Linear Algebra
MAT 331

Wai Yan Pong

September 19, 2017

Contents

Chapter 1

Linear Equations

1.1 Systems of Linear Equations

We begin our investigation by looking at the simplest kind of linear equation:

$$ax = b. \tag{1.1}$$

Whether the equation in (1.1) has a solution depends what kind of solutions are allowed. For example, $6x = 2$ has no solutions if we consider only the integers. However, $1/3$ is the solution if the rationals are allowed. And if we are talking about integers modulo 10, then it has two solutions namely $2, 7 \pmod{10}$. This shows that the question of whether (1.1) is solvable is more about the "number system" in consideration than the nature of the equation. To make things simple, we assume (1.1) has a solution whenever $a \neq 0$. More precisely, we assume our "number system" is a **field** (see Appendix A). For now, it is enough to keep in mind that rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$ are all fields with their usual addition and multiplication. Also keep in mind that there are fields with only finitely many elements, like $\mathbb{Z}/p\mathbb{Z}$ where p is a prime number. We simply use $\mathbb{K}$ to denote a field. The key point of having a field for us is that:

Every non-zero element of a field has a multiplicative inverse.

One can also shows that if b and c are multiplicative inverses of a then $b = c$. In other words, if $a \neq 0$, then a has a unique multiplicative inverse and we use either a^{-1} or $1/a$ to denote the multiplicative inverse of a. Morally speaking, if $a \in \mathbb{K}$ is non-zero then we can "divide by a" by multiplying its inverse.

In general, a system of m linear equations in n unknowns looks like

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.2}$$

We call a_{ij}'s $(1 \leq i \leq n, 1 \leq j \leq m)$ the **coefficients** and b_i's the **constants** of (1.2). System (1.2) can be expressed as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{1.3}$$

or even more succinctly as $A\mathbf{x} = \mathbf{b}$ where $A = (a_{ij})$ is a rectangular array of elements of elements of $\mathbb{K}$, or a **matrix** over $\mathbb{K}$. We call it the **coefficient matrix** of System (1.2). We call $\mathbf{b} = (b_i)$ the **constant vector** of (1.2). A system of linear equations is **consistent** (or **solvable**) if it has a solution [1]. A system is **homogeneous** if all its constants are 0. A homogeneous system must be consistent since $x_1 = \cdots = x_n = 0$ is clearly a solution.

Before diving into an algorithm of solving linear systems, let us observe that if $\mathbf{v}_1, \mathbf{v}_2$ are two solutions, then

$$\begin{aligned} A\mathbf{v}_1 - A\mathbf{v}_2 &= \mathbf{b} - \mathbf{b} \\ A(\mathbf{v}_1 - \mathbf{v}_2) &= \mathbf{0}. \end{aligned}$$

That means $\mathbf{h} := \mathbf{v}_1 - \mathbf{v}_2$ is a solution of

$$A\mathbf{x} = \mathbf{0}. \tag{1.4}$$

In other words, any two solutions of (1.2) differs by a solution of (1.4). System (1.4) is called the **homogeneous system** associated to System (1.2). On the other hand, if $\mathbf{v}_0$ is a particular solution of (1.2) and $\mathbf{h}$ is a solution of the corresponding homogeneous system then

$$A(\mathbf{v}_0 + \mathbf{h}) = A\mathbf{v}_0 + A\mathbf{h} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

so $\mathbf{v}_0 + \mathbf{h}$ is a solution of (1.2). Putting these observations together, we proved

Theorem 1.1.1. *Suppose $\mathbf{v}_0$ is a solution of $A\mathbf{x} = \mathbf{b}$. Then its solutions set is*

$$\{\mathbf{v}_0 + \mathbf{h} \colon A\mathbf{h} = \mathbf{0}\} = \mathbf{v}_0 + H$$

where is H is a solution set of $A\mathbf{x} = \mathbf{0}$.

Example 1.1.1. Consider the system of linear equations

$$\begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{1.5}$$

[1] i.e. an element of $\mathbb{K}^n$ that satisfying the equation.

That is

$$\begin{aligned} x_1 + 2x_2 - x_4 &= -1 \\ x_3 - 2x_4 &= 1 \end{aligned}$$

In this form, the variables x_1 and x_3 (correspond to the first non-zero entry of the rows) are readily expressed in terms of the other variables, i.e. x_2 and x_4.

$$\begin{aligned} x_1 &= -1 - 2x_2 + x_4 \\ x_3 &= 1 + 2x_4 \end{aligned}$$

From this, the solution sets of (1.5) can be readily expressed in the form $\mathbf{v}_0 + H$:

$$\begin{aligned} &\left\{ \begin{pmatrix} -1 - 2x_2 + x_4 \\ x_2 \\ 1 + 2x_4 \\ x_4 \end{pmatrix} : x_2, x_4 \in \mathbb{K} \right\} \\ &= \left\{ \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -2x_2 + x_4 \\ x_2 \\ 2x_4 \\ x_4 \end{pmatrix} : x_2, x_4 \in \mathbb{K} \right\} \\ &= \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \left\{ x_2 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : x_2, x_4 \in \mathbb{K} \right\} \end{aligned}$$

as described in Theorem 1.1.1.

1.2 Gauss-Jordan Elimination

In this section, we describe the most fundamental algorithm of linear algebra—the **Gauss-Jordan elimination**. A row (or a column) of a matrix is **non-zero** if it has at least one non-zero entry. The **pivot** of a non-zero row is the first non-zero entry of the row. The **augmented matrix** of a linear system is the matrix obtained by appending its constant vector to its coefficient matrix as the last column[2]:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array} \right).$$

For example, the augmented matrix of (1.5) is

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & -1 \\ 0 & 0 & 1 & -2 & 1 \end{array} \right).$$

[2]The vertical bar is not part of the matrix but just a typographical trick to remind us that the last column holds the constants.

This matrix has the special form:

Definition 1.2.1. A matrix is in **reduced row echelon form** (or **rref** in short) if

1. All nonzero rows are above any zero row;
2. The pivot of a nonzero row is always strictly to the right of the pivot of the row above it; and
3. Every pivot is 1 and is the only nonzero entry in its column.

As illustrated in Example 1.1.1, the solutions set of a system can be easily described if its augmented matrix is in rref. In this sense a system of linear equations is "solved" if its augmented matrix is in rref. The Gauss-Jordan elimination can be viewed as an algorithm that transforms a given matrix into an equivalent reduced row echolen (rre) matrix, here "equivalent" means their corresponding linear systems have the same solution sets.

Each step in the Gauss-Jordan elimination is an **elementary row operations** (or **ERO** in short) which is an operation of one of the following forms:

RS $R_k \leftrightarrow R_\ell$. Swapping the k-th and the ℓ-th row.

RM $R_k \to cR_k$. Multiply the k-row by a non-zero scalar c.

RR $R_k \to cR_\ell + R_k$ $(k \neq \ell)$. Add c times the ℓ-th row to the k-th row.

An **elementary matrix** is a matrix obtained by applying an ERO to an identity matrix. Suppose ρ is an ERO, denote by $E_\rho(m)$ the elementary matrix obtained by applying ρ to I_m. We often simplify the notation to just E_ρ. For instance, when we write $E_\rho A$, we understood that $E_\rho = E_\rho(m)$ where m is the number of rows of A. The follow are few key facts about elementary matrices.

1. Every elementary matrix is invertible (every elementary row operation is reversible). The inverse of an elementary matrix is also an elementary matrix (Exercise 1.1).
2. $E_\rho A$ is the matrix obtained by applying ρ to A.

An important consequence of these observations is:

Theorem 1.2.1. *EROs preserve solutions of linear systems.*

Proof. It is plain that every solution of $A\mathbf{x} = \mathbf{b}$ is a solution of $E_\rho A\mathbf{x} = E_\rho\mathbf{b}$. The latter system is, by the second observation, obtained by applying ρ to the first one. By multiplying E_ρ^{-1}, whose existence is asserted by the first fact, we see that the converse holds as well. □

Finally, we are in the position to state the Gauss-Jordan Elimination.

Round I: Elimination

1. If the matrix is a zero matrix (i.e. every entry is 0), stop and proceed to the next round.
2. Pick a row with a pivot in the first non-zero column and make it the 1st row by an RS.
3. Make the pivot of the first row 1 by an RM.
4. Make every entry underneath the pivot of the 1st row 0 by applying RR as many times as needed.
5. If the matrix has only one row, stop and proceed to the next round. Otherwise go through the steps again on the matrix obtained by deleting the first row.

Round II: Backward Substitution

1. Take the output of Round 1. If the matrix is zero, stop, otherwise make all the entries above the pivot of the last non-zero row 0 by applying RR as many times as needed.
2. Repeat the steps to the submatrix obtained by deleting the last row until the submatrix contains only a single row.

Example 1.2.1. Consider the system of linear equations

$$\begin{aligned} -x_1 + 2x_3 - x_4 &= -1 \\ -14x_1 + x_2 - 2x_4 &= 1 \\ 2x_1 - x_2 + 16x_3 + x_4 &= -3 \end{aligned} \tag{1.6}$$

So the coefficient matrix and the constant vector of the system are

$$\begin{pmatrix} -1 & 0 & 2 & -1 \\ -14 & 1 & 0 & -2 \\ 2 & -1 & 16 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 \\ 1 \\ -3 \end{pmatrix}$$

respectively. And the augmented matrix of the system is:

$$A = \left(\begin{array}{cccc|c} -1 & 0 & 2 & -1 & -1 \\ -14 & 1 & 0 & -2 & 1 \\ 2 & -1 & 16 & 1 & -3 \end{array}\right)$$

Now we apply Gauss-Jordan elimination to A. The first pass of Round I, performs the following EROs.

1. $A_1 = R1$

```
# R1 --> R1*(-1)
   A1 = A.with_rescaled_row(0,-1)
# R2 --> R1*(14) + R2
   A2 = A1.with_added_multiple_of_row(1,0,14)
# R3 --> R1*(-2) + R3
   A3 = A2.with_added_multiple_of_row(2,0,-2)
```

$$A3 = \left(\begin{array}{cccc|c} 1 & 0 & -2 & 1 & 1 \\ 0 & 1 & -28 & 12 & 15 \\ 0 & -1 & 20 & -1 & -5 \end{array}\right)$$

The submatrix obtained by deleting the first row is

```
A4 = A3.submatrix(1,0)
```

$$A4 = \left(\begin{array}{ccccc} 0 & 1 & -28 & 12 & 15 \\ 0 & -1 & 20 & -1 & -5 \end{array}\right)$$

Applying the steps in Part I to the submatrix yields a submatrix

```
A5 = A4.with_added_multiple_of_row(1,0,1)
```

$$A5 = \left(\begin{array}{ccccc} 0 & 1 & -28 & 12 & 15 \\ 0 & 0 & -8 & 11 & 10 \end{array}\right)$$

Again, the submatrix obtained by deleting the first row is non-zero:

```
A6 = A5.submatrix(1,0)
```

$$A6 = \left(\begin{array}{ccccc} 0 & 0 & -8 & 11 & 10 \end{array}\right)$$

Pass this through the steps in Part I again, we get

```
# R1 --> R1 *(-1/8)
A7 = A6.with_rescaled_row(0,-1/8)
```

$$A7 = \left(\begin{array}{ccccc} 0 & 0 & 1 & -\frac{11}{8} & -\frac{5}{4} \end{array}\right)$$

Thus the output of passing A to Round I is

$$A8 = \left(\begin{array}{ccccc} 1 & 0 & -2 & 1 & 1 \\ 0 & 1 & -28 & 12 & 15 \\ 0 & 0 & 1 & -\frac{11}{8} & -\frac{5}{4} \end{array}\right)$$

Now use $A8$ as the input for Round II.

```
A9 = A8.with_added_multiple_of_row(1,2,28)
A10 = A9.with_added_multiple_of_row(0,2,2)
```

$$A10 = \left(\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{7}{4} & -\frac{3}{2} \\ 0 & 1 & 0 & -\frac{53}{2} & -20 \\ 0 & 0 & 1 & -\frac{11}{8} & -\frac{5}{4} \end{array}\right)$$

The matrix is now in **reduce row echelon form** and the next two passes of Round II yields the same matrix. And so the algorithm stops and this is the output for the Gauss-Jordan elimination on the original matrix A.

The observant reader will note that the steps in Gauss-Jordan elimination are not unique. However,

Theorem 1.2.2. *The rref of a matrix is unique.*

In other words, every matrix is row equivalent to a unique rre matrix. In fact, it is not hard to see that if two rre matrices (of the same shape) are row equivalent then they must be equal (see Exercise 1.3)

Exercise

Exercise 1.1. Elementary Matrices

Exercise 1.2. Find the multiplicative inverse of the complex number $3 + 2\mathbf{i}$ as follows: suppose

$$(3 + 2\mathbf{i})(x + \mathbf{i}y) = 1$$

Then by expanding the left-hand side of the above equation and equating the real part and the imaginary part of both sides one get two linear equations in x, y. Find x, y by solving that system. More generally, if $a + b\mathbf{i} \neq 0$, find the multiplicative inverse of $a + b\mathbf{i}$ using this method.

Exercise 1.3. Show that if two rre matrices of the same shape are row equivalent then they are equal.

Chapter 2

Linear Independence

2.1 Vector Spaces

Let $\mathbb{K}$ be a field. A **vector space over** $\mathbb{K}$ (or a $\mathbb{K}$-vector space) is a set V together with two operations:

1. $+: V \times V \to V$ (addition)
2. $\cdot: \mathbb{K} \times V \to V$ (scalar multiplication)

which satisfy certain rules (see Appendix B). We call elements of V **vectors** and elements of $\mathbb{K}$ **scalars**. Unless otherwise stated, V always denote a $\mathbb{K}$-vector space. The following are some common examples of vector spaces. More examples are given in Exercise 2.1.

Example 2.1.1. The set $\mathbb{K}^n$ consisting of n-tuples of elements of $\mathbb{K}$ equipped with coordinate-wise addition and scalar multiplication:

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

Many familiar vector spaces, like $\mathbb{R}^2, \mathbb{R}^3$, as well as 2^n the space of n-bits are this form. Here $\mathbb{K} = 2 = \{0, 1\}$ is the field of two elements.

Example 2.1.2. The set $M_{m\times n}(\mathbb{K})$ of $m \times n$-matrices over $\mathbb{K}$ $(m, n \in \mathbb{N})$ with matrix addition i.e. $(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$ and scalar multiplication, i.e. $\lambda(a_{ij}) = (\lambda a_{ij})$.

Example 2.1.3. Fix an $n \geq 0$. The set of polynomials over $\mathbb{K}$ of degree $\leq n$, denoted by $\mathbb{K}_n[x]$, is a $\mathbb{K}$ vector space with polynomial addition and the usual scalar multiplication. As $\mathbb{K}$-vector spaces $\mathbb{K}_n[x]$ can be identified with $\mathbb{K}^{n+1}$, via the map $\sum_{k=0}^{n} a_k x^k \mapsto (a_0, a_1, \ldots, a_n)$.

Definition 2.1.1. A **subspace** of a vector space V is a *non-empty subset* of V that is closed under the restrictions of the operations of V. In other words, W is a subspace of V, denoted by $W \leq V$, if

1. $\emptyset \neq W \subseteq V$.
2. For any $\mathbf{w}_1, \mathbf{w}_2 \in W$, $\mathbf{w}_1 + \mathbf{w}_2 \in W$.
3. For any $a \in \mathbb{K}$ and $\mathbf{w} \in W$, $a\mathbf{w} \in W$.

Definition 2.1.2. Let X be a subset of V. A **linear combination of vectors in** X is an expression of the form

$$\sum_{\mathbf{v} \in X} a_{\mathbf{v}} \mathbf{v} \tag{2.1}$$

where the $a_{\mathbf{v}}$'s scalar and only finitely many of them are non-zero.

The scalar $a_{\mathbf{v}}$ is the **coefficient of** $\mathbf{v}$ in the linear combination (2.1). A linear combination is **trivial** if all its coefficients are zero. The empty linear combination (i.e. when $X = \emptyset$) is defined to be the zero vector and it has *no coefficients*. Since every linear combination, is a sum of finitely many non-zero vectors, we often write it as $a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m$ with the understanding that the $\mathbf{v}_i$'s are distinct. Also if $\mathbf{w}$ is the sum of the non-zero vectors of a linear combination, we also say that $\mathbf{w}$ is **expressed** (or **represented**) by the linear combination. Note that different linear combinations may well express the same vector.

Definition 2.1.3. The **span** of X denoted by $\langle X \rangle$, is the set of vectors in V that can be expressed by a linear combination of vectors in X.

It is easy to verify that $\langle X \rangle$ is a subspace of V and is the smallest subspace of V containing X, i.e. that every subspace of V that contains X also contains $\langle X \rangle$.

Example 2.1.4. Let $A \in M_{m\times n}(\mathbb{K})$. The span of the rows (resp. the columns) of A is called the **row space** (resp. the **column space**) of A. The row space of A is a subspace of $\mathbb{K}^n$ and the column space of A is a subspace of $\mathbb{K}^m$.

Example 2.1.5. Let $A \in M_{m\times n}(\mathbb{K})$. The solution set of the homogeneous system $A\mathbf{x} = \mathbf{0}$, i.e.

$$\{\mathbf{v} \in \mathbb{K}^n \colon A\mathbf{v} = \mathbf{0}\},$$

form a subspace of $\mathbb{K}^n$. We call it the **nullspace** (or the **kernel**) of A.

Example 2.1.6. Given a vector $\mathbf{v}_0 \in \mathbb{K}^n$, the set of matrices that "maps" it to $\mathbf{0} \in \mathbb{K}^m$, i.e.

$$\{A \in M_{m\times n}(\mathbb{K}) \colon A\mathbf{v}_0 = \mathbf{0}\}$$

is a subspace of $M_{m\times n}(\mathbb{K})$.

Example 2.1.7. The **diagonal** of a square matrix A consists of entries with the same row and column index, i.e. entries of the form a_{ii}. We define $\operatorname{tr}(A)$ the **trace** of A as the sum its diagonal entries. It is easy to check the set

$$\{A \in M_n(\mathbb{K}) \colon \operatorname{tr}(A) = 0\}$$

is a subspace of $M_n(\mathbb{K})$.

2.2 Linear Dependence

The key concept of linear algebra is linear dependence.

Definition 2.2.1. A subset X of a vector space is **linearly dependent** if the zero vector can be expressed as a non-trivial linear combination of vectors in X. We say that X is **linearly independent** if X is not linearly dependent.

In other words, X is linearly dependent if and only if there exist $a_1, \ldots, a_n \in \mathbb{K}$ not all zero and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in X$ such that $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$.

Proposition 2.2.1. *X is linearly dependent if and only if there are two distinct linear combinations of vectors in X having the same sum.*

Proof. Suppose $\sum_{\mathbf{v}\in X} a_{\mathbf{v}}\mathbf{v}$ and $\sum_{\mathbf{v}\in X} b_{\mathbf{v}}\mathbf{v}$ are two distinct linear combinations of vectors in X. Then there are still only finitely many $\mathbf{v} \in X$ such that $a_{\mathbf{v}} - b_{\mathbf{v}}$ is non-zero and at least one of them is since the linear combinations are assumed to be different. Therefore,

$$\sum_{\mathbf{v}\in X} (a_{\mathbf{v}} - b_{\mathbf{v}})\mathbf{v} = \sum_{\mathbf{v}\in X} a_{\mathbf{v}}\mathbf{v} - \sum_{\mathbf{v}\in X} b_{\mathbf{v}}\mathbf{v} = \mathbf{0}$$

is a non-trivial linear combination of vectors in X expressing $\mathbf{0}$. Thus X is linearly dependent. Conversely suppose X is linearly dependent, then $\mathbf{0}$ can be expressed as a non-trivial linear combination of vectors in X. So that linear combination and the trivial linear combination of vectors in X are distinct and have the same sum (namely $\mathbf{0}$). □

We leave the proof of the following proposition as an exercise.

Proposition 2.2.2. *X is linearly independent if and only if no vector in X can be expressed as a linear combination of the other vectors in X.*

Example 2.2.1. The empty set is linearly independent.

Example 2.2.2. Any set that contains the zero vector is linearly dependent since $1 \cdot \mathbf{0} = \mathbf{0}$.

Example 2.2.3. The subset $X = \{(1, 5, 3), (1, 9, 1), (3, 3, 1)\}$ of $\mathbb{R}^3$ is linearly independent. To see this, note that

$$a(1, 5, 3) + b(1, 9, 1) + c(3, 3, 1) = (0, 0, 0)$$

exactly means that (a, b, c) is a solution of the homogeneous system

$$\begin{pmatrix} 1 & 1 & 3 \\ 5 & 9 & 3 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

So X is linearly independent if and only if the system above has only trivial solution and this condition can be determined by the rref of the coefficient matrix A. In our case, the rref of A is

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

Thus our original system indeed only has trivial solution, therefore X is linearly independent.

Example 2.2.4. More generally, a finite subset $X = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of $\mathbb{K}^n$ is linearly independent if and only if the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only trivial solution where A is the $n \times m$-matrix with $\mathbf{v}_i$ as its i-th column ($1 \leq i \leq m$). The last fact is equivalent to the rref of A has the form I_m the identity $m \times m$-matrix followed by some, if any, zero rows. A consequence of this observation is that if m, the size of X, is larger than n, then X must be linearly dependent. This is because in that case, A will have more columns than rows and hence its rref must contain a column with no pivot and that means $A\mathbf{x} = \mathbf{0}$ has a free-variable and hence a non-trivial (i.e. non-zero) solution. Note that this observation is just a restatement of the following fundamental fact (some call it the *fundamental theorem of linear algebra*):

Theorem 2.2.3. *A homogeneous linear system with more unknowns than equations has a non-trivial solution.*

Definition 2.2.2. A subset B of a vector space V is a **spanning set** if $\langle B \rangle = V$. An independent spanning set is called a **basis**. An **ordered basis** is a basis together with a well order on it.

Example 2.2.5. Let $\mathbf{e}_i$ be the vector $(0, \ldots, 1, \ldots, 0) \in \mathbb{K}^n$ where the 1 appears at the i-th component. The ordered set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is called the **standard basis** of $\mathbb{K}^n$.

Example 2.2.6. Let $E_{ij} \in M_{n\times m}$ be the matrix with $e_{ij} = 1$ be the only non-zero entry. One checks readily that the set

$$\{E_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\}$$

is a basis of $M_{n\times m}(\mathbb{K})$.

Example 2.2.7. The powers of x, $1, x, x^2, \ldots$, form a basis of the vector space of polynomials $\mathbb{K}[x]$. The first $n+1$ of them, i.e. $1, x \ldots, x^n$ form a basis of P_n.

Every vector space has a basis. To establish this fact in full generality requires some assumptions[1] in dealing with infinite sets. We will prove this here under the assumption that the vector space V has *a finite spanning set* and indicate how that assumption can be removed in the exercises. First we need a lemma:

[1] namely Zorn's lemma or equivalently the axiom of choice.

Lemma 2.2.4. *Let U be an linearly independent subset of V and $\mathbf{w} \in V$. If $\mathbf{w} \notin \langle U \rangle$, then $U \cup \{\mathbf{w}\}$ is linearly independent as well.*

Proof. Suppose $U \cup \{\mathbf{w}\}$ is linearly dependent then some non-trivial linear combination $\sum_{\mathbf{u} \in U} a_{\mathbf{u}} \mathbf{u} + a_{\mathbf{w}} \mathbf{w} = \mathbf{0}$. Hence $-a_{\mathbf{w}} \mathbf{w} \in \langle U \rangle$ and by dividing $-a_{\mathbf{w}}$, $\mathbf{w} \in \langle U \rangle$ as well. The last step can be justified because if $a_{\mathbf{w}}$ were 0 then the linear combination involves only vectors in U but that contradicts the linear independence of U. □

Theorem 2.2.5. *Suppose U is a linearly independent subset of V and W is a finite spanning set of V. Then there exists a basis B of V such that $U \subseteq B \subseteq W$.*

Proof. If $W \subseteq \langle U \rangle$ then $V = \langle W \rangle \subseteq \langle \langle U \rangle \rangle = \langle U \rangle$. Thus U itself is a spanning set and hence a basis of V. Otherwise, some $\mathbf{w} \in W$ is not in $\langle U \rangle$ and by Lemma 2.2.4, $U \cup \{\mathbf{w}\} \subseteq W$ is linearly independent. So the same argument shows that either $U \cup \{\mathbf{w}\}$ is a basis of V or else can be extended by an element of W to a linearly independent subset of W. Since W is finite, this process eventually has to stop and we obtained a basis B of V such that $U \subseteq B \subseteq W$. □

Since the empty set is linearly independent, we conclude that if V has a finite spanning set then V has a finite basis.

Theorem 2.2.6. *Suppose U is a linearly independent subset of V and W is a finite spanning set of V then $|U| \leq |W|$.*

Proof. Suppose $W = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ and there are $\mathbf{u}_1, \ldots, \mathbf{u}_{n+1}$ distinct vectors in U. Since W is a spanning set, for each $1 \leq j \leq n+1$, $\mathbf{u}_j = \sum_{i=1}^{n} a_{ij} \mathbf{w}_i$. Since the matrix $\mathbf{A} = (a_{ij})$ has n rows and $n+1$ columns, by the fundamental theorem of linear algebra (Theorem 2.2.3) $\mathbf{A}\mathbf{v} = \mathbf{0}$ for some non-zero vector $\mathbf{v}$. Thus,

$$\sum_{j=1}^{n+1} v_j \mathbf{u}_j = \sum_{j=1}^{n+1} v_j \sum_{i=1}^{n} a_{ij} \mathbf{w}_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n+1} a_{ij} v_j \right) \mathbf{w}_i = \sum_{i=1}^{n} 0 \mathbf{w}_i = \mathbf{0}.$$

But this contradicts the assumption that U is linearly independent. □

It follows from Theorem 2.2.6 that if V has a finite spanning set then every basis of V is finite. Moreover any two bases of V have the same size. This justifies the following definition.

Definition 2.2.3. The **dimension** of V, denoted by $\dim(V)$, is the common cardinality of its bases.

Here are the dimensions of some vector spaces that we have encountered:

- The dimension of $\mathbb{K}^n$ is n since the standard basis has size n.
- The dimension of $M_{m \times n}(\mathbb{K})$ is mn.
- The dimension of $\mathbb{K}[x]$ is the cardinality of the set of natural numbers.
- The dimension of P_n is $n+1$.

2.3 Determinant

There is a more geometric way of thinking about linearly dependence in $\mathbb{R}^2$, namely two vectors in $\mathbb{R}^2$ are linearly independent if and only if they are the adjacent sides of a parallelogram with positive area. Likewise, three vectors in $\mathbb{R}^3$ are linearly independent if and only if the parallelepiped that they span has positive volume. Is there a natural generalization of this criteria of linear independence to an arbitrary finite dimensional vector space? The first thing to realize is that we have to be flexible on the idea of "volume" since an element of an arbitrary field needs not be a "number" and whether an element of a field is positive needs not make sense in general. However, we can still require our "volume" to be a function from the set of n-tuples of vectors from $\mathbb{K}^n$, or equivalently from $M_n(\mathbb{K})$ to $\mathbb{K}$ such that its value on A is non-zero if and only if the rows [2] of A are linearly independent. So what other properties should this function possess? In the case of $\mathbb{R}^2$, if we relax the requirement so that "area" can take any real number as value then it is not hard to see that for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$,

1. $\text{area}(\lambda\mathbf{u}, \mathbf{v}) = \lambda \text{ area}(\mathbf{u}, \mathbf{v}) = \text{area}(\mathbf{u}, \lambda\mathbf{v})$;
2. $\text{area}(\mathbf{u} + \mathbf{v}, \mathbf{w}) = \text{area}(\mathbf{u}, \mathbf{w}) + \text{area}(\mathbf{v}, \mathbf{w})$;
   $\text{area}(\mathbf{w}, \mathbf{u} + \mathbf{v}) = \text{area}(\mathbf{w}, \mathbf{u}) + \text{area}(\mathbf{w}, \mathbf{v})$; and
3. $\text{area}(\mathbf{u}, \mathbf{u}) = 0$.

The last property is clear: a collapsed parallelogram should have no area. The first property is also clear if we allow "area" to take arbitrary real number as value. The second property also follows from this relaxation and the observation that a parallelogram and a rectangle with the same base and height have the same area (so we can assume that $\mathbf{u}, \mathbf{v}$ are perpendicular to $\mathbf{w}$). In general, a function from a finite Cartesian product of vector spaces over $\mathbb{K}$ to $\mathbb{K}$ respect the vector space operations at each of is argument is called a **multilinear form**. We will discuss multilinear forms again in Chapter 5. We say that a multilinear form $f\colon V^n \to \mathbb{K}$ is **alternating** if it vanishes whenever two of its arguments are the same. This implies the value of the form is multiplied by -1 whenever two of its arguments are swapped, i.e.

$$f(\cdots, \mathbf{u}, \cdots, \mathbf{v}, \cdots) = -f(\cdots, \mathbf{v}, \cdots, \mathbf{u}, \cdots).$$

Moreover, these two conditions are equivalent if $1 \neq -1$ in $\mathbb{K}$ (Exercise). Finally it is reasonable to require our generalized "volume" to take the value 1 on the standard basis of $\mathbb{K}^n$ (i.e. the identity matrix I_n). Interesting, there is only one alternating multilinear form that fits the bill. More precisely,

Theorem 2.3.1. *There is a unique alternating multilinear form on $M_n(\mathbb{K})$, regarding the rows of a matrix as the arguments of the form, which takes value 1 on the identity matrix I_n.*

[2] one can also identify an n-tuple of vectors in $\mathbb{K}^n$ with the columns of an $n \times n$-matrix over $\mathbb{K}$

Proof. By multilinearity, any such form is determined by its values on matrices whose rows are the $\mathbf{e}_i$ $(1 \leq i \leq n)$. Being an alternating form, its value at such a matrix will be 0 if the matrix has some rows repeated, or else the matrix is simply obtained from I_n by permuting its rows, say s times[3], and so the value of the form at that matrix is $(-1)^s$. □

We denote this form[4] by det and call its value at A the **determinant** of A.

Example 2.3.1. It is instructive to compute the determinant of a 2×2 matrix.

$$\begin{aligned}\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \det(a\mathbf{e_1} + b\mathbf{e_2}, c\mathbf{e}_1 + d\mathbf{e_2}) \\ &= a\det(\mathbf{e}_1, c\mathbf{e}_1 + d\mathbf{e_2}) + b\det(\mathbf{e}_2, c\mathbf{e}_1 + d\mathbf{e_2}) \\ &= ac\det(\mathbf{e}_1, \mathbf{e}_1) + ad\det(\mathbf{e_1}, \mathbf{e}_2) + bc\det(\mathbf{e}_2, \mathbf{e}_1) + bd\det(\mathbf{e}_2, \mathbf{e}_2) \\ &= (ad - bc)\det(\mathbf{e_1}, \mathbf{e_2}) = ad - bc.\end{aligned}$$

With some book-keeping, it is not hard to generalize the previous computation to an arbitrary $n \times n$ matrix A and obtained the *Leibniz formula* for determinant:

$$\begin{aligned}\det(A) &= \det\left(\sum_j a_{1j}\mathbf{e}_j, \ldots, \sum_j a_{nj}\mathbf{e}_j\right) \\ &= \sum_\sigma a_{1\sigma(1)} \ldots a_{n\sigma(n)} \det\left(\mathbf{e}_{\sigma(1)}, \ldots, \mathbf{e}_{\sigma(n)}\right) \\ &= \sum_\sigma \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}\end{aligned}$$

where σ runs through the permutations of $\{1, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ is either 1 or -1 depending on whether σ is an even or odd permutation[5].

Proposition 2.3.2. *Let $E, M \in M_n(\mathbb{K})$ and E is elementary, then*

1. $\det(E) \neq 0$.
2. $\det(EM) = \det(E)\det(M)$.

We give a proof for the case when $E = E_\rho$ where ρ is the operation that adds λ times of the i-th row to the j-row (wlog $i < j$).

$$\det(E_\rho M) = \det \begin{pmatrix} \vdots \\ \mathbf{m}_i \\ \vdots \\ \lambda\mathbf{m}_i + \mathbf{m}_j \\ \vdots \end{pmatrix} = \lambda \det \begin{pmatrix} \vdots \\ \mathbf{m}_i \\ \vdots \\ \mathbf{m}_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ \mathbf{m}_i \\ \vdots \\ \mathbf{m}_j \\ \vdots \end{pmatrix} = \det M.$$

[3] The number s is certainly not uniquely determined by such matrix but its parity is.
[4] strictly speaking, one determinant form for each n.
[5] Every permutation is a product of transpositions. It is a fact that the parity of the number of transpositions needed depends only on the permutation.

By taking $M = I_n$, we get $\det(E_\rho) = 1$ and so $\det(E_\rho M) = \det(M) = \det(E_\rho)\det(M)$. We leave the proofs for the other two cases as an exercise to the reader.

Proposition 2.3.3. *For any $A, B \in M_n(\mathbb{K})$, $\det(AB) = \det(A)\det(B)$.*

Proof. We have $A = E_1 \cdots E_k R_A$ for some elementary matrices E_i ($1 \leq i \leq k$). By Proposition 2.3.2 (2), $\det(A) = \det(E_1) \cdots \det(E_n)\det(R_A)$. Since $\det(E_i) \neq 0$ (Proposition 2.3.2 (1)). Thus $\det(A) = 0$ if and only if $\det(R_A) = 0$. Since R_A is a row reduced echelon square matrix, if the rows of R_A are non-zero then $R_A = I_n$ which has determinant 1. Thus $\det(R_A) = 0$ implies its last row must be zero. But then the last row of $R_A B$ is zero as well and so $\det(R_A B) = 0$. Thus if $\det(A) = 0$ then we conclude from Proposition 2.3.2 (2) that

$$\det(A)\det(B) = 0 = \det(E_1) \cdots \det(E_k)\det(R_A B) = \det(AB)$$

Now if $\det(A) \neq 0$, then by the same argument R_A must be I_n and so $A = E_1 \cdots E_k$. Therefore, it follows from Proposition 2.3.2 (2) again that

$$\begin{aligned} \det(AB) &= \det(E_1 \cdots E_k B) = \det(E_1) \cdots \det(E_k)\det(B) \\ &= \det(E_1 \cdots E_k)\det(B) = \det(A)\det(B). \end{aligned}$$

□

Theorem 2.3.4. *For any $A \in M_n(\mathbb{K})$, The following conditions are equivalent:*

1. $\det(A) = 0$
2. *The rows of A are linearly dependent.*
3. *$A\mathbf{x} = \mathbf{0}$ has a non-trivial solution.*

Proof. $\det(A) \neq 0$ if and only if $R_A = I_n$ if and only if $\operatorname{rank}(R_A) = n$. Since $\operatorname{rank}(A) = \operatorname{rank}(R_A)$, so we conclude (1) and (2) are equivalent. Since $A\mathbf{v} = \sum v_i \mathbf{a}_i$, the equivalence of (2) and (3) follows. □

2.4 Direct Sums and Quotients

Definition 2.4.1. Given vector spaces V and W over the same field $\mathbb{K}$. The **direct sum** of V and W, denoted by $V \oplus W$, is the vector space on the Cartesian product $V \times W$ equipped with the operations:

- $(\mathbf{v}, \mathbf{w}) + (\mathbf{v}', \mathbf{w}') = (\mathbf{v} + \mathbf{v}', \mathbf{w} + \mathbf{w}')$
- $\alpha(\mathbf{v}, \mathbf{w}) = (\alpha\mathbf{v}, \alpha\mathbf{w})$

We call the two maps $\iota_1 \colon \mathbf{v} \mapsto (\mathbf{v}, \mathbf{0}_W)$ and $\iota_2 \colon \mathbf{w} \mapsto (\mathbf{0}_V, \mathbf{w})$ the **canonical embedding** of V and W, respectively. Clearly $\iota_1(V)$ is a copy V and $\iota_2(W)$ is a copy of W sitting inside $V \oplus W$. Note that the intersection of these two subspaces is the zero space.

It is clear that if $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of V and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ is a basis of W, then

$$\{\iota_1(\mathbf{v}_1), \dots, \iota_1(\mathbf{v}_n), \iota_2(\mathbf{w}_1), \dots \iota_2(\mathbf{w}_m)\}$$

is a basis of $V \oplus W$. In particular,

Proposition 2.4.1. *For finite dimensional spaces V and W,* $\dim(V \oplus W) = \dim V + \dim W$.

Example 2.4.1. $\mathbb{K}^n \oplus \mathbb{K}^m$ is isomorphic to K^{n+m}.

Let U be a subspace of V. We say that $\mathbf{v}, \mathbf{v}' \in V$ are U-equivalent if their difference $\mathbf{v} - \mathbf{v}'$ is in U. One checks easily this is indeed an equivalent relation and that the equivalent classes are of the form

$$\mathbf{v} + U = \{\mathbf{v} + \mathbf{u} \colon \mathbf{u} \in U\}.$$

Definition 2.4.2. Let U be a subspace of V. The **quotient space of V by U** is the set $V/U = \{\mathbf{v} + U \colon \mathbf{v} \in V\}$ of U-equivalent classes equipped with the operations:

- $(\mathbf{v} + U) + (\mathbf{v}' + U) = (\mathbf{v} + \mathbf{v}') + U$.
- $\alpha(\mathbf{v} + U) = (\alpha\mathbf{v}) + U$

Note that these operations on the equivalent classes are defined through the operations on the representatives and so one needs to check the independence of such choices. We refer the map

$$\pi_U \colon V \to V/U, \quad \mathbf{v} \overset{\pi_U}{\longmapsto} \mathbf{v} + U.$$

as the **canonical surjection** (associated to U). Clearly, π_U is a surjection and preserves vector space operations, i.e.

$$\pi_U(\mathbf{v} + \mathbf{v}') = \pi_U(\mathbf{v}) + \pi_U(\mathbf{v}')$$
$$\pi_U(\alpha\mathbf{v}) = \alpha\pi_U(\mathbf{v})$$

We will show that

Proposition 2.4.2. *Suppose U is a subspace of a finite dimensional space V. Then*

$$\dim V/U = \dim V - \dim U.$$

Example 2.4.2. In general, it is hard to "see" a quotient. However, in some cases, e.g. $V = \mathbb{R}^3$ and $U = xy$-plane, one can visualize the quotient V/U as a subspace of V consisting of one representative for each class $\mathbf{v} + U$. However, the choice of such subspace is still far from unique. In the case mentioned above, one can identify the quotient V/U with the z-axis. But in fact, any line ℓ passing through the origin and is not contained in the xy-plane will do.

Exercises

Exercise 2.1. Show that each of the following sets forms a $\mathbb{K}$-vector space with the given operations.

1. The set $K[x]$ of polynomials over $\mathbb{K}$ with polynomial addition and scalar multiplication.
2. The set $K^{\mathbb{N}}$ of sequences over $\mathbb{K}$ with addition and scalar multiplication defined by
   - $(x_k) + (y_k) = (x_k + y_k)$ and;
   - $a(x_k) = (ax_k)$
3. The set $K^{\mathbb{K}}$ of functions from $\mathbb{K}$ to $\mathbb{K}$ with function addition and scalar multiplication, i.e.
   - $(f + g)(x) = f(x) + g(x)$;
   - $(af)(x) = af(x)$.
4. Field extension: Let $\mathbb{L}/\mathbb{K}$ be a field extension, i.e. $\mathbb{K}$ is a subfield of $\mathbb{L}$. Then $\mathbb{L}$ has a natural $\mathbb{K}$-vector space structure, namely the addition and multiplication in $\mathbb{L}$. For example, $\mathbb{L} = \mathbb{Q}(\sqrt{2})$ and $\mathbb{K} = \mathbb{Q}$. Then for $a, b, r, s \in \mathbb{Q}$,
   - $(a + b\sqrt{2}) + (r + s\sqrt{2}) = (a + r) + (b + s)\sqrt{2}$
   - $a(r + s\sqrt{2}) = ar + as\sqrt{2}$

Exercise 2.2. Let α be a element in some extension of $\mathbb{K}$. Show that the set of polynomials over $\mathbb{K}$ vanish at α:

$$\{p(x) \in \mathbb{K}[x] \colon p(\alpha) = 0\}$$

is a subspace of $\mathbb{K}[x]$.

Exercise 2.3. Let c be the set of convergence real sequences. Show that c form a subspace of the space of real sequences.

Exercise 2.4. Show that the set of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ form a subspace of the space of all real functions of one variable.

Exercise 2.5. Show that the intersection of two subspaces of V is a subspace of V. More generally, show that the intersection of any family of subspaces of V is a subspace of V. Deduce that $\langle X \rangle$ is the intersection of the subspaces of V that contains X.

Exercise 2.6. Give an example that the union of two subspaces is not a subspace.

Exercise 2.7. Show that a subset X of a vector space is a subspace if and only if $\langle X \rangle = X$. Deduced that for any subset X of a vector space, $\langle \langle X \rangle \rangle = \langle X \rangle$.

Exercise 2.8. Show that a singleton $\{\mathbf{v}\}$ is linearly independent unless $\mathbf{v}$ is $\mathbf{0}$.

Exercise 2.9. Show that if the char $\mathbb{K} \neq 2$, then the subset

$$X = \{(0,1),(1,1),(1,-1)\}$$

of $\mathbb{K}^2$ is linearly dependent. What if char $\mathbb{K} = 2$?

Exercise 2.10. Prove Proposition 2.2.2.

Exercise 2.11. A set $\mathcal{C}$ of subsets of a given set is a **chain** if $\mathcal{C}$ is totally ordered by set inclusion. The union of a chain is defined to be the union of its elements. Show that

1. the union of a chain of subspaces is a subspace.
2. the union of a chain of linearly independent subsets is a linearly independent subset.

Exercise 2.12. Show that a maximal linearly independent subset of a vector space is also a spanning set (hence a basis). Here a linearly independent subset is maximal if it not properly contained in any linearly independent subset of the same vector space.

Exercise 2.13. Use Zorn's lemma, Exercise 2.11, Exercise 2.12 and Lemma 2.2.4 then the assuming of existence of finite spanning set can be removed.

Exercise 2.14 (Rank of a matrix). The **row** (resp. **column**) **rank** of a matrix is the dimension of its row (resp. column) space.

(i) Show that the row (resp. column) space of BA is a subspace of the row space of A (resp. the column space of B) whenever BA is defined.

(ii) Deduce that two row equivalent matrices have the same row space and column space.

(iii) Show that the row and the column rank of a rre matrix are the same.

(iv) Deduce that the row and the column rank of any matrix are the same.

This common value is called the **rank** of the matrix.

Exercise 2.15. Check that U-equivalence is an equivalence relation.

Exercise 2.16. Let U, W be subspaces of a vector space V. We define their **sum** to be

$$U + W = \{\mathbf{u} + \mathbf{w} \colon \mathbf{u} \in U, \mathbf{w} \in W\}$$

1. Show that $U + W$ is a subspace of V.
2. Let X, Y be subsets of V show that $\langle X \rangle + \langle Y \rangle = \langle X \cup Y \rangle$.

Exercise 2.17. Suppose $\dim V = n$ and $U \leq V$. Let $\mathcal{D} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a basis of U. And let

$$\mathcal{B} = \mathcal{D} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_{n-m}\}$$

be a basis of V extending $\mathcal{D}$. Show that $\mathcal{C} = \{\mathbf{v}_j + U \colon 1 \leq j \leq n-m\}$ is a basis of the quotient space V/U. Hence $\dim V/U = n - m = \dim V - \dim U$.

Chapter 3

Linear Maps

3.1 Linear Maps

An interesting map between two vector spaces should respect the vector space structures of its domain and codomain otherwise it will be just a map between sets.

Definition 3.1.1. A map $T\colon V \to W$ between two $\mathbb{K}$-vector spaces is **$\mathbb{K}$-linear** (or simply **linear**) if

- For any $\mathbf{v}_1, \mathbf{v}_2 \in V$, $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$.
- For any $\alpha \in \mathbb{K}$ and $\mathbf{v} \in V$, $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$.

So linear map (a.k.a **linear transform** or **homomorphism**) is a map between vector spaces that preserves the vector space operations between its domain and codomain. By a **monomorphism** we mean an injective linear map and a **epimorphism** we mean a surjective linear map. We say that a linear map $T\colon V \to W$ is an **isomorphism** if it is invertible, i.e. there exist a linear map $S\colon W \to V$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$. We will see that a linear map is an isomorphism if and only if it is both a monomorphism and an epimorphism (i.e. if and only if it is bijective as a map between sets). Two vector spaces are **isomorphic** if there is an isomorphism between them.

Example 3.1.1. The map sending every vector in V to the zero vector of W is linear. We call it the **zero map** from V to W.

As a result, $\mathrm{Hom}_{\mathbb{K}}(V, W)$ (or simply $\mathrm{Hom}(V, W)$), the set of linear maps from V to W is always non-empty. One checks directly that sum of two linear maps is linear and a scalar multiple of a linear map is linear (Exercise 3.2). Moreover, $\mathrm{Hom}(V, W)$ equipped with these two operations is a vector space over $\mathbb{K}$ with the zero map as its zero vector. If $V = W$, we write $\mathrm{End}(V)$ instead of $\mathrm{Hom}(V, V)$ and call its elements **endomorphisms** of V.

Example 3.1.2. The **identity map** id_V from V to itself is an endomorphism of V.

One checks directly that a composition of linear maps is linear and that $(\mathrm{End}(V), \mathbf{0}, \mathrm{id}_V)$ is a ring with function addition and composition (Exercise 3.3). The group of units of this ring, i.e. the group of isomorphisms from V to itself is called the **general linear group of** V and is denoted by $\mathrm{GL}(V)$. The identity element of $\mathrm{GL}(V)$ is id_V.

Example 3.1.3. Let V be a vector space. The inclusion map of a subspace of V into V is a monomorphism. In fact, a subset U of V is a subspace if and only if the inclusion of U into V is linear.

Example 3.1.4. The canonical embeddings of V and W into $V \oplus W$ are monomorphisms.

Example 3.1.5. The canonical projection $\pi_U : V \to V/U$ is an epimorphism.

Since linear maps preserve vector space operations, it is clear that

Proposition 3.1.1. *Let $T : V \to W$ be linear then $T(\langle X \rangle) = \langle T(X) \rangle$ for any $X \subseteq V$.*

Proposition 3.1.2. *The image (pre-image) of a subspace under a linear map is a subspace.*

Consequently,

Proof. Suppose $T : V \to W$ is linear and V' is a subspace of V. Since V' is a subspace, $\langle V' \rangle = V'$. So according to Proposition 3.1.1

$$T(V') = T(\langle V' \rangle) = \langle T(V') \rangle.$$

Therefore, $T(V')$ is a subspace of W. To see that the pre-image of a subspace is also a subspace, let W' be a subspace of W. So $W' = \langle W' \rangle$ and by Proposition 3.1.1

$$T(\langle T^{-1}(W') \rangle) = \langle TT^{-1}(W') \rangle \subseteq \langle W' \rangle = W'$$

Thus $\langle T^{-1}(W') \rangle \subseteq T^{-1}(W')$. But we the reverse inclusion always holds. Thus $T^{-1}(W') = \langle T^{-1}(W') \rangle$ is a subspace of V. □

Definition 3.1.2. The **kernel** of a linear map $T : V \to W$ is the set

$$\ker T = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}.$$

The **image** of T, $\mathrm{im}\, T$, is the set $T(V) = \{T(\mathbf{v}) : \mathbf{v} \in V\}$.

It follows from Proposition 3.1.2 that $\ker T$ is a subspace of V and $\mathrm{im}\, T$ is a subspace of W. Moreover, $T(\{\mathbf{0}_V\})$ is a subspace of W and is a singleton hence must be $\{\mathbf{0}_W\}$. This shows that $T(\mathbf{0}_V) = \mathbf{0}_W$. That means *a linear map must send the zero of its domain to the zero of its codomain.*

Proposition 3.1.3. *A linear map $T\colon V \to W$ is injective if and only if* $\ker T = \{\mathbf{0}_V\}$.

Proof. If T is injective, then $\ker T = T^{-1}(\mathbf{0}_W)$ is a singleton. Since it is also a subspace of V, so it must be $\{\mathbf{0}_V\}$. Conversely, suppose $\ker T = \{\mathbf{0}_V\}$ and that $T(\mathbf{v}_1) = T(\mathbf{v}_2)$. Then $\mathbf{v}_1 - \mathbf{v}_2 \in \ker T$ because

$$T(\mathbf{v}_1 - \mathbf{v}_2) = T(\mathbf{v}_1) - T(\mathbf{v}_2) = \mathbf{0}_W.$$

So $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}_V$, i.e. $\mathbf{v}_1 = \mathbf{v}_2$. Therefore, T is injective. □

Example 3.1.6. Let X be a linearly independent subset of V. Then any map T from the set X to a vector space W extends uniquely to a linear map $\widetilde{T}$ from $\langle X\rangle$ to W: for $\mathbf{v} \in \langle X\rangle$,

$$\widetilde{T}(\mathbf{v}) = \sum_{\mathbf{x}\in X} c_{\mathbf{x}} T(\mathbf{x})$$

where $\sum_{\mathbf{x}\in X} c_{\mathbf{x}}\mathbf{x}$ is the unique linear combination of vectors in X expressing $\mathbf{v}$. It is straight-forward to check that $\widetilde{T}$, defined as above, is linear and extends T to $\langle X\rangle$. For uniqueness, suppose $\widetilde{T}'$ is another linear extension of T to $\langle X\rangle$, then

$$\begin{aligned}\widetilde{T}'(\mathbf{v}) = \widetilde{T}'\left(\sum_{\mathbf{x}\in X} c_{\mathbf{x}}\mathbf{x}\right) &= \sum_{\mathbf{x}\in X} c_{\mathbf{x}}\widetilde{T}'(\mathbf{x}) \\ = \sum_{\mathbf{x}\in X} c_{\mathbf{x}} T(\mathbf{x}) &= \sum_{\mathbf{x}\in X} c_{\mathbf{x}}\widetilde{T}(\mathbf{x}) = \widetilde{T}(\mathbf{v})\end{aligned}$$

We call $\widetilde{T}$ the **extension of T by linearity**. Since this extension is canonical, from now on we use the same letter to denote a map and its linear extension to the span of its domain. In particular, every map from a basis X of V to a vector space W extends uniquely to a linear map from V to W. Consequently, a linear map T is completely determined by its action on a basis of its domain.

Definition 3.1.3. Let $T\colon V \to W$ be an epimorphism. A **section** of T is a linear map $S\colon W \to V$ such that $T \circ S = \mathrm{id}_W$.

Proposition 3.1.4. *Every epimorphism has a section.*

Proof. Pick a basis of $\mathcal{D}$ of W. Since T is surjective, for each $\mathbf{w} \in \mathcal{D}$, there exists $\mathbf{v}_{\mathbf{w}} \in V$, depending on $\mathbf{w}$, such that $T(\mathbf{v}_{\mathbf{w}}) = \mathbf{w}$. Now let S be the linear extension of the map which sends $\mathbf{w} \mapsto \mathbf{v}_{\mathbf{w}}$. Clearly we have

$$T \circ S(\mathbf{w}) = T(\mathbf{v}_{\mathbf{w}}) = \mathbf{w}, \qquad \forall \mathbf{w} \in \mathcal{D}.$$

And so by uniqueness of extension, we conclude that $T \circ S = \mathrm{id}_W$. □

Many statements in linear algebra (more generally, in algebra) can be conveniently phrased in terms of exact sequence. We give a brief discussion of this concept here.

Definition 3.1.4. A sequence of linear maps

$$U \xrightarrow{S} V \xrightarrow{T} W$$

is exact at V if $\ker T = \operatorname{im} S$. We say that a sequence of linear maps

$$\cdots V_{i-1} \longrightarrow V_i \longrightarrow V_{i+1} \cdots$$

is **exact** if it is exact at each V_i.

For example, instead of saying that a linear map $\iota\colon U \to V$ is injective, we can say that the sequence

$$0 \longrightarrow U \xrightarrow{\iota} V$$

is exact. This is because the image of the first map is simply the zero subspace of U so exactness at U means $\ker \iota = \{\mathbf{0}_U\}$ which precisely means that ι is injective (Proposition 3.1.3). Likewise, instead of say the linear map $\pi\colon V \to W$ is surjective, we can say that the sequence

$$V \xrightarrow{\pi} W \longrightarrow 0$$

is exact.

Example 3.1.7. Let $T\colon V \to W$ be a linear map. The sequence

$$0 \longrightarrow \ker T \longrightarrow V \xrightarrow{T} \operatorname{im} T \longrightarrow 0$$

is exact. We call it the **sequence exact associate to** T.

Theorem 3.1.5. *Suppose the sequence of linear maps*

$$0 \longrightarrow U \xrightarrow{\iota} V \xrightarrow{\pi} W \longrightarrow 0$$

is exact then $V \simeq U \oplus W$.

Proof. Since π is surjective, it has a section, say σ (Proposition 3.1.4). We claim that the linear map $\phi\colon U \oplus W \to V$ defined by

$$(\mathbf{u}, \mathbf{w}) \xmapsto{\phi} \iota(\mathbf{u}) + \sigma(\mathbf{w})$$

is an isomorphism. First we argue that $\mathbf{v} - \sigma\pi(\mathbf{v}) \in \operatorname{im} \iota$ for any $\mathbf{v} \in V$. To see this, apply π to this vector. Since σ is a section of π, we have

$$\pi(\mathbf{v} - \sigma\pi(\mathbf{v})) = \pi(\mathbf{v}) - \pi\sigma\pi(v) = \pi(\mathbf{v}) - \mathrm{id}_W(\pi(\mathbf{v})) = \mathbf{0}_W.$$

Thus $\mathbf{v} - \sigma\pi(\mathbf{v}) \in \ker \pi$ which is $\operatorname{im} \iota$ by exactness at V. By exactness at U, ι is injective, so there is a unique $\mathbf{u} \in U$ such that $\iota(\mathbf{u}) = \mathbf{v} - \sigma\pi(\mathbf{v})$. We leave it to the reader to check that the map γ defined by

$$\mathbf{v} \xmapsto{\gamma} (\mathbf{u}, \pi(\mathbf{v}))$$

is linear and is the inverse of ϕ. □

It follows from Theorem 3.1.5 and Example 3.1.7 that

Proposition 3.1.6. *For any linear map $T\colon V \to W$, $V \simeq \ker T \oplus \operatorname{im} T$.*

Corollary 3.1.7. *A bijective linear map is an isomorphism.*

Proof. Let $T\colon V \to W$ be a bijective linear map. Then $\operatorname{im} T = W$ and $\ker T = \{\mathbf{0}_V\}$. And the isomorphism given in Proposition 3.1.6 is

$$\mathbf{v} \longmapsto (\mathbf{0}_V, T(\mathbf{v})).$$

Note that $\{\mathbf{0}_V\} \oplus W \simeq W$ via $(\mathbf{0}_V, \mathbf{w}) \mapsto \mathbf{w}$. This completes the proof since T is the composition of these isomorphisms. □

Definition 3.1.5. The **rank** of a linear map T, denoted by $\operatorname{rank} T$, is the dimension of its image. The **nullity** of T, denoted by $\operatorname{nullity} T$, is the dimension of its kernel.

Now it follows from Proposition 3.1.6 and Proposition 2.4.1 that

Corollary 3.1.8. *Let T be a linear map with domain a finite dimensional vector space V, then*

$$\dim V = \operatorname{nullity} T + \operatorname{rank} T$$

Corollary 3.1.9. *Let U be a subspace of a finite dimensional space V, then*

$$\dim V/U = \dim V - \dim U.$$

Proof. It follows from Corollary 3.1.8 when $T = \pi_U$. □

Lemma 3.1.10. *Let $T\colon V \to W$ be linear and $X \subseteq V$. If $T(X)$ is linearly independent then so is X.*

Proof. Suppose $\sum_{i=1}^{k} a_i \mathbf{x}_i$ is a linear combination of vectors in X expressing $\mathbf{0}_V$, then

$$\mathbf{0}_W = T(\mathbf{0}_V) = T\left(\sum_{i=1}^{k} a_i \mathbf{x}_i\right) = \sum_{i=1}^{k} a_i T(\mathbf{x}_i).$$

Since $T(X)$ is linearly independent, we conclude that $a_i = 0$ for each $1 \leq i \leq k$. This shows that X is linearly independent. □

The converse of Lemma 3.1.10 is not true, i.e. $T(X)$ need not be independent even if X is. For example, take X to be a basis of V and T to be the zero map. We end this section by showing that dimension is the only isomorphic invariant of vector spaces.

Theorem 3.1.11. *Two vector spaces are isomorphic if and only if they have the same dimension.*

Proof. Suppose $T: V \to W$ is an isomorphism and $\mathcal{B}$ is a basis of V. Then by Proposition 3.1.1,

$$\langle T(\mathcal{B})\rangle = T(\langle \mathcal{B}\rangle) = T(V) = W.$$

So $T(\mathcal{B})$ spans W. Moreover, since $T^{-1}(T(\mathcal{B})) = \mathcal{B}$ is linearly independent and T^{-1} is linear, by Lemma 3.1.10, $T(\mathcal{B})$ is linearly independent as well. Thus $T(\mathcal{B})$ is a basis of W. Thus V and W have the same dimension.

Now suppose V and W have the same dimension. Let τ be a bijection from a basis of V to a basis of W. Let σ be the set theoretic inverse of τ. Let T be the extension of τ to V and S be the extension of σ to W. Since both $S \circ T$ and id_V are linear extensions of $\sigma \circ \tau$, by uniqueness $S \circ T = \mathrm{id}_V$. The same argument with the role of τ and σ reversed shows that $T \circ S = \mathrm{id}_W$. Thus V and W are isomorphic. □

Note that the proof of Theorem 3.1.11 actually works for arbitrary $\mathbb{K}$-vector spaces, finite dimensional or not.

Example 3.1.8. Let V be an n-dimensional vector space over $\mathbb{K}$. A well-order on a basis $\mathcal{B}$ of V can be viewed as a bijection from $\mathcal{B}$ to the standard basis of $\mathbb{K}^n$ and vice versa (the i-th element of $\mathcal{B}$ maps to $\mathbf{e}_i$ $(1 \le i \le n)$). Thus an ordered basis $\mathcal{B}$ of V determines, according to Theorem 3.1.11, a unique isomorphism $\Phi^{\mathcal{B}}$ from V to $\mathbb{K}^n$. The image of $\mathbf{v} \in V$ under $\Phi^{\mathcal{B}}$, denoted by $(\mathbf{v})^{\mathcal{B}}$, is called the **coordinate vector** (or simply the **coordinates**) of $\mathbf{v}$ **with respects to the ordered basis** $\mathcal{B}$.

Example 3.1.9. Since matrix multiplication distributes over addition and respects scalar multiplication, any $A \in M_{m\times n}(\mathbb{K})$ determines a linear map m_A from $\mathbb{K}^n$ to $\mathbb{K}^m$ defined by $m_A(\mathbf{v}) = A\mathbf{v}$. In fact, every linear map from $\mathbb{K}^n$ to $\mathbb{K}^m$ is a matrix multiplication: suppose T is a linear map from $\mathbb{K}^n$ to $\mathbb{K}^m$, then for any $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{K}^n$,

$$T(\mathbf{c}) = T\left(\sum c_i \mathbf{e}_i\right) = c_1 T(\mathbf{e}_1) + \ldots + c_n T(\mathbf{e}_n).$$

So $T = m_{(T)}$ where (T) is the $m \times n$ matrix with $T(\mathbf{e}_i) \in \mathbb{K}^m$ $(1 \le i \le n)$ as its ith column. In fact, one checks readily that $T \mapsto m_{(T)}$ is an isomorphism between $\mathrm{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ and $M_{m\times n}(\mathbb{K})$.

3.2 Matrix Representation of Linear Maps

In this section, we will show that every linear map $T: V \to W$ between finite dimensional vector spaces, after choosing ordered bases for V and W, is of the form m_A for some matrix A. **From now on when we talk about bases of vector spaces we always assumed they are ordered bases and we simply refer them as bases.**

Suppose $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{D} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ are bases of V and W, respectively. Given $\mathbf{v} \in V$, let $(\mathbf{v})^{\mathcal{B}} = (c_1, \ldots c_n) \in K^n$ be its coordinates with

respect to $\mathcal{B}$. By linearity of T, we have

$$T(\mathbf{v}) = T\left(\sum_{j=1}^{n} c_j \mathbf{v}_j\right) = \sum_{j=1}^{n} c_j T(\mathbf{v}_j).$$

For each $1 \leq j \leq n$, $T(\mathbf{v}_j)$ can be expressed uniquely as a linear combination $\sum_{i=1}^{m} a_{ij}\mathbf{w}_i$ of vectors in $\mathcal{D}$, (note that $(T(\mathbf{v}_j)^{\mathcal{D}} = (a_{1j}, \ldots, a_{mj})$), so

$$T(\mathbf{v}) = \sum_{j=1}^{n} c_j T(\mathbf{v}_j) = \sum_{j=1}^{n} c_j \sum_{i=1}^{m} a_{ij}\mathbf{w}_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}c_j\right) \mathbf{w}_i$$

By identifying the vectors with their coordinates with respect to these bases, we see that T is the map

$$\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \longmapsto \begin{pmatrix} \sum_{j=1}^{n} a_{1j}c_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}c_j \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

We call the matrix (a_{ij}) the **representation of T with respect to $\mathcal{B}$ and $\mathcal{D}$** and denote it by $(T)_{\mathcal{B}}^{\mathcal{D}}$. To be more precise, $(T)_{\mathcal{B}}^{\mathcal{D}}$ is the unique matrix that makes the follow diagram commutes:

$$\begin{array}{ccc} V & \xrightarrow{\quad T \quad} & W \\ {\scriptstyle \Phi^{\mathcal{B}}}\downarrow & & \downarrow{\scriptstyle \Phi^{\mathcal{D}}} \\ K^n & \xrightarrow[m_{(T)_{\mathcal{B}}^{\mathcal{D}}}]{} & K^m \end{array}$$

Definition 3.2.1. Let $\mathcal{B}$ and $\mathcal{B}'$ be two bases of V. The matrix $(\mathrm{id}_V)_{\mathcal{B}}^{\mathcal{B}'}$ is called the **change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$**. (see P.236 of [2]).

Note that the change of basis matrix $(\mathrm{id}_V)_{\mathcal{B}}^{\mathcal{B}'}$ simply sends (by left multiplication) the coordinates of a vector with respect to $\mathcal{B}$ to its coordinates with respect to $\mathcal{B}'$, i.e.

$$(\mathbf{v})^{\mathcal{B}'} = (\mathrm{id}_V)_{\mathcal{B}}^{\mathcal{B}'} (\mathbf{v})^{\mathcal{B}}$$

The following facts about change of basis matrices are clear:

Proposition 3.2.1. *Let V be an n-dimensional vector space, $\mathcal{B}, \mathcal{B}', \mathcal{B}''$ are bases of V.*

1. $(\mathrm{id}_V)_{\mathcal{B}}^{\mathcal{B}} = I_n$.
2. $(\mathrm{id}_V)_{\mathcal{B}'}^{\mathcal{B}''} (\mathrm{id}_V)_{\mathcal{B}}^{\mathcal{B}'} = (\mathrm{id}_V)_{\mathcal{B}}^{\mathcal{B}''}$.

Example 3.2.1. Consider the bases $\mathcal{B} = \{1, 1+x, 1+x+x^2\}$ and $\mathcal{D} = \{1, x-1, x^2-x\}$ of $\mathbb{R}_2[x]$. Let us find the change of basis matrix from $\mathcal{B}$ to $\mathcal{D}$. Since

$$\begin{aligned} 1 &= (1)1 + (0)(x-1) + (0)(x^2-x) \\ 1+x &= (2)1 + (1)(x-1) \\ 1+x+x^2 &= (3)1 + (2)(x-1) + (x^2-x) \end{aligned}$$

Therefore,

$$(1)^{\mathcal{D}} = (1,0,0), \quad (1+x)^{\mathcal{D}} = (2,1,0), \quad \left(1+x+x^2\right)^{\mathcal{D}} = (3,2,1)$$

and so the change of basis matrix from $\mathcal{B}$ to $\mathcal{D}$ is

$$(\mathrm{id})_{\mathcal{B}}^{\mathcal{D}} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

Example 3.2.2. From the parallelogram law of vector addition it is easy to see that R_α rotation by an angle α about the origin is a linear map from $\mathbb{R}^2$ to itself. To find the matrix representing R_α with respect to the standard basis $\mathcal{E} = \{(1,0),(0,1)\}$ of $\mathbb{R}^2$, we compute the coordinates of the images of $(1,0)$ and $(0,1)$ with respect to $\mathcal{E}$.

$$R_\alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix} \qquad R_\alpha \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin(\alpha) \\ \cos(\alpha) \end{pmatrix}$$

And so

$$(R_\alpha)_{\mathcal{E}}^{\mathcal{E}} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

The map $\rho_{\mathcal{B}}^{\mathcal{D}}$ sending T to $(T)_{\mathcal{B}}^{\mathcal{D}}$ is in fact an isomorphism from $\mathrm{Hom}_{\mathbb{K}}(V,W)$ to $M_{m\times n}(\mathbb{K})$. Consequently, for any finite dimensional vector spaces V and W,

$$\dim \mathrm{Hom}(V,W) = \dim V \cdot \dim W.$$

Furthermore, an isomorphism of the form $\rho_{\mathcal{B}}^{\mathcal{D}}$ gives us a dictionary between linear maps (abstract) and matrices (concrete). For example, image of a linear map corresponds to the column-space of a matrix, kernel corresponds to nullspace. As a result, Corollary 3.1.8, is translated into: For any $A \in M_{m\times n}(\mathbb{K})$,

$$\mathrm{nullity}\, A + \mathrm{rank}\, A = n.$$

Since composition of linear maps corresponds to multiplication of matrices (see Exercise 3.5), when $V = W$ and $\mathcal{B} = \mathcal{D}$ then $\rho_{\mathcal{B}}^{\mathcal{B}}$ (or simply $\rho_{\mathcal{B}}$) is a $\mathbb{K}$-algebra isomorphism from $\mathrm{End}(V)$ to $M_n(\mathbb{K})$. And $\mathrm{GL}(V)$, the invertible elements of $\mathrm{End}(V)$ maps to the invertible elements of $M_n(\mathbb{K})$. This leads to the following definition:

Definition 3.2.2. A square matrix $A \in M_n(\mathbb{K})$ is **invertible** if there exists a matrix $B \in M_n(\mathbb{K})$ such that $AB = I_n = BA$.

Note that the inverse of A, if exists, is unique. It is because if B and B' are inverses of A, then

$$B = BI_n = B(AB') = (BA)B' = I_nB' = B'.$$

We denote the inverse of A, if exists, by A^{-1}.

By setting $\mathcal{B}'' = \mathcal{B}$ in Proposition 3.2.1, we see that a change of basis matrix must be invertible: $(\mathrm{id}_V)^{\mathcal{B}}_{\mathcal{B}'}$ and $(\mathrm{id}_V)^{\mathcal{B}'}_{\mathcal{B}}$ are inverses of each other. The matrix $(\mathrm{id}_V)^{\mathcal{B}}_{\mathcal{B}'}$ is also called the **transition matrix** from $\mathcal{B}$ to $\mathcal{B}'$ (it is the matrix $P_{\mathcal{B},\mathcal{B}'}$ on p.11 [1]).

3.3 Equivalence

By choosing different bases, we have seen how different matrices can represent the same linear transformation. So it is natural to make the following definition:

Definition 3.3.1. Two matrices are **equivalent** if they represent the same linear transformation from V to W.

The relation defined above is clearly an equivalence relation on the space of matrices of fixed shape. To define it completely in terms of matrices, suppose A and $A' \in M_{m\times n}(\mathbb{K})$ represent the same linear map T, i.e.

$$A = (T)^{\mathcal{D}}_{\mathcal{B}}, \quad \text{and} \quad A' = (T)^{\mathcal{D}'}_{\mathcal{B}'}$$

for some bases $\mathcal{B}, \mathcal{B}'$ of V and $\mathcal{D}, \mathcal{D}'$ of W. Let $P = (\mathrm{id}_V)^{\mathcal{B}}_{\mathcal{B}'}$ and $Q = (\mathrm{id}_W)^{\mathcal{D}}_{\mathcal{D}'}$ be the corresponding transition matrices. Since

$$(T)^{\mathcal{D}'}_{\mathcal{B}'} = (\mathrm{id}_W)^{\mathcal{D}'}_{\mathcal{D}}(T)^{\mathcal{D}}_{\mathcal{B}}(\mathrm{id}_V)^{\mathcal{B}}_{\mathcal{B}'},$$

and transition matrices are invertible, we have

$$A' = Q^{-1}AP. \tag{3.1}$$

Conversely, since every invertible matrix can be regards as a transition matrix, so if there exists $P \in \mathrm{GL}_n(\mathbb{K})$ and $Q \in \mathrm{GL}_m(\mathbb{K})$ such that $A' = Q^{-1}AP$, then A and A' represent the same linear map[1]. Thus we conclude that

> *Two matrices $A, A' \in M_{m\times n}(\mathbb{K})$ are equivalent if and only if there exists $P \in \mathrm{GL}_n(\mathbb{K})$ and $Q \in \mathrm{GL}_m(\mathbb{K})$ such that $A' = Q^{-1}AP$.*

The next question is: How can we tell whether two matrices $A, A' \in M_{m\times n}(\mathbb{K})$ are equivalent? Suppose $\operatorname{rank} A = r$ then the nullity of A is $n-r$. Pick an basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_{n-r}\}$ of $\ker A$ and extends it to a basis

$$\mathcal{B}_A = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{u}_1, \ldots, \mathbf{u}_{n-r}\}$$

[1] Since Q is invertible, so is Q^{-1}, hence the condition is certainly the same as the existence of invertible Q and P such that $A' = QAP$

of $\mathbb{K}^n$. We claim that $A\mathbf{v}_1, \dots A\mathbf{v}_r$ are linearly independent. Suppose

$$\mathbf{0}_W = \alpha_1 A\mathbf{v}_1 + \dots \alpha_r A\mathbf{v}_r = A\left(\sum_i \alpha_i \mathbf{v}_i\right)$$

for some α_i's in $\mathbb{K}$. Then $\sum_i \alpha_i \mathbf{v}_i \in \ker A$ and so there exists β_j's in $\mathbb{K}$ such that

$$\sum_{i=1}^{r} \alpha_i \mathbf{v}_i = \sum_{j=1}^{n-r} \beta_j \mathbf{u}_j.$$

That means

$$\sum_{i=1}^{r} \alpha_i \mathbf{v}_i - \sum_{j=1}^{n-r} \beta_j \mathbf{u}_j = \mathbf{0}_V.$$

Since $\mathcal{B}_A$ is a basis of V, we must be $\alpha_i = 0$ for each i. Hence the $A\mathbf{v}_i$ form a linearly independent subset of $\mathbb{K}^m$. Now extends this set to an basis

$$\mathcal{D}_A = \{A\mathbf{v}_1, \dots, A\mathbf{v}_r, \mathbf{w}_1, \dots, \mathbf{w}_{m-r}\}$$

of $\mathbb{K}^m$. It is clear that

$$(\mathrm{id}_W)_{\mathcal{E}_m}^{\mathcal{D}_A} A (\mathrm{id}_V)_{\mathcal{B}_A}^{\mathcal{E}_n} = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{3.2}$$

where the $\mathbf{0}$'s represent blocks of zeros of appropriate shapes. Equation (3.2) shows that A is equivalent to, $I_{m,n,r}$, the matrix on its right-hand side. Since $A \in M_{m\times n}(\mathbb{K})$ is arbitrary, we conclude that any $m \times n$-matrices of rank r is equivalent to $I_{m,n,r}$. Consequently,

Theorem 3.3.1. *Two matrices of the same shape are equivalent if and only if they have the same rank.*

Example 3.3.1. Consider the 4×3-matrix $M = \begin{pmatrix} 6 & 1 & -2 \\ 4 & 0 & 1 \\ -2 & 1 & -4 \\ 14 & -3 & 14 \end{pmatrix}$. Its reduced row echelon form is

$$\begin{pmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & -\frac{7}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so M is of rank 2. We will find $Q \in \mathrm{GL}_4(\mathbb{K})$ and $P \in \mathrm{GL}_3(\mathbb{K})$ such that

$$Q^{-1} M P = \begin{pmatrix} I_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

First, from the rref of M, we know that its kernel (or nullspace) is generated by $\mathbf{u}_1 = (-1/4, 7/2, 1)^T$. We extend it to a basis

$$\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{u}\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{4} \\ \frac{7}{2} \\ 1 \end{pmatrix} \right\}.$$

So we can take P to be $(\mathrm{id})_{\mathcal{B}}^{\mathcal{E}_3} = \begin{pmatrix} 1 & 0 & -\frac{1}{4} \\ 0 & 1 & \frac{7}{2} \\ 0 & 0 & 1 \end{pmatrix}$. From the previous discussion, we know that

$$M\mathbf{e}_1 = \begin{pmatrix} 6 \\ 4 \\ -2 \\ 14 \end{pmatrix} \text{ and } M\mathbf{e}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ -3 \end{pmatrix}$$

are linearly independent (in fact, form a basis of the column space of M). We extend this to a basis of $\mathbb{K}^4$ by appending two vectors at the end:

$$\mathcal{D} = \left\{ \begin{pmatrix} 6 \\ 4 \\ -2 \\ 14 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

It may not be completely clear that $\mathcal{D}$ is a basis of $\mathbb{K}^4$ but we can check that by computing the rref of

$$Q := (\mathrm{id})_{\mathcal{D}}^{\mathcal{E}_4} = \begin{pmatrix} 6 & 1 & 1 & 0 \\ 4 & 0 & 0 & 1 \\ -2 & 1 & 0 & 0 \\ 14 & -3 & 0 & 0 \end{pmatrix}$$

which is

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

So $\mathcal{D}$ is indeed a basis of $\mathbb{K}^4$. Lastly, we compute the inverse of Q by finding the rref of the matrix $(Q \mid I_4)$,

$$\begin{pmatrix} 6 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 14 & -3 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{\text{row reduce}}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \frac{3}{8} & \frac{1}{8} \\ 0 & 1 & 0 & 0 & 0 & 0 & \frac{7}{4} & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 1 & 0 & -4 & -1 \\ 0 & 0 & 0 & 1 & 0 & 1 & -\frac{3}{2} & -\frac{1}{2} \end{pmatrix}$$

Finally, we can check our computation

$$Q^{-1}MP = \begin{pmatrix} 0 & 0 & \frac{3}{8} & \frac{1}{8} \\ 0 & 0 & \frac{7}{4} & \frac{1}{4} \\ 1 & 0 & -4 & -1 \\ 0 & 1 & -\frac{3}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 6 & 1 & -2 \\ 4 & 0 & 1 \\ -2 & 1 & -4 \\ 14 & -3 & 14 \end{pmatrix} \begin{pmatrix} 1 & 0 & -\frac{1}{4} \\ 0 & 1 & \frac{7}{2} \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Exercises

Exercise 3.1. Show that the map sending f to its derivative f' is a linear map on the space of smooth functions.

Exercise 3.2. Show the sum of two linear maps is linear and that any scalar multiple of a linear map is linear.

Exercise 3.3. Show that if $T \in Hom(U, V)$ and $S \in \mathrm{Hom}(V, W)$ then $S \circ T \in \mathrm{Hom}(U, W)$. Verify that $\mathrm{End}(V)$ is a ring (in fact a $\mathbb{K}$-algebra) with 0 the zero map and 1 the identity map id_V under function addition and composition.

Exercise 3.4. Let $T\colon V \to W$ be linear. Show that

1. $T^{-1}(U)$ is a subspace of V for any U subspace of W.
2. $T(Z)$ is a subspace of W for any Z subspace of V.

Exercise 3.5. Suppose $\dim(V) = n$ and $\dim(W) = m$. Let $\mathcal{B}$ be a basis of V and $\mathcal{D}$ be a basis of W. Show that

(a) the map $\rho_{\mathcal{B}}^{\mathcal{D}}\colon \mathrm{Hom}(V, W) \to M_{m\times n}(\mathbb{K})$

$$\rho_{\mathcal{B}}^{\mathcal{D}}\colon T \longmapsto (T)_{\mathcal{B}}^{\mathcal{D}}$$

is an isomorphism of vector spaces.

(b) if $V = W$ and $\mathcal{B} = \mathcal{D}$ then $\rho_{\mathcal{B}}^{\mathcal{D}}$ is a $\mathbb{K}$-algebra isomorphism between $\mathrm{End}(V)$ and $M_n(\mathbb{K})$. In particular, the units of $\mathrm{End}(V)$ corresponds to the invertible matrices.

Exercise 3.6. Verify that $\mathrm{GL}_n(\mathbb{K})$ the set of invertible $(n \times n)$-matrices over $\mathbb{K}$ is a group under matrix multiplication with I_n as its identity element.

Exercise 3.7. Compute the change of basis matrix $(\mathrm{id})_{\mathcal{D}}^{\mathcal{B}}$ in Example 3.2.1.

Exercise 3.8. Using the fact that $R_{\alpha+\beta} = R_\beta R_\alpha$,

(a) show that

$$\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \sin(\beta)\cos(\alpha)$$
$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta).$$

(b) express $\cos(3x)$ in terms for $\sin(x)$ and $\cos(x)$.

Exercise 3.9. Consider the bases

$$\mathcal{B} = (1, 1-x, 1-x+x^2), \quad \mathcal{D} = (x^2+x, 2x+1, -1), \quad \mathcal{E} = (1, x, x^2)$$

of $\mathbb{R}_2[x]$ and the linear transform $T\colon \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ given by

$$\begin{aligned} 1 &\longmapsto 3x + 3x^2 \\ x &\longmapsto -3 - 3x + 3x^2 \\ x^2 &\longmapsto 4 - 6x \end{aligned}$$

Compute

(a) $(\mathrm{id})^{\mathcal{E}}_{\mathcal{B}}$, $(\mathrm{id})^{\mathcal{B}}_{\mathcal{E}}$ and $(\mathrm{id})^{\mathcal{E}}_{\mathcal{D}}$.

(b) The product of $(\mathrm{id})^{\mathcal{E}}_{\mathcal{B}}(\mathrm{id})^{\mathcal{B}}_{\mathcal{E}}$.

(c) The matrix representation of T with respect to the basis $\mathcal{E}$ and $\mathcal{E}$, i.e. the matrix $(T)^{\mathcal{E}}_{\mathcal{E}}$.

(d) The matrix representation of T with respect to the basis $\mathcal{B}$ and $\mathcal{D}$, i.e. the matrix $(T)^{\mathcal{D}}_{\mathcal{B}}$.

(e) $(\mathrm{id})^{\mathcal{E}}_{\mathcal{D}}(T)^{\mathcal{D}}_{\mathcal{B}}(\mathrm{id})^{\mathcal{B}}_{\mathcal{E}}$. How does this matrix compare to the one in Part (c)?

Exercise 3.10. Show that every linear map $T\colon V \to W$ factors as an epimorphism $\pi_T\colon V \to V/\ker T$ followed by an monomorphism $\iota_T\colon V/\ker T \to W$.

Exercise 3.11. Let

$$0 \to V_1 \to V_2 \to \cdots \to V_{n-1} \to V_n \to 0$$

be a sequence of vector spaces. The alternating sum of the dimensions of the V_i's, i.e. $\sum_{i=1}^{n}(-1)^i \dim(V_i) = 0$, is called the **Euler-Poincaré Characteristic** of the sequence. Show that the Euler-Poincaré characteristic of an exact sequence is 0. Note that this generalizes Corollary 3.1.8.

Chapter 4

Endomorphisms

In this chapter we focus on the endomorphisms of a vector space. Recall that the endomorphisms of V forms a ring $\mathrm{End}(V)$ with respect to function addition and composition (as multiplication). The 0 and 1 of $\mathrm{End}(V)$ are the zero map and the identity map of V, respectively.

4.1 Similarity

There is a subtle but important different between the general representation theory of linear maps and endomorphisms. To represent a linear map T by a matrix, one chooses ordered bases, one for its domain and one for its codomain. These bases can be different even when the vector spaces are the same. However, in representing of an endomorphism, we regard the linear map as a map from the structure $(V, \mathcal{B})$ to itself. In other words, the ordered bases chosen for its domain and codomain are the same. So we simply write $(T)_{\mathcal{B}}$ instead of $(T)^{\mathcal{B}}_{\mathcal{B}}$ for the matrix that represents $T \in \mathrm{End}(V)$ with respect to the basis $\mathcal{B}$ of V. The impact of this requirement on the change of basis formula is as follows: Suppose $\mathcal{B}'$ is another basis of V, then $(T)_{\mathcal{B}'}$ and $(T)_{\mathcal{B}}$ are related by

$$(T)_{\mathcal{B}'} = (I_V)^{\mathcal{B}'}_{\mathcal{B}} (T)_{\mathcal{B}} (I_V)^{\mathcal{B}}_{\mathcal{B}'} \tag{4.1}$$

So the transition matrices $(I_V)^{\mathcal{B}'}_{\mathcal{B}}$ and $(I_V)^{\mathcal{B}}_{\mathcal{B}'}$ involved are inverse of each other. This naturally leads to the following definition:

Definition 4.1.1. We say that $A, A' \in M_n(\mathbb{K})$ are **similar** if there exists $P \in \mathrm{GL}_n(K)$ such that $A' = P^{-1}AP$.

Clearly, similarity is an equivalence relation. Since invertible matrices can be regarded as transition matrices, thus *two matrices are similar if and only if they represent the same endomorphism.* Two natural questions arise:

1. Can we decide whether two square matrices are similar?
2. Is there any particularly simple element in each similar class?

4.2 Diagonalization

Given $\tau \in \mathrm{End}(V)$, it is desirable to find a basis of V so that the representation of τ is as simple as possible. As we seen, every square matrix is equivalent to a diagonal matrix. However, as we will see, not every square matrix is similar to a diagonal matrix. Indeed,

Theorem 4.2.1. *A necessary and sufficient condition for an endomorphism τ of V to be representable by a diagonal matrix is the existence of a basis $\mathcal{B}$ of V such that for every $\mathbf{v} \in \mathcal{B}$, there is a $\lambda_{\mathbf{v}} \in \mathbb{K}$ such that $\tau(\mathbf{v}) = \lambda_{\mathbf{v}}\mathbf{v}$.*

Proof. If such a basis $\mathcal{B}$ of V exists, then $(\tau)_{\mathcal{B}}$ is clearly a diagonal matrix with the $\lambda_{\mathbf{v}}$'s on its diagonal. Conversely, if for some basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V, $(\tau)_{\mathcal{B}}$ is a diagonal matrix with λ_i the i-th diagonal entry, then $\tau(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ for each $1 \leq i \leq n$. □

Example 4.2.1. It follows immediately from Theorem 4.2.1 that any non-trivial rotation R_θ of $\mathbb{R}^2$ cannot be represented by a diagonal matrix.

Definition 4.2.1. We say that an endomorphism τ of V is **diagonalizable** if τ is represented by a diagonal matrix with respect to some basis of V.

The scalars $\lambda_{\mathbf{v}}$ appear in Theorem 4.2.1 have a special name:

Definition 4.2.2. A scalar $\lambda \in \mathbb{K}$ is an **eigenvalue** of an endomorphism $\tau \in \mathrm{End}(V)$ if the solution space

$$\tau(\mathbf{x}) = \lambda\mathbf{x}$$

is non-trivial. The solution space, in this case, is called the **λ-eigenspace**. Any nonzero vector in the λ-eigenspace is called an **λ-eigenvector**.

With this terminology, Theorem 4.2.1 can be restated as:

> *an endomorphism τ of V is diagonalizable if and only if there is a basis of V consists of eigenvectors of τ.*

To find the eigenvalues of τ, let A be a representation (with respect to some basis of V) of τ. Then λ is an eigenvalue of τ if and only if the solution space of the equation

$$A\mathbf{x} = \lambda\mathbf{x}$$

is non-trivial. In other words, $(\lambda I_n - A)\mathbf{x} = \mathbf{0}$ has a non-trivial solution, i.e. λ is a root of the **characteristic polynomial of τ**

$$\chi_\tau(x) := \det(xI_n - A). \tag{4.2}$$

Note that $\chi_\tau(x)$ is a monic polynomial of degree n over K. Moreover, χ_τ only depends on τ but not its representations. To see this, suppose A' is another representation of τ then $A = P^{-1}AP$ for some invertible P and so

$$\begin{aligned}\det(xI_n - A') &= \det(P^{-1}(xI_n - A)P) \\ &= \det(P^{-1})\det(xI_n - A)\det(P) = \det((P)^{-1})\det(P)\det(xI_n - A) \\ &= \det(xI_n - A).\end{aligned}$$

Example 4.2.2. Consider $R_{\frac{\pi}{2}}$ the rotation by $\pi/2$ on $\mathbb{R}^2$. Its representation with respect to the standard basis of $\mathbb{R}^2$ is

$$R_{\frac{\pi}{2}} = \begin{pmatrix} \cos(\pi/2) & -\sin(\pi/2) \\ \sin(\pi/2) & \cos(\pi/2) \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

The characteristic polynomial of $\chi(x)$ of $R_{\frac{\pi}{2}}$ is

$$\det(xI_2 - R_{\frac{\pi}{2}}) = \det \begin{pmatrix} x & 1 \\ -1 & x \end{pmatrix} = x^2 + 1$$

which has no real solution.

This example shows a subtle point: while the coefficients of a characteristic polynomial is always in $\mathbb{K}$, its roots may not be. In this example, $\mathbb{K} = \mathbb{R}$ but if we think of $R_{\frac{\pi}{2}}$ as an endomorphism of $\mathbb{C}^2$, then i and $-i$ are its eigenvalues.

As we can identify an endomorphism of a finite dimensional vector space with a similarity class of square matrices, and we define the concepts of eigenvalue, eigenspace, etc for similarity classes accordingly.

Example 4.2.3. Consider the matrix $M = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. The characteristic polynomial of M is

$$\det(xI_2 - M) = \det \begin{pmatrix} x & -1 \\ -1 & x-1 \end{pmatrix} = x(x-1) - 1 = x^2 - x - 1.$$

Hence the eigenvalues of M are $\lambda_1 = (1+\sqrt{5})/2, \lambda_2 = (1-\sqrt{5})/2$. By solving the system

$$\begin{pmatrix} \frac{1+\sqrt{5}}{2} & -1 \\ -1 & \frac{\sqrt{5}-1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

we see that eigenspace of M corresponding to the eigenvalue $\dfrac{1+\sqrt{5}}{2}$ is

$$\left\{ x_2 \begin{pmatrix} \frac{\sqrt{5}-1}{2} \\ 1 \end{pmatrix} : x_2 \in K \right\}$$

and is generated by $\begin{pmatrix} -\lambda_2 \\ 1 \end{pmatrix}$. Likewise, the eigenspace of M corresponding to the eigenvalue $\dfrac{1-\sqrt{5}}{2}$ is generated by $\begin{pmatrix} -\lambda_1 \\ 1 \end{pmatrix}$. These eigenvectors are clearly independent and hence form a basis, say $\mathcal{B}$ of $\mathbb{R}^2$. The representation of m_M with respect to $\mathcal{B}$ is a diagonal matrix:

$$P := (I)^{\mathcal{E}}_{\mathcal{B}} = \begin{pmatrix} -\lambda_2 & -\lambda_1 \\ 1 & 1 \end{pmatrix}$$

and its inverse is

$$P^{-1} = (I)_{\mathcal{E}}^{\mathcal{B}} = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} 1 & \lambda_1 \\ -1 & -\lambda_2 \end{pmatrix}.$$

The representation of m_M with respect to $\mathcal{B}$ is

$$(m_M)_{\mathcal{B}} = P^{-1}MP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

The fact that the eigenvectors in Example 4.2.3 are linearly independent is not a coincidence.

Proposition 4.2.2. *Eigenvectors of distinct eigenvalues are linearly independent.*

Proof. We prove this by induction on k the number of eigenvectors. By definition an eigenvector is non-zero so the case $k = 1$ follows. Suppose the statement is true for some $k = r$ and $\mathbf{v}_i$ is a λ_i-eigenvectors $(1 \leq i \leq r+1)$ of an endomorphism τ with distinct λ_i's. Suppose $\sum_{i=1}^{r+1} a_i \mathbf{v}_i = \mathbf{0}$ then by applying τ to this relation, we get

$$\begin{aligned} &\tau\left(a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r + a_{r+1}\mathbf{v}_{r+1}\right) \\ &\quad = a_1\lambda_1\mathbf{v}_1 + \cdots + a_r\lambda_r\mathbf{v}_r + a_{r+1}\lambda_{r+1}\mathbf{v}_{r+1} = \mathbf{0} \end{aligned} \tag{4.3}$$

On the other hand, multiplying λ_{r+1} to the relation, we get

$$a_1\lambda_{r+1}\mathbf{v}_1 + \cdots + a_r\lambda_r\mathbf{v}_r + a_{r+1}\lambda_{r+1}\mathbf{v}_{r+1} = \mathbf{0} \tag{4.4}$$

Subtracting Equation (4.4) from Equation (4.3), we get

$$a_1(\lambda_1 - \lambda_{r+1})\mathbf{v}_1 + \cdots + a_r(\lambda_r - \lambda_{r+1})\mathbf{v}_r = \mathbf{0}.$$

The $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent by the induction hypothesis, hence

$$a_i(\lambda_i - \lambda_{r+1}) = 0 \quad (1 \leq i \leq r).$$

But the $\lambda_i - \lambda_{r+1} \neq 0$, so $a_i = 0$ for all $1 \leq i \leq r$ and that forces a_{r+1} to be zero as well. □

Corollary 4.2.3. *An endomorphism of V having $\dim V$ distinct eigenvalues is diagonalizable.*

However, the converse of the corollary is not true.

Example 4.2.4. The matrix $M = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ has only two eigenvalues namely 1 and 2. However, the 1-eigenspace has dimension 2 and the 2-eigenspace has dimension 1 and so it is diagonalizable.

We conclude this section with an application of diagonalization. The $(F_n)_n$ sequence of numbers defined recursively by $F_0 = 0, F_1 = 1$,

$$F_{n+1} = F_n + F_{n-1} \quad (n \geq 1)$$

is called the **Fibonacci sequence**. The first few terms of this sequence are

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$$

The recurrence can be expressed by the following matrix equation

$$\begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} F_{n-1} \\ F_n \end{pmatrix} \tag{4.5}$$

So F_n $(n \geq 1)$ is given by

$$\begin{pmatrix} F_{n-1} \\ F_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{n-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{4.6}$$

So it remains to compute the power of $M = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Finding high powers of a matrix can be computationally expensive. However, if the matrix is diagonalizable, like in our case (Example 4.2.3), $P^{-1}MP = D$ is a diagonal matrix for some invertible P then

$$D^2 = (P^{-1}MP)(P^{-1}MP) = P^{-1}MPP^{-1}MP = P^{-1}M^2P$$

and so inductively,

$$D^n = P^{-1}M^nP \quad (n \geq 1).$$

That means $M^n = PD^nP^{-1}$. Since powers of a diagonal matrix is easy to compute, this give us an easy way to compute M^n. In our case,

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \qquad P = \begin{pmatrix} -\lambda_2 & -\lambda_1 \\ 1 & 1 \end{pmatrix}$$

where $\lambda_1 = (1 + \sqrt{5})/2$ and $\lambda_2 = (1 - \sqrt{5})/2$. So

$$\begin{aligned} \begin{pmatrix} F_{n-1} \\ F_n \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^{n-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} -\lambda_2 & -\lambda_1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1^{n-1} & 0 \\ 0 & \lambda_2^{n-1} \end{pmatrix} \begin{pmatrix} 1 & \lambda_1 \\ -1 & -\lambda_2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} * & * \\ * & \lambda_1^n - \lambda_2^n \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned}$$

Therefore

$$F_n = \frac{\lambda_1^n - \lambda_2^n}{\lambda_1 - \lambda_2} = \frac{\lambda_1^n - \lambda_2^n}{\sqrt{5}}. \tag{4.7}$$

This is the **binet formula** for the Fibonacci number. It is quite remarkable since, a piror, there is no reason for the right-hand side to be a natural number.

4.3 Minimal Polynomials

The dimension of $\mathrm{End}(V)$ is the square of the dimension of V. So for any endomorphism τ of an n-dimensional vector space V,

$$1, \tau, \tau^2, \ldots, \tau^{n^2}$$

must be linearly dependent, i.e. there exist $a_0, \ldots, a_{n^2}$ not all zeros such that

$$a_0 + a_1\tau + \cdots + a_{n^2}\tau^{n^2} = 0.$$

In other words, there exists a non-zero polynomial $p(x)$ over K with degree at most n^2 such that $p(\tau) = 0$. The following theorem tells us one such polynomial. See Section 2.6 of [1] for a proof.

Theorem 4.3.1 (Cayley-Hamilton)**.** $\chi_\tau(\tau) = 0$.

Definition 4.3.1. The **minimal polynomial** of τ, denoted by $m_\tau(x)$, is the monic polynomial over $\mathbb{K}$ of smallest degree such that $m_\tau(\tau) = 0$.

It follows from the Cayley-Hamilton Theorem and the division algorithm of polynomials in one variable over a field that m_τ divides χ_τ. Consequently, every root of m_τ is an eigenvalue of τ. The converse is also true:

Proposition 4.3.2. *Every eigenvalue of τ is a root of $m_\tau(x)$.*

Proof. Let λ be an eigenvalue of τ and $\mathbf{v}$ be an eigenvector of λ. Then $p(\tau)(\mathbf{v}) = p(\lambda) \cdot \mathbf{v}$ for any $p \in \mathbb{K}[x]$. In particular,

$$m_\tau(\lambda) \cdot \mathbf{v} = m_\tau(\tau)(\mathbf{v}) = 0(\mathbf{v}) = \mathbf{0}.$$

So we must have $m_\tau(\lambda) = 0$ since $\mathbf{v} \neq \mathbf{0}$. □

Theorem 4.3.3. *An endomorphism $\tau \in \mathrm{End}(V)$ is diagonalizable if and only if $m_\tau(x)$ is the product of distinct linear factors (over $\mathbb{K}$).*

Proof. First suppose τ is diagonalizable. Let $\lambda_1, \ldots, \lambda_r$ be its eigenvalues. Let $\mathcal{B}$ be a basis of V consisting of eigenvectors of τ. Consider the monic polynomial

$$p(x) = (x - \lambda_1) \cdots (x - \lambda_r).$$

By Proposition 4.3.2, every roots of $p(x)$ is a root of $m_\tau(x)$ and since all roots of p is of multiplicity 1 that implies $p(x)$ divides $m_\tau(x)$. On the other hand,

$$p(\tau) = (\tau - \lambda_1 I) \cdots (\tau - \lambda_r I).$$

Here the multiplication is in $\mathrm{End}(V)$ (i.e. composition of functions). Since I, the identity map of V, commutes with every endomorphism, so the factors appeared above commute with each other. Consequently, $p(\tau)(\mathbf{v}) = \mathbf{0}$ for each eigenvector $\mathbf{v}$ of τ. Thus $p(\tau)$ must be the zero map since V has a basis consists of eigenvectors of τ. Thus $m_\tau(x)$ divides $p(x)$ as well and so $p(x) = m_\tau(x)$

because they are both monic. This shows that $m_\tau(x)$ factors into distinct linear factors over $\mathbb{K}$.

Conversely, suppose $m_\tau(x) = \prod_{i=1}^{r}(x - \lambda_i)$ is a product of distinct linear polynomials over $\mathbb{K}$. For each $1 \leq i \leq r$, let $g_i(x) = m_\tau(x)/(x - \lambda_i)$. Then the g.c.d of the g_i's is 1. Consequently, there exist $k_i(x) \in \mathbb{K}[x]$ $(1 \leq i \leq r)$ such that

$$1 = \sum_{i=1}^{r} g_i(x)k_i(x) \tag{4.8}$$

Let $V_i = \mathrm{im}(g_i(\tau))$. We claim that V_i is a subspace of the λ_i-eigenspace of τ. To see this, suppose $\mathbf{u} \in V_i$, say $\mathbf{u} = g_i(\tau)(\mathbf{v})$ for some $\mathbf{v}$. Then

$$(\tau - \lambda_i I)(\mathbf{u}) = (\tau - \lambda_i)g_i(\tau)(\mathbf{v}) = m_\tau(\tau)(\mathbf{v}) = \mathbf{0}.$$

Moreover, it follows from Equation (4.8) that $I = \sum_{i=1}^{r} g_i(\tau)k_i(\tau)$ and so

$$\mathbf{v} = I(\mathbf{v}) = \sum_{i=1}^{r} g_i(\tau)(k_i(\tau)(\mathbf{v})), \quad \text{for any } \mathbf{v} \in V.$$

In particular, the eigenvectors of τ is a spanning set of V and so there is a basis of V consisting of eigenvectors of τ. Therefore, τ is diagonalizable. □

In order for Theorem 4.3.3 to be useful, we need a practical way of finding the minimal polynomial of an endomorphism. And luckily, we have one: go through the products of linear factors of the χ_τ in ascending order of degree, the first one that makes τ vanishes is its minimal polynomial.

Example 4.3.1. The characteristic polynomial of $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is $(x-1)^2$. Since $A - I_2 \neq 0$, its minimial polynomial is not $(x-1)$ and hence must be $(x-1)^2$. In particular, this shows that A is not diagonalizable.

4.4 Jordan Normal Form

Our discussion of endomorphism will not be completed without mentioning *Jordan form*. As we have seen, not every endomorphism can be diagonalized so the best that we can hope for is to find in every similarity class a representative that is of particular simple form.

Definition 4.4.1. A **Jordan block** $J(n, \lambda)$ is a $n \times n$ matrix of the form

$$\begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ & & \ddots & & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}$$

where the diagonal entries are all λ, the entries immediately above the diagonal are all 1 and the remaining entries are all 0.

A matrix is in **Jordan form** if it has Jordan blocks on its diagonal and zero entries elsewhere.

We state without proof of the following theorem

Theorem 4.4.1. *Over the complex field $\mathbb{C}$, any matrix is similar to one in Jordan form. Moreover, the Jordan form of a matrix is unique up to permuting the Jordan blocks on its diagonal.*

Exercises

Exercise 4.1. Find a 2×2 matrix over the complex numbers that diagonalize the matrix in Example 4.2.2, i.e. find $P \in M_2(\mathbb{C})$ such that $P^{-1}R_{\frac{\pi}{2}}P$ is a diagonal matrix.

Exercise 4.2. Let A be the matrix $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$

(a) Find the characteristic polynomial of A.

(b) Find the minimal polynomial of A.

(c) Is A diagonalizable?

(d) Find a basis for each eigenspace of A.

Exercise 4.3. Diagonalize the matrix M in Example 4.2.4 and use that to find M^5 the 5-th power of M.

Chapter 5

Forms

5.1 Linear Forms

A **linear form** is a map from a $\mathbb{K}$-vector space V to $\mathbb{K}$. These maps form a $\mathbb{K}$-vector space $\mathrm{Hom}(V, \mathbb{K})$ under function addition and scalar multiplication. We call it the **dual space** of V and denote it by V^*.

Example 5.1.1. A linear form α on $\mathbb{K}^n$ is simply given by left multiplication of an $1 \times n$ matrix:

$$(a_1, \ldots, a_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = a_1 x_1 + \cdots + a_n x_n.$$

Definition 5.1.1. The **natural pairing** of V and V^* is the map

$$\langle\ ,\ \rangle \colon V \times V^* \to K, \quad \langle \mathbf{v}, \phi \rangle = \phi(\mathbf{v})$$

We also call this the **evaluation map**.

Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V. For each $1 \leq i \leq n$, let ϕ_i to be the linear form defined by

$$\langle \mathbf{v}_j, \phi_i \rangle = \delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

We call δ_{ij} the **Kronecker delta** symbol. Let $\Delta_{\mathcal{B}}$ be the linear map from V to V^* given by $\mathbf{v}_i \mapsto \phi_i$.

Proposition 5.1.1. *$\Delta_{\mathcal{B}}$ is an isomorphism. In particular* $\dim V = \dim V^*$.

Proof. Suppose $\sum a_i \phi_i$ is the zero function. Then for each $1 \leq j \leq n$,

$$0 = \left(\sum_i a_i \phi_i \right) (\mathbf{v_j}) = \sum_i a_i \phi_i(\mathbf{v}_j) = \sum_i a_i \delta_{ij} = a_j.$$

Therefore, the ϕ_i's are linearly independent. Next, we argue that they generate V^*. For any $\phi \in V^*$, it is straight-forward to check that $\sum_i \phi(\mathbf{v}_i)\phi_i$ and ϕ agree on each $\mathbf{v}_i$ in $\mathcal{B}$ and hence must be the same map by linearity. □

We call the ordered basis $\{\phi_1, \ldots, \phi_n\}$ of V^* the **dual basis** of $\mathcal{B}$. It is worth pointing out that Proposition 5.1.1 is not true for infinite dimensional vector spaces (Exercise 5.1). Also, the isomorphism $\Delta_{\mathcal{B}}$, as the notation suggests, depends on the choice of $\mathcal{B}$ (Exercise 5.2). On the other hand, for any $\mathbf{v} \in V$, one checks easily that the map $\mathbf{v}^{**} := \langle \mathbf{v}, \ \rangle$ is a linear form on V^* hence an element of V^{**}.

Theorem 5.1.2. *The map η from V to V^{**} defined by $\mathbf{v} \mapsto \mathbf{v}^{**}$ is injective. In particular, it is an isomorphism when V is finite dimensional.*

Proof. One checks readily that η is linear. We need to show that its kernel is trivial. Suppose $\mathbf{v}^{**}$ is the zero form. We argue that $\mathbf{v}$ must be $\mathbf{0}$. If not, $\{\mathbf{v}\}$ is linearly independent hence there exists $\phi_{\mathbf{v}} \in V^*$ such that $\phi_{\mathbf{v}}(\mathbf{v}) = 1$. But since

$$\mathbf{v}^{**}(\phi_{\mathbf{v}}) = \langle \mathbf{v}, \phi_{\mathbf{v}} \rangle = \phi_{\mathbf{v}}(\mathbf{v}),$$

so $\mathbf{v}^{**}(\phi_{\mathbf{v}}) = 1$, contradicting the assumption that $\mathbf{v}^{**}$ is the zero form. If V is finite dimension, then by Proposition 5.1.1 $\dim(V^{**}) = \dim(V^*) = \dim(V)$. □

The map η in Theorem 5.1.2 is called the **natural isomorphism** from V to its double dual V^{**}. This isomorphism is independent of the choice of bases. The construction of dual is actually "functorial": given a linear map $\tau \colon V \to W$, we define its **adjoint** (or **dual**) to be the linear map $\tau^* \colon W^* \to V^*$ given by

$$\tau^*(\phi) = \phi \circ \tau, \qquad (\phi \in W^*).$$

This can be expressed using the natural pairing by saying for any $\mathbf{v} \in V$,

$$\langle \mathbf{v}, \tau^*(\phi) \rangle = \langle \tau(\mathbf{v}), \phi \rangle. \tag{5.1}$$

Next we relate the matrix representations of τ and its adjoint. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V and $\mathcal{D} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be a basis of W. Let $\mathcal{B}^* = \{\phi_1 \ldots, \phi_m\}$ and $\mathcal{D}^* = \{\gamma_1, \ldots, \gamma_m\}$ be the corresponding dual basis, respectively. Let $A = (a_{ij})$ be the matrix representing τ with respect to $\mathcal{B}$ and $\mathcal{D}$ and $B = (b_{ij})$ be the matrix representing τ^* with respect to $\mathcal{B}^*$ and $\mathcal{D}^*$. Then for each $1 \leq i \leq n$ and $1 \leq j \leq m$,

$$b_{ij} = \langle \mathbf{v}_j, \tau^*(\phi_i) \rangle = \langle \tau(\mathbf{v}_j), \phi_i \rangle = a_{ij}$$

That means $(\tau^*)^{\mathcal{D}^*}_{\mathcal{B}^*} = B = A^T = ((\tau)^{\mathcal{D}}_{\mathcal{B}})^T$.

5.2 Bilinear and Quadratic Forms

Definition 5.2.1. Let $V_1, \dots, V_n$ and W be $\mathbb{K}$-vector spaces. A map $\Phi\colon V_1 \times \cdots \times V_n \to W$ is **multilinear** if Φ is linear on each of the argument, i.e. for each $1 \le i \le n$, the map

$$\Phi(\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \_\_, \mathbf{v}_{i+1}, \dots, \mathbf{v}_n)$$

is a linear map from V_i to W for any choice of $\mathbf{v}_j \in V_j$ $(j \neq i)$. When $n = 2$, say that Φ is **bilinear** instead.

Example 5.2.1. Determinant is a multilinear form on $\mathbb{K}^{n\times n}$ $(n \ge 1)$.

Example 5.2.2. The natural pairing $\langle\ ,\ \rangle$ is a bilinear form on $V \times V^*$.

Example 5.2.3. The **dot product**) on $\mathbb{R}^n \times \mathbb{R}^n$ defined by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_i v_i w_i$$

is bilinear.

Example 5.2.4. More generally, given a square matrix $B \in M_n(\mathbb{K})$. The map on $\mathbb{K}^n \times \mathbb{K}^n$ defined by (think of elements of $\mathbb{K}^n$ as column vectors)

$$(\mathbf{v}, \mathbf{w}) \longmapsto \mathbf{v}^T B \mathbf{w}$$

is bilinear. We get the standard inner product on $\mathbb{R}^n$ when $\mathbb{K} = \mathbb{R}$ and $B = I_n$.

Definition 5.2.2. Let β be a bilinear form on V and $\mathcal{B}$ be a basis of V. We say that a matrix B **represents** β **with respect a basis** $\mathcal{B}$ if for all $\mathbf{v}, \mathbf{v}' \in V$,

$$\beta(\mathbf{v}, \mathbf{v}') = (\mathbf{v})_{\mathcal{B}}^T B (\mathbf{v}')_{\mathcal{B}}.$$

If $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, then

$$b_{ij} = \mathbf{e}_i^T B \mathbf{e}_j = \beta(\mathbf{v}_i, \mathbf{v}_j). \tag{5.2}$$

Definition 5.2.3. A bilinear form $\beta\colon V \times V \to \mathbb{K}$ is **symmetric** if

$$\beta(\mathbf{v}, \mathbf{v}') = \beta(\mathbf{v}', \mathbf{v})$$

for all $\mathbf{v}, \mathbf{v}' \in V$. It follows from (5.2) that any matrix representing a symmetric bilinear form must be a symmetric matrix.

Next we study the effect of change of basis on the representation of bilinear form. Suppose B and B' are matrices representing a bilinear form β on V with respect to bases $\mathcal{B}$ and $\mathcal{B}'$, respectively, then for any $\mathbf{v}, \mathbf{v}' \in V$,

$$(\mathbf{v})_{\mathcal{B}}^T B (\mathbf{v}')_{\mathcal{B}} = \beta(\mathbf{v}, \mathbf{v}') = (\mathbf{v})_{\mathcal{B}'}^T B' (\mathbf{v}')_{\mathcal{B}'}. \tag{5.3}$$

For $\mathbf{w} \in V$, since $(\mathbf{w})_{\mathcal{B}} = (\mathrm{id})_{\mathcal{B}'}^{\mathcal{B}} (\mathbf{w})_{\mathcal{B}'}$, it follows from (5.3) and (5.2) that

$$B' = P^T B P, \qquad \text{where } P = (\mathrm{id})_{\mathcal{B}'}^{\mathcal{B}}.$$

This leads us naturally to the following definition.

Definition 5.2.4. Two matrices $B, B' \in M_n(\mathbb{K})$ are **congruent** if there exist $P \in \mathrm{GL}_n(\mathbb{K})$ such that $B' = P^T B P$.

As shown in the above discussion, *two square matrices are congruent if and only if they represent the same bilinear form.*

Given a symmetric bilinear form β, we associate to it a form

$$q\colon V \to K, \quad q(\mathbf{v}) = \beta(\mathbf{v}, \mathbf{v}),$$

and call it the **quadratic form** associated to β. The rationale behind the terminology is that

$$q(c\mathbf{v}) = c^2 q(\mathbf{v}) \qquad \forall c \in \mathbb{K} \text{ and } \mathbf{v} \in V.$$

If the characteristic of $\mathbb{K}$ is not 2, i.e. 2 is invertible in $\mathbb{K}$, then we can recover β from q, by the so call **polarizing formula**

$$\beta(\mathbf{v}, \mathbf{w}) = \frac{1}{2}(q(\mathbf{v} + \mathbf{w}) - q(\mathbf{v}) - q(\mathbf{w})). \tag{5.4}$$

This shows how quadratic forms arise from symmetric bilinear forms. On the other hand, we can start with quadratic forms:

Definition 5.2.5. A form $q\colon V \to \mathbb{K}$ is a **quadratic form** if

- $q(c\mathbf{v}) = c^2 q(\mathbf{v})$ for all $c \in \mathbb{K}$ and $\mathbf{v} \in V$ and;
- the function β defined by the polarizing formula is a bilinear form. (Note that it is necessarily symmetric).

Roughly speaking, symmetric bilinear forms and quadratic forms are the same kind of objects and if a basis is chosen then we can think of them as symmetric matrices.

Example 5.2.5. Consider the symmetric matrix $B = \begin{pmatrix} 1 & \frac{1}{2} & -1 \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ -1 & -\frac{1}{2} & -1 \end{pmatrix}$.
The quadratic form q associated to B is given by

$$\begin{aligned} q(x, y, z) &= \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} & -1 \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ -1 & -\frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= x^2 + xy + y^2 - 2\,xz - yz - z^2 \end{aligned}$$

Let β be the bilinear form given by B. We can verify the formula (5.2), for example,

$$b_{23} = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} & -1 \\ \frac{1}{2} & 1 & -\frac{1}{2} \\ -1 & -\frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -\frac{1}{2}.$$

Again it is natural and useful to ask for any given bilinear form (or quadratic form), does it exist a special basis so that the matrix representation of β with respect to that basis has a particular simple form. The answer is "yes" and we phrase our answer in terms of matrices as the following theorem.

Theorem 5.2.1. *Any symmetric matrix $B \in M_n(\mathbb{K})$ over a field of characteristics $\neq 2$ is congruent to a diagonal matrix.*

A proof of this theorem can be found in [1] p.60. Basically, it is a fancy way of completing squares in several variables. Here we will simply illustrate the theorem by two examples.

Example 5.2.6. Consider the matrix B in Example 5.2.5. The corresponding quadratic form is $q(x, y, z) = x^2 + xy + y^2 - 2\,xz - yz - z^2$. To complete the squares, we deal any variable that has a square term, say x, first. Let

$$u = x + \frac{1}{2}\,y - z$$

Note that the coefficients of x, y, z are just the entries of the first row of B.

$$u^2 = x^2 + xy + \frac{1}{4}\,y^2 - 2\,xz - yz + z^2$$

and so

$$u^2 + \frac{3}{4}y^2 - 2z^2 = q(x, y, z).$$

In terms of matrices, we have

$$\begin{pmatrix} u \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

and so

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ y \\ z \end{pmatrix}$$

One checks that

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

The next example shows how to deal with the case when none of the variable's square appear.

Example 5.2.7. Consider the symmetric matrix $B = \begin{pmatrix} 0 & \frac{1}{2} & -1 \\ \frac{1}{2} & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ and the corresponding quadratic form $q(x, y, z) = xy - 2\,xz + 2\,yz$. This time none

of the variables have their squares appear. So we cannot apply the method in the pervious example directly. Instead, we make square terms appear by using *difference of two squares*: for example, take the term $2yz$, by letting $u = y + z$ and $v = y - z$, we have

$$yz = \frac{1}{4}((y+z)^2 - (y-z)^2) = \frac{1}{4}(u^2 - v^2)$$

and $y = (u+v)/2, z = (u-v)/2$. Thus

$$\begin{aligned} q &= x\left(\frac{u+v}{2}\right) - 2x\left(\frac{u-v}{2}\right) + 2\frac{1}{4}(u^2 - v^2) \\ &= \frac{1}{2}\,u^2 - \frac{1}{2}\,v^2 - \frac{1}{2}\,ux + \frac{3}{2}\,vx. \end{aligned}$$

Now square terms appear. So the method in the previous example applies. In matrix term, the transformation $x = x$, $u = y + z$ and $v = y - z$ corresponds to

$$\begin{pmatrix} x \\ u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

and so

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} x \\ u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ u \\ v \end{pmatrix}.$$

Indeed,

$$\begin{aligned} &\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & -1 \\ \frac{1}{2} & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ =&\begin{pmatrix} x & u & v \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & -1 \\ \frac{1}{2} & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ u \\ v \end{pmatrix} \\ =&\begin{pmatrix} x & u & v \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{4} & \frac{3}{4} \\ -\frac{1}{4} & \frac{1}{2} & 0 \\ \frac{3}{4} & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ u \\ v \end{pmatrix} \end{aligned}$$

This completes the first step. Now since square terms appear, the method in the pervious example applies. We leave the rest as an exercise.

If the field of scalars is $\mathbb{R}$ or $\mathbb{C}$, further reductions apply.

Example 5.2.8. We have seen in Example 5.2.6 the quadratic form $q(x, y, z) = x^2 + xy - 2xz + y^2 - yz - z^2$ is congruent to the diagonal form $x^2 + \frac{3}{4}y^2 - 2z^2$.

If we take the field to be $\mathbb{R}$, then since every positive number is a square, so we can make the following change of coordinates

$$u = x, v = \frac{\sqrt{3}}{2}y, w = \sqrt{2}z.$$

Then the $x^2 + \frac{3}{4}y^2 - 2z^2 = u^2 + v^2 - w^2$. In matrix terms,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{4} & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

Certainly, if the field in discussion is algebraically closed, i.e. $\mathbb{C}$, then every number is a square, including -1. Hence, the quadratic form in our example is congruent to the form $x^2 + y^2 + z^2$.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{i}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{4} & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{i}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

We conclude this section by the following two results. They tell us that what we have seen in the examples are in fact general phenomena. Their proofs can be found in [1] P.64–65.

Theorem 5.2.2. *Every symmetric complex matrix A is congruent to a matrix of the form $\begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ where $r = \operatorname{rank} A$.*

Theorem 5.2.3 (Sylvester Law of Inertia)**.** *Every symmetric real matrix A is congruent to a matrix of the form $\begin{pmatrix} I_s & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -I_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$ where $s + t = \operatorname{rank} A$. Moreover, s and t are independent of the reduction. The number $s - t$ is called the* **inertia** *of A.*

Exercises

Exercise 5.1. Let V be an infinite dimensional vector space.

(a) Let $\mathcal{B}$ be a basis of V. For each $\mathbf{v} \in \mathcal{B}$, let $\phi_{\mathbf{v}}$ be the linear form defined by $\phi_{\mathbf{v}}(\mathbf{v}) = 1$ and $\phi_{\mathbf{v}}(\mathbf{w}) = \mathbf{0}$ for all $\mathbf{w} \in \mathcal{B}, \mathbf{w} \neq \mathbf{v}$. Show that

$$\mathcal{B}^* := \{\phi_{\mathbf{v}} \colon \mathbf{v} \in \mathcal{B}\}$$

is linearly independent.

(b) Let $\phi \in V^*$ be the linear form defined by $\phi(\mathbf{v}) = 1$ for all $\mathbf{v} \in \mathcal{B}$. Show that ϕ is not in the span of $\mathcal{B}^*$.

(c) Deduced that $\dim V^* > \dim V$.

Exercise 5.2. Show that if $\mathcal{B}_1, \mathcal{B}_2$ are two distinct basis of V. Show that $\Delta_{\mathcal{B}_1}$ and $\Delta_{\mathcal{B}_2}$ are two distinct isomorphisms from V to V^*.

Exercise 5.3. Redo Example 5.2.6, this time start by completing the square with the variable y instead of x.

Exercise 5.4. Finish the computation in Example 5.2.7 by finding the P such that $P^T BP$ is a diagonal matrix.

Chapter 6

Inner Products

6.1 Inner products

In this chapter, we consider vector spaces over $\mathbb{R}$ with an extra structure, given by a special kind of bilinear form, called an inner product. A bilinear form $\langle\ ,\ \rangle\colon V \times V \to \mathbb{R}$ is **semi-positive definite** if $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$ for all $\mathbf{v} \in V$ and is **positive-definite** if, in addition, $\langle \mathbf{v}, \mathbf{v}\rangle = 0$ if and only if $\mathbf{v} = 0$.

Definition 6.1.1. An **inner product** on a vector space V is a positive definite symmetric bilinear form. An **inner product space** is a real vector space equipped with an inner product.

It is immediate that the standard dot product on $\mathbb{R}^n$ (Example 5.2.3) is an inner product. Note that $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$ where $\|\mathbf{v}\|$ is the Euclidean distance from the origin of $\mathbb{R}^n$ to the point $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$. More generally, the **norm** (or **length**) of a vector $\mathbf{v}$ in an inner product space, denoted by $\|\mathbf{v}\|$, is defined to be the square root of $\langle \mathbf{v}, \mathbf{v}\rangle$. In fact, we often write $\langle \mathbf{v}, \mathbf{w}\rangle$ as $\mathbf{v} \cdot \mathbf{w}$ and use the words "dot product" and "inner product" inter-changeably.

Example 6.1.1. The bilinear form β on $C[0, 1]$ given by

$$\beta(f, g) = \int_0^1 f(x)g(x)\ dx$$

is an inner product.

The fundamental result governing inner product is the following inequality.

Theorem 6.1.1 (Cauchy-Schwarz Inequality)**.** *For any $\mathbf{v}, \mathbf{w}$ in an inner product space,*

$$(\mathbf{v} \cdot \mathbf{w})^2 \leq \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 .$$

Moreover, the equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ are linearly dependent.

Proof. If $\mathbf{v}$ and $\mathbf{w}$ are linearly dependent, then one verifies readily that the equality holds. It remains to show the strict inequality holds when $\mathbf{v}$ and $\mathbf{w}$ are linearly independent. In that case, neither of them can be $\mathbf{0}$ and for any $t \in \mathbb{R}$, $\mathbf{v} - t\mathbf{w} \neq \mathbf{0}$. Thus,

$$0 < (\mathbf{v} - t\mathbf{w})(\mathbf{v} - t\mathbf{w}) = t^2 (\mathbf{w} \cdot \mathbf{w}) - 2t (\mathbf{v} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{v}) . \tag{6.1}$$

Therefore, the discriminant of the above quadratic polynomial (since $\mathbf{w} \cdot \mathbf{w} > 0$) must be negative and so $(\mathbf{v} \cdot \mathbf{w})^2 < (\mathbf{v} \cdot \mathbf{v}) (\mathbf{w} \cdot \mathbf{w}) = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$. □

An immediate consequence of the Cauchy-Schwarz inequality is that for any nonzero $\mathbf{v}$ and $\mathbf{w}$, $-1 \leq \dfrac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} \leq 1$ and so $\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta$ for some unique $\theta \in [0, \pi]$. We call this unique θ the **angle between** $\mathbf{v}$ and $\mathbf{w}$. This choice is natural in the sense that it allows a proof of the *Law of Cosine* via the use of inner product (see exercise). The angle between a vector and the zero vector is undefined. We say that two vectors (zero or not) **orthogonal** if there inner product is zero. So if neither of them is zero then that means the angle between them is $\pi/2$.

Example 6.1.2. The function $\sin(2\pi x)$ and $\cos(2\pi x)$ are orthogonal with respect to the inner product given in Example 6.1.1.

6.2 Orthonormal bases

A set of vectors X in an inner product space is **orthonormal** if 1) each of its members are all unit vectors (i.e. vectors of length 1) and 2) distinct members of X are orthogonal. An **orthonormal basis** is also called a **frame**.

Proposition 6.2.1. *A set of nonzero vectors whose members are orthogonal to each other is linearly independent. In particular, any orthonormal set of vectors must be linearly independent.*

Proof. Suppose X is a set of non-zero vectors whose members are mutually orthogonal. Let $\mathbf{0}$ be some linear combination of members of X, say $\sum_{i=1}^{n} c_i \mathbf{x_i}$. Then for $1 \leq j \leq n$,

$$0 = \mathbf{0} \cdot \mathbf{x_j} = \left(\sum_{i=1}^{n} c_i \mathbf{x}_i \right) \cdot \mathbf{x}_j = \sum_{i=1}^{n} c_i \delta_{ij} \mathbf{x}_i \cdot \mathbf{x}_j = c_j \|\mathbf{x}_j\|^2 .$$

Since $\mathbf{x}_j \neq 0$ by assumption, therefore $\|\mathbf{x}_j\| \neq 0$ and so the equation above implies that $c_j = 0$. □

By definition, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij}$ whenever $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a frame and so the matrix representation of the inner product with respect to a frame is simply the identity matrix with rank the same as the dimension. In other words, with respect to a suitably choose basis (a frame, that is) any inner product of an n-dimensional space is computed as the dot product of $\mathbb{R}^n$. Now we establish the existence of frames.

Theorem 6.2.2. *Every (non-trivial, finite dimensional) inner product space* $(V, \langle\ ,\ \rangle)$ *has a frame.*

Proof. We prove this by induction on the dimension of V. The result is clear if V is the span of a non-zero vector $\mathbf{v}_1$. In that case, then $\mathbf{u}_1 := \mathbf{v}_1/\|\mathbf{v}_1\|$ forms a frame of V. Now suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V. By the induction hypothesis, V_{n-1}, the span of $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ together with the restriction of the inner product onto it, has a frame, say $\{\mathbf{u}_1, \ldots \mathbf{u}_{n-1}\}$. Let

$$\mathbf{w}_n = \mathbf{v}_n - \sum_{i=1}^{n-1} \langle \mathbf{u}_i, \mathbf{v}_n \rangle \mathbf{u}_i \tag{6.2}$$

Then $\mathbf{w}_n \neq \mathbf{0}$ otherwise $\mathbf{v}_n \in V_{n-1}$ contradicting the $\mathbf{v}_i$'s are linearly independent. Moreover, $\mathbf{w}_n$ is orthogonal to each the $\mathbf{u}_i$'s. This is because for each $1 \leq j \leq n-1$,

$$\begin{aligned}
\langle \mathbf{w}_n, \mathbf{u}_j \rangle &= \langle \mathbf{v}_n, \mathbf{u}_j \rangle - \sum_{i=1}^{n-1} \langle \mathbf{u}_i, \mathbf{v}_n \rangle \langle \mathbf{u}_i, \mathbf{u}_j \rangle \\
&= \langle \mathbf{v}_n, \mathbf{u}_j \rangle - \sum_{i}^{n-1} \langle \mathbf{u}_i, \mathbf{v}_n \rangle \delta_{ij} \\
&= \langle \mathbf{v}_n, \mathbf{u}_j \rangle - \langle \mathbf{v}_n, \mathbf{u}_j \rangle = 0.
\end{aligned}$$

So $\mathbf{u}_n := \mathbf{w}_n/\|\mathbf{w}_n\|$ together with $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}, \mathbf{u}_n$ still form an orthonormal set. Its span contains V_{n-1} and by Equation 6.2 contains $\mathbf{v}_n$ as well. Hence it is equal to the whole V. Finally, by Proposition 6.2.1 $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is also linearly independent and so is a frame of V. □

The process of obtaining an orthonormal basis for an arbitrary one as given in the proof above is known as the **Gram-Schmidt process**.

6.3 Projections

Given a set of vectors X of an inner product space V, the **orthogonal complement** of X, denoted by $X^\perp$ is the set vectors that are orthogonal to every vector in X, i.e.

$$X^\perp = \{\mathbf{v} \in V : \mathbf{v} \cdot \mathbf{x} = 0 \text{ for all } \mathbf{x} \in X\}.$$

One checks readily that $X^\perp$ is always a subspace of V and that $X^{\perp\perp} \supseteq^X$.

Proposition 6.3.1. $V = W \oplus W^\perp$ *for any subspace of* V

Proof. Extend $\{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$ a base of W to a base of V. The Gram-Schmidt process applied to that base produce an orthonormal base $\mathcal{B} = \{\mathbf{u}_1 \ldots, \mathbf{u}_n\}$ of V. Since $\mathbf{u}_1, \ldots, \mathbf{u}_r$ are linear combinations of $\mathbf{w}_1, \ldots, \mathbf{w}_r$ so they still form a base of W. The remaining vectors $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n$ are in $W^\perp$ and are linearly

independent (since $\mathcal{B}$ is an orthonormal basis). They also span $W^\perp$ as well, take any $\mathbf{w} = \sum_{i=1}^{n} c_i \mathbf{u}_i$ in $W^\perp$. Then by orthogonality of the $\mathbf{u}_i$'s, $c_j = \langle \mathbf{w}, \mathbf{u}_j \rangle$ which is 0 if $1 \leq j \leq r$ since for those j's $\mathbf{u}_j \in W$ and is orthogonal to $\mathbf{w}$. Clearly, V is a sum of W and $W^\perp$ and it is a direct sum since if $\mathbf{v} \in W \cap W^\perp$, then $\mathbf{v}$ is orthogonal to itself hence $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = 0$ and so $\mathbf{v} = \mathbf{0}$. □

Exercise

Exercise 6.1. Give a proof of the Law of Cosine in $\mathbb{R}^n$ using inner product, i.e.

$$\|\mathbf{c}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\|\mathbf{a}\|\|\mathbf{b}\|\cos\theta$$

where $\mathbf{a}, \mathbf{b}$ are nonzero vectors in $\mathbb{R}^n$, $\mathbf{c} = \mathbf{a} - \mathbf{b}$ and θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. Note that this generalizes the Law of Cosine in $\mathbb{R}^2$ since if $\mathbf{a}$ and $\mathbf{b}$ represent two sides of a triangle then $\mathbf{c} := \mathbf{a} - \mathbf{b}$ is the third side opposite to the angle between $\mathbf{a}$ and $\mathbf{b}$.

Exercise 6.2. Show that two non-zero vectors in an inner product space are linearly dependent if and only if the angle between them is either 0 or π. In the first case, we say that the two vectors are in the *same direction* and in the latter case, we say that they are in *opposite direction*.

Chapter 7

The Spectral Theorem

7.1 The Spectral Theorem

Theorem 7.1.1. *The eigenspaces of a self-adjoint operator α on an inner product space V form an orthogonal decomposition of V.*

Proof. We prove this by induction on n the dimension of V. If V is 1-dimensional, then V itself is an eigenspace of α. So we assume the theorem holds for all $(n-1)$-dimensional spaces for some $n \geq 2$. Since $\mathbb{C}$ is algebraically closed, α has a potentially complex eigenvalue λ. Let $\mathbf{v}$ (again potentially complex) be an λ-eigenvector. Since α is self-adjoint, the matrix A representing α with respect to an orthonormal basis of V is a real symmetric. Therefore,

$$\begin{aligned}\overline{\lambda}\left\|\mathbf{v}\right\|^2 &= \overline{\lambda\mathbf{v}}^T\mathbf{v} \\ &= \overline{A\mathbf{v}}^T\mathbf{v} = A\overline{\mathbf{v}}^T\mathbf{v} = \overline{\mathbf{v}}^T A\mathbf{v} \\ &= \overline{\mathbf{v}}^T\lambda\mathbf{v} = \lambda\left\|\mathbf{v}\right\|^2.\end{aligned}$$

Since $\|\mathbf{v}\| \neq 0$, $\lambda = \overline{\lambda}$. This shows that λ is real. Consequently, we can find a real λ-eigenvector, $\mathbf{u}$ and by dividing by the $\|\mathbf{u}\|$, we can further assume $\mathbf{u}$ is a unit vector. Let $W = \langle\mathbf{u}\rangle^\perp$. Then W is a subspace of V of dimensional $n-1$. We claim that α maps W into W as well. Take $\mathbf{w} \in W$, then

$$\langle\mathbf{u}, \alpha(\mathbf{w})\rangle = \langle\alpha(\mathbf{u}), \mathbf{w}\rangle = \lambda\langle\mathbf{u}, \mathbf{w}\rangle = 0$$

Therefore, $\alpha(\mathbf{w})$ is still orthogonal to $\mathbf{u}$ and hence in W. Now by the induction hypothesis W has an orthonormal basis consisting of eigenvectors of α. All of them are orthogonal to $\mathbf{u}$ and so $\mathbf{u}$ together with them form an orthonormal basis of V. □

Chapter 8

Singular Value Decomposition

We come full circle, in this last chapter, discussing how to solve a system of linear equations $A\mathbf{x} = \mathbf{b}$ but this time we will handle the case when it is *not solvable*, that is when $\mathbf{b}$ is not in the column space of A. First, we introduce a decomposition of matrices that has much more versatile usage then when we just solving this problem.

8.1 Singular value decomposition

Theorem 8.1.1. *An $m \times n$ matrix A can be factored as $A = Q_1 \Sigma Q_2^*$ where Q_1, Q_2 are orthogonal matrices and Σ is a $m \times n$ diagonal matrix.*

The non-zero diagonal entries of Σ are called the **singular values** of A. They depend only on A but not by the choice of Q_1 and Q_2.

Proof. The symmetric matrix A^*A, according to the spectral theorem, has a full set of orthonormal eigen-vectors, say $\mathbf{v}_1, \ldots, \mathbf{v}_n$. They form the columns of Q_2. Let λ_j $(1 \leq j \leq n)$ be the eigenvalue of $\mathbf{v}_j$, then

$$\lambda_j = \lambda_j \mathbf{v}_j^* \mathbf{v}_j = \mathbf{v}_j^* (A^* A \mathbf{v}_j) = (A\mathbf{v}_j)^* (A\mathbf{v}_j) = \|A\mathbf{v}_j\|^2 \geq 0.$$

By reindexing, if necessary, $\lambda_1, \ldots, \lambda_r > 0$ and the rest are 0. Let $\mathbf{u}_j = A\mathbf{v}_j/\sigma_j$ where $\sigma_j = \sqrt{\lambda_j}$ $(1 \leq j \leq r)$. Then

$$\mathbf{u}_i^* \mathbf{u}_j = \frac{\mathbf{v}_i^* A^* A \mathbf{v}_j}{\sigma_i \sigma_j} = \frac{\lambda_j \mathbf{v}_i^* \mathbf{v}_j}{\sigma_i \sigma_j} = \delta_{ij}.$$

Thus the $\mathbf{u}_i$ $(1 \leq i \leq r)$ are orthonormal as well. Extend them (by the Gram-Schmidt process) to a complete orthonormal bases $\mathbf{u}_i$ $(1 \leq i \leq m)$. Let Q_1 be the $m \times m$ matrix with $\mathbf{u}_i$ as columns. Clearly both Q_1 and Q_2 are orthogonal

matrices. It remains to show that $\Sigma := Q_1^* A Q_2$ is diagonal and it is verified as follows:

$$(\Sigma)_{ij} = \mathbf{u}_i^* A \mathbf{v}_j = \begin{cases} \mathbf{u}_i^* \mathbf{0} = 0 & j > r \\ \mathbf{u}_i^* \sigma_j \mathbf{u}_j = \sigma_j \delta_{ij} & 1 \leq j \leq r. \end{cases}$$

□

The svd decomposition of A can be viewed as choosing suitable bases for the four fundamental vector spaces associate to A. It is clear that the rank of A is r. And since $A^*(A\mathbf{v}_j) = \lambda_j \mathbf{v}_j$ so the first r columns of Q_2 forms an orthonormal basis of the column space of A^*, i.e. the row space of A and its last $n-r$ columns forms an orthonormal basis of the nullspace of A. Similary, since $\mathbf{u}_j = A\mathbf{v}_j/\sigma_j$ $(1 \leq j \leq r)$, the first r columns of Q_1 forms an orthonormal basis of the column space of A and since the $\mathbf{u}_i$ are orthonormal, the last $m - r$ columns of Q_1 are orthogonal to the column space of A and hence form an orthonormal basis of the left nullspace of A.

8.2 Pseudoinverse of a matrix

Theorem 8.2.1. *For any $m \times n$ matrix A, there is a unique $n \times m$ matrix A^+ which satisfies:*

1. $AA^+A = A$
2. $A^+AA^+ = A^+$
3. AA^+ *and* A^+A *are both orthogonal (Hermitian if it's over* $\mathbb{C}$*)*

The matrix A^+ is called the **pseudoinverse** of A.

Uniqueness of Pseudoinverse. Suppose B and C are two matrices that satisfy the three conditions in Theorem 8.2.1. First, we have

$$\begin{aligned} AB &\overset{(3)}{=} (AB)^* = B^*A^* \overset{(1)}{=} B^*(ACA)^* \\ &= B^*A^*C^*A^* = (AB)^*(AC)^* \overset{(3)}{=} ABAC = AC. \end{aligned}$$

Likewise, we conclude that $BA = CA$ and so

$$B \overset{(2)}{=} BAB = BAC = CAC = C.$$

□

Next we establish the existence of pseudoinverse as an application of svd. First it is clear that for a diagonal matrix Σ, Σ^+ is easy to verify that the matrix obtained by replacing the non-zero entries of Σ^* by their reciprocals is a matrix that satisfies the properties in the theorem above and hence it is Σ^+. For a

general matrix A, let $A = Q_1 \Sigma Q_2^*$ be a svd of A, then it is straight-forward to check that $Q_2 \Sigma^+ Q_1^*$ satisfies the properties in Theorem 8.2.1 hence is A^+ the pseudoinverse of A.

We now argue that the "best solution" to a general system $A\mathbf{x} = \mathbf{b}$ is $A^+ b$. It is the best solution the following sense: since $\mathbf{b}$ need not be in the column space of A, so all we can hope for is to find $\mathbf{x}$ that minimizes $\|A\mathbf{x} - b\|$. Thus we are looking for solution of $A\mathbf{x} = \mathbf{p}$ where $\mathbf{p}$ is the projection of $\mathbf{b}$ to the column space of A. Among the solutions to this later system, the declare the shortest one, call it $\mathbf{x}^+$, to be the best. It remains to show that $\mathbf{x}^+ = A^+\mathbf{b}$.

The result is easy when $A = \Sigma$ is a diagonal matrix. Let r be the rank (so the number of non-zero rows) of Σ. Then $\mathbf{p} := (b_1, \ldots, b_r, 0, \ldots, 0)^*$ is the projection of $\mathbf{b}$ to the column space of Σ. It is then clear that

$$\Sigma^+ \mathbf{p} = \Sigma^+ \mathbf{b} = \begin{pmatrix} b_1/\sigma_1 \\ \vdots \\ b_r/\sigma_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is the solution with the shortest length.

As for the general case, let $Q_1 \Sigma Q_2^*$ be a svd of A. Since Q_1 is orthogonal (pre-serving dot-product),

$$\|A\mathbf{x} - \mathbf{b}\| = \|\Sigma Q_2^* \mathbf{x} - Q_1^* \mathbf{b}\|$$

for any $\mathbf{x}$. So the problem becomes minimizing $\|\Sigma \mathbf{y} - Q_1^* \mathbf{b}\|$ after the (invertible) substitution $\mathbf{y} = Q_2^* \mathbf{x}$. Since the system now is diagonal, we know that the solution with the least length is $\mathbf{y}^+ = \Sigma^+ Q_1^* \mathbf{b}$. Finally since Q_2 is orthogonal, so the length of $\mathbf{x}^+ := Q_2 \mathbf{y}^+$ and $\mathbf{y}^+$ are the same. Therefore $x^+ = Q_2 \Sigma^+ Q_1^* \mathbf{b} = A^+ \mathbf{b}$ is the shortest length vector such that $\|A\mathbf{x}^+ - \mathbf{b}\|$ is minimal.

Appendix A

Groups, Rings and Fields

Groups, rings and fields are three fundamental algebra structures.

A.1 Groups

Definition A.1.1. A **group** is a set G together with an operation $\circ\colon G\times G\to G$ satisfying:

1. $(g\circ h)\circ k = g\circ(h\circ k)$ for any $g,h,k\in G$. (associative)
2. there exists $e\in G$ such that $e\circ g = g = g\circ e$ for all $g\in G$. (existence of identity)
3. for all $g\in G$, there exists $h\in G$ such that $g\circ h = e = h\circ g$. (existence of inverse)

One can show that there is a unique element $e\in G$ satisfying condition (2), it is called the **identity** of G. Also, give any $g\in G$, the h appears in Condition (3) is uniquely determined by g. It is called the **inverse** of g. If $(G,\circ)$ satisfies only Condition (1), we call it a **semi-group**. If $(G,\circ)$ satisfies both Condition (1) and (2), we call it a **monoid**.

An important *"dividing line"* for groups is whether its elements commute with each other:

Definition A.1.2. A group $(G,\circ)$ is **abelian** if for any $g,h\in G$, $g\circ h = h\circ g$.

We often write an abelian group *additively* (i.e. thinking of the group operation as some sort of "addition"). In that case, we use 0 to denote the identity element and write the inverse of $g\in G$ as $-g$. For groups that are not necessary abelian, we often write it *multiplicatively*. In that case, the identity is written as 1 and the inverse of g is written as g^{-1}.

We often think of a group as a collection of *symmetries*, roughly speaking symmetries are maps that preserve the *"structures"* of an object. Here are some examples.

Example A.1.1. For any $n \geq 1$, consider a set X of size n (to be concrete, for instance, one can take $X = \{1, 2, \ldots, n\}$) with no structures. A symmetry of X is simply a **permutation of** X, i.e. a bijection of X to itself, since there are no structures to preserve. It is straight-forward to check that the bijections of X form a group under composition of functions. It is called the **symmetric group on** n **letters** and is denoted by S_n. Clearly, the identity of S_n is the identity map of X.

Example A.1.2. The set of integers $(\mathbb{Z}, +)$ together with the usual addition is a group with 0 as the identity. One can view it as the group of symmetries of the ordered set $(\mathbb{Z}, \leq)$: view $n \in \mathbb{Z}$ as the map $m \mapsto n + m$. So it is a "right-shift" if n is positive and a "left-shift" is n is negative.

Example A.1.3. As we have seen, $\mathrm{GL}(V)$ the set of invertible linear map from a vector space V to itself is a group under composition of functions. It is the group of symmetries of the bases of V. If V is finite dimensional over $\mathbb{K}$, then a choice of basis of V identifies $\mathrm{GL}(V)$ with $\mathrm{GL}(\mathbb{K}^n)$, the group of $n \times n$ matrices over $\mathbb{K}$ with matrix multiplication.

Exercises

Exercise A.1. Let $(G, \circ)$ be a group.

1. Show that the identity of G is unique. In other words, show that if e, e' are elements of G such that $e \circ g = g = g \circ e$ and $e' \circ g = g = g \circ e'$ for all $g \in G$, then $e = e'$.
2. Show the uniqueness of inverse, i.e. show that for any $g \in G$ if h, h' are elements of G such that $g \circ h = e = h \circ g$ and $g \circ h' = e = h' \circ g$, then $h = h'$.

Exercise A.2. Let $(G, \circ)$ be a semigroup. Suppose e is an element of G such that $g \circ e = g$ for all $g \in G$ and for all $g \in G$, there exists $h \in G$ such that $g \circ h = e$. Show that

1. $h \circ g = e$ where h is an element of G such that $g \circ h = e$.
2. $e \circ g = g$ for all $g \in G$.
3. Deduce that $(G, \circ)$ is a group.

Exercise A.3. Give an example of a semigroup $(G, \circ)$ satisfying

1. there exists $e \in G$ such that $e \circ g = g$ for all $g \in G$; and
2. for all $g \in G$, there exists $h \in G$ such that $g \circ h = e$

and yet $(G, \circ)$ is not a group. (Hint: consider the set of real-matrices $\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$ $(a \neq 0)$ under matrix multiplication.)

Appendix B

Axioms of Vector Space

Let $\mathbb{K}$ be a field. A *vector space over* $\mathbb{K}$ is a structure $(V, +, \cdot, \mathbf{0})$ where V is a set and the functions $+\colon V \times V \to V$ (addition) and $\cdot\colon \mathbb{K} \times V \to V$ (scalar multiplication) satisfying the following conditions:

1. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$.
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for any $\mathbf{u}, \mathbf{v} \in V$.
3. There exists an element of V, usually denoted by $\mathbf{0}$, such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for any $\mathbf{v} \in V$.
4. For every $\mathbf{v} \in V$, there exists an element of V, denoted by $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
5. $1 \cdot \mathbf{v} = \mathbf{v}$ for any $\mathbf{v} \in V$ where 1 is the multiplicative identity of $\mathbb{K}$.
6. $(\lambda\mu) \cdot \mathbf{v} = \lambda(\mu \cdot \mathbf{v})$ for any $\lambda, \mu \in \mathbb{K}$ and $\mathbf{v} \in V$.
7. $(\lambda + \mu) \cdot \mathbf{v} = \lambda \cdot \mathbf{v} + \mu \cdot \mathbf{v}$ for any $\lambda, \mu \in \mathbb{K}$ and $\mathbf{v} \in V$.
8. $\lambda \cdot (\mathbf{u} + \mathbf{v}) = \lambda \cdot \mathbf{u} + \lambda \cdot \mathbf{v}$ for any $\lambda \in \mathbb{K}$, $\mathbf{u}, \mathbf{v} \in V$.

The first four conditions express the fact that $(V, +\mathbf{0})$ forms an abelian group and if $\mathbb{K}$ is simply a ring, the same eight conditions define a structure that we call an $\mathbb{K}$-module.

Index

Bibliography

[1] Peter J. Cameron, Notes on Linear Algebra

[2] Jim Hefferon, Linear Algebra, 3rd edition