \end{pmatrix} \implies A^n\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{2}{5}A^n\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{5}A^n\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 2\cdot 5^n - 2\cdot 15^n \\ 4\cdot 5^n + 15^n \end{pmatrix}.$$

The answer can be interpreted as the general formula for the sequence defined by the recursive relation

$$x_0 = 0, \quad y_0 = 1, \quad x_n = 13x_{n-1} - 4y_{n-1}, \quad y_n = -4x_{n-1} + 7y_{n-1}.$$

7.1 Eigenspace

The ideal case for a linear operator $L\colon V \to V$ is that the vector space is decomposed into a direct sum $V = E_1 \oplus \cdots \oplus E_k$, such that L simply multiplies a number λ_i to vectors in E_i. Then $L(\vec{v})$ for general $\vec{v}$ can be understood by decomposing the vector according to the direct sum

$$\vec{v} = \vec{x}_1 + \cdots + \vec{x}_k,\ \vec{x}_i \in E_i \implies L(\vec{v}) = \lambda_1\vec{x}_1 + \cdots + \lambda_k\vec{x}_k.$$

Definition 7.1. A number λ is an *eigenvalue* of a linear operator $L\colon V \to V$ if the linear operator $\lambda I - L$ is not invertible. The *eigenspace* associated to the eigenvalue is

$$\{\vec{x}\colon L(\vec{x}) = \lambda\vec{x}\} = \operatorname{Ker}(\lambda I - L).$$

For finite dimensional V, the non-invertiblity of $\lambda I - L$ is equivalent to that the eigenspace is not the zero space. Any nonzero vector in the eigenspace is an *eigenvector*.

Proposition 7.2. *The sum of eigenspaces with distinct eigenvalues is a direct sum.*

Proof. Suppose E_i is an eigenspace of L with eigenvalue λ_i. Suppose

$$\vec{x}_1 + \cdots + \vec{x}_k = \vec{0}, \quad \vec{x}_i \in E_i.$$

Then applying L gives

$$\lambda_1\vec{x}_1 + \cdots + \lambda_k\vec{x}_k = \vec{0},$$

and we get

$$(\lambda_1 - \lambda_k)\vec{x}_1 + \cdots + (\lambda_{k-1} - \lambda_k)\vec{x}_{k-1} = \lambda_1\vec{x}_1 + \cdots + \lambda_k\vec{x}_k - \lambda_k(\vec{x}_1 + \cdots + \vec{x}_k) = \vec{0}.$$

This can be viewed as

$$\vec{x}'_1 + \cdots + \vec{x}'_{k-1} = \vec{0}, \quad \vec{x}'_i = (\lambda_i - \lambda_k)\vec{x}_i \in E_i.$$

So an inductive argument can be carried out. Assume the proposition is alredy proved for $k-1$ eigenspaces. Then we get $\vec{x}'_i = \vec{0}$ for $i = 1, \dots, k-1$. By $\lambda_i \neq \lambda_k$, we get $\vec{x}_i = \vec{0}$ for $i = 1, \dots, k-1$. By $\vec{x}_1 + \cdots + \vec{x}_k = \vec{0}$, this further implies $\vec{x}_k = \vec{0}$. □

For the given linear operator, we hope to find enough eigenspaces, so that the whole vector space can be decomposed into a sum of eigenspaces. We will see that this may or may not always happen. Even in the non-ideal case, it would be helpful to have the following.

Definition 7.3. A subspace $H \subset V$ is an *invariant subspace* of a linear operator $L\colon V \to V$ if $\vec{h} \in H$ implies $L(\vec{h}) \in H$.

The restriction of the operator to an invariant subspace H can be regarded as an operator of H. If $V = H_1 \oplus \cdots \oplus H_k$ is a direct sum of invariant subspaces, and the restriction of L on H_i is $L_i\colon H_i \to H_i$, then

$$\vec{v} = \vec{h}_1 + \cdots + \vec{h}_k \implies L(\vec{v}) = L_1(\vec{h}_1) + \cdots + L_k(\vec{h}_k).$$

In blocked notation, this means

$$L = \begin{pmatrix} L_1 & & & O \\ & L_2 & & \\ & & \ddots & \\ O & & & L_k \end{pmatrix} = L_1 \oplus L_2 \oplus \cdots \oplus L_k.$$

By Proposition 5.4, the non-invertibility can be detected by the vanishing of the determinant. Therefore the eigenvalues of L can be calculated as the roots of the polynomial $\det(\lambda I - L)$, called the *characteristic polynomial* of L. By Theorem 6.5, we have the following result.

Proposition 7.4. *Any linear operator $L\colon V \to V$ has eigenvalue.*

Example 7.3. For the matrix in Example 7.2, we have

$$\det(\lambda I - A) = \det\begin{pmatrix}\lambda-13 & 4\\ 4 & \lambda-7\end{pmatrix} = (\lambda-13)(\lambda-7)-16 = \lambda^2-20\lambda+75 = (\lambda-5)(\lambda-15).$$

This gives two possible eigenvalues $\lambda_1 = 5, \lambda_2 = 15$.

The eigenspace $\mathrm{Ker}(5I - A)$ for λ_1 is the solutions of

$$(5I - A)\vec{x} = \begin{pmatrix}5-13 & 4\\ 4 & 5-7\end{pmatrix}\vec{x} = \begin{pmatrix}-8 & 4\\ 4 & -2\end{pmatrix}\vec{x} = \vec{0}.$$

The solution is $\mathrm{Ker}(5I - A) = \mathbb{R}(1\ 2)^T$.

The eigenspace $\mathrm{Ker}(15I - A)$ for λ_2 is the solutions of

$$(15I - A)\vec{x} = \begin{pmatrix}15-13 & 4\\ 4 & 15-7\end{pmatrix}\vec{x} = \begin{pmatrix}2 & 4\\ 4 & 8\end{pmatrix}\vec{x} = \vec{0}.$$

The solution is $\mathrm{Ker}(15I - A) = \mathbb{R}(2\ -1)^T$.

The direct sum $\mathrm{Ker}(5I - A) \oplus \mathrm{Ker}(15I - A) = \mathbb{R}(1\ 2)^T \oplus \mathbb{R}(2\ -1)^T$ is 2-dimensional and therefore must be $\mathbb{R}^2$. This means that $\{(1\ 2)^T, (2\ -1)^T\}$ is a basis of eigenvectors.

Example 7.4. For the matrix in Example 7.1, we have

$$\det(\lambda I - A) = \det\begin{pmatrix}\lambda-1 & 1\\ -1 & \lambda-1\end{pmatrix} = (\lambda-1)^2+1 = (\lambda-1-i)(\lambda-1+i).$$

This gives two possible eigenvalues $\lambda_1 = 1+i, \lambda_2 = 1-i$.

The eigenspace $\mathrm{Ker}((1+i)I - A)$ for λ_1 is the solutions of

$$((1+i)I - A)\vec{x} = \begin{pmatrix}(1+i)-1 & 1\\ -1 & (1+i)-1\end{pmatrix}\vec{x} = \begin{pmatrix}i & 1\\ -1 & i\end{pmatrix}\vec{x} = \vec{0}.$$

The solution is scalar multiples of $\vec{v} = (1\ -i)^T$. Since $\vec{v}$ is a complex vector (and the eigenvalue is complex), we should view all the calculations to happen inside $\mathbb{C}^2$, and the eigenspace is $\mathrm{Ker}((1+i)I - A) = \mathbb{C}(1\ -i)^T$.

The calculation for the eigenspace $\mathrm{Ker}((1-i)I - A)$ for λ_2 is conjugation of the calculation for λ_1. We get $\mathrm{Ker}((1-i)I - A) = \mathbb{C}(\bar{1}\ \overline{-i})^T = \mathbb{C}(1\ i)^T$.

The direct sum $\mathrm{Ker}((1+i)I - A) \oplus \mathrm{Ker}((1-i)I - A) = \mathbb{C}(1\ -i)^T \oplus \mathbb{C}(1\ i)^T$ is complex 2-dimensional and therefore must be $\mathbb{C}^2$. This means that $\{(1\ -i)^T, (1\ i)^T\}$ is a basis of eigenvectors.

Example 7.5. For the matrix

$$A = \begin{pmatrix}1 & 3 & -3\\ -3 & 7 & -3\\ -6 & 6 & -2\end{pmatrix},$$

we have

$$\begin{aligned}
\det(\lambda I - A) &= \det\begin{pmatrix}\lambda-1 & -3 & 3\\ 3 & \lambda-7 & 3\\ 6 & -6 & \lambda+2\end{pmatrix} = \det\begin{pmatrix}\lambda-1 & 0 & 3\\ 3 & \lambda-4 & 3\\ 6 & \lambda-4 & \lambda+2\end{pmatrix}\\
&= \det\begin{pmatrix}\lambda-1 & 0 & 3\\ 3 & \lambda-4 & 3\\ 3 & 0 & \lambda-1\end{pmatrix} = (\lambda-4)\det\begin{pmatrix}\lambda-1 & 3\\ 3 & \lambda-1\end{pmatrix}\\
&= (\lambda-4)[(\lambda-1)^2-3^2] = (\lambda-4)^2(\lambda+2).
\end{aligned}$$

For the eigenvalues $\lambda_1 = 4$ and $\lambda_2 = -2$, we have

$$4I - A = \begin{pmatrix}3 & -3 & 3\\ 3 & -3 & 3\\ 6 & -6 & 6\end{pmatrix}, \quad -2I - A = \begin{pmatrix}-3 & -3 & 3\\ 3 & -9 & 3\\ 6 & -6 & 0\end{pmatrix}.$$

The eigenspaces are $\mathrm{Ker}(4I-A) = \mathbb{R}(1\ 1\ 0)^T \oplus \mathbb{R}(1\ 0\ -1)^T$ and $\mathrm{Ker}(-2I-A) = \mathbb{R}(1\ 1\ 2)^T$. By Proposition 7.2, the sum of eigenspaces is always direct. By $\dim\mathrm{Ker}(4I-A) + \dim\mathrm{Ker}(-2I-A) = 2+1 = \dim\mathbb{R}^3$, the sum is $\mathbb{R}^3$. In other words, $\{(1\ 1\ 0)^T, (1\ 0\ -1)^T, (1\ 1\ 2)^T\}$ is a basis of eigenvectors.

Example 7.6. For the matrix

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

we have

$$\begin{aligned}
\det(\lambda I - A) &= \det\begin{pmatrix}\lambda-3 & -1 & 3\\ 1 & \lambda-5 & 3\\ 6 & -6 & \lambda+2\end{pmatrix} = \det\begin{pmatrix}\lambda-4 & -1 & 3\\ \lambda-4 & \lambda-5 & 3\\ 0 & -6 & \lambda+2\end{pmatrix}\\
&= \det\begin{pmatrix}\lambda-4 & -1 & 3\\ 0 & \lambda-4 & 0\\ 0 & -6 & \lambda+2\end{pmatrix} = (\lambda-4)^2(\lambda+2).
\end{aligned}$$

For the eigenvalues $\lambda_1 = 4$ and $\lambda_2 = -2$, we have

$$4I - A = \begin{pmatrix}1 & -1 & 3\\ 1 & -1 & 3\\ 6 & -6 & 6\end{pmatrix}, \quad -2I - A = \begin{pmatrix}-5 & -1 & 3\\ 1 & -7 & 3\\ 6 & -6 & 0\end{pmatrix}.$$

The eigenspaces are $\mathrm{Ker}(4I-A) = \mathbb{R}(1\ 1\ 0)^T$ and $\mathrm{Ker}(-2I-A) = \mathbb{R}(1\ 1\ 2)^T$. Since $\dim\mathrm{Ker}(4I-A) + \dim\mathrm{Ker}(-2I-A) = 2 < \dim\mathbb{R}^3$, there is no basis of eigenvectors.

Example 7.7. Consider the derivative linear transformation $D(f) = f'$ on the space of all real valued smooth functions $f(t)$ on $\mathbb{R}$. The eigenspace of eigenvalue λ consists of all functions f

satisfying $f' = \lambda f$. This means exactly $f(t) = ce^{\lambda t}$ is a constant multiple of the exponential function $e^{\lambda t}$. Therefore any real number λ is an eigenvalue, and the eigenspace $\mathrm{Ker}(\lambda I - D) = \mathbb{R}e^{\lambda t}$.

Example 7.8. We may also consider the derivative linear transformation on the space V of all complex valued smooth functions $f(t)$ on $\mathbb{R}$ of period 2π. In this case, the eigenspace $\mathrm{Ker}^{\mathbb{C}}(\lambda I - D)$ still consists of $ce^{\lambda t}$, but c and λ can be complex numbers. For the function to have period 2π, we further need $e^{\lambda 2\pi} = e^{\lambda 0} = 1$. This means $\lambda = in \in i\mathbb{Z}$. Therefore the (relabeled) eigenspaces are $\mathrm{Ker}^{\mathbb{C}}(nI - D) = \mathbb{C}e^{int}$. The "eigenspace decomposition" $V = \oplus_n \mathbb{C}e^{int}$ essentially means Fourier series. More details will be given in Example 7.15.

Example 7.9. To find the general formula for the Fibonacci numbers, we introduce

$$\vec{x}_n = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}, \quad \vec{x}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \vec{x}_{n+1} = \begin{pmatrix} F_{n+1} \\ F_{n+1} + F_n \end{pmatrix} = A\vec{x}_n, \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Then the n-th Fibonacci number F_n is the first coordinate of $\vec{x}_n = A^n \vec{x}_0$.

The characteristic polynomial $\det(\lambda I - A) = \lambda^2 - \lambda - 1$ has two roots

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \quad \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

By

$$\begin{pmatrix} \lambda_1 & -1 \\ -1 & \lambda_1 - 1 \end{pmatrix} = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & -1 \\ -1 & \frac{-1+\sqrt{5}}{2} \end{pmatrix},$$

we get the eigenspace $E_1 = \mathbb{R}(1 \ \frac{1+\sqrt{5}}{2})^T$. If we substitute $\sqrt{5}$ by $-\sqrt{5}$, then we get the second eigenspace $E_2 = \mathbb{R}(1 \ \frac{1-\sqrt{5}}{2})^T$. To find $\vec{x}_n$, we decompose $\vec{x}_0$ according to the basis of eigenvectors

$$\vec{x}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix} - \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix}.$$

The two coefficients can be obtained by solving a system of linear equations. Then

$$\vec{x}_n = \frac{1}{\sqrt{5}} A^n \begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix} - \frac{1}{\sqrt{5}} A^n \begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix} = \frac{1}{\sqrt{5}} \lambda_1^n \begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix} - \frac{1}{\sqrt{5}} \lambda_2^n \begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix}.$$

Picking the first coordinate, we get

$$F_n = \frac{1}{\sqrt{5}}(\lambda_1^n - \lambda_2^n) = \frac{1}{2^n\sqrt{5}}\left[(1+\sqrt{5})^n - (1-\sqrt{5})^n\right].$$

Exercise 7.1. Let L_1, L_2 be linear operators. Prove that the characteristic polynomials of blocked linear operators

$$\begin{pmatrix} L_1 & * \\ O & L_2 \end{pmatrix}, \quad \begin{pmatrix} L_1 & O \\ * & L_2 \end{pmatrix}$$

are the product of the characteristic polynomials of L_1 and L_2.

Exercise 7.2. For a complex square matrix A, prove that $A\vec{x} = \lambda\vec{x}$ if and only if $\bar{A}\bar{\vec{x}} = \bar{\lambda}\bar{\vec{x}}$. This means the following equality for eigenspaces

$$\mathrm{Ker}^{\mathbb{C}}(\bar{\lambda}I - \bar{A}) = \overline{\mathrm{Ker}^{\mathbb{C}}(\lambda I - A)}.$$

Exercise 7.3. Suppose $\vec{v}$ is an eigenvector of L of eigenvalue λ. Prove that $\vec{v}$ is an eigenvector of $L^2 + aL + b$ of eigenvalue $\lambda^2 + a\lambda + b$. Extend the property to polynomials of linear operators.

Exercise 7.4. Suppose a linear operator satisfies $L^2 + 3L + 2 = O$. What can you say about the eigenvalues of L?

Exercise 7.5. Find eigenvalues and eigenvectors ....

7.2 Diagonalisation

The examples in Section 7.1 show how to find all the eigenspaces. We first find all the eigenvalues as all the distinct roots λ_i of the charateristic polynomial $\det(\lambda I - L)$. Then for each λ_j, we find a basis of the corresponding eigenspace $E_j = \mathrm{Ker}(\lambda_j I - L)$. The process gives all the possible eigenspaces, and we get the ideal case $V = E_1 \oplus \cdots \oplus E_k$ if and only if $\dim E_1 + \cdots + \dim E_k = \dim V$. In other words, the union of bases of the eigenspaces form a basis of V consisting of only eigenvectors. This is a *basis of eigenvectors*.

Suppose $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ is a basis of eigenvectors, with corresponding eigenvalues $d_1, \dots, d_n$. The numbers d_j are the eigenvalues λ_i, perhaps with some repetition. For example, the basis obtained at the end of Example 7.5 has $d_1 = 4$, $d_2 = 4$, $d_3 = -2$. The basis of eigenvectors simply means

$$L(\vec{v}_j) = d_j\vec{v}_j.$$

This also means that the matrix of L with respect to the basis α is *diagonal*

$$[L]_{\alpha\alpha} = \begin{pmatrix} \ddots & & O \\ & d_j & \\ O & & \ddots \end{pmatrix} = D.$$

For this reason, we say L is *diagonalisable* if it has a basis of eigenvectors. Therefore the linear transformations in Example 7.3, 7.4, 7.5 are diagonalisable, and the linear transformation in Example 7.6 is not diagonalisable.

For the special case $L\colon \mathbb{R}^n \to \mathbb{R}^n$, we have the standard basis $\epsilon = \{\vec{e}_1, \dots, \vec{e}_n\}$. Then

$$[I]_{\epsilon\alpha} = (\vec{v}_1 \ \cdots \ \vec{v}_n) = P, \quad [L]_{\epsilon\epsilon} = PDP^{-1}.$$

Therefore a matrix is diagonalisable if and only if it can be written as PDP^{-1} for invertible P and diagonal D. The columns of P are the eigenvectors and the diagonal entries of D are the eigenvalues. From the examples in Section 7.1, we have

$$\begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 15 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}^{-1},$$
$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}^{-1},$$
$$\begin{pmatrix} 1 & 3 & -3 \\ -3 & 7 & -3 \\ -6 & 6 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix}^{-1}.$$

Exercise 7.6. Suppose L is diagonalisable, with $V = E_1 \oplus \cdots \oplus E_k$ for eigenspaces E_j of distinct eigenvalues λ_j. Prove that any invariant subspace of L is of the form $H_1 \oplus \cdots \oplus H_k$, with H_j being a subspace of E_j.

In the second example above (see Example 7.4), the real matrix is diagonalised by using complex matrices. We wish to know what this means in terms of only real matrices.

Suppose λ is an eigenvalue of a real $n \times n$ matrix A. If $\lambda \in \mathbb{R}$, then the complex eigenspace $\text{Ker}^{\mathbb{C}}(\lambda I - A)$ consists of $\vec{x} + i\vec{y}$, with $\vec{x}, \vec{y} \in \mathbb{R}^n$ and satisfying

$$A\vec{x} + iA\vec{y} = A(\vec{x} + i\vec{y}) = \lambda(\vec{x} + i\vec{y}) = \lambda\vec{x} + i\lambda\vec{y}.$$

Since $A\vec{x}, A\vec{y}, \lambda\vec{x}, \lambda\vec{y}$ are real vectors, this implies $A\vec{x} = \lambda\vec{x}$ and $A\vec{y} = \lambda\vec{y}$. Therefore $\vec{x}, \vec{y} \in \text{Ker}^{\mathbb{R}}(\lambda I - A)$ are real vectors, and $\text{Ker}^{\mathbb{C}}(\lambda I - A)$ is the complexification of $\text{Ker}^{\mathbb{R}}(\lambda I - A)$.

The second case is $\lambda \notin \mathbb{R}$. This means $\lambda = a + ib$ with $a \in \mathbb{R}$ and $b \in \mathbb{R} - 0$. Then we have the complex eigenspace $\text{Ker}^{\mathbb{C}}(\lambda I - A)$ but no corresponding real eigenspace. On the other hand, $\bar{\lambda} = a - ib$ is also an eigenvalue of A, with $\text{Ker}^{\mathbb{C}}(\bar{\lambda} I - A) = \overline{\text{Ker}^{\mathbb{C}}(\lambda I - A)}$ (see Exercises 6.19 and 7.2). When it comes to a non-real eigenvalue λ, therefore, we should really consider the conjugate pair $\lambda, \bar{\lambda}$ of eigenvalues and the direct sum (by Proposition 7.2) $\text{Ker}^{\mathbb{C}}(\lambda I - A) \oplus \text{Ker}^{\mathbb{C}}(\bar{\lambda} I - A)$ of eigenspaces. Proposition 6.8 can be applied by taking the direct sum to be V and taking $\text{Ker}^{\mathbb{C}}(\lambda I - A)$ to be H.

Let $\alpha = \{\vec{u}_1 - i\vec{w}_1, \dots, \vec{u}_m - i\vec{w}_m\}$, $\vec{u}_j, \vec{w}_j \in \mathbb{R}^n$, be a basis of $\text{Ker}^{\mathbb{C}}(\lambda I - A)$. Then $\bar{\alpha} = \{\vec{u}_1 + i\vec{w}_1, \dots, \vec{u}_m + i\vec{w}_m\}$ is a basis of $\text{Ker}^{\mathbb{C}}(\bar{\lambda} I - A)$, and the direct sum $\text{Ker}^{\mathbb{C}}(\lambda I - A) \oplus \text{Ker}^{\mathbb{C}}(\bar{\lambda} I - A)$ has the union $\alpha \cup \bar{\alpha}$ as a basis. The set $\beta = \{\vec{u}_1, \vec{w}_1, \dots, \vec{u}_m, \vec{w}_m\}$ of real vectors and the set $\alpha \cup \bar{\alpha}$ can be expressed as each other's complex linear combinations. Since the two sets have the same number of vectors, the real set β is also a complex basis of the direct sum $\text{Ker}^{\mathbb{C}}(\lambda I - A) \oplus \text{Ker}^{\mathbb{C}}(\bar{\lambda} I - A)$. Then we may introduce real subspaces

$$E = \text{Span}^{\mathbb{R}}\{\vec{u}_1, \dots, \vec{u}_m\}, \quad E^\dagger = \text{Span}^{\mathbb{R}}\{\vec{w}_1, \dots, \vec{w}_m\}.$$

Since β is linearly independent, we have direct sum $E \oplus E^\dagger$, and we also have an isomorphism $\dagger\colon E \cong E^\dagger$ defined by $\vec{u}_j^\dagger = \vec{w}_j$ and characterised by

$$A(\vec{u} - i\vec{w}) = \lambda(\vec{u} - i\vec{w}) \iff \vec{w} = \vec{u}^\dagger.$$

The equality

$$A\vec{u} - iA\vec{u}^\dagger = A(\vec{u} - i\vec{u}^\dagger) = (a+ib)(\vec{u} - i\vec{u}^\dagger) = (a\vec{u} + b\vec{u}^\dagger) + i(b\vec{u} - a\vec{u}^\dagger)$$

means

$$A\vec{u} = a\vec{u} + b\vec{u}^\dagger, \quad A\vec{u}^\dagger = -b\vec{u} + a\vec{u}^\dagger.$$

For $\vec{u} \neq \vec{0}$, the 2-dimensional subspace $\mathbb{R}\vec{u} \oplus \mathbb{R}\vec{u}^\dagger$ is an invariant subspace of A. If $a + ib = re^{i\theta}$, then the restriction of A has the following matrix with respect to the basis $\{\vec{u}, \vec{u}^\dagger\}$

$$[A|_{\mathbb{R}\vec{u}\oplus\mathbb{R}\vec{u}^\dagger}] = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = r\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In other words, under the isomorphism $\mathbb{R}\vec{u} \oplus \mathbb{R}\vec{u}^\dagger \cong \mathbb{R}^2$ given by the basis $\{\vec{u}, \vec{u}^\dagger\}$, the restriction of A to the invariant subspace is the combination of the rotation by θ and scalar multiplication by r.

For Example 7.4, we have

$$\lambda = 1 + i = \sqrt{2}e^{i\frac{\pi}{4}}, \quad \vec{v} = \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - i\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \vec{u} - i\vec{u}^\dagger.$$

The basis $\{\vec{u}, \vec{u}^\dagger\}$ is therefore the standrd basis of $\mathbb{R}^2$, and the linear operator is rotation by $\frac{\pi}{4}$ and scalar multiplication by $\sqrt{2}$, as originally given in Example 7.1.

We remark that the rotation does not mean there is an inner product. In fact, the concepts of eigenvalue, eigenspace and eigenvectors do not require inner product. The rotation only means that, if we "pretend" $\{\vec{u}, \vec{u}^\dagger\}$ is an orthonormal basis (which may not be the case with respect to th eusual dot product, see Example 7.10), then the linear transformation is the rotation by θ and scalar multiplication by r.

Proposition 7.5. *Suppose a real $n \times n$ matrix A is complex diagonalisable, with complex basis of eigenvectors*

$$\dots, \vec{v}_j, \dots, \vec{u}_k - i\vec{u}_k^\dagger, \vec{u}_k + i\vec{u}_k^\dagger, \dots,$$

and corresponding eigenvalues

$$\dots, d_j, \dots, a_k + ib_k, a_k - ib_k, \dots.$$

Then

$$A = PDP^{-1}, \quad P = (\cdots\ \vec{v}_j\ \cdots\ \vec{u}_k\ \vec{u}_k^\dagger\ \cdots), \quad D = \begin{pmatrix} \ddots & & & & & \\ & d_j & & & O & \\ & & \ddots & & & \\ & & & a_k & -b_k & \\ & O & & b_k & a_k & \\ & & & & & \ddots \end{pmatrix}.$$

Example 7.10. Consider

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

We have

$$\begin{aligned}\det(\lambda I-A)&=\begin{pmatrix}\lambda-1 & 2 & -4\\ -2 & \lambda-2 & 1\\ 0 & -3 & \lambda\end{pmatrix}\\ &=\lambda(\lambda-1)(\lambda-2)-24+4\lambda+3(\lambda-1)\\ &=\lambda^3-3\lambda^2+9\lambda-27=(\lambda-3)(\lambda^2+9).\end{aligned}$$

For the eigenvalue 3, we easily find the eigenspace $\mathrm{Ker}(3I-A)=\mathbb{R}(1\ 1\ 1)^T$. For the conjugate pair of eigenvalues $3i,-3i$, we have

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

For complex vector $\vec{x}\in\mathrm{Ker}(3iI-A)$, the last equation tells us $x_2=ix_3$. Substituting into the second equation, we get

$$0=-2x_1+(3i-2)x_2+x_3=-2x_1+(3i-2)ix_3+x_3=-2x_1-(2+2i)x_3.$$

Therefore $x_1=-(1+i)x_3$, and we get the complex eigenspace $\mathrm{Ker}(3iI-A)=\mathbb{C}\vec{v}$ with

$$\vec{v}=\begin{pmatrix}-1-i\\ i\\ 1\end{pmatrix}=\begin{pmatrix}-1\\ 0\\ 1\end{pmatrix}-i\begin{pmatrix}1\\ -1\\ 0\end{pmatrix}=\vec{u}-i\vec{u}^{\dagger}.$$

We conclude

$$\begin{aligned}A&=\begin{pmatrix}1 & -1-i & -1+i\\ 1 & i & -i\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}3 & 0 & 0\\ 0 & 3i & 0\\ 0 & 0 & -3i\end{pmatrix}\begin{pmatrix}1 & -1-i & -1+i\\ 1 & i & -i\\ 1 & 1 & 1\end{pmatrix}^{-1}\\ &=\begin{pmatrix}1 & -1 & 1\\ 1 & 0 & -1\\ 1 & 1 & 0\end{pmatrix}\begin{pmatrix}3 & 0 & 0\\ 0 & 0 & -3\\ 0 & 3 & 0\end{pmatrix}\begin{pmatrix}1 & -1 & 1\\ 1 & 0 & -1\\ 1 & 1 & 0\end{pmatrix}^{-1}.\end{aligned}$$

Geometrically, the operator first fixes $(1\ 1\ 1)^T$ and rotate by 90° with respect to $\{(-1\ 0\ 1)^T,(1\ -1\ 0)^T\}$, and then multiply the whole space by 3.

Example 7.11. Consider the linear operator L on $\mathbb{R}^3$ that flips $\vec{v}=(1\ 1\ 1)^T$ to its negative and rotate the plane $H=(\mathbb{R}\vec{v})^{\perp}=\{x+y+z=0\}$ by 90°. Here the rotation of H is compatible with the normal direction $\vec{v}$ of H by the right hand rule.

In Example 4.14, we obtained an orthogonal basis $\vec{u} = (1\ -1\ 0)^T$, $\vec{w} = (1\ 1\ -2)^T$ of H. By

$$\det(\vec{u}\ \vec{w}\ \vec{v}) = \det\begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix} = 6 > 0,$$

the rotation from $\vec{u}$ to $\vec{w}$ is incompatible with $\vec{v}$. Therefore the compatible rotation of H by $90°$ is from $\vec{w}$ to $\vec{u}$, and we have

$$L\left(\frac{\vec{u}}{\|\vec{u}\|}\right) = \frac{\vec{w}}{\|\vec{w}\|}, \quad L\left(\frac{\vec{w}}{\|\vec{w}\|}\right) = -\frac{\vec{u}}{\|\vec{u}\|}, \quad L(\vec{v}) = -\vec{v}.$$

This means

$$L(\sqrt{3}\vec{u}) = \vec{w}, \quad L(\vec{w}) = -\sqrt{3}\vec{u}, \quad L(\vec{v}) = -\vec{v}.$$

Therefore the matrix of L is (taking $P = (\sqrt{3}\vec{u}\ \vec{w}\ \vec{v})$)

$$\begin{pmatrix} \sqrt{3} & 1 & 1 \\ -\sqrt{3} & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 1 & 1 \\ -\sqrt{3} & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix}^{-1} = \frac{1}{3}\begin{pmatrix} -1 & \sqrt{3}-1 & \sqrt{3}-1 \\ \sqrt{3}-1 & -1 & -\sqrt{3}-1 \\ -\sqrt{3}-1 & \sqrt{3}-1 & 1 \end{pmatrix}.$$

From the meaning of the linear operator, we know the 4-th power of the matrix is the identity.

Example 7.12. If $A = PDP^{-1}$, then $A^n = PD^nP^{-1}$. It is easy to see that D^n simply means taking the n-th power of the diagonal entries. Therefore

$$\begin{aligned}\begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}^n &= \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 5^n & 0 \\ 0 & 15^n \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}^{-1} \\ &= \frac{1}{5}\begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 5^n & 0 \\ 0 & 15^n \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} = 5^{n-1}\begin{pmatrix} 1+4\cdot 3^n & 2-2\cdot 3^n \\ 2-2\cdot 3^n & 4+3^n \end{pmatrix}.\end{aligned}$$

By using Taylor expansion, we also have $f(A) = Pf(D)P^{-1}$, where the function f is applies to each diagonal entry of D. For example, we have

$$e^{\begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} e^5 & 0 \\ 0 & e^{15} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}^{-1}.$$

7.3 Normal Operator

Let V be a complex vector space with Hermitian inner product $\langle\ ,\ \rangle$. A linear transformation $L\colon V \to V$ is *normal* if

$$L^*L = LL^*.$$

The commutativity implies that

$$(L^k(L^*)^l)(L^m(L^*)^n) = L^{k+m}(L^*)^{l+n}.$$

Therefore all the polynomials of L and L^*, or all the complex linear combinations of $L^m(L^*)^n$, $m, n \geq 0$, form a *commutative algebra* $\mathbb{C}[L, L^*]$. In other words, the addition and multiplication of operators in $\mathbb{C}[L, L^*]$ still lie in $\mathbb{C}[L, L^*]$. Moreover, we have

$$(L^m(L^*)^n)^* = L^n(L^*)^m,$$

so that the algebra is also closed under adjoint. This implies that all operators in $\mathbb{C}[L, L^*]$ are normal.

Proposition 7.6. *A linear operator on a complex inner product space has an orthonormal basis of eigenvectors if and only if it is normal.*

In terms of matrix, the proposition says that A is normal (i.e., satisfying $A^*A = AA^*$) if and only if $A = UDU^{-1} = UDU^*$ for diagonal D and unitary U. This is the unitary diagonalisation of A.

Proof of Sufficiency of Proposition 7.6. Suppose L has an orthonormal basis of eigenvectors $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$

$$L(\vec{v}_j) = d_j\vec{v}_j.$$

Then by (6.1), we have

$$[L]_{\alpha\alpha} = \begin{pmatrix} \ddots & & O \\ & d_i & \\ O & & \ddots \end{pmatrix}, \quad [L^*]_{\alpha\alpha} = ([L]_{\alpha\alpha})^* = \begin{pmatrix} \ddots & & O \\ & \bar{d}_j & \\ O & & \ddots \end{pmatrix}.$$

In other words, we have

$$L^*(\vec{v}_j) = \bar{d}_j\vec{v}_j.$$

This implies $L^*L = LL^*$ by both sides being equal on basis vectors

$$L^*L(\vec{v}_j) = L_*(d_j\vec{v}_j) = d_jL_*(\vec{v}_j) = d_j\bar{d}_j\vec{v}_j, \quad LL^*(\vec{v}_j) = L(\bar{d}_j\vec{v}_j) = \bar{d}_jL(\vec{v}_j) = \bar{d}_jd_j\vec{v}_j. \qquad \square$$

The necessity follows from a more general result.

Theorem 7.7. *Suppose a set $\mathcal{A}$ of operators on a (finite dimensional) Hermitian inner product space V has the following property.*

- *If $L, K \in \mathcal{A}$, then $aL + bK, LK \in \mathcal{A}$.*
- *If $L \in \mathcal{A}$, then $L^* \in \mathcal{A}$.*
- *If $L, K \in \mathcal{A}$, then $LK = KL$.*

Then there is a basis of eigenvectors for all the linear operators in $\mathcal{A}$.

The first condition means that $\mathcal{A}$ is an algebra. The first two conditions mean that $\mathcal{A}$ is a $*$-algebra. The three condition together mean that $\mathcal{A}$ is a *commutative* $*$*-algebra*.

For any $L \in \mathcal{A}$, we have $L^* \in \mathcal{A}$ by the second condition and then $L^*L = LL^*$ by the third condition. Therefore all operators in $\mathcal{A}$ are normal.

Conversely, for any normal operator L, the set $\mathbb{C}[L, L^*]$ of polynomials of L and L^* satisfies the three conditions. However, the theorem also includes commuting pairs (or triples, and so on) of normal operators K, L, so that the set $\mathbb{C}[L, K, L^*, K^*]$ of polynomials also satisfies the three conditions. The the theorem says that L and K have simultaneous orthogonal basis of eigenvectors.

The proof of Theorem 7.7 will be based on two observations. Both do not require normal operator.

Lemma 7.8. *If $LK = KL$, then eigenspaces of L are K-invariant.*

Proof. The lemma follows from

$$\begin{aligned}\vec{x} \in \operatorname{Ker}(\lambda I - L) &\iff L(\vec{x}) = \lambda\vec{x} \\ &\implies L(K(\vec{x})) = K(L(\vec{x})) = K(\lambda\vec{x}) = \lambda K(\vec{x}) \\ &\iff K(\vec{x}) \in \operatorname{Ker}(\lambda I - L).\end{aligned}$$

□

Lemma 7.9. *If H is L-invariant, then $H^\perp$ is L^*-invariant.*

Proof. Let $\vec{v} \in H^\perp$. Then for any $\vec{h} \in H$, we have $L(\vec{h}) \in H$. Therefore $\vec{v} \perp L(\vec{h})$, and we have

$$\langle \vec{h}, L^*(\vec{v}) \rangle = \langle L(\vec{h}), \vec{v} \rangle = 0.$$

Since this holds for all $\vec{h} \in H$, we get $L^*(\vec{v}) \in H^\perp$. □

Proof of Theorem 7.7. If all operators in $\mathcal{A}$ are constant multiples on the whole V, then any basis is a basis of eigenvectors for all operators in $\mathcal{A}$. So we assume an operator $L \in \mathcal{A}$ is not a constant multiple. By Proposition 7.4, L has an eigenvalue λ, and the corresponding eigenspace $E = \operatorname{Ker}(\lambda I - L)$ is neither zero nor V.

For any $K \in \mathcal{A}$, we have $KL = LK$ by the second condition. Then Lemma 7.8 says that E is K-invariant. Therefore by Lemma 7.9, $E^\perp$ is K^*-invariant.

Since $K \in \mathcal{A}$ implies $K^* \in \mathcal{A}$ by the third condition, if we apply the argument above to $K^* \in \mathcal{A}$, we conclude that $E^\perp$ is also K-invariant.

Since both E and $E^\perp$ are K invariant for any $K \in \mathcal{A}$. We may restrict $\mathcal{A}$ to the two subspaces and get two sets of operators $\mathcal{A}_E$ and $\mathcal{A}_{E^\perp}$ of the respective Hermitian inner product spaces E and $E^\perp$. Since E is neither zero nor V, both E and $E^\perp$ have strictly smaller dimension than V. This sets up an inductive process. By induction on $\dim V$, both E and $E^\perp$ have simultaneous orthogonal bases of eigenvectors for $\mathcal{A}_E$ and $\mathcal{A}_{E^\perp}$. Then the union of the two orthogonal bases is a simultaneous orthogonal bases of eigenvectors for $\mathcal{A}$.

It remains to justify the beginning of the induction, which is when $\dim V = 1$. But any linear operator is a constant multiple in this case, and the conclusion follows trivially. □

Theorem 7.7 can be greatly extended. The infinite dimensional version of the theorem is the theory of commutative C^*-algebra.

Exercise 7.7. For a normal operator L, use Exercise 4.26 to prove that $\text{Ker}L = \text{Ker}L^*$. Then prove that the eigenspace of L of eigenvalue λ is the same as the eigenspace of L^* of eigenvalue $\bar{\lambda}$.

Exercise 7.8. Suppose L is normal. Prove that any L-invariant subspace is also L^*-invariant.

7.4 Hermitian Operator

An operator L is *Hermitian* if $L = L^*$. This means

$$\langle L(\vec{v}), \vec{w}\rangle = \langle \vec{v}, L(\vec{w})\rangle \text{ for all } \vec{v}, \vec{w}.$$

Therefore we also say L is *self adjoint*. Hermitian operators are normal.

Proposition 7.10. *A linear operator on a complex inner product space is Hermitian if and only if all the eigenvalues are real and it has an orthonormal basis of eigenvectors.*

In terms of matrix, the proposition says that A is Hermitian (i.e., satisfying $A^* = A$) if and only if $A = UDU^{-1} = UDU^*$ for real diagonal D and unitary U.

Proof. By Proposition 7.6, we need to show that a normal operator is Hermitian if and only if all the eigenvalues are real.

For the sufficiency, let $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ be an orthonormal basis of eigenvectors of L, with real number eigenvalues d_j. Then

$$\langle L(\vec{v}_j), \vec{v}_k\rangle = \langle d_j\vec{v}_j, \vec{v}_k\rangle = d_j\langle \vec{v}_j, \vec{v}_k\rangle = \begin{cases} 0 & \text{if } j \neq k, \\ d_j & \text{if } j = k; \end{cases}$$
$$\langle \vec{v}_j, L(\vec{v}_k)\rangle = \langle \vec{v}_j, d_k\vec{v}_k\rangle = \bar{d}_k\langle \vec{v}_j, \vec{v}_k\rangle = d_k\langle \vec{v}_j, \vec{v}_k\rangle = \begin{cases} 0 & \text{if } j \neq k, \\ d_k & \text{if } j = k. \end{cases}$$

This verifies the equality $\langle L(\vec{v}), \vec{w}\rangle = \langle \vec{v}, L(\vec{w})\rangle$ on basis vectors. Then it is easy to show the equality holds for all vectors $\vec{v}$ and $\vec{w}$.

For the necessity, let $L(\vec{v}) = d\vec{v}$, $\vec{v} \neq \vec{0}$. Then

$$\langle L(\vec{v}), \vec{v}\rangle = \langle d\vec{v}, \vec{v}\rangle = d\langle \vec{v}, \vec{v}\rangle,$$
$$\langle \vec{v}, L(\vec{v})\rangle = \langle \vec{v}, d\vec{v}\rangle = \bar{d}\langle \vec{v}, \vec{v}\rangle.$$

Since the left are equal and $\langle \vec{v}, \vec{v}\rangle \neq 0$, we get $d = \bar{d}$. □

We may apply Proposition 7.10 to linear operators of real inner product spaces. The self adjoint property means that the matrix with respect to an orthonormal basis is symmetric. Therefore we also call real self adjoint operators *symmetric* operators. Then Proposition 7.10 says that a real linear operator is symmetric if and only if all the eigenvalues are real and it has an orthonormal basis of eigenvectors. In terms of matrix, this means that any symmetric matrix is of the form $UDU^{-1} = UDU^T$ for a real diagonal matrix D and an orthogonal matrix U.

Example 7.13. The matrix

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

is Hermitian. The characteristic polynomial

$$\det \begin{pmatrix} \lambda - 2 & -1-i \\ -1+i & \lambda - 3 \end{pmatrix} = (\lambda-2)(\lambda-3) - (1^2+1^2) = \lambda^2 - 5\lambda + 4 = (\lambda-1)(\lambda-4).$$

By

$$I - A = \begin{pmatrix} -1 & -1-i \\ -1+i & -2 \end{pmatrix}, \quad 4I - A = \begin{pmatrix} 2 & -1-i \\ -1+i & 1 \end{pmatrix},$$

we get $\text{Ker}(I - A) = \mathbb{C}(1+i \;\; -1)^T$ and $\text{Ker}(4I - A) = \mathbb{C}(1 \;\; 1-i)^T$. Then we get the diagonalisation

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

If you want to use unitary matrix, then we need to divide the columns by the lengths $\|(1+i \;\; -1)^T\| = \sqrt{3} = \|(1 \; 1-i)^T\|$ and get

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

Example 7.14. The matrix

$$\begin{pmatrix} 1 & \sqrt{2} & \pi \\ \sqrt{2} & 2 & \sin 1 \\ \pi & \sin 1 & e \end{pmatrix}$$

is numerically too complicated to calculate the eigenvalues and eigenspaces. Yet by Proposition 7.10, the matrix is diagonalisable, with an orthogonal basis of eigenvectors.

Example 7.15 (Fourier Series)**.** We extend Example 4.9 to the vector space V of complex valued smooth functions $f(t)$ of single real variable t, such that $f(t+2\pi) = f(t)$. The real inner product in Example 4.9 is extended to the Hermitian inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t)\overline{g(t)}dt.$$

By the calculation in Example 4.9, the derivative operator $D(f) = f'$ on V satisfies $D^* = -D$. By multiplying i, the operator $L = iD$ is then Hermitian.

As pointed out earlier, Theorem 7.7 and Proposition 7.10 can be extended to inifnite dimensional spaces (more precisely to the so called *Hilbert spaces*). Then the operator $L = iD$ should have an orthogonal basis of eigenvectors with real eigenvalues. Indeed we have found in Example 7.8 that the eigenvalues of L are precisely integers n, and the eigenspace $\text{Ker}(nI - L) = \mathbb{C}e^{int}$. By general theory, we already know the eigevectors e^{int} are orthogonal. The following is a direct verification that they are orthonormal

$$\langle e^{imt}, e^{int} \rangle = \frac{1}{2\pi} \int_0^{2\pi} e^{imt} e^{-int} dt = \begin{cases} \frac{1}{2\pi i(m-n)} e^{i(m-n)t}|_0^{2\pi} = 0 & \text{if } m \neq n, \\ \frac{1}{2\pi} 2\pi = 1 & \text{if } m = n. \end{cases}$$

The general theory also says that the eigenvectors should form a basis. This means that any periodic function of period 2π should be expressed as

$$f(t) = \sum_{n \in \mathbb{Z}} c_n e^{int} = c_0 + \sum_{n=1}^{+\infty} (c_n e^{int} + c_{-n} e^{-int}) = a_0 + \sum_{n=1}^{+\infty} (a_n \cos nt + b_n \sin nt).$$

Here we use

$$e^{int} = \cos nt + i \sin nt, \quad a_n = c_n + c_{-n}, \quad b_n = c_n - c_{-n}.$$

We conclude that the Fourier series grows naturally out of the diagonalisation of the derivative operator, as the orthogonal basis of eigenvectors. If we apply the same kind of thinking to the derivative operator on the second vector space in Example 4.9 and use the eigenvectors in Example 7.7, then we get the Fourier transform.

Example 7.16 (Legendre Polynomial)**.** The *Legendre polynomials* P_n are obtained by applying the Gram-Schmidt process to the polynomials $1, t, t^2, \dots$ with respect to the inner product $\langle f, g \rangle = \int_{-1}^{1} f(t)g(t)dt$. The following steps show that

$$P_n = \frac{1}{2^n n!} \frac{d^n}{dt^n} [(t^2 - 1)^n].$$

1. Prove $[(1 - t^2)P_n']' + n(n+1)P_n = 0$.
2. Prove $\sum_{n=0}^{\infty} P_n x^n = \frac{1}{\sqrt{1 - 2tx + x^2}}$.
3. Prove Bonnet's recursion formula $(n+1)P_{n+1} = (2n+1)tP_n - nP_{n-1}$.
4. Prove that $\int_{-1}^{1} P_m P_n dt = \frac{2\delta_{m,n}}{2n+1}$. [eigenvectors of hermitian operator used]

   ....

Exercise 7.9. Use (**??**) to derive the formula for the coefficients of the Fourier series.

Exercise 7.10. Prove that an operator is Hermitian if and only if $\langle L(\vec{v}), \vec{v} \rangle$ is real for all $\vec{v}$.

Exercise 7.11. Prove that a Hermitian operator L satisfies

$$2\langle L(\vec{u}), \vec{v}\rangle = \langle L(\vec{u}+\vec{v}), \vec{u}+\vec{v}\rangle - \langle L(\vec{u}), \vec{u}\rangle - \langle L(\vec{v}), \vec{v}\rangle.$$

Then use the equality to prove that $\langle L(\vec{v}), \vec{v}\rangle = 0$ for all $\vec{v}$ implies $L = O$.

Exercise 7.12. Prove that the determinant of a Hermitian operator is real.

Exercise 7.13. Prove that a linear operator L is normal if and only if $\|L(\vec{v})\| = \|L^*(\vec{v})\|$ for all $\vec{v}$.

Exercise 7.14. A linear operator L on a complex inner product space is *skew Hermitian* if $L^* = -L$. Prove that L is skew Hermitian if and only if all the eigenvalues are imaginary and it has an orthonormal basis of eigenvectors.

Exercise 7.15. Prove for any linear operator L, there are unique Hermitian L_1 and skew Hermitian L_2, such that $L = L_1 + L_2$. Moreover, L is normal if and only if L_1 and L_2 commute. In fact, the algebras $\mathbb{C}[L, L^*]$ and $\mathbb{C}[L_1, L_2]$ are equal.

Exercise 7.16. A linear operator L on a real inner product space is *skew symmetric* if $L^* = -L$. What can you say about the diagonalisation of skew symmetric operators?

Exercise 7.17. Prove that the orthogonal projection P to a subspace satisfies $P^2 = P$ and $P^* = P$. Then prove the converse in the following steps.

1. If $P^2 = P$, then $V = \mathrm{Ker}P \oplus \mathrm{Ker}(I - P)$.
2. If $P^2 = P$ and $P^* = P$, then $\mathrm{Ker}P$ and $\mathrm{Ker}(I - P)$ are orthogonal.

7.5 Unitary Operator

An operator is *unitary* if it is an isometric isomorphism. This means $L^*L = I$. Since $L^*L = I$ implies $L^{-1} = L^*$ (for linear transformations between vector spaces of the same dimension), we have $LL^* = I = L^*L$ and therefore unitary operators are normal.

Proposition 7.11. *A linear operator on a complex inner product space is unitary if and only if all the eigenvalues have length* 1 *and it has an orthonormal basis of eigenvectors.*

Proof. By Proposition 7.6, we need to show that a normal operator is unitary if and only if all the eigenvalues have length 1.

For the sufficiency, let $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ be an orthonormal basis of eigenvectors of L, with eigenvalues d_j satisfying $|d_j| = 1$. Then $L(\alpha) = \{d_1\vec{v}_1, \dots, d_n\vec{v}_n\}$ is still orthogonal and normal

$$\langle L(\vec{v}_j), L(\vec{v}_k)\rangle = \langle d_j\vec{v}_j, d_k\vec{v}_k\rangle = d_j\bar{d}_k\langle \vec{v}_j, \vec{v}_k\rangle = \begin{cases} d_jd_k0 = 0 & \text{if } j \neq k, \\ d_j\bar{d}_k = |d_j|^2 = 1 & \text{if } j = k. \end{cases}$$

Therefore L takes an orthonormal basis to another orthonormal basis. Such an operator is an isometry.

For the necessity, let $L(\vec{v}) = d\vec{v}$, $\vec{v} \neq \vec{0}$. Then the isometry implies $|d|\,\|\vec{v}\| = \|L(\vec{v})\| = \|\vec{v}\|$. Since $\vec{v} \neq \vec{0}$, we get $|d| = 1$. $\square$

We may apply Proposition 7.11 to orthogonal operators of real inner product spaces. The complex eigenvalues of an orthogonal operator are 1, -1 or $e^{i\theta}$ ($\theta \notin \pi\mathbb{Z}$). Let $e^{i\theta}$ have eigenvector $\vec{u} - i\vec{u}^\dagger$, $\vec{u}$ and $\vec{u}^\dagger$ real. We should also consider the conjugate eigenvalue $e^{-i\theta}$ have eigenvector $\vec{u} + i\vec{u}^\dagger$. The two eigenvectors are orthogonal with respect to the Hermitian inner product

$$\begin{aligned}
0 &= \langle \vec{u} - i\vec{u}^\dagger, \vec{u} + i\vec{u}^\dagger \rangle \\
&= \langle \vec{u}, \vec{u} \rangle + \bar{i}\langle \vec{u}, \vec{u}^\dagger \rangle - i\langle \vec{u}^\dagger, \vec{u} \rangle - i\bar{i}\langle \vec{u}\vec{u}^\dagger, \vec{u}\vec{u}^\dagger \rangle \\
&= \|\vec{u}\|^2 - \|\vec{u}^\dagger\|^2 - 2i\langle \vec{u}, \vec{u}^\dagger \rangle.
\end{aligned}$$

Here the inner product in the second and third lines are real, and therefore $\langle \vec{u}, \vec{u}^\dagger \rangle = \langle \vec{u}^\dagger, \vec{u} \rangle$. The equality shows that $\vec{u}$ and $\vec{u}^\dagger$ are orthogonal and have the same length. By dividing the same length, the two real vectors become orthonormal. Then by Proposition 7.5, the complex orthogonal basis of eigenvectors for the orthogonal operator can be modified into a real orthonormal basis, such that the matrix with respect to the basis is

$$\begin{pmatrix}
\ddots & & & & & & & O \\
& 1 & & & & & & \\
& & \ddots & & & & & \\
& & & -1 & & & & \\
& & & & \ddots & & & \\
& & & & & \cos\theta & -\sin\theta & \\
& & & & & \sin\theta & \cos\theta & \\
O & & & & & & & \ddots
\end{pmatrix}.$$

Let $E_\pm$ be the eigenspaces of eigenvalues ± 1. Let $H_j = \mathbb{R}\vec{u} \oplus \mathbb{R}\vec{u}^\dagger$ be the 2-dimensional subspace corresponding to the eigenvector pair $\vec{u} \mp i\vec{u}^\dagger$ of eigenvalues $e^{\pm i\theta}$. Then we have pairwise orthogonal direct sum $V = E_+ \oplus E_- \oplus H_1 \oplus \cdots \oplus H_k$, such that the orthogonal operator is identity I on E_+, antipodal $-I$ on E_-, and rotations on the 2-dimensional subspaces H_j.

Proposition 7.12. *Any orthogonal operator on a real inner product space is the orthogonal direct sum of identity and flipping (on lines) and rotations (on planes).*

Example 7.17. Let $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ be an orthogonal operator. If all eigenvalues of L are real, then L is one of the following.

1. $L = \pm I$, the identity or the antipodal.

2. L fixes a line and is antipodal on the plane orthogonal to the line.

3. L flips a line and is identity on the plane orthogonal to the line.

If L has no real eigenvalue, then L is one of the following.

1. L fixes a line and is rotation on the plane orthogonal to the line.
2. L flips a line and is rotation on the plane orthogonal to the line.

Exercise 7.18. Describe all the orthogonal operators of $\mathbb{R}^2$.

Finally, we note that an orthogonal operator has determinant $\det U = (-1)^{\dim E_-}$. In case U preserves orientation, we have $\det U > 0$, so that $\dim E_-$ is even. This means that the -1 entries in the diagonal can be grouped into pairs

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos\pi & -\sin\pi \\ \sin\pi & \cos\pi \end{pmatrix}.$$

Then each pair is a rotation by 180°. Therefore an orientation preserving orthogonal operator is an orthogonal combination of identities and rotations. For example, an orientation preserving orthogonal operator on $\mathbb{R}^3$ is always a rotation around a fixed axis.

8 Spectral Theory

The spectral theory is the general theory of the structure of linear operators. The ingredients in describing the structure are intrinsic to the linear operator. In other words, they are independent of the matrices of the linear operator with respect to different bases. By (**??**), a quantity $f(A)$ about square matrices is also defined for linear operators if and only if $f(PAP^{-1}) = f(A)$ for any invertible P. We say the quantity is *similarity invariant*, and the quantity is an *invariant* $f(L)$ of linear operators. Examples of invariants are the rank and $\dim \mathrm{Ker} L = n - \mathrm{rank} L$.

A basic invariant of linear operators is the *spectrum*, which is the collection of all eigenvalues (together with the multiplicities of the eigenvalues). Other invariants are the determinant and the *trace* (see Exercise 8.1)

$$\mathrm{tr}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = a_{11} + a_{22} + \cdots + a_{nn}.$$

In fact, the determinant and the trace can be derived from the spectrum, and the spectrum can be derived from the characteristic polynomial. In some sense, the characteristic polynomial is the most primary invariant.

Exercise 8.1. Prove that $\mathrm{tr}AB = \mathrm{tr}BA$. Then use this to show that $\mathrm{tr}PAP^{-1} = \mathrm{tr}A$.

We will first find all the numerical invariants of linear operators. Then with the help of invariants, we find the structure of general linear operators.

8.1 Invariants of Linear Operator

In the discussion about normal operators, we saw the power of thinking in terms of algebra of operators instead of a single operator. Given a linear operator L and a polynomial $p(t) = a_nt^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0$, we denote

$$p(L) = a_nL^n + a_{n-1}L^{n-1} + \cdots + a_1L + a_0I.$$

Then we construct the algebra of all polynomials of L

$$\mathbb{C}[L] = \{p(L) \colon p \text{ is a complex polynomial}\}.$$

The algebra is clearly commutative. The following shows that an eigenspace of L is also an eigenspace of all the operators in $\mathbb{C}[L]$.

Proposition 8.1. *$L(\vec{v}) = \lambda\vec{v}$ implies $p(L)(\vec{v}) = p(\lambda)\vec{v}$.*

Exercise 8.2. A linear operator L is *nilpotent* if it satisfies $L^n = O$ for some n. Show that the derivative operator from P_n to itself is nilpotent. What are the eigenvalues of nilpotent operators?

Exercise 8.3. Prove that $p(PAP^{-1}) = Pp(A)P^{-1}$.

The most important polynomial associated to a linear operator is the characteristic polynomial. It is a monic polynomial of degree $n = \dim V$ and can be completely decomposed using complex roots

$$\begin{aligned}\det(tI - L) &= t^n - \sigma_1t^{n-1} + \cdots + (-1)^{n-1}\sigma_{n-1}t + (-1)^n\sigma_n \\ &= (t-\lambda_1)^{n_1}\cdots(t-\lambda_k)^{n_k}, \quad n_1 + \cdots + n_k = n \\ &= (t-d_1)\cdots(t-d_n).\end{aligned} \tag{8.1}$$

Here $\lambda_1, \dots, \lambda_k$ are all the distinct eigenvalues, and n_j is the *algebraic multiplicity* of λ_j. Moreover, $d_1, \dots, d_n$ are all the eigenvalues repeated in their multiplicities (i.e., λ_1 repeated n_1 times, λ_2 repeated n_2 times, etc.). The (unordered) set $\{d_1, \dots, d_n\}$ of all roots of the characteristic polynomial is the *spectrum* of L.

Since $\det(tI - L)$ is defined for linear operators, the coefficients σ_j, the eigenvalues λ_j with their multiplicities n_j, and the spectrum are all invariants of L. All three invariants are different ways of describing the same invariant $\det(tI - L)$, and therefore determine each other. The dictionary between the three viewpoints is given by *Vieta's formula*

$$\sigma_k = \sum_{1\le i_1<\cdots<i_k\le n} d_{i_1}\cdots d_{i_k}. \tag{8.2}$$

Moreover, σ_k is a homogeneous polynomial of L of degree k. The explicit formula is given in Exercise 8.5.

Exercise 8.4. Prove that $\sigma_1 = d_1 + \cdots + d_n = \mathrm{tr}L$ and $\sigma_n = d_1 \cdots d_n = \det L$. In particular, for 2×2 matrix, we have

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathrm{tr}A = a + d, \quad \det A = ad - bc, \quad \det(tI - A) = t^2 - (\mathrm{tr}A)t + \det A.$$

Exercise 8.5. For an $n \times n$ matrix A and $1 \le i_1 < \cdots < i_k \le n$, let $A(i_1, \ldots, i_k)$ be the $k \times k$ submatrix of A of the $i_1, \ldots, i_k$ rows and $i_1, \ldots, i_k$ columns. Prove that the coefficient of the characteristic polynomial is

$$\sigma_k = \sum_{1 \le i_1 < \cdots < i_k \le n} \det A(i_1, \ldots, i_k).$$

Suppose a function of $f(A)$ of square matrices A is an invariant of linear operators. If A is diagonalisable, then

$$A = PDP^{-1}, \quad D = \begin{pmatrix} d_1 & & O \\ & \ddots & \\ O & & d_n \end{pmatrix}, \quad f(A) = f(PDP^{-1}) = f(D).$$

Therefore $f(A)$ is actually a function of the unordered set $\{d_1, \ldots, d_n\}$. The order does not affect the value because changing order of d_j is the same of exchanging columns of P. This shows that the invariant is really a function $f(d_1, \ldots, d_n)$ satisfying $f(d_{i_1}, \ldots, d_{i_n}) = f(d_1, \ldots, d_n)$ for any permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$. We call such functions *symmetric*.

By (8.2), each σ_k is symmetric in all the eigenvalues. In fact, they are *elementary* symmetric polynomials in the sense that any other symmetric function is a function of the elementary ones. For example,

$$\begin{aligned} d_1^3 + d_2^3 + d_3^3 &= (d_1 + d_2 + d_3)^3 - 3(d_1 + d_2 + d_3)(d_1 d_2 + d_1 d_3 + d_2 d_3) + 3 d_1 d_2 d_3 \\ &= \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3. \end{aligned}$$

We note that $\sigma_1, \ldots, \sigma_n$ are defined for n variables, and

$$\sigma_k(d_1, \ldots, d_n) = \sigma_k(d_1, \ldots, d_n, 0, \ldots, 0), \quad k \le n.$$

The following shows that the characteristic polynomial (8.1) (or the equivalent data) are the primary invariant (at least for polynomial invariants) for diagonalisable linear operators.

Theorem 8.2. *Any symmetric polynomial $f(d_1, \ldots, d_n)$ is a unique polynomial of the elementary symmetric polynomials $\sigma_1, \ldots, \sigma_n$.*

Note that the rank of a linear operator is the number of nonzero eigenvalues in $\{d_1, \ldots, d_n\}$. Although the rank is an invariant, it is not a polynomial of the eigenvalues.

Proof. We prove by induction on n. If $n = 1$, then $f(d_1)$ is a function of the only symmetric polynomial $\sigma_1 = d_1$. Suppose the theorem is proved for $n-1$. Then $g(d_1, \dots, d_{n-1}) = f(d_1, \dots, d_{n-1}, 0)$ is a symmetric polynomial of $n-1$ variables. By induction, we have $g(d_1, \dots, d_{n-1}) = p(\sigma'_1, \dots, \sigma'_{n-1})$ for a polynomial p and elementary symmetric polynomials $\sigma'_1, \dots, \sigma'_{n-1}$ of $d_1, \dots, d_{n-1}$. Now consider

$$h(d_1, \dots, d_n) = f(d_1, \dots, d_n) - p(\sigma_1, \dots, \sigma_{n-1}),$$

where $\sigma_1, \dots, \sigma_{n-1}$ are the elementary symmetric polynomials of $d_1, \dots, d_n$. The polynomial $h(d_1, \dots, d_n)$ is still symmetric. By $\sigma_k(d_1, \dots, d_{n-1}, 0) = \sigma'_k(d_1, \dots, d_{n-1})$, we have

$$h(d_1, \dots, d_{n-1}, 0) = f(d_1, \dots, d_{n-1}, 0) - p(\sigma'_1, \dots, \sigma'_{n-1}) = 0.$$

This means that all the monomial terms of $h(d_1, \dots, d_n)$ have a d_n factor. By symmetry, all the monomial terms of $h(d_1, \dots, d_n)$ also have a d_j factor for every j. Therefore all the monomial terms of $h(d_1, \dots, d_n)$ have $\sigma_n = d_1 \cdots d_n$ factor. This implies that

$$h(d_1, \dots, d_n) = \sigma_n k(d_1, \dots, d_n),$$

for a symmetric polynomial $k(d_1, \dots, d_n)$. Since $k(d_1, \dots, d_n)$ has strictly lower total degree than f, a double induction on the total degree of h can be used. This means that we may assume $k(d_1, \dots, d_n) = q(\sigma_1, \dots, \sigma_n)$ for a polynomial q. Then we get

$$f(d_1, \dots, d_n) = p(\sigma_1, \dots, \sigma_{n-1}) + \sigma_n q(\sigma_1, \dots, \sigma_n).$$

For the uniqueness, we need to prove that, if $p(\sigma_1, \dots, \sigma_n) = 0$ as a polynomial of $d_1, \dots, d_n$, then $p = 0$ as a polynomial of $\sigma_1, \dots, \sigma_n$. Again we induct on n. By taking $d_n = 0$, we have $\sigma'_n = \sigma_n(d_1, \dots, d_{n-1}, 0) = 0$ and $p(\sigma'_1, \dots, \sigma'_{n-1}, 0) = 0$ as a polynomial of $d_1, \dots, d_{n-1}$. Then by induction, we get $p(\sigma'_1, \dots, \sigma'_{n-1}, 0) = 0$ as a polynomial of $\sigma'_1, \dots, \sigma'_{n-1}$. This implies that $p(\sigma_1, \dots, \sigma_{n-1}, 0) = 0$ as a polynomial of $\sigma_1, \dots, \sigma_{n-1}$ and further implies that $p(\sigma_1, \dots, \sigma_{n-1}, \sigma_n) = \sigma_n q(\sigma_1, \dots, \sigma_{n-1}, \sigma_n)$ for some polynomial q. Since $\sigma_n \neq 0$ as a polynomial of $d_1, \dots, d_n$, the assumption $p(\sigma_1, \dots, \sigma_n) = 0$ as a polynomial of $d_1, \dots, d_n$ implies that $q(\sigma_1, \dots, \sigma_n) = 0$ as a polynomial of $d_1, \dots, d_n$. Since q has strictly lower degree than p, a further double induction on the degree of p implies that $q = 0$ as a polynomial of $\sigma_1, \dots, \sigma_n$. □

Exercise 8.6 (Newton's Identity). Consider the symmetric polynomial

$$s_k = d_1^k + \cdots + d_n^k.$$

Explain that for $i = 1, \dots, n$, we have

$$d_i^n - \sigma_1 d_i^{n-1} + \cdots + (-1)^{n-1}\sigma_{n-1} d_i + (-1)^n \sigma_n = 0.$$

Then use the equalities to derive

$$s_n - \sigma_1 s_{n-1} + \cdots + (-1)^{n-1}\sigma_{n-1} s_1 + (-1)^n n \sigma_n = 0.$$

Theorem 8.2 says that polynomial invariants of diagonalisable linear operators are exactly polynomials of the coefficients of characteristic polynomials. To extend the result to general not necessarily diagonalisable linear operators, we establish the fact that any linear operator is a limit of diagonalisable linear operators. First we establish the following.

Proposition 8.3. *If the characteristic polynomial of a linear operator has no repeated root, then the linear operator is diagonalisable.*

Proof. Each d_j has nontrivial eigenspace $\operatorname{Ker}(d_j I - L)$, which contains a nonzero vector $\vec{v}_j$. If the characteristic polynomial has no repeated root, then $d_1, \dots, d_n$ are distinct. By Proposition 7.2, we have n linearly independent eigenvectors $\vec{v}_1, \dots, \vec{v}_n$. Since $n = \dim V$, this is a basis of eigenvectors. $\square$

Example 8.1. The matrix .... has five distinct eigenvalues and is therefore digonalisable.

Proposition 8.4. *Any linear operator is the limit of a sequence of diagonalisable linear operators.*

Proof. By Proposition 8.3, the proposition is the consequence of the claim that any linear operator is approximated by a linear operator such that the characteristic polynomial has no repeated root. We will prove the claim by inducting on the dimension of the vector space. The claim is clearly true for linear operators on 1-dimensional vector spaces.

By Proposition 7.4, any linear operator L has an eigenvalue λ. Let $E = \operatorname{Ker}(\lambda I - L)$ be the corresponding eigenspace. Then we have $V = H \oplus E$ for some subspace H, and in the blocked form, we have

$$L = \begin{pmatrix} K & O \\ * & \lambda I \end{pmatrix}, \quad K \colon H \to H,\ I \colon E \to E.$$

By induction on dimension, K is approximated by an operator K', such that $\det(tI - K')$ has no repeated root. Moreover, we may approximate λI by the diagonal matrix

$$T = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{pmatrix}, \quad r = \dim E,$$

such that λ_i are very close to λ, are distinct, and are not roots of $\det(tI - K')$. Then

$$L' = \begin{pmatrix} K' & O \\ * & T \end{pmatrix}$$

approximates L, and by Exercise 7.1, has $\det(tI - L') = (t - \lambda_1) \cdots (t - \lambda_r) \det(tI - K')$. By our setup, the characteristic polynomial of L' has no repeated root. $\square$

Since polynomials are continuous, by using Proposition 8.4 and taking the limit, we may extend Theorem 8.2 to all linear operators.

Theorem 8.5. *Polynomial invariants of linear operators are exactly polynomial functions of the coefficients of the characteristic polynomial.*

Note that a key ingredient in the proof of the theorem is the continuity of the invariant. The theorem cannot be applied to invariants such as the rank because it is not continuous

$$\operatorname{rank}\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} = \begin{cases} 1 & \text{if } a \neq 0, \\ 0 & \text{if } a = 0. \end{cases}$$

Exercise 8.7. Identify the traces of powers L^k of linear operators with the symmetric functions s_n in Exercises 8.6. Then use Theorem 8.5 and Newton's identity to show that polynomial invariants of linear operators are exactly polynomials of $\mathrm{tr}L^k$, $k = 1, 2, \dots$.

8.2 Cayley-Hamilton Theorem

Let $p(t) = \det(tI - L)$ be the characteristic polynomial of L. Let $q(t) = (t - \lambda_1)\cdots(t - \lambda_k)$ be the product of monomials for all the distinct eigenvalues. Then $q(t)$ divides $p(t)$.

Suppose L is diagonalisable. Then we have $L(\vec{v}_i) = \lambda_i \vec{v}_i$ for a basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$. By Propsoition 8.1, this implies $q(L)(\vec{v}_i) = q(\lambda_i)\vec{v}_i = 0\vec{v}_i = \vec{0}$. Since $q(L)$ vanishes on a basis, we get $q(L) = O$. Since $q(t)$ divides $p(t)$, we get $p(L) = O$.

Theorem 8.6 (Cayley-Hamilton Theorem)**.** *Let $p(t)$ be the characteristic polynomial of a linear operator L. Then $p(L) = O$.*

For 2×2 matrix, the theorem says

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + (ad - bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

On can imagine that a direct calculational proof woiuld be very complicated.

Proof. We just proved

$$p(L) = L^n - \sigma_1(L)L^{n-1} + \cdots + (-1)^{n-1}\sigma_{n-1}(L)L + (-1)^n\sigma_n(L)I = O$$

for diagonalisable L. Since $\sigma_k(L)$ are polynomials of L and are therefore continuous in L, by Proposition 8.4 and taking the limit of the equality above, we find that the equality still holds for any linear operator L. □

The proof above makes use of analysis and requires Proposition 8.4. Since the Cayley-Hamilton Theorem is a purely algebraic statement that is supposed to be true even in some general context where the limit is not defined, the proof is not quite satisfactory. A more satisfactory proof that does not use limit (but still uses complex numbers) is the following[1].

[1] For a more general proof that does not use complex numbers, see *A combinatorial proof of the Cayley-Hamilton theorem* by Howard Straubing, in Discrete Mathematics, vol 43, p273-279, 1983.

Alternative proof. In the proof of Proposition 8.4, we start with an eigenspace $E = \mathrm{Ker}(\lambda I - L)$ of L. Then with respect to a direct sum decomposition $V = H \oplus E$, we have

$$L = \begin{pmatrix} K & O \\ * & \lambda I \end{pmatrix}, \quad K : H \to H,\ I : E \to E.$$

By Exercise 7.1, we have $p(t) = (t-\lambda)^{\dim E} q(t)$, with $q(t) = \det(tI - K)$.

Since H has smaller dimension than V, by induction, we have $q(K) = O$. This means

$$q(L) = \begin{pmatrix} q(K) & O \\ * & q(\lambda I) \end{pmatrix} = \begin{pmatrix} O & O \\ * & q(\lambda) I \end{pmatrix},$$

and implies $q(L)(\vec{v}) \in E$ for any $\vec{v} \in V$. Since E is the eigenspace with eigenvalue λ, we get

$$(\lambda I - L) q(L)(\vec{v}) = \vec{0}.$$

Therefore

$$p(L)(\vec{v}) = (\lambda I - L)^{\dim E} q(L)(\vec{v}) = \vec{0}.$$

□

The equality $p(t) = (t-\lambda)^{\dim E} q(t)$ in the proof above shows that $(t-\lambda)^{\dim E}$ is a factor of $p(t)$. This implies $\dim E$ is no more than the multiplicity of λ in the characteristic polynomial $p(t)$. We call $\dim E = \dim \mathrm{Ker}(\lambda I - L)$ the *geometric multiplicity* of λ.

Proposition 8.7. *The dimension of an eigenspace is no more than the multiplicity of the eigenvalue in the characteristic polynomial. In other words, the geometric multiplicity is no more than the algebraic multiplicity.*

The sum of algebraic multiplicities of all the eigenvalues is the dimension n of the vector space. On the other hand, the linear operator is diagonalisable if and only if the sum of the dimensions of all the eigenspaces is n. By Proposition, therefore, a linear operator is diagonalisable if and only if the geometric multiplicity and the algebraic multiplicity of any eigenvalue are the same.

8.3 Jordan Canonical Form

We study a linear operator L by thinking in terms of the algebra $\mathbb{C}[L]$ of polynomials of L. We will use the algebra of polynomials over fields to help our study. The key fact we will use is the factors of polynomials. Specifically, a collection of polynomials $p_1(t), \dots, p_m(t)$ have greatest common divisor $q(t)$, which is characterised by the property that a polynomial divides each of $p_1(t), \dots, p_m(t)$ if and only if it divides $q(t)$. This fact is analogous to the greatest common divisor of integers. Similar to integers, the greatest common divisor can be expressed as

$$q(t) = p_1(t)u_1(t) + \cdots + p_m(t)u_m(t),$$

for some polynomials $u_1(t), \dots, u_m(t)$. One way to find the greatest common divisor is by the Euclidean algorithm that repeatedly divides among the given polynomials. The process can also be used to find the expression above. A special case is when $p_1(t),\dots,p_m(t)$ are coprime, or $q(t)=1$ is the greatest common divisor. In this case, we get

$$1 = p_1(t)u_1(t) + \cdots + p_m(t)u_m(t). \tag{8.3}$$

The characteristic polynomial of a (complex) linear operator $L\colon V \to V$ has complex root decomposition

$$p(t) = \det(tI - L) = (t-\lambda_1)^{n_1}\cdots(t-\lambda_k)^{n_k}, \quad \lambda_1,\dots,\lambda_k \text{ distinct}.$$

Let

$$p_j(t) = \frac{p(t)}{(t-\lambda_j)^{n_j}} = (t-\lambda_1)^{n_1}\dots(t-\lambda_{j-1})^{n_{j-1}}(t-\lambda_{j+1})^{n_{j+1}}\dots(t-\lambda_j)^{n_k}.$$

Since the polynomials $p_1(t),\dots,p_k(t)$ do not share common root, they are coprime and we get (8.3) for suitable polynomials $u_j(t)$. Applying the equality to L, we get

$$I = p_1(L)u_1(L) + \cdots + p_k(L)u_k(L),$$

or any $\vec{v}$ has decomposition

$$\vec{v} = p_1(L)u_1(L)(\vec{v}) + \cdots + p_k(L)u_k(L)(\vec{v}).$$

By the Cayley-Hamilton Theorem (Theorem 8.6), we have

$$(\lambda_j I - L)^{n_j}[p_j(L)u_j(L)(\vec{v})] = (-1)^{n_j}p(L)u_j(L)(\vec{v}) = \vec{0}.$$

Therefore $p_j(L)u_j(L)(\vec{v}) \in \operatorname{Ker}(\lambda_j I - L)^{n_j}$, and we just proved

$$V = \operatorname{Ker}(\lambda_1 I - L)^{n_1} + \cdots + \operatorname{Ker}(\lambda_k I - L)^{n_k}.$$

Next we prove that the sum is direct. This generalises Proposition 7.2. Suppose

$$\vec{v}_1 + \cdots + \vec{v}_k = \vec{0}, \quad \vec{v}_j \in \operatorname{Ker}(\lambda_j I - L)^{n_j}.$$

For any $i \neq j$, $p_j(t)$ has factor $(t-\lambda_i)^{n_i}$. Therefore $p_j(L)$ has factor $(\lambda_i I - L)^{n_i} = (-1)^{n_i}(L - \lambda_i I)^{n_i}$, and $\vec{v}_i \in \operatorname{Ker}(\lambda_i I - L)^{n_i}$ implies $p_j(L)(\vec{v}_i) = \vec{0}$. Applying $p_j(L)$ to the equality above then gives

$$p_j(L)(\vec{v}_j) = \vec{0}.$$

On the other hand, we also know $(\lambda_j I - L)^{n_j}(\vec{v}_j) = \vec{0}$. Since $p_j(t)$ and $(t-\lambda_j)^{n_j}$ are coprime (because of no common root), we have $u(t)p_j(t) + v(t)(t-\lambda_j)^{n_j} = 1$ for some polynomials $u(t), v(t)$. Then we conclude

$$\vec{v}_j = u(L)p_j(L)(\vec{v}_j) + v(L)(L-\lambda_j I)^{n_j}(\vec{v}_j) = \vec{0}.$$

This completes the proof of the direct sum.

By taking $(\lambda_j I - L)^{n_j}$ and L as L and K in Lemma 7.8 (and taking eigenvalue 0), we find that $\text{Ker}(\lambda_j I - L)^{n_j}$ is an invariant subspace of L. Then with respect to the direct sum

$$V = \text{Ker}(\lambda_1 I - L)^{n_1} \oplus \cdots \oplus \text{Ker}(\lambda_k I - L)^{n_k},$$

we have

$$L = \begin{pmatrix} \lambda_1 I + T_1 & & O \\ & \ddots & \\ O & & \lambda_k I + T_k \end{pmatrix}, \tag{8.4}$$

where $T_j = L - \lambda_j I$ is an operator on $\text{Ker}(\lambda_1 I - L)^{n_1}$ satisfying $T^{n_j} = O$.

Definition 8.8. A linear operator $T\colon V \to V$ is *nilpotent* if $T^m = O$ for some m.

Exercise 8.8. Prove that the only eigenvalue of a nilpotent linear operator is 0. Then prove that it is always possible to have $m \le \dim V$.

Exercise 8.9. Show that the matrix shifts the coordinates by one position

$$A = \begin{pmatrix} 0 & 1 & & & O \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ O & & & & 0 \end{pmatrix} : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} \to \begin{pmatrix} 0 \\ x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{pmatrix}.$$

Then explain that $A^n = O$ and $A^{n-1} \ne O$.

The decomposition (8.4) reduces the study of the structure of a general linear operator to the study of a nilpotent operator $T\colon V \to V$.

We always have $\text{Ker}T^{s+1} \supset \text{Ker}T^s$. If $\text{Ker}T^{s+1} = \text{Ker}T^s$, then

$$\vec{v} \in \text{Ker}T^{s+2} \iff T(\vec{v}) \in \text{Ker}T^{s+1} = \text{Ker}T^s \iff \vec{v} \in \text{Ker}T^{s+1}.$$

This shows that, if m is the smallest number such that $T^m = O$, then we have a *filtration* of subspaces

$$V = V_m = \text{Ker}T^m \supsetneq V_{m-1} = \text{Ker}T^{m-1} \supsetneq \cdots \supsetneq V_1 = \text{Ker}T \supsetneq \{\vec{0}\}. \tag{8.5}$$

We also note that $T^s(V_l) \subset V_{l-s}$.

Let $G_m \subset V_m$ be a subspace such that $V_m = G_m \oplus V_{m-1}$ (i.e., $G_m = V_m \ominus V_{m-1}$ is the "gap" between V_m and V_{m-1}, see Exercise **??**). Then we have $T^s(G_m) \subset V_{m-s}$. We claim the following properties for $1 \le s \le m-1$.

1. $T^s\colon G_m \to T^s(G_m)$ is an isomorphism.
2. $T^s(G_m) + V_{m-s-1}$ is a direct sum in V_{m-s}.

Therefore we get isomorphic copies of G_m inside the gaps between consecutive subspaces in the filtration (8.5)

$$G_m \cong T(G_m) \cong \cdots \cong T^{m-1}(G_m), \quad T^s(G_m) \oplus V_{m-s-1} \subset V_{m-s}.$$

For the first property, we note that the linear transformation is onto. The linear transformation is also one-to-one because it has trivial kernel

$$\text{Ker}(T^s \colon G_m \to T^s(G_m)) = G_m \cap \text{Ker}T^s \subset G_m \cap \text{Ker}T^{m-1} = \{\vec{0}\}.$$

Here the last equality uses the direct sum $V_m = G_m \oplus V_{m-1} = G_m \oplus \text{Ker}T^{m-1}$. The second property follows from

$$T^s(G_m) \cap V_{m-s-1} = T^s(G_m) \cap \text{Ker}T^{m-s-1} = T^s(G_m \cap \text{Ker}T^{m-1}) = T^s(\{\vec{0}\}) = \{\vec{0}\}.$$

Here the second equality uses Exercise 3.32, and the third equality uses the direct sum $G_m \oplus \text{Ker}T^{m-1}$.

We remove the first gap G_m in the filtration (8.5) and its isomorphic copies $T^s(G_m)$ between other consecutive spaces in the filtration. Then we consider the remaining gap between V_{m-1} and $T(G_m) \oplus V_{m-2}$ by introducing a subspace $G_{m-1} \subset V_{m-1}$ satisfying $V_{m-1} = G_{m-1} \oplus T(G_m) \oplus V_{m-2}$. Then we establish similar properties for $2 \le s \le m-1$.

1. $T^{s-1} \colon G_{m-1} \to T^{s-1}(G_{m-1})$ is an isomorphism.

2. $T^{s-1}(G_{m-1}) + (T^s(G_m) \oplus V_{m-s-1})$ is a direct sum in V_{m-s}.

Therefore we get isomorphic copies of G_{m-1} inside the remaining gaps between consecutive subspaces in the filtration (8.5)

$$G_{m-1} \cong T(G_{m-1}) \cong \cdots \cong T^{m-2}(G_{m-1}), \quad T^{s-1}(G_{m-1}) \oplus T^s(G_m) \oplus V_{m-s-1} \subset V_{m-s}.$$

The proof of the first property is the same as before

$$\text{Ker}(T^{s-1} \colon G_{m-1} \to T^{s-1}(G_{m-1})) = G_{m-1} \cap \text{Ker}T^{s-1} \subset G_{m-1} \cap \text{Ker}T^{m-2} = \{\vec{0}\}.$$

For the second property, we consider $\vec{v} \in G_{m-1}$ and $T^{s-1}(\vec{v}) \in T^s(G_m) \oplus V_{m-s-1}$. We have

$$T^{s-1}(\vec{v}) = T^s(\vec{w}) + \vec{u}, \quad \vec{w} \in G_m, \quad \vec{u} \in V_{m-s-1} = \text{Ker}T^{m-s-1}.$$

Then

$$T^{m-2}(\vec{v} - T(\vec{w})) = T^{m-s-1}(T^{s-1}(\vec{v}) - T^s(\vec{w})) = T^{m-s-1}(\vec{u}) = \vec{0}.$$

This implies $\vec{v} - T(\vec{w}) \in V_{m-2}$, or $\vec{v} \in T(G_m) + V_{m-2}$. Since $\vec{v} \in G_{m-1}$ and $G_{m-1} \oplus T(G_m) \oplus V_{m-2}$ is a direct sum, we conclude that $\vec{v} = \vec{0}$, so that $T^s(\vec{v}) = \vec{0}$.

We may continue the construction by finding G_{m-2} satisfying $V_{m-2} = G_{m-2} \oplus T(G_{m-1}) \oplus T^2(G_m) \oplus V_{m-3}$, and so on. At the end, we get the following arrays of subspaces.

$$\begin{array}{cccccc}
V = V_m \supsetneq V_{m-1} \supsetneq & V_{m-2} \supsetneq & \cdots \supsetneq & V_1 \supsetneq & \{\vec{0}\} \\
G_m \quad T(G_m) & T^2(G_m) & \cdots\ T^{m-2}(G_m) & T^{m-1}(G_m) & \\
G_{m-1} & T(G_{m-1}) & \cdots\ T^{m-3}(G_{m-1}) & T^{m-2}(G_{m-1}) & \\
 & G_{m-2} & \cdots\ T^{m-4}(G_{m-2}) & T^{m-3}(G_{m-2}) & \\
 & & \vdots & \vdots & \\
 & & G_2 & T(G_2) & \\
 & & & G_1 &
\end{array}$$

The spaces in each row are isomorphic by powers of T, and $T = O$ on the subspaces in the last column (the direct sum of subspaces in the last column is exactly $V_1 = \text{Ker}T$). The subspaces in each column form a direct sum that fills the gap between consecutive spaces in the filtration

$$V_s = G_s \oplus T(G_{s+1}) \oplus T^2(G_{s+2}) \oplus \cdots \oplus T^{m-s-1}(G_{m-1}) \oplus T^{m-s}(G_m) \oplus V_{m-s-1}, \quad 1 \le s \le m.$$

The whole space is the direct sum of all the subspaces in the array

$$V = \oplus_{0 \le r < s \le m} T^r(G_s).$$

Pick one basis α_s for each G_s. The union $\beta = \alpha_1 \cup \cdots \cup \alpha_m$ is a basis of $G_1 \oplus \cdots \oplus G_m$. Moreover, for any $\vec{v} \in \alpha_s$, we get linear independent vectors

$$\vec{v},\ T(\vec{v}),\ \ldots,\ T^{s-1}(\vec{v}). \tag{8.6}$$

The union of the collection (8.6) for all $\vec{v} \in \alpha_k$ form a basis $\alpha_s \cup T(\alpha_s) \cup \cdots \cup T^{s-1}(\alpha_s)$ of $G_s \oplus T(G_s) \oplus \cdots \oplus T^{s-1}(G_s)$. The union of (8.6) for all $\vec{v} \in \beta$ form a basis of V

$$\gamma = \cup_{s=1}^m (\alpha_s \cup T(\alpha_s) \cup \cdots \cup T^{s-1}(\alpha_s)) = \cup_{0 \le r < s \le m} T^r(\alpha_s).$$

The subspace spanned by (8.6) is T-invariant, and the matrix of T with respect to the basis (8.6) of the subspace is

$$[T]_{\{\vec{v}, T(\vec{v}), \ldots, T^{s-1}(\vec{v})\}} = \begin{pmatrix} 0 & 1 & & & O \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ O & & & & 0 \end{pmatrix}.$$

The matrix of T with respect to the basis γ of V is diagonal the matrix above as diagonal blocks

$$[T]_\gamma = \oplus_{\vec{v} \in \beta} [T]_{\{\vec{v}, T(\vec{v}), \ldots, T^{s-1}(\vec{v})\}}.$$

Exercise 8.10. What is $\mathrm{Ker}T^2$? What is $\mathrm{Ran}T$?

We apply the structure of a nilpotent operator to one block in the decomposition (8.4) of a general linear operator L. In this case, the nilpotent operator is the restriction of $T = L - \lambda_j I$ on $V = \mathrm{Ker}(\lambda_j I - L)^{n_j}$. Then vectors (8.6) become

$$\vec{v},\ (L - \lambda_j I)(\vec{v}),\ \ldots,\ (L - \lambda_j I)^{s-1}(\vec{v}). \tag{8.7}$$

The vectors span an invariant subspace of L, and the matrix of L with respect to the basis (8.7) of the invariant subspace is

$$[L] = [\lambda_j I + T] = \begin{pmatrix} \lambda_j & 1 & & & O \\ & \lambda_j & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_j & 1 \\ O & & & & \lambda_j \end{pmatrix}. \tag{8.8}$$

The matrix is called a *Jordan block*.

Theorem 8.9 (Jordan Canonical Form). *Any complex linear operator has the matrix*

$$[L] = \begin{pmatrix} J_1 & & O \\ & \ddots & \\ O & & J_l \end{pmatrix}, \quad J_* \text{ is a Jordan block,}$$

with respect to a basis.

Exercise 8.11. In terms of Jordan canonical form, what is the condition for diagonalisability?

Exercise 8.12. Prove that $\dim \mathrm{Ker}(\lambda_j I - L)^{n_j} = n_j$ is the algebraic multiplicity of λ_j.

Exercise 8.13. Prove that the geometric multiplicity $\dim \mathrm{Ker}(\lambda_j I - L)$ is the number of Jordan blocks with eigenvalue λ_j.

Exercise 8.14. Prove that A and A^T are similar.

Exercise 8.15. Compute the powers of a Jordan block. Then compute the exponential of a Jordan block.

8.4 Minimal Polynomial

If $p(L) = O$ and $q(L) = O$, then the greatest common divisor $r(t) = p(t)u(t) + q(t)v(t)$ for some polynomials $u(t), v(t)$, and we get

$$r(L) = p(L)u(L) + q(L)v(L) = O.$$

This implies that there is a unique monic polynomial, called *minimal polynomial*

$$m(t) = t^m + a_{m-1}t^{m-1} + \cdots + a_1 t + a_0,$$

such $p(L) = O$ if and only if $m(t)$ divides $p(t)$. The degree m is the smallest among polynomials $p(t)$ satisfying $p(L) = O$.

The Cayley-Hamilton Theorem (Theorem 8.6) says that the minimal polynomial divides the characteristic polynomial, so that

$$m(t) = (t - \lambda_1)^{m_1} \cdots (t - \lambda_k)^{m_k}, \quad 1 \le m_j \le n_j. \tag{8.9}$$

For a blocked diagonal matrix of linear operators, we have

$$L = \begin{pmatrix} L_1 & & \\ & \ddots & \\ & & L_k \end{pmatrix}, \quad p(L) = \begin{pmatrix} p(L_1) & & \\ & \ddots & \\ & & p(L_k) \end{pmatrix}.$$

Therefore $p(L) = O$ if and only if $p(L_1) = O, \dots, p(L_k) = O$. In other words, $p(t)$ is divisible by all the minimal polynomials of $L_1, \dots, L_k$. Therefore the minimal polynomial of L is the least common multiple of the minimal polynomials of $L_1, \dots, L_k$.

For the special case L is diagonalisable, the discussion about shows that the minimal polynomial is

$$m(t) = (t - \lambda_1) \cdots (t - \lambda_k), \quad \text{or all } m_j = 1.$$

By Exercise 8.9, each $m \times m$ Jordan block of eignevalue λ has minimal polynomial $(t-\lambda)^m$. By the Jordan canonical form of a general linear operator L, therefore, the exponent m_j in the minimal polynomial of L is the size of the largest Jordan block of eigenvalue λ_j. Since this size is the length of the filtration (8.5)

$$\cdots \supsetneq \mathrm{Ker}(\lambda_j I - L)^l \supsetneq \mathrm{Ker}(\lambda_j I - L)^{l-1} \supsetneq \cdots \supsetneq \mathrm{Ker}(\lambda_j I - L) \supsetneq \{\vec{0}\},$$

the exponent m_j is the smallest number such that $\mathrm{Ker}(\lambda_j I - L)^{m_j+1} = \mathrm{Ker}(\lambda_j I - L)^{m_j}$.

8.5 Rational Canonical Form

The construction of Jordan canonical form makes critical use of the complete decomposition (8.1) using complex roots. For a real eigenvalue λ of a real matrix A, the subspace $\mathrm{Ker}(\lambda I - A)^m$ is real, and we may choose vectors in (8.7) to be real. Then the corresponding Jordan block is a real matrix.

For a complex eigenvalue $\lambda = a + ib$, $b \neq 0$, corresponding to complex vectors in 8.7

$$\vec{v} = \vec{u}_1 - i\vec{w}_1,\ (A - \lambda I)(\vec{v}) = \vec{u}_2 - i\vec{w}_2,\ \dots,\ (A - \lambda I)^{s-1}(\vec{v}) = \vec{u}_s - i\vec{w}_s,$$

we have conjugate complex eigenvalue $\bar{\lambda} = a - ib$ and the corresponding complex vectors

$$\bar{\vec{v}} = \vec{u}_1 + i\vec{w}_1,\ (A - \bar{\lambda} I)(\bar{\vec{v}}) = \vec{u}_2 + i\vec{w}_2,\ \dots,\ (A - \bar{\lambda} I)^{s-1}(\bar{\vec{v}}) = \vec{u}_s + i\vec{w}_s.$$

The two sets of vectors together are complex linearly equivalent (i.e., written as linear combination of each other) to the set of real vectors

$$\vec{u}_1,\ \vec{w}_1,\ \vec{u}_2,\ \vec{w}_2,\ \dots,\ \vec{u}_s,\ \vec{w}_s. \tag{8.10}$$

By $(A-\lambda I)(\vec{u}_r - i\vec{w}_r) = \vec{u}_{r+1} - i\vec{w}_{r+1}$, we have

$$\begin{aligned} A\vec{u}_r - iA\vec{w}_r &= (a+ib)(\vec{u}_r - i\vec{w}_r) + \vec{u}_{r+1} - i\vec{w}_{r+1} \\ &= (a\vec{u}_r + b\vec{w}_r + \vec{u}_{r+1}) - i(-b\vec{u}_r + a\vec{w}_r + \vec{w}_{r+1}). \end{aligned}$$

This shows that the *real Jordan block* of A with respect to the vectors (8.10) is

$$J = \begin{pmatrix} a & -b & 1 & 0 & & & & & & \\ b & a & 0 & 1 & & & & & & \\ & & a & -b & 1 & 0 & & & & \\ & & b & a & 0 & 1 & & & & \\ & & & & \ddots & & \ddots & & & \\ & & & & & a & -b & 1 & 0 \\ & & & & & b & a & 0 & 1 \\ & & & & & & & a & -b \\ & & & & & & & b & a \end{pmatrix}.$$

The real Jordan canonical form of the real matrix A is then a direct sum of the usual Jordan block (8.8) for real eigenvalues and the real Jordan block above for conjugate pairs of complex eigenvalues.

Jordan canonical form is not the only canonical form.

.... cyclic subspace ....

9 Tensor

9.1 Dual Space

A *linear functional* on a vector space is a number valued linear transformation $l\colon V \to \mathbb{R}$. In other words, l is a function satisfying

$$l(\vec{u}+\vec{v}) = l(\vec{u}) + l(\vec{v}), \quad l(a\vec{u}) = a\, l(\vec{u}).$$

Since addition and scalar multiplication of linear functionals are still linear, all the linear functionals form a vector space V^*, called the *dual space*.

Example 9.1. Linear functionals on $\mathbb{R}^n$ are of the form (using the $1 \times n$ matrix of l)

$$l(\vec{x}) = (a_1\ \cdots\ a_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = a_1x_1 + \cdots + a_nx_n = \vec{a}\cdot\vec{x}.$$

The one-to-one correspondence $l \leftrightarrow \vec{a}$ is an isomorphism between the dual space $(\mathbb{R}^n)^*$ and the Euclidean space itself. We say the Euclidean space is *self dual*.

Example 9.2. Consider the space of functions on $\mathbb{R}$. The evaluation at a point $f \mapsto f(t_0)$ is a linear functional. The integration on a fixed interval $f \mapsto \int_a^b f(t)dt$ is also a linear functional.

The function space is infinite dimensional. The dual space needs to take into account of the topology. For the vector space of power p integrable functions,

$$L_p[a,b] = \{f(t)\colon \int_a^b |f(t)|^p dt < \infty\}, \quad p \geq 1,$$

the *continuous* linear functionals $L_p[a,b]$ are of the form

$$l(f) = \int_a^b f(t)g(t)dt$$

for a function $g(t)$ satisfying

$$\int_a^b |g(t)|^q dt < \infty, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

This shows that the dual space $L_p[a,b]^*$ of all continuous linear functionals on $L_p[a,b]$ is $L_q[a,b]$. In particular, the *Hilbert space* $L_2[a,b]$ of square integrable functions is self dual.

Exercise 9.1. Prove that $(V \oplus W)^* = V^* \oplus W^*$.

A basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ of V has a *dual basis* $\alpha^* = \{\vec{v}_1^*, \dots, \vec{v}_n^*\}$ of V^*, by taking $\vec{v}_i^*$ to be the i-th α-coordinate

$$\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n \mapsto \vec{v}_i^*(\vec{x}) = x_i.$$

In other words, the dual basis is characterised by

$$\vec{x} = \vec{v}_1^*(\vec{x})\vec{v}_1 + \cdots + \vec{v}_n^*(\vec{x})\vec{v}_n, \quad [\vec{x}]_\alpha = (\vec{v}_1^*(\vec{x}) \ \cdots \ \vec{v}_n^*(\vec{x}))^T.$$

Moreover, α^* is indeed a basis of V^* because any linear functional l has unique linear expression

$$l = l(\vec{v}_1)\vec{v}_1^* + \cdots + l(\vec{v}_n)\vec{v}_n^*, \quad [l]_{\alpha^*} = (l(\vec{v}_1) \ \cdots \ l(\vec{v}_n))^T.$$

In particular, we conclude

$$\dim V^* = \dim V.$$

Example 9.3. For the standard basis $\epsilon = \{\vec{e}_1, \dots, \vec{e}_n\}$ of $\mathbb{R}^n$. The dual basis $\epsilon^* = \{\vec{e}_1^*, \dots, \vec{e}_n^*\}$ of $\mathbb{R}^n$ is given by $\vec{e}_i^*(\vec{x}) = x_i$. In other words, $\vec{e}_i^*$ is the i-th coordinate.

Example 9.4. A polynomial p of degree $\le n$ is determined by its values $\epsilon_i(p) = p(t_i)$ at $n+1$ distinct places $t_0, t_1, \dots, t_n$ (see Exercise 5.8). To find the explicit formula, we note that the evaluations form a basis $\beta = \{\epsilon_0, \epsilon_1, \dots, \epsilon_n\}$ of P_n^* (see Exercise 9.2). The basis would be the dual basis $\alpha = \{p_0, p_1, \dots, p_n\}$ of P_n if

$$p_i(t_j) = \epsilon_j(p_i) = \begin{cases} 0 & \text{if } i \ne j, \\ 1 & \text{if } i = j. \end{cases}$$

The polynomials

$$p_i(t) = \frac{(t-t_0)\cdots(t-t_{i-1})(t-t_{i+1})\cdots(t-t_n)}{(t_i-t_0)\cdots(t_i-t_{i-1})(t_i-t_{i+1})\cdots(t_i-t_n)}$$

have degree $\le n$ and clearly satisfies the requirement. The dual basis gives the following *interpolation formula* for polynomials of degree $\le n$

$$\begin{aligned} p(t) &= \epsilon_0(p)p_0(t) + \epsilon_1(p)p_1(t) + \cdots + \epsilon_n(p)p_n(t) \\ &= \sum_{i=0}^{n} \frac{(t-t_0)\cdots(t-t_{i-1})(t-t_{i+1})\cdots(t-t_n)}{(t_i-t_0)\cdots(t_i-t_{i-1})(t_i-t_{i+1})\cdots(t_i-t_n)} p(t_i). \end{aligned}$$

Exercise 9.2. Prove that linear functionals $l_1, \dots, l_n$ form a basis of the dual space V^* if and only if the related linear transformation

$$L(\vec{x}) = (l_1(\vec{x}) \ \cdots \ l_n(\vec{x}))^T : V \to \mathbb{R}^n$$

is an isomorphism. What is the condition for the linear functionals to be linearly independent?

Theorem 9.1. *Suppose V is a finite dimensional vector space. Then*

$$\vec{x} \mapsto \vec{x}^{**}, \quad \vec{x}^{**}(l) = l(\vec{x})$$

is an isomorphism $V \cong (V^*)^*$.

The key idea for the theorem is that the value $l(\vec{x})$ is the combination of two ingredients l and $\vec{x}$, just like $\sin\pi = 0$ means that 0 is the combination of sin and π. We may consider both ingredients as variables, say both sin and π are variables in the combination $\sin\pi$. In this way, we may also fix one variable π and view the number as a "function" of the the other variable sin, by writing $\pi^{**}(\sin) = \sin\pi$. The "function" π can also be applied to cos and gives $\pi^{**}(\cos) = \cos\pi = -1$. This thinking foreshadows the upcoming concept of bilinear functions.

Proof. The following shows that $\vec{x}^{**}$ is indeed a linear functional on V^* and therefore belongs to the double dual $(V^*)^*$

$$\vec{x}^{**}(al + bl') = (al + bl')(\vec{x}) = al(\vec{x}) + bl'(\vec{x}) = a\vec{x}^{**}(l) + b\vec{x}^{**}(l').$$

The following shows that $\vec{x}^{**}$ is linear

$$(a\vec{x} + b\vec{y})^{**}(l) = l(a\vec{x} + b\vec{y}) = al(\vec{x}) + bl(\vec{y}) = a\vec{x}^{**}(l) + b\vec{y}^{**}(l) = (a\vec{x}^{**} + b\vec{y}^{**})(l).$$

The following shows that $\vec{x} \mapsto \vec{x}^{**}$ is injective

$$\vec{x}^{**} = 0 \implies l(\vec{x}) = \vec{x}^{**}(l) = 0 \text{ for all } l \in V^* \implies \vec{x} = \vec{0}.$$

In the second implication, if $\vec{x} \neq \vec{0}$, then we can expand $\vec{x}$ to a basis and then construct a linear functional l such that $l(\vec{x}) \neq 0$.

Since $\dim (V^*)^* = \dim V^* = \dim V$, the injective linear transformation $\vec{x} \mapsto \vec{x}^{**}$ is an isomorphism. $\square$

A linear transformation $L\colon V \to W$ induces the *dual transformation* in the opposite direction by simply sending $l \in W^*$ to $l \circ L \in V^*$

$$L^*\colon V^* \leftarrow W^*, \quad L^*(l)(\vec{x}) = l(L(\vec{x})).$$

It is easy to show that $L^*(l)$ is still linear, and the dual transformation satisfies

$$(aL + bL')^* = aL^* + bL'^*, \quad (L \circ L')^* = L'^* \circ L^*.$$

Example 9.5. The evaluation of functions at three locations is a linear transformation

$$L(f) = (f(1)\ f(2)\ f(3))^T\colon V \to \mathbb{R}^3.$$

The dual map $L^*\colon (\mathbb{R}^3)^* \to V^*$ is given by

$$l(x\ y\ z)^T = ax + by + cz \mapsto L^*(l)(f) = l(f(1)\ f(2)\ f(3))^T = af(1) + bf(2) + cf(3).$$

A linear combination of values of function at several places is still a linear functional.

Exercise 9.3. Prove that $(L^*)^*$ corresponds to L under the isomorphism given by Theorem 9.1.

Exercise 9.4. Prove that L is injective if and only if L^* is surjective, and L is surjective if and only if L^* is injective.

Exercise 9.5. Prove that $[L^*]_{\alpha^*\beta^*} = [L_{\beta\alpha}]^T$.

9.2 Bilinear Function

A *bilinear function* is a function $b\colon V \times W \to \mathbb{R}$ that is linear in both variables

$$b(\lambda\vec{v} + \mu\vec{v}', \vec{w}) = \lambda b(\vec{v}, \vec{w}) + \mu b(\vec{v}', \vec{w}), \quad b(\vec{v}, \lambda\vec{w} + \mu\vec{w}') = \lambda b(\vec{v}, \vec{w}) + \mu b(\vec{v}, \vec{w}').$$

Any inner product on a (real) vector space V is a bilinear form on $V \times V$.

Let $\alpha = \{\vec{v}_1, \ldots, \vec{v}_m\}$ and $\beta = \{\vec{w}_1, \ldots, \vec{w}_n\}$ be bases of V and W. Then the quadratic form can be expressed in terms of the α-coordinate $[\vec{x}]_\alpha = (x_1\ \cdots\ x_m)^T$ and the β-coordinate $[\vec{y}]_\beta = (y_1\ \cdots\ y_n)^T$

$$b(\vec{x}, \vec{y}) = \sum_{i,j} b(\vec{v}_i, \vec{w}_j) x_i y_j = [\vec{x}]_\alpha^T B [\vec{y}]_\beta.$$

This gives a one-to-one correspondence between bilinear functions and matrices

$$[b]_{\alpha\beta} = B = (b(\vec{v}_i, \vec{w}_j))_{i,j=1}^{m,n}.$$

In a bi-function, fixing one variable gives a function of the other variable. Therefore a bilinear function induces two linear transformations

$$V \to W^*\colon \vec{v} \mapsto b(\vec{v}, \cdot), \quad W \to V^*\colon \vec{w} \mapsto b(\cdot, \vec{w}). \tag{9.1}$$

The fact $b(\vec{v}, \cdot) \in W^*$ makes use of the linearity of b in the second variable, and the linearity of $V \to W^*$ makes use of the linearity of b in the first variable.

Conversely, a linear transformation $L\colon V \to W^*$ gives a bilinear form $b(\vec{v}, \vec{w}) = L(\vec{v})(\vec{w})$. Here $L(\vec{v})$ is a linear functional on W and is applied to a vector $\vec{w}$ in W. We have a three way correspondence between bilinear functions on $V \times W$, linear transformations $V \to W^*$, and linear transformations $W \to V^*$.

Exercise 9.6. How is the matrix of a bilinear function changed when the bases are changed?

Exercise 9.7. Prove that the two linear transformations (9.1) are dual to each other, after applying Theorem 9.1.

Definition 9.2. A bilinear function $b\colon V \times W \to \mathbb{R}$ is a *dual pairing* if both induced linear transformations $V \to W^*$ and $W \to V^*$ are isomorphisms.

By Exercise 9.7, $V \to W^*$ if and only if $W \to V^*$ is isomorphic.

A basis α of V and a basis β of W are *dual bases* with respect to the dual pairing if α is mapped to the basis β^* of W^*. This is equivalent to that β is mapped to the basis α^* of V^*, and is also equivalent to

$$b(\vec{v}_i, \vec{w}_j) = \delta_{ij}, \text{ or } [b]_{\alpha\beta} = I.$$

This implies the expression of any $\vec{x} \in V$ in terms of α

$$\vec{x} = b(\vec{x}, \vec{w}_1)\vec{v}_1 + b(\vec{x}, \vec{w}_2)\vec{v}_2 + \cdots + b(\vec{x}, \vec{w}_n)\vec{v}_n,$$

and the similar expression of vectors in W in terms of β.

Example 9.6. The *evaluation pairing* is

$$\langle \vec{x}, l \rangle = l(\vec{x})\colon V \times V^* \to \mathbb{R}.$$

The first isomorphism in (9.1) is the isomorphism in Theorem 9.1. The second isomorphism is the identity on V^*. For a basis α of V, the dual basis with respect to the evaluation pairing is the dual basis α^* in Section 9.1.

Example 9.7. An inner product $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{R}$ is a dual pairing. The induced isomorphism

$$V \cong V^* \colon \vec{a} \mapsto \langle \vec{a}, \cdot \rangle$$

makes V into a *self dual* vector space. The special case of Euclidean space is given by Example 9.1.

In a self dual vector space, it makes sense to ask for self dual basis. This means that a basis α of V is the dual of itself with respect to the inner product. By the definition, this means

$$\langle \vec{v}_i, \vec{v}_i \rangle = 1, \quad \langle \vec{v}_i, \vec{v}_j \rangle = 0 \text{ for } i \neq j.$$

Therefore a self dual basis of an inner product space is exactly an orthonormal basis.

Exercise 9.8. Prove that, if both transformations in (9.1) are injective, then both transformations are isomorphic.

Exercise 9.9. Let α and β be bases of V and W. Let α^* and β^* be dual bases of V^* and W^*. Then for any bilinear function on $V \times W$, we have the matrix of the function, the matrix of linear transformation $V \to W^*$, and the matrix of linear transformation $W \to V^*$, with respect to the relevant bases. How are these matrices related?

Now we additionally assume that both V and W are inner product spaces. Using the isomorphism in Example 9.7, the bilinear function b induces a linear transformation

$$L \colon V \to W^* \cong W, \quad \vec{v} \mapsto b(\vec{v}, \cdot) = \langle L(\vec{v}), \cdot \rangle.$$

This means

$$b(\vec{v}, \vec{w}) = \langle L(\vec{v}), \vec{w} \rangle \text{ for all } \vec{v} \in V, \ \vec{w} \in W. \tag{9.2}$$

Conversely, any linear transformation L gives a bilinear function by the formula above. We have a one-to-one correspondence between bilinear functions and linear transformations.

The relation between b and L above only makes use of the inner product on W. Using the inner product in V, we have the adjoint transformation $\langle L(\vec{v}), \vec{w} \rangle = \langle \vec{v}, L^*(\vec{w}) \rangle$. This shows that the second linear transformation in (9.1) can be identified with the adjoint $L^* \colon W \to V^* \cong V$. The adjoint can also be interpreted as the combination of dual transformation and the isomorphism in Example 9.7

$$L^* \colon W \cong W^* \xrightarrow{L^*} V^* \cong V.$$

9.3 Quadratic Form

A *quadratic form* on a vector space V is $q(\vec{x}) = b(\vec{x}, \vec{x})$, where b is some bilinear function on $V \times V$. By replacing b with the symmetric bilinear function $\frac{1}{2}(b(\vec{x}, \vec{y}) + b(\vec{y}, \vec{x}))$, we may always take b to be symmetric without affecting q. The symmetric b can be recovered from q by *polarisation*

$$b(\vec{x}, \vec{y}) = \frac{1}{4}(q(\vec{x} + \vec{y}) - q(\vec{x} - \vec{y})) = \frac{1}{2}(q(\vec{x} + \vec{y}) - q(\vec{x}) - q(\vec{y})).$$

The matrix of q with respect to a basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ is the matrix of b with respect to the basis, and is symmetric

$$[q]_\alpha = [b]_{\alpha\alpha} = B = (b(\vec{v}_i, \vec{v}_j)).$$

Then the quadratic form can be expressed in terms of α-coordinate $[\vec{x}]_\alpha = (x_1 \ \cdots \ x_n)^T$

$$\begin{aligned} q(\vec{x}) = [\vec{x}]_\alpha^T [q]_\alpha [\vec{x}]_\alpha &= \sum_{1\le i\le n} b_{ii} x_i^2 + 2 \sum_{1\le i<j\le n} b_{ij} x_i x_j \\ &= b_{11}x_1^2 + b_{22}x_2^2 + \cdots + b_{nn}x_n^2 + 2b_{12}x_1x_2 + 2b_{13}x_1x_3 + \cdots + 2b_{(n-1)n}x_{n-1}x_n. \end{aligned}$$

For another basis β, the β-coordinate $[\vec{x}]_\beta = [I]_{\beta\alpha}[\vec{x}]_\alpha$, and

$$q(\vec{x}) = [\vec{x}]_\beta^T [q]_\beta [\vec{x}]_\beta = [\vec{x}]_\alpha^T [I]_{\beta\alpha}^T [q]_\beta [I]_{\beta\alpha} [\vec{x}]_\alpha.$$

Therefore we get

$$[q]_\alpha = [I]_{\beta\alpha}^T [q]_\beta [I]_{\beta\alpha}. \tag{9.3}$$

Since the matrix $[I]_{\beta\alpha}$ for changing basis is invertible, we get $\text{rank}[q]_\alpha = \text{rank}[q]_\beta$. This shows that the *rank* of a quadratic is independent of the basis and is therefore well defined.

Exercise 9.10. Prove that a quadratic form is homogeneous of second order

$$q(c\vec{x}) = c^2 q(\vec{x}),$$

and satisfies the *parallelogram law*

$$q(\vec{x} + \vec{y}) + q(\vec{x} - \vec{y}) = 2q(\vec{x}) + 2q(\vec{y}).$$

If V is an inner product space, then we can use (9.2) to get a one-to-one correspondence between quadratic functions and symmetric linear operators

$$q(\vec{x}) = \langle L(\vec{x}), \vec{y}\rangle = \langle \vec{x}, L(\vec{y})\rangle.$$

By Proposition 7.10 and the subsequent discussion for real symmetric operators, we know L has an orthonormal basis α of eigenvectors. With respect to the basis, we have

$$\langle L(\vec{v}_i), \vec{v}_j\rangle = \langle d_i\vec{v}_i, \vec{v}_j\rangle = \begin{cases} 0 & \text{for } i \ne j, \\ d_i & \text{for } i = j. \end{cases}$$

Then the expression of q with respect to α has no cross terms

$$q(\vec{x}) = d_1 x_1^2 + \cdots + d_n x_n^2. \tag{9.4}$$

The coefficients are the eigenvalues of L.

Proposition 9.3. *Suppose $L \colon V \to V$ is a symmetric operator on an inner product space. Then the maximal eigenvalue of L is $\max_{\|\vec{x}\|=1}\langle L(\vec{x}), \vec{x}\rangle$ and the minimum eigenvalue of L is $\min_{\|\vec{x}\|=1}\langle L(\vec{x}), \vec{x}\rangle$.*

Let $\lambda_{\max}$ and $\lambda_{\min}$ be the maximal and minimal eigenvalues of L. The proposition says

$$\lambda_{\min}\|\vec{x}\|^2 \le \langle L(\vec{x}), \vec{x}\rangle \le \lambda_{\max}\|\vec{x}\|^2.$$

Moreover, the equality is reached at exactly the eigenvectors with $\lambda_{\max}$ or $\lambda_{\min}$ as eigenvalue.

Proof. Let $d_{i_0} = \max\{d_1, \dots, d_n\}$ in the formula (9.4) with respect to an orthonormal basis of eigenvalues. Then we have

$$\langle L(\vec{x}), \vec{x}\rangle \le d_{i_0}(x_1^2 + \cdots + x_n^2) = d_{i_0}\|\vec{x}\|^2.$$

On the other hand, the equality happens when $\vec{x} = \vec{v}_{i_0}$. Therefore

$$d_{i_0} = \max \frac{\langle L(\vec{x}), \vec{x}\rangle}{\|\vec{x}\|^2} = \max\left\langle L\left(\frac{\vec{x}}{\|\vec{x}\|}\right), \frac{\vec{x}}{\|\vec{x}\|}\right\rangle = \max_{\|\vec{x}\|=1}\langle L(\vec{x}), \vec{x}\rangle.$$

An alternative proof is given by the Lagrange multiplier method. We try to find the maximum of the function $f(\vec{x}) = \langle L(\vec{x}), \vec{x}\rangle = \langle \vec{x}, L(\vec{x})\rangle$ subject to the constraint $g(\vec{x}) = \langle \vec{x}, \vec{x}\rangle = 1$. By

$$\begin{aligned} f(\vec{x}_0 + \vec{\delta}) &= \langle L(\vec{x}_0 + \vec{\delta}), \vec{x}_0 + \vec{\delta}\rangle = \langle L(\vec{x}_0), \vec{x}_0\rangle + \langle L(\vec{x}_0), \vec{\delta}\rangle + \langle L(\vec{\delta}), \vec{x}_0\rangle + \langle L(\vec{\delta}), \vec{\delta}\rangle \\ &= f(\vec{x}_0) + \langle L(\vec{x}_0), \vec{\delta}\rangle + \langle \vec{\delta}, L(\vec{x}_0)\rangle + o(\|\vec{\delta}\|) = f(\vec{x}_0) + 2\langle L(\vec{x}_0), \vec{\delta}\rangle + o(\|\vec{\delta}\|), \end{aligned}$$

we get $f'(\vec{x}_0) = 2\langle L(\vec{x}_0), \cdot\rangle$. By the similar reason, we have $g'(\vec{x}_0) = 2\langle \vec{x}_0, \cdot\rangle$. If the extreme of f subject to the constraint $g = 1$ happens at $\vec{x}_0$, then the linear functional $f'(\vec{x}_0)$ is parallel to the linear functional $g'(\vec{x}_0)$. This means that $L(\vec{x}_0) = \lambda\vec{x}_0$ for a scalar λ. Moreover, the maximum is

$$f(\vec{x}_0) = \langle L(\vec{x}_0), \vec{x}_0\rangle = \langle \lambda\vec{x}_0, \vec{x}_0\rangle = \lambda\langle \vec{x}_0, \vec{x}_0\rangle = \lambda g(\vec{x}_0) = \lambda.$$

For any other eigenvalue μ, we have $L(\vec{v}) = \mu\vec{v}$ for a unit length vector $\vec{v}$ and get

$$\mu = \mu\langle \vec{v}, \vec{v}\rangle = \langle \mu\vec{v}, \vec{v}\rangle = \langle L(\vec{v}), \vec{v}\rangle \le \lambda.$$

This proves that λ is the maximum eigenvalue. □

Exercise 9.11. Prove that Proposition 9.3 also holds for Hermitian operators on complex inner product spaces. Moreover, can you find a similar result for normal operators?

The diagonalisation of a symmetric matrix is usually not so easy to calculate. To eliminate the cross terms, we only need to find a basis satisfying $\langle L(\vec{v}_i), \vec{v}_j\rangle = 0$ for $i \ne j$. The simpler requirement can be achieved by the more efficient method of *completing the square* (which is a version of Gaussian elimination).

The formula (9.3) shows that, if B is the matrix of a quadratic form q with respect to a basis, then the matrix with respect to another basis is P^TBP. If we use orthogonal diagonalisation $B = UDU^{-1}$, then $U^{-1} = U^T$, and $U^TBU = D$ is diagonal. This is the reason we get (9.4).

The method of completing the square gives a non-orthogonal matrix P (often upper triangular), such that P^TBP is diagonal.

Example 9.8. For $q(x,y,z) = x^2 + 13y^2 + 14z^2 + 6xy + 2xz + 18yz$, we gather together all the terms involving x and complete the square

$$\begin{aligned} q &= x^2 + 6xy + 2xz + 13y^2 + 14z^2 + 18yz \\ &= [x^2 + 2x(3y+z) + (3y+z)^2] + 13y^2 + 14z^2 + 18yz - (3y+z)^2 \\ &= (x+3y+z)^2 + 4y^2 + 13z^2 + 12yz. \end{aligned}$$

The remaining terms involve only y and z. Gathering all the terms involving y and completing the square, we get $4y^2 + 13z^2 + 12yz = (2y+3z)^2 + 4z^2$ and

$$q = (x+3y+z)^2 + (2y+3z)^2 + (2z)^2 = u^2 + v^2 + w^2.$$

In terms of matrix, the process gives

$$\begin{pmatrix} 1 & 3 & 1 \\ 3 & 13 & 9 \\ 1 & 9 & 14 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}.$$

Geometrically, the original variables x, y, z are the coordinates with respect to the standard basis. The new variables u, v, w are supposed to be the coordinates with respect to a new basis.

The original expression of the quadratic form is with respect to the standard basis ϵ. The coordinate with respect to the new basis α is

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} x+3y+z \\ 2y+3z \\ 2z \end{pmatrix} = \begin{pmatrix} 1 & 3 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

The matrix above is then the matrix $[I]_{\alpha\epsilon}$ of changing from the standard basis ϵ to the new basis α. By the second part of Exercise **??**, the basis α is the columns of the inverse matrix

$$\begin{pmatrix} 1 & 3 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -\frac{3}{2} & \frac{7}{4} \\ 0 & \frac{1}{2} & -\frac{3}{4} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Example 9.9. The cross terms in the quadratic form $q = 4x_1^2 + 19x_2^2 - 4x_4^2 - 4x_1x_2 + 4x_1x_3 -$

$8x_1x_4 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4$ can be eliminated as follows.

$$
\begin{aligned}
q &= 4[x_1^2 - x_1x_2 + x_1x_3 - 2x_1x_4] + 19x_2^2 - 4x_4^2 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4 \\
&= 4\left[x_1^2 + 2x_1\left(-\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right) + \left(-\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right)^2\right] \\
&\quad + 19x_2^2 - 4x_4^2 + 10x_2x_3 + 16x_2x_4 + 12x_3x_4 - 4\left(-\frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right)^2 \\
&= 4\left(x_1 - \frac{1}{2}x_2 + \frac{1}{2}x_3 - x_4\right)^2 + 18\left[x_2^2 + \frac{2}{3}x_2x_3 + \frac{2}{3}x_2x_4\right] - x_3^2 - 8x_4^2 + 16x_3x_4 \\
&= (2x_1 - x_2 + x_3 - 2x_4)^2 + 18\left[x_2^2 + 2x_2\left(\frac{1}{3}x_3 + \frac{1}{3}x_4\right) + \left(\frac{1}{3}x_3 + \frac{1}{3}x_4\right)^2\right] \\
&\quad - x_3^2 - 8x_4^2 + 16x_3x_4 - 18\left(\frac{1}{3}x_3 + \frac{1}{3}x_4\right)^2 \\
&= (2x_1 - x_2 + x_3 - 2x_4)^2 + 18\left(x_2 + \frac{1}{3}x_3 + \frac{1}{3}x_4\right)^2 - 3(x_3^2 - 4x_3x_4) - 10x_4^2 \\
&= (2x_1 - x_2 + x_3 - 2x_4)^2 + 2(3x_2 + x_3 + x_4)^2 \\
&\quad - 3[x_3^2 + 2x_3(-2x_4) + (-2x_4)^2] - 10x_4^2 + 3(-2x_4)^2 \\
&= (2x_1 - x_2 + x_3 - 2x_4)^2 + 2(3x_2 + x_3 + x_4)^2 - 3(x_3 - 2x_4)^2 + 2x_4^2 \\
&= y_1^2 + 2y_2^2 - 3y_3^3 + 2y_4^2.
\end{aligned}
$$

The new variables y_1, y_2, y_3, y_4 are the coordinates with respect to the basis of the columns of

$$
\begin{pmatrix} 2 & -1 & 1 & -2 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & -\frac{2}{3} & -\frac{1}{2} \\ 0 & \frac{1}{3} & -\frac{1}{3} & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
$$

Example 9.10. The quadratic form $q = 4xy + y^2$ has no x^2 term. We may complete the square by using the y^2 term and get $q = (y + 2x)^2 - 4x^2 = u^2 - 4v^2$.

The quadratic form $q = xy + yz$ has no square term. We may eliminate the cross terms by introducing $x = x_1 + y_1$, $y = x_1 - y_1$, so that $q = x_1^2 - y_1^2 + x_1z - y_1z$. Then we complete the square and get $q = \left(x_1 - \frac{1}{2}z\right)^2 - \left(y_1 + \frac{1}{2}z\right)^2 = \frac{1}{4}(x + y - z)^2 - \frac{1}{4}(x - y + z)^2$.

Exercise 9.12. Eliminate the cross terms.

1. $x^2 + 4xy - 5y^2$.

2. $2x^2 + 4xy$.

3. $4x_1^2 + 4x_1x_2 + 5x_2^2$.

4. $x^2 + 2y^2 + z^2 + 2xy - 2xz$.

5. $-2u^2 - v^2 - 6w^2 - 4uw + 2vw$.

6. $x_1^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_1x_4 + 2x_3x_4$.

Exercise 9.13. Eliminate the cross terms in the quadratic form $x^2 + 2y^2 + z^2 + 2xy - 2xz$ by first completing a square for terms involving z, then completing for terms involving y.

Now we study the process of completing the square in general. Let $q(\vec{x}) = \vec{x}^T B\vec{x}$ for $\vec{x} \in \mathbb{R}^n$ and a symmetric $n \times n$ matrix B. The *leading principal minors* of B are the determinants of the square submatrices formed by the entries in the first k rows and first k columns of B

$$d_1 = b_{11},\ d_2 = \det \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix},\ d_3 = \det \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix},\ \dots,\ d_n = \det B.$$

If $d_1 \neq 0$, then eliminating all the cross terms involving x_1 gives

$$\begin{aligned} q(\vec{x}) = b_{11} &\left[x_1^2 + 2x_1 \frac{1}{b_{11}}(b_{12}x_2 + \cdots + b_{1n}x_n) + \frac{1}{b_{11}^2}(b_{12}x_2 + \cdots + b_{1n}x_n)^2 \right] \\ &+ b_{22}x_2^2 + \cdots + b_{nn}x_n^2 + 2b_{23}x_2x_3 + 2b_{24}x_2x_4 + \cdots + 2b_{(n-1)n}x_{n-1}x_n \\ &- \frac{1}{b_{11}}(b_{12}x_2 + \cdots + b_{1n}x_n)^2 \\ = d_1 &\left[x_1 + \frac{b_{12}}{d_1}x_2 + \cdots + \frac{b_{1n}}{d_1}x_n \right]^2 + q_2(\vec{x}_2), \end{aligned}$$

where q_2 is a quadratic form of the truncated vector $\vec{x}_2 = (x_2 \ \cdots \ x_n)^T$. The symmetric coefficient matrix B_2 for q_2 is obtained as follows. For each $2 \le i \le n$, adding $-\dfrac{b_{1i}}{b_{11}}$ multiple of the first row of B to the i-th row makes the i-th entry in the first column to become zero. Then we get a matrix $\begin{pmatrix} d_1 & * \\ \vec{0} & B_2 \end{pmatrix}$. In fact, doing the same operations to the rows also eliminates the entries in the first row except the first one, and gives a symmetric matrix $\begin{pmatrix} d_1 & 0 \\ \vec{0} & B_2 \end{pmatrix}$. Since the operations do not change the determinant of the matrix (and all the leading principal minors), the principal minors $d_1^{(2)}, \dots, d_{n-1}^{(2)}$ of B_2 are related to the principal minors $d_1^{(1)} = d_1, \dots, d_n^{(1)} = d_n$ of B_1 by $d_{k+1}^{(1)} = d_1 d_k^{(2)}$.

The discussion sets up an inductive argument. Assume $d_1, \dots, d_k$ are all nonzero. Then we may complete the squares in k steps and obtain

$$\begin{aligned} q(\vec{x}) = &\, d_1^{(1)}(x_1 + c_{12}x_2 + \cdots + c_{1n}x_n)^2 + d_1^{(2)}(x_2 + c_{23}x_3 + \cdots + c_{2n}x_n)^2 \\ &+ \cdots + d_1^{(k)}(x_k + c_{k(k+1)}x_{k+1} + \cdots + c_{kn}x_n)^2 + q_{k+1}(\vec{x}_{k+1}), \end{aligned}$$

with

$$d_1^{(i)} = \frac{d_2^{(i-1)}}{d_1^{(i-1)}} = \frac{d_3^{(i-2)}}{d_2^{(i-2)}} = \cdots = \frac{d_i^{(1)}}{d_{i-1}^{(1)}} = \frac{d_i}{d_{i-1}},$$

and the coefficient of x_{k+1}^2 in q_{k+1} is $d_1^{(k+1)} = \dfrac{d_{k+1}}{d_k}$.

Proposition 9.4 (Lagrange-Beltrami Identity). *Suppose $q(\vec{x}) = \vec{x}^T B\vec{x}$ is a quadratic form of rank r. If all the leading principal minors $d_1, \dots, d_r$ of the symmetric coefficient matrix B are nonzero, then there is an upper triangular change of variables*

$$y_k = x_k + c_{k(k+1)}x_{k+1} + \cdots + c_{kn}x_n, \quad k = 1, \dots, r,$$

such that $q = d_1 y_1^2 + \dfrac{d_2}{d_1}y_2^2 + \cdots + \dfrac{d_r}{d_{r-1}}y_r^2$.

Example 9.10 shows that the nonzero condition on leading principal minors may not always be satisfied.

Exercise 9.14. Suppose $q(\vec{x}) = \langle L(\vec{x}), \vec{x}\rangle = c_1x_1^2 + \cdots + c_rx_r^2$ with respect to a basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$, and $c_i \neq 0$. Use

$$\langle L(\vec{v}_i), \vec{v}_j\rangle = \begin{cases} 0 & \text{for } i \neq j, \\ c_i & \text{for } i = j, \end{cases}$$

to prove that $\text{Span}\{\vec{v}_{r+1}, \dots, \vec{v}_n\} = \text{Ker}L$. Then prove $r = \text{rank}L$.

9.4 Signature

By rearranging the order of vectors in basis, we may assume that the "diagonalisation" (9.4) of a quadratic form is

$$q(\vec{x}) = d_1x_1^2 + \cdots + d_nx_n^2, \quad d_1, \dots, d_s > 0, \quad d_{s+1}, \dots, d_{s+t} < 0, \quad d_{s+t+1} = \cdots = d_n = 0.$$

Then

$$q(\vec{x}) = (\sqrt{d_1}x_1)^2 + \cdots + (\sqrt{d_s}x_s)^2 - (\sqrt{-d_{s+1}}x_{s+1})^2 - \cdots - (\sqrt{-d_{s+t}}x_{s+t})^2.$$

By using $\sqrt{\pm d_j}x_j$ as new coordinates (which means rescaling the vectors in the basis), we have

$$q(\vec{x}) = x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_{s+t}^2 \tag{9.5}$$

with respect to a basis. This is the *canonical form* of the quadratic form. In fact, both s and t are independent of the choice of special basis.

Theorem 9.5 (Sylvester's Law). *After eliminating the cross terms in a quadratic form, the number of positive coefficients, the number of negative coefficients, and the number of zero coefficients are independent of the choice of basis.*

Proof. We choose an inner product on V and write $q(\vec{x}) = \langle L(\vec{x}), \vec{x}\rangle$ for a symmetric operator L. We also denote the associated bilinear form $b(\vec{x}, \vec{y}) = \langle L(\vec{x}), \vec{y}\rangle$.

Suppose q has the canonical form (9.5) with respect to a basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$, and has another canonical form with respect to another basis $\beta = \{\vec{w}_1, \dots, \vec{w}_n\}$

$$q(\vec{x}) = y_1^2 + \cdots + y_{s'}^2 - y_{s'+1}^2 - \cdots - y_{s'+t'}^2.$$

We prove that the vectors $\vec{v}_1, \dots, \vec{v}_s, \vec{w}_{s'+1}, \dots, \vec{w}_n$ are linearly independent. Suppose

$$x_1\vec{v}_1 + \cdots + x_s\vec{v}_s + y_{s'+1}\vec{w}_{s'+1} + \cdots + y_n\vec{w}_n = \vec{0}.$$

Then

$$x_1\vec{v}_1 + \cdots + x_s\vec{v}_s = -y_{s'+1}\vec{w}_{s'+1} - \cdots - y_n\vec{w}_n.$$

Applying the quadratic function q to both sides, we get

$$x_1^2 + \cdots + x_s^2 = -y_{s'+1}^2 - \cdots - y_{s'+t'}^2 \le 0.$$

This implies $x_1 = \cdots = x_s = 0$. Then $y_{s'+1}\vec{w}_{s'+1} + \cdots + y_n\vec{w}_n = \vec{0}$ and by the linear independence of $\vec{w}_{s'+1}, \dots, \vec{w}_n$, we get $y_{s'+1} = \cdots = y_n = 0$. This completes the proof of the linear independence.

The linear independence implies that $s + (n - s') \le n$, or $s \le s'$. By exchanging α and β, we also get $s' \le s$. Then $s = s'$. Similar argument shows $t = t'$. $\square$

The number $s - t$ of positive coefficients subtracting the number of negative coefficients is called the *signature* of the quadratic form. The quadratic forms in Examples 9.8, 9.9, 9.10 have signatures 3, 2, 0.

If we use the diagonalisation of symmetric operator to get (9.4), then we find that signature of a quadratic form is the number of positive eigenvalues subtracting the number of negative eigenvalues (the repetition of the same eigenvalue is counted in the number). On the other hand, under the assumption of Proposition 9.4, we also know how to calculate the signature from the determinants of leading principal minors.

Exercise 9.15. Two quadratic forms $p(\vec{x})$ and $q(\vec{x})$ are *equivalent* if there is an invertible operator L, such that $p(\vec{x}) = q(L(\vec{x}))$. Prove that two (real) quadratic forms are equivalent if and only if they have the same rank and same signature.

9.5 Positive Definite Operator

A quadratic form is *positive definite* if $q(\vec{x}) > 0$ for any $\vec{x} \neq 0$. It is *negative definite* if $q(\vec{x}) < 0$ for any $\vec{x} \neq 0$. If equalities are allowed, then we get *positive semi-definite* or *negative semi-definite* forms. The final possibility is *indefinite* quadratic forms, for which the value can be positive as well as negative. The quadratic form in Example 9.8 is positive definite. The quadratic forms in Examples 9.9 and 9.10 are indefinite.

Since quadratic forms on an inner product space are in one-to-one correspondence with symmetric operators by $q(\vec{x}) = \langle L(\vec{x}), \vec{x}\rangle$, the concepts of positive (semi-)definite, negative (semi-)definite and indefinite can also be applied to symmetric operators. For example, a symmetric operator L is positive definite if $\langle L(\vec{x}), \vec{x}\rangle > 0$ for any $\vec{x} \neq 0$. By the formula (9.4) with respect to an orthonormal basis of eigenvectors for L, we see that L is positive definite if and only if all its eigenvalues are positive. We have similar characterisations of the other (in)definiteness in terms of signs of eigenvalues. Proposition 9.4 also gives a criterion for (in)definiteness in terms of determinants of leading principal minors.

Proposition 9.6 (Sylvester's Criterion). *Suppose B is a symmetric matrix of rank r. Suppose the leading principal minors $d_1, \dots, d_r$ are nonzero.*

1. *If $d_1 > 0, d_2 > 0, \dots, d_r > 0$, then B is positive semi-definite. If we further have $r = n$, then B is positive definite.*

2. *If $-d_1 > 0, d_2 > 0, \dots, (-1)^r d_r > 0$, then B is negative semi-definite. If we further have $r = n$, then B is negative definite.*

3. *Otherwise B is indefinite.*

If the condition that all $d_1, \dots, d_r$ are nonzero is not satisfied, then the matrix cannot be positive or negative definite. The criterion for the other possibilities is a little more complicated[2].

Exercise 9.16. Suppose in a quadratic form, the coefficient of a square term is 0. Prove that the quadratic form is indefinite.

Exercise 9.17. Prove that if a quadratic form $q(\vec{x})$ is positive definite, then $q(\vec{x}) \geq c\|\vec{x}\|^2$ for any $\vec{x}$ and a constant $c > 0$. What is the maximum of such c?

Exercise 9.18. Prove that positive definite and negative operators are invertible.

Exercise 9.19. Suppose L and L' are positive definite symmetric operators.

1. Prove that $L + L'$ is positive definite.

2. For $a > 0$, prove that $a\,L$ is positive definite.

3. Prove that L^n is positive definite for any integer n.

4. Prove that $L^2 - L + I$ is positive definite. Is $L^2 + aL + bI$ always positive definite?

[2] Sylvester's Minorant Criterion, Lagrange-Beltrami Identity, and Nonnegative Definiteness, by Sudhir R. Ghorpade, Balmohan V. Limaye, in The Mathematics Student. Special Centenary Volume (2007), 123-130

Proposition 9.7. *For a symmetric operator L, the following are equivalnet.*

1. *L is positive semi-definite.*
2. *$L = K^*K$ for a linear transformation K.*
3. *$L = K^2$ for a symmetric operator K.*

Moreover, K is positive definite if and only if K is injective.

Proof. The second statement implies the first by

$$\langle L(\vec{x}), \vec{x}\rangle = \langle K^*K(\vec{x}), \vec{x}\rangle = \langle K(\vec{x}), K(\vec{x})\rangle = \|K(\vec{x})\|^2 \geq 0.$$

Moreover, if we want L to be positive definite, then we need $\|K(\vec{x})\| = 0$ implying $\vec{x} = \vec{0}$. This means the injectivity of K.

The third statement clearly implies the second. It remains to prove that the first implies the third. So we assume that L has an orthonormal basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ of eigenvectors

$$L(\vec{v}_i) = d_i\vec{v}_i.$$

If L is positive semi-definite, then $d_i \geq 0$ and we may introduce a symmetric operator K by

$$K(\vec{v}_i) = \sqrt{d_i}\vec{v}_i.$$

The operator satisfies $L = K^2 = K^*K$. □

Definition 9.8. The *square root* operator $\sqrt{L}$ of a positive semi-definite symmetric operator L is the positive semi-definite symmetric operator K commuting with L and satisfying $K^2 = L$.

By applying Theorem 7.7 to the commutative $*$-algebra (the $*$-operation is trivial) generated by L and K, we find that L and K have simultaneous orthonormal basis of eigenvectors. Therefore K is the unique symmetric operator constructed in the proof above. If L is positive definite, then $\sqrt{L}$ is also positive definite.

In matrix form, we have

$$A = U\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}U^T,\ d_i > 0,\ U^TU = I \implies \sqrt{A} = U\begin{pmatrix} \sqrt{d_1} & & \\ & \ddots & \\ & & \sqrt{d_n} \end{pmatrix}U^T.$$

Suppose $L\colon V \to W$ is an injective linear transformation between inner product spaces. Then L^*L is a positive definite operator, and $P = \sqrt{L^*L}$ is also a positive definite operator. Since positive definite operators are invertible (because of full rank), we may introduce $U = LP^{-1}\colon V \to W$. Then

$$U^*U = P^{-1}L^*LP^{-1} = P^{-1}P^2P^{-1} = I.$$

Therefore U is an isometry. Another way to see that U is an isometry is by

$$\begin{aligned}\|U((\vec{v}))\|^2 &= \|L(\vec{v})\|^2 = \langle L(\vec{v}), L(\vec{v})\rangle = \langle \vec{v}, L^*L(\vec{v})\rangle = \langle \vec{v}, P^2(\vec{v})\rangle \\ &= \langle \vec{v}, P^*P(\vec{v})\rangle = \langle P(\vec{v}), L(\vec{v})\rangle = \|P(\vec{v})\|^2\end{aligned}$$

and the fact that any vector in V is of the form $P(\vec{v})$ (since P is invertible).

The decomposition $L = UP$ is comparable to the polar decomposition of complex numbers

$$z = e^{i\theta}r, \quad r = |z| = \sqrt{\bar{z}z}.$$

and is called the *polar decomposition* of L.

******* note: the calculation of square root can use any basis of eigenvectors.

9.6 Complex Functionals and Forms

The dual space V^* of a complex vector space V is the vector space of all complex linear transformations $l\colon V \to \mathbb{C}$, again called linear functionals. A basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ of V has a *dual basis* $\alpha^* = \{\vec{v}_1^*, \dots, \vec{v}_n^*\}$ of V^* characterised by

$$\vec{x} = \vec{v}_1^*(\vec{x})\vec{v}_1 + \cdots + \vec{v}_n^*(\vec{x})\vec{v}_n, \quad l = l(\vec{v}_1)\vec{v}_1^* + \cdots + l(\vec{v}_n)\vec{v}_n^*.$$

This implies $\dim V^* = \dim V$. Theorem 9.1 also holds for complex vector spaces, and we have canonical $V \cong (V^*)^*$. Moreover, a linear transformation $L\colon V \to W$ induces the dual linear transformation $L^*\colon W^* \to V^*$.

All the discussion in Section 9.1 about the dual space extends to complex vector spaces.

There are two extensions of bilinear functions from real to complex numbers. The first is still *bilinear*, which means $b\colon V \times W \to \mathbb{C}$ satisfying

$$b(\lambda\vec{v} + \mu\vec{v}', \vec{w}) = \lambda b(\vec{v}, \vec{w}) + \mu b(\vec{v}', \vec{w}), \quad b(\vec{v}, \lambda\vec{w} + \mu\vec{w}') = \lambda b(\vec{v}, \vec{w}) + \mu b(\vec{v}, \vec{w}').$$

The second is *sesquilinear*, which means $b\colon V \times W \to \mathbb{C}$ satisfying

$$b(\lambda\vec{v} + \mu\vec{v}', \vec{w}) = \lambda b(\vec{v}, \vec{w}) + \mu b(\vec{v}', \vec{w}), \quad b(\vec{v}, \lambda\vec{w} + \mu\vec{w}') = \bar{\lambda} b(\vec{v}, \vec{w}) + \bar{\mu} b(\vec{v}, \vec{w}').$$

The complex inner product is not bilinear, but sesquilinear.

Half of the discussion in Section 9.2 can be applied to the first extension, and the other half can be applied to the second extension.

A bilinear function induces complex linear transformations $V \to W^*$ and $W \to V^*$, that are dual to each other after applying the complex version of Theorem 9.1. When the linear transformations are isomorphic, then we have a dual pairing like a real case. The evaluation pairing

$$\langle \vec{x}, l\rangle = l(\vec{x})\colon V \times V^* \to \mathbb{C}.$$

is the canonical example of dual pairing. The dual bases of V and W with respect to a dual pairing can also be defined as before.

The major problem with complex bilinear functions is that the concept is incompatible with complex inner product. On inner product spaces, there is no correspondence between complex bilinear functions and linear transformations.

A sesquilinear function b is compatible with complex inner product, in the sense that if V and W are complex inner product spaces, then b corresponds to a unique linear transformation $L\colon V \to W$ by $b(\vec{v}, \vec{w}) = \langle L(\vec{v}), \vec{w}\rangle$. On the other hand, by the linearity in the first vector, a sesquilinear function b induces a map $L_b(\vec{w}) = b(\cdot, \vec{w})\colon W \to V^*$. However, by the conjugate linearity in the second vector, the map is not linear, but conjugate linear

$$L_b(\lambda\vec{w} + \mu\vec{w}') = \bar{\lambda}L_b(\vec{w}) + \bar{\mu}L_b(\vec{w}').$$

On the other hand, ${}_bL(\vec{v}) = b(\vec{v}, \cdot)$ is linear in $\vec{v}$. However, ${}_bL(\vec{v})$ is not a linear functional, but a conjugate linear functional, and belongs to the *conjugate dual space*

$$\bar{W}^* = \{l\colon W \to \mathbb{C} \text{ satisfying } l(\lambda\vec{w} + \mu\vec{w}') = \bar{\lambda}l(\vec{w}) + \bar{\mu}l(\vec{w}')\}.$$

Therefore the extension of the discussion about induced "linear transformations" $V \to W^*$ and $W \to V^*$ requires some set up about conjugate linearity.

Next we move on to complex quadratic form. The two extensions of real bilinear functions means two extensions of real quadratic form.

The first type is *symmetric bilinear form* $b\colon V \times V \to \mathbb{C}$, which is a bilinear function satisfying $b(\vec{y}, \vec{x}) = b(\vec{x}, \vec{y})$. The *quadratic form* associated to b is $q(\vec{z}) = b(\vec{z}, \vec{z})$, and b can be recovered from the quadratic form by polarisation

$$b(\vec{x}, \vec{y}) = \frac{1}{4}(q(\vec{x}+\vec{y}) - q(\vec{x}-\vec{y})) = \frac{1}{2}(q(\vec{x}+\vec{y}) - q(\vec{x}) - q(\vec{y})).$$

Moreover, for any basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ of V, we have the corresponding complex symmetric matrix $[q]_\alpha = B = (b(\vec{v}_i, \vec{v}_j))$ by

$$q(\vec{z}) = [\vec{z}]_\alpha^T B[\vec{z}]_\alpha = \sum_{1\le i\le n} b_{ii}z_i^2 + 2\sum_{1\le i<j\le n} b_{ij}z_iz_j, \quad \vec{z} = z_1\vec{v}_1 + \cdots + z_n\vec{v}_n.$$

The change of basis formula (9.3) $[q]_\alpha = [I]_{\beta\alpha}^T[q]_\beta[I]_{\beta\alpha}$ still holds. However, there is no diagonalisation theorem about complex symmetric matrices. Instead, we may still use the method of completing the square, and the Lagrange-Beltrami Identity (Proposition 9.4) still holds. Therefore by a suitable change of basis, we may still eliminate all the cross terms and get

$$q(\vec{z}) = d_1\tilde{z}_1^2 + \cdots + d_n\tilde{z}_n^2.$$

Here $\tilde{z}_1, \dots, \tilde{z}_n$ are the coordinates with respect to a new basis, and d_i has no meaning as some eigenvalue. By changing the order of vectors in the new basis, we may further assume

$$q(\vec{z}) = d_1\tilde{z}_1^2 + \cdots + d_r\tilde{z}_r^2 = (\sqrt{d_1}\tilde{z}_1)^2 + \cdots + (\sqrt{d_r}\tilde{z}_r)^2, \quad d_1, \dots, d_r \neq 0.$$

This means that, if we further divide the vectors in the new basis by $\sqrt{d_i}$, then the quadratic form becomes the following *canonical form*

$$q(\vec{z}) = \tilde{z}_1^2 + \cdots + \tilde{z}_r^2.$$

In terms of matrix, this means that, for any complex symmetric matrix B, there is an invertible matrix P, such that

$$P^TBP = \begin{pmatrix} I_r & O \\ O & O_{n-r} \end{pmatrix}.$$

Evidently $r = \text{rank}B$.

Example 9.11. For $q(x, y, z) = x^2 + iy^2 + 3z^2 + 2(1 + i)xy + 4yz$, we have

$$\begin{aligned} q &= [x^2 + 2(1+i)xy + ((i+1)y)^2] + iy^2 + 3z^2 + 4yz - (i+1)^2y^2 \\ &= (x + (1+i)y)^2 - iy^2 + 3z^2 + 4yz \\ &= (x + (1+i)y)^2 - i\left[y^2 + 4iyz + (2iz)^2\right] + 3z^2 + i(2i)^2z^2 \\ &= (x + (1+i)y)^2 - i(y + 2iz)^2 + (3 - 4i)z^2. \end{aligned}$$

This eliminates the cross terms. We may further use (we pick one of the possible complex square roots)

$$\sqrt{-i} = \sqrt{e^{-i\frac{\pi}{2}}} = e^{-i\frac{\pi}{4}} = \frac{1-i}{\sqrt{2}}, \quad \sqrt{3-4i} = \sqrt{(2-i)^2} = 2 - i,$$

to get

$$q = (x + (1+i)y)^2 + \left(\frac{1-i}{\sqrt{2}}y + \sqrt{2}(1+i)z\right)^2 + ((2-i)z)^2.$$

In terms of matrix, the process gives

$$\begin{pmatrix} 1 & 1+i & 0 \\ 1+i & i & 2 \\ 0 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 1+i & 0 \\ 0 & \frac{1-i}{\sqrt{2}} & \sqrt{2}(1+i) \\ 0 & 0 & 2-i \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1+i & 0 \\ 0 & \frac{1-i}{\sqrt{2}} & \sqrt{2}(1+i) \\ 0 & 0 & 2-i \end{pmatrix}.$$

The new variables $x + (1+i)y$, $\frac{1-i}{\sqrt{2}}y + \sqrt{2}(1+i)z$, $(2-i)z$ are the coordinates with respect to some new basis. By the row operation

$$\begin{aligned} &\begin{pmatrix} 1 & 1+i & 0 & 1 & 0 & 0 \\ 0 & \frac{1-i}{\sqrt{2}} & \sqrt{2}(1+i) & 0 & 1 & 0 \\ 0 & 0 & 2-i & 0 & 0 & 1 \end{pmatrix} \xrightarrow[\frac{2+i}{5}R_3]{\frac{1+i}{\sqrt{2}}R_2} \begin{pmatrix} 1 & 1+i & 0 & 1 & 0 & 0 \\ 0 & 1 & 2i & 0 & \frac{1+i}{\sqrt{2}} & 0 \\ 0 & 0 & 1 & 0 & 0 & \frac{2+i}{5} \end{pmatrix} \\ \xrightarrow{R_2 - 2iR_3} &\begin{pmatrix} 1 & 1+i & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{1+i}{\sqrt{2}} & \frac{-2+4i}{5} \\ 0 & 0 & 1 & 0 & 0 & \frac{2+i}{5} \end{pmatrix} \xrightarrow{R_1 - (1+i)R_2} \begin{pmatrix} 1 & 0 & 0 & 1 & \sqrt{2}i & \frac{-6+2i}{5} \\ 0 & 1 & 0 & 0 & \frac{1+i}{\sqrt{2}} & \frac{-2+4i}{5} \\ 0 & 0 & 1 & 0 & 0 & \frac{2+i}{5} \end{pmatrix}, \end{aligned}$$

the new basis is the last three columns of last matrix above.

Exercise 9.20. In the complex version of Exercise 9.15, we say two complex quadratic forms (associated to complex symmetric bilinear forms) $p(\vec{x})$ and $q(\vec{x})$ are *equivalent* if there is an invertible operator L, such that $p(\vec{x}) = q(L(\vec{x}))$. Prove that two such quadratic forms are equivalent if and only if they have the same rank.

The second type of complex quadratic form is associated to a *Hermitian bilinear form* $b\colon V \times V \to \mathbb{C}$, which is a sesquilinear function satisfying $b(\vec{y},\vec{x}) = \overline{b(\vec{x},\vec{y})}$. The associated *quadratic form* is $q(\vec{z}) = b(\vec{z},\vec{z})$, and b can be recovered from the quadratic form by polarisation

$$\begin{aligned} b(\vec{x},\vec{y}) &= \frac{1}{4}(q(\vec{x}+\vec{y}) - q(\vec{x}-\vec{y}) + iq(\vec{x}+i\vec{y}) - iq(\vec{x}-i\vec{y})) \\ &= \frac{1}{2}(q(\vec{x}+\vec{y}) + iq(\vec{x}+i\vec{y}) - (1+i)q(\vec{x}) - (1+i)q(\vec{y})). \end{aligned}$$

Moreover, for any basis $\alpha = \{\vec{v}_1, \dots, \vec{v}_n\}$ of V, we have the corresponding Hermitian matrix $[q]_\alpha = B = (b(\vec{v}_i,\vec{v}_j))$ satisfying

$$q(\vec{z}) = [\vec{z}]_\alpha^T B \overline{[\vec{z}]}_\alpha = \sum_{1\le i\le n} b_{ii}|z_i|^2 + 2\sum_{1\le i<j\le n} b_{ij} z_i \bar{z}_j, \quad \vec{z} = z_1\vec{v}_1 + \cdots + z_n\vec{v}_n.$$

The change of basis formula (9.3) becomes $[q]_\alpha = [I]_{\beta\alpha}^*[q]_\beta[I]_{\beta\alpha}$.

If V has a complex inner product, then b and q correspond to a Hermitian operator L by

$$b(\vec{x},\vec{y}) = \langle L(\vec{x}),\vec{y}\rangle, \quad q(\vec{z}) = \langle L(\vec{z}),\vec{z}\rangle.$$

Then we may use an orthonormal basis $\alpha = \{\vec{v}_1, \dots \vec{v}_n\}$ of eigenvectors of L to get

$$\begin{aligned} q(\vec{z}) &= \langle L(\vec{z}),\vec{z}\rangle \\ &= \langle d_1 z_1\vec{v}_1 + \cdots + d_n z_n \vec{v}_n, z_1\vec{v}_1 + \cdots + z_n\vec{v}_n\rangle \\ &= d_1|z_1|^2 + \cdots + d_n|z_n|^2. \end{aligned}$$

Here d_i are the eigenvalues of L and are real numbers. Alternatively, we may also complete the square.

Example 9.12. For $q(x,y,z) = x\bar{x} - y\bar{y} + 2z\bar{z} + (1+i)x\bar{y} + (1-i)y\bar{x} + 3iy\bar{z} - 3iz\bar{y}$, we have

$$\begin{aligned} q &= \left[x\bar{x} + (1+i)x\bar{y} + (1-i)y\bar{x} + (1-i)y\overline{(1-i)y}\right] - y\bar{y} + 2z\bar{z} - |1-i|^2 y\bar{y} + 3iy\bar{z} - 3iz\bar{y} \\ &= (x+(1-i)y)\overline{(x+(1-i)y)} - 3y\bar{y} + 2z\bar{z} + 3iy\bar{z} - 3iz\bar{y} \\ &= |x+(1-i)y|^2 - 3\left[y\bar{y} - iy\bar{z} + iz\bar{y} + iz\overline{iz}\right] + 2z\bar{z} + 3|i|^2 z\bar{z} \\ &= |x+(1-i)y|^2 - 3|y+iz|^2 + 5|z|^2. \end{aligned}$$

In terms of matrix, the process gives

$$\begin{pmatrix} 1 & 1+i & 0 \\ 1-i & -1 & 3i \\ 0 & -3i & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1-i & 0 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{pmatrix}^* \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 1-i & 0 \\ 0 & 1 & i \\ 0 & 0 & 1 \end{pmatrix}.$$

The new variables $x+(1-i)y, y+iz, z$ are the coordinates with respect to the basis of the columns of the inverse matrix

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

The Sylvester's Law (Theorem 9.5) can be similarly proved, and the rank and signature can be defined. This leads to the *canonical form* with respect to a suitable basis, similar to the real case

$$q(\vec{z}) = |z_1|^2 + \cdots + |z_p|^2 - |z_{p+1}|^2 - \cdots - |z_{p+q}|^2.$$

We may further define the (in)definiteness of a quadratic form and the associated Hermitian operator. For example, a hermitian operator L is positive if $\langle L(\vec{z}), \vec{z}\rangle > 0$ for all $\vec{z} \neq \vec{0}$. The *square root* operator $\sqrt{L}$ of a positive semi-definite Hermitian operator L is the positive semi-definite Hermitian operator K commuting with L and satisfying $K^2 = L$. Then we also have complex version of polar decomposition of injective linear transformations between complex inner product spaces.

Exercise 9.21. Extend Exercise 9.15 to quadratic forms associated to Hermitian forms.

Exercise 9.22. Extend Sylvester's criterion (Proposition 9.6) to quadratic forms associated to Hermitian forms.

9.7 Tensor of Vector Spaces

9.8 Exterior Algebra