简明线性代数

孟国武
mameng@ust.hk
香港科技大学

香港, May 6, 2006

前言

线性代数是大学生们最基础的同时也是最容易的一门课。说它基础大概没有问题，但说它容易或许会有争论。这本小书的一个目的就是试图去证明它的确容易。

本书主要是为国内的正常聪明的理工科大学生们写的，但风格同现有的课本可能大不一样。它的特色如下：

- 同许多国内课本相比，主要区别在于更注重于一些概念的几何来源。本书强调线性空间与线性映射。几何与物理应当是与坐标选取无关，所以刻划他们的量或关系应当也是与坐标选取无关。基于这点，我没有采用计算型的定义。

- 本书极为简练，尤其是同我所知的美国课本相比。但它的内容却可能比一般非数学系课本的内容要多一点[1], 譬如正定矩阵的判据, Cayley-Hamilton 定理, 同时对角化等等。美国的教材通常都很厚但内容并不多，而且时常更新，大概是有商业的动机才会这样的吧。

- 本书没有难题，但有很多练习。那些练习是本书整体的一部分，对概念清楚的人来说几乎是一目了然；线性代数本来就是容易的嘛。

本书对一些材料的处理与传统教科书的很不一样。我倒不是为了标新立异，而是觉得这样处理更自然。我做学生时觉得线性代数的有些概念很神秘，譬如矩阵乘法，行列式的定义，秩的定义，Hermitian 内积的定义; 也对本来是很清楚的关系不知道如何去证明，譬如$\operatorname{Det} AB = \operatorname{Det} A \operatorname{Det} B$, AB的秩不大于A或B的秩。同时我对线性空间的定义也不满意，因为我记不住那十条公理。我想大多数中国理工科的学生有同样的感觉。我深知每个人的成长过程不一样，所以我觉得不自然的东西别人可能会觉得自然，反之亦然。尽管如此，我还是斗胆来做这么一次尝试。

本书的最大缺陷是对读者的要求可能高了一点。我假定读者熟悉有关集合与映射的基本概念以及有关一元多项式的基本概念，也假定他们熟悉反证法和归纳法。另外，我假定读者不害怕尝一点点现代数学的味道并勇于做练习。

阅读建议

习题是本书不可缺少的部分，阅读时不要漏掉，因为后面有可能要用到。对难一点的习题我已尽可能地把它拆成几个容易的部分或提供提示。有些习题太简单，其作用是加深对某些概念或关系的印象。相对于国内的课本，本书的习题又容易又不多，但是如果你们能把它们全做了，我相信你们的线性代数比大部分我见到的人要好。我不赞成做技术上很难的题，因为我们的目的是要掌握基本精神并融会贯通而不是竞赛。我也不赞成做大量的计算题，因为聪明的人最需要的是好的思想，在有需要时他自己会算而且会算得很好。

[1]对非数学系的学生来说，第3.5节和第6章可以略去。

线性代数是平直空间的几何的一部分, 所以你们要尽量多画图。书中没有图是因为电脑上作图太麻烦。

鸣谢

WAN Yanyi 仔细阅读了本书并做了全部习题和指正了书中的一些typing errors。他也提了一些很好的建议。在此我非常感谢。

Contents

Part I

Elementary Part

Chapter 1

Preliminaries(2 weeks)

The principal objects in Linear Algebra are vector spaces and linear maps. If we forget the linear structures, then we get the underlying objects: sets and set maps. It is then not a surprise that many concepts in Linear Algebra are actually the concepts in set theory. Therefore it is a good idea for us to record some basic set-theoretical terminologies and notations that we will use throughout this book. This is done in section 1.1.

Matrices are generalizations of numbers. But why do they take a rectangular shape rather than other shapes such as a triangular shape? Also, why is the matrix multiplication defined in that peculiar way? These questions are naturally answered by emphasizing that matrices are actually the algebraic or coordinate representations of linear maps. This is explained in section 1.2.

In the remaining sections the high school problem of solving systems of linear equations by **Gauss eliminations** is reviewed, but in the efficient and compact language of matrices. The reformulation of this high school problem as a problem for linear maps is an important step and this conforms to the spirit of modern mathematics. Here, the fundamental technique of Linear Algebra — the one that reduces a matrix to its row reduced echelon form, plus some related concepts, are developed. Let us emphasize that the solution of many problems in Linear Algebra (either theoretical or computational) eventually boils down to how well you really understand this technique and its related concepts.

We end this chapter with a list of exercise problems. Please make sure to do all of them because some of them will be used later again and again.

Remark 1.1. For the sake of having geometric intuitions and avoiding dull and unnecessary repetitions, we explain the basic concepts here in the real case, i.e., the ground field is taken to be $\mathbb{R}$—the collection of real numbers. However, everything goes through if the ground field is replaced by $\mathbb{C}$—the collection of complex numbers.

1.1 Sets and Set Maps

A **set** is just a collection of objects and is often denoted by capital Latin letters, such as S, X, etc.. The objects of a set are called **elements** of the set and are often denoted by small Latin letters. The set that contains no element is called the **empty set**, and is denoted by ϕ. The set of all real numbers is denoted by $\mathbb{R}$ and the set of all complex numbers is denoted by $\mathbb{C}$. Of course, $\mathbb{R}$ is a proper subset of $\mathbb{C}$, here 'proper' means that $\mathbb{R} \neq \mathbb{C}$. The empty set is a proper subset of any nonempty set.

By convention, a finite set could be an empty set. If S is a finite set, we use $|S|$ to denote the cardinality of S, i.e., the number of elements in S. For example, the set of first n positive integers, denoted by $\{1, 2, \ldots, n\}$, has the cardinality equal to n.

If x is an element of X, we write $x \in X$. If X is a subset of Y, we write $X \subset Y$, and the complement of X in Y is written as $X \setminus Y$. If X is a proper subset of Y, we write $X \subsetneq Y$. Let X, Y be two sets, to prove $X = Y$ we just need to prove that $X \subset Y$ and $Y \subset X$. If A, B are subsets of X, the union of them is denoted by $A \cup B$ and the intersection of them is denoted by $A \cap B$.

If X, Y are two sets, then the Cartesian product of X with Y is denoted by $X \times Y$. Recall that

$$X \times Y = \{(x, y) \,|\, x \in X, \, y \in Y\} \,.$$

The Cartesian product of n copies of X is denoted by X^n.

Intuitively speaking, a map is just a gun. Recall that if $f\colon X \longrightarrow Y$ is a (set) map, then Y is called the **target** and X is called the **domain**. If $x \in X$, then $f(x)$ is called the **image** of x under f. For $y \in Y$, we use $f^{-1}(y)$ to denote the solution set of equation $f(x) = y$ — the collection of all $x_0 \in X$ such that $f(x_0) = y$. For $S \subset X$, we use $f(S)$ to denote the set of images under f of the elements in S. $f(X)$ is called the **range** or **image** of f, denoted by $\mathrm{Im}(f)$. We use $I\colon X \longrightarrow X$ to denote the identity map in X: $x \mapsto x$ for any x. If W is a subset of X, then we use $f|_W$ to denote the restriction of f to W.

We say f is **one-to-one** if $f^{-1}(y)$ has at most one element for any y, we say f is **onto** if $f(X) = Y$. By definition, the **graph** of f (denoted by $\Gamma(f)$) is this subset of $X \times Y$:

$$\Gamma(f) := \{(x, f(x)) | x \in X\}.$$

We say f is a **one-one correspondence** if f is both one-to-one and onto; such a f is also called **invertible**. It is clear that f is invertible means that there is a unique map $g\colon Y \longrightarrow X$ such that both gf and fg are the identity map on the respective set; such unique g is called the **inverse** of f, written as f^{-1}.

Exercise 1.1. Let $f\colon X \longrightarrow Y$, $g\colon Y \longrightarrow Z$ be set maps, $gf\colon X \longrightarrow Z$ be their composition which maps $x \in X$ to $g(f(x))$. Show that, 1) gf is onto implies that g is onto; 2) gf is one-to-one implies that f is one-to-one; 3) If f and g are invertible, then gf is invertible and $(gf)^{-1} = f^{-1}g^{-1}$; 4) For maps f, g and h, we have $(fg)h = f(gh)$ if either side is defined. (So we simply write fgh, rarely write $(fg)h$ or $f(gh)$.)

If $f\colon X \longrightarrow Y$ is a map, the induced map from X^n to Y^n which sends $(x_1, \ldots, x_n)$ to $(f(x_1), \ldots, f(x_n))$ will be denoted by $f^{\times n}$. However, $\mathbb{R}^n$ denotes both the set of n-tuple of real numbers and the set of real column matrices with n entries. Likewise, $\mathbb{C}^n$ denotes both the set of n-tuple of complex numbers and the set of complex column matrices with n entries.

We say that square

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] & Y \arrow[d, "g"] \\ Z \arrow[r, "k"'] & W \end{tikzcd}$$

is commutative if $gf = kh$. Similarly, we have the notion of commutative triangles.

1.1.1 Equivalence Relations

Let X be a set. A **relation** on X is just a subset R of $X \times X$. A map $f\colon X \longrightarrow X$ defines a relation on X, i.e., $\Gamma(f)$, but not every relation on X arises this way. We say R is an **equivalence relation** if it satisfies the following three axioms:

1. Reflexive: (x, x) is in R for any $x \in X$.
2. Symmetry: (x, y) is in $R \implies (y, x)$ is in R.
3. Transitive: (x, y), (y, z) are in $R \implies (x, z)$ is in R.

If R is an equivalence relation on X, it is customary to write $(x, y) \in R$ as $x \sim_R y$ or simply $x \sim y$. So the axioms can be written as 1) $x \sim x$ for any $x \in X$, 2) $x \sim y$ implies that $y \sim x$, 3) $x \sim y$ and $y \sim z$ imply that $x \sim z$.

Let R be an equivalence relation on X, and S be a subset of X. We say that S is a R-**equivalence class** if any two elements of S are R-equivalent and S is maximal: if $S \subset T$ and any two elements of T are R-equivalent, then $S = T$. It is not hard to see that the R-equivalence classes are mutually disjoint, and X is the union of them. In other words, R partitions X into disjoint union of subsets. It is also not hard to see that an equivalence relation on X is nothing but a partition of X into disjoint union of subsets. The set of R-equivalence classes on X is denoted by X/R or $X/\sim$. If $x \in X$, we write $[x]$ for the equivalence class containing x, and call x a representative of the equivalence class.

Example 1.1. Let $\mathbb{Z}$ be the set of integers. Let

$$R = \{(m, n) \in \mathbb{Z} \times \mathbb{Z} \mid m - n \text{ is divisible by } 2\}.$$

Then R is an equivalence relation on $\mathbb{Z}$. There are two equivalence classes: $[0]$ (the set of even integers) and $[1]$ (the set of odd integers). The set of this equivalence classes is customarily denoted by $\mathbb{Z}_2$.

1.2 Matrices and Linear Maps

A real (complex) **matrix** is just a rectangular array of real (complex) numbers. If the array has m rows and n columns, we say it is an $m \times n$-matrix. For example,

$$\begin{matrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{matrix}$$

is a 2×3-matrix. It is customary to write the above matrix as

$$\begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{pmatrix}$$

for a good reason.

Matrices are not introduced artificially; they arise as the coordinates for linear maps between linear spaces. If we take the linear spaces as $\mathbb{R}^n$ (the set of n-tuple of real numbers) and $\mathbb{R}^m$, then by definition, a map T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ is called a **linear map** if each component of T is a homogeneously linear polynomial in $(x_1, \ldots, x_n) \in \mathbb{R}^n$, i.e., if

$$T(x_1, \ldots, x_n) = (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n,\ \ldots,\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n) \tag{1.1}$$

for some real numbers a_{ij}'s. For simplicity we may write x for $(x_1, \ldots, x_n)$ and $T(x)$ for $T(x_1, \ldots, x_n)$. We let

$$T_i(x) = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n$$

and call function T_i the i-th component of T. It is clear that the linear map T in equation (1.1) is uniquely determined by the following real $m \times n$-matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{1.2}$$

— the **coordinate matrix** of the linear map T. In fact, this assignment of a matrix to a linear map defines a one-one correspondence:

$$\boxed{\text{linear maps from } \mathbb{R}^n \text{ to } \mathbb{R}^m \leftrightarrow m \times n\text{-matrices.}}$$

For simplicity, we often write matrix in (1.2) as

$$[a_{ij}].$$

Exercise 1.2. Let T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a map and for each $1 \le i \le m$ we let T_i be the i-th component of T. Show that T is a linear map $\iff$ for each $1 \le i \le m$ and each number c, we have 1) $T_i(x_1 + y_1, \ldots, x_n + y_n) = T_i(x_1, \ldots, x_n) + T_i(y_1, \ldots, y_n)$; 2) $T_i(cx_1, \ldots, cx_n) = cT_i(x_1, \ldots, x_n)$. (Here x_i's and y_i's are variables)

Suppose that T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$, S: $\mathbb{R}^p \longrightarrow \mathbb{R}^n$ are linear maps, then TS is also a linear map. One can show that the composition of linear maps, when translated into the language of matrices, becomes the multiplication of matrices.

Exercise 1.3. Derive the matrix multiplication rule for yourself. If you know this rule already, verify it.

Suppose that T, S: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ are two linear maps, then $T+S$: $x \mapsto (T_1(x)+S_1(x), \ldots, T_m(x)+S_m(x))$ is also a linear map. One can show that the sum of linear maps, when translated into the language of matrices, become the sum of matrices.

Exercise 1.4. Derive the matrix addition rule for yourself. If you know this rule already, verify it.

Suppose that $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear map, c is a real number, then $cT\colon x \mapsto (cT_1(x), \ldots, cT_m(x))$ is also a linear map. One can show that the multiplication of linear maps by numbers (called scalar multiplication), when translated into the language of matrices, become the scalar multiplication of matrices.

Exercise 1.5. Derive the scalar multiplication rule for yourself. If you know this rule already, verify it.

The zero linear maps correspond to the zero matrices and the identity maps corresponds to the identity matrices. Note that an identity matrix must be a square matrix.

Exercise 1.6. Write down the 3×3-identity matrix.

1.3 Linear Systems and their Solution Sets

A linear system is just a system of linear equations. An equation in (real or complex) variables x_1, x_2, ..., x_n is an expression of the form

$$f(x_1, x_2, \ldots, x_n) = 0.$$

This equation is called **linear** if f is a linear polynomial in variables x_1, x_2, ..., x_n, i.e., if it is of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n - b = 0$$

for some numbers (real or complex) a_1, a_2, ..., a_n and b. We prefer to put the preceeding equation in the following equivalent form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

Therefore, a **linear system** of m equations in n variables x_1, x_2, ..., x_n is of the form

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

where a_{ij}'s and b_k's are numbers. (1.3) is called **homogeneous** if all b_k's are zero, and is called **inhomogeneous** otherwise.

By definition, a **solution** of (1.3) is an n-tuple of numbers $(s_1, s_2, \ldots, s_n)$ such that when plugging $x_1 = s_1$, $x_2 = s_2$, ..., $x_n = s_n$ into (1.3), the two sides of (1.3) agree numerically.

The primary question concerning linear system (1.3) is its **solution set** — the collection of all solutions of (1.3). What does the solution set of the linear system (1.3) look like? Well, in the real case, geometrically, a solution of (1.3) is just a point in $\mathbb{R}^n$ and the solution set of an equation in (1.3) could be empty, $\mathbb{R}^n$, or an $(n-1)$-dimensional flat space inside $\mathbb{R}^n$ (called **hyperplane**)[1].

[1]To see this, we observe that, up to a change of the coordinate system, a linear equation can be equivalently put in the form $0 = 1$, or $0 = 0$ or $x_1 = ?$.

Note that the solution set of (1.3) is the intersection of the solution sets of equations in (1.3), therefore, it is either empty or $\mathbb{R}^n$ or the non-empty intersection of at most m hyperplanes in $\mathbb{R}^n$. In any case, the solution set of (1.3) is a k-dimensional flat space inside $\mathbb{R}^n$ with $-1 \le k \le n$. Here, by convention, an empty set has dimension -1.

Note that the information on linear system (1.3) is completely encoded in its **augmented coefficient matrix**:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix} \tag{1.4}$$

which is often compactly written as $[A, \vec{b}]$ where

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

Therefore, we may just work with (1.4) and keep the correspondence between linear system (1.3) and matrix (1.4) in mind.

Remark 1.2. Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then, in case you already know how to do matrix multiplication, (1.3) can be compactly rewritten as this matrix equation

$$\boxed{A\vec{x} = \vec{b}\ .}$$

In the rest of this chapter we will show you how to obtain the solution set of (1.3) in a systematic way. Before we start, let us present some simple examples.

Example 1.2. Assume $m = n = 1$, i.e., the linear system is just one linear equation in one variable x. In the simplest situation, it is of the form $1 \cdot x = ?$, or $0 \cdot x = 0$, or $0 \cdot x = 1$ and the corresponding augmented coefficient matrices are $[1 \quad ?]$, $[0 \quad 0]$ and $[0 \quad 1]$. (Here ? means an arbitrary number.) ∎

We can quickly read off the solution set in any of these three cases: they are $\{?\}$, $\mathbb{R}$ and ϕ (the empty set) respectively.

Example 1.3. Assume $m = n = 2$. In the simplest situation, the augmented coefficient matrix is of one of the following forms:

$$\begin{bmatrix} 1 & 0 & ? \\ 0 & 1 & ? \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & ? & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & ? \\ 0 & 0 & 0 \end{bmatrix},$$

$$\begin{bmatrix} 1 & ? & ? \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

∎

We can quickly read off the solution set in any of above cases. For example, the bottom left matrix corresponds to linear system

$$\begin{cases} 1 \cdot x_1 + b \cdot x_2 = c \\ 0 \cdot x_1 + 0 \cdot x_2 = 0 \end{cases}.$$

Let $x_2 = r$, from the first equation of the above linear system, we have $x_1 = -bx_2 + c = -br + c$, so the solution set is $\{(-br + c, r) \mid r \text{ is arbitrary}\}$.

I hope the above examples have convinced you that if the augmented coefficient matrix is simple enough, we can find the solution set of the linear system very quickly. What about the general situation? It turns out that any general linear system is equivalent to a unique simplest possible linear system, here "equivalent" means having the same set of solutions. Therefore, to show you how to solve a general linear system, I need to tell you what a simplest possible linear system (or matrix) is and how to reduce a linear system (or matrix) to its simplest possible form. That will be the subject of the next section.

1.4 Elementary Operations and Echelon Forms

The simplest possible form of a matrix mentioned in the preceding section is technically called the reduced row echelon form. By definition, a matrix is in **row echelon form** if it satisfies the following conditions[2]:

1) *All zero rows must be in the bottom*;

2) *The left most nonzero entry of the nonzero rows must be positioned this way: suppose that the first k rows are the nonzero rows and the left most nonzero entries of those rows are a_{1j_1}, ..., a_{kj_k} respectively, then $j_1 < j_2 < \cdots < j_k$* . (Intuitively, if you view these entries as your stepping stones, when you jump down one row, you always jump right at least one column.)

These left most nonzero entries are called the **pivotal entries**. Note that if a matrix is in row echelon form, then each of its nonzero rows contains exactly one pivotal entry, and each of its columns contains at most one pivotal entry. A column which contains a pivotal entry is called a **pivotal column**. It is clear that, for a matrix in row echelon form, the number of pivotal entries

[2]Please draw a picture for yourself.

is equal to the number of pivotal columns or nonzero rows, and is bounded by the number of rows and also by the number of columns. We record this as follows: for a matrix in row echelon form,

$$\boxed{\begin{array}{c}\#\text{ of pivotal entries} = \#\text{ of pivotal columns}\\ = \#\text{ of nonzero rows} \le \#\text{ of rows or }\#\text{ of columns.}\end{array}} \tag{1.5}$$

By definition, a matrix is in **reduced row echelon form** if it is in row echelon form and also satisfies the following conditions:
1) *Each pivotal entry is* 1;
2) *Each pivotal column has only one nonzero entry (i.e., the pivotal entry).*

One can check that the matrices listed in example 1.3 in the previous section are all in reduced echelon form.

Exercise 1.7. Show that a square matrix in reduced row echelon form is an identity matrix if and only if its columns are all pivotal.

Next, we would like to point out that *any matrix can be reduced to a unique matrix in reduced row echelon form by doing finitely many elementary row operations*[3]. Therefore, the notion of pivotal column can be naturally extended: the i-th column of a matrix is called a pivotal column if the i-th column of its reduced row echelon form is a pivotal column. From now on, if A is a matrix, then its reduced row echelon form will be denoted by $\bar{A}$.

There are three types of elementary operations: *switching two rows, multiplying one row by a* **nonzero** *number and adding a multiple of one row to another row.* It is clear that elementary row operations correspond to the operations used in the Gauss eliminations and are all invertible operations.

Let A, B be two matrices, we say $A \sim_r B$ if after doing finitely many elementary row operations one can turn A into B. One can check that $\sim_r$ defines an equivalence relation (called **row equivalence**) on the set of all matrices and equivalent matrices must have the same sizes. Moreover, two linear systems have the same set of solutions if their augmented coefficient matrices are row equivalent.

We are now ready to state the steps for solving linear systems. Of course, these are the actual steps used in the Gauss eliminations.

Step 1 Write down the augmented coefficient matrix A and then reduce A to its reduced row echelon form $\bar{A}$ by doing elementary row operations.

Step 2 Write down the solution set S.

Case 1 The last column of $\bar{A}$ is a pivotal column. $S = \phi$, because the rows that contains the pivotal entry in the last column corresponds to equation $0 = 1$.

[3]It is not hard to imagine how a matrix can be turned into a matrix in reduced row echelon form. The difficulty is the uniqueness part. If you can not prove it, that is OK, because you will be able to prove it easily (see Ex. 2.13) after we develop enough concepts, and the uniqueness is never used at least before Ex. 2.13.

Case 2 The last column of $\bar{A}$ is a non-pivotal column. Set each non-pivotal variable as a free parameter. For each nonzero rows of $\bar{A}$, write down the corresponding equation and then solve it.

Here, a variable is called a **pivotal variable** if the corresponding column is a pivotal column, otherwise it is called a non-pivotal variable. Note that each of those equations in Case 2 contains exactly one pivotal variable whose coefficient is always 1, so we can solve it for this pivotal variable by simply moving all other terms (they may contain free parameters) to the right hand side.

Example 1.4. Solve

$$\begin{cases} x_1 + 2x_2 = 1 \\ 2x_1 + 3x_2 = 3 \end{cases}$$

Solution. Step 1.

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 3 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 0 & 3 \\ 0 & -1 & 1 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix}.$$

(The three elementary row operations used here are: add -2 multiple of row 1 to row 2, add 2 multiple of row 2 to row 1, multiply row 2 by -1. The choice of these operation is not unique, however the final form is the same.)

Step 2. All variables are pivotal variables. The two nonzero rows correspond to two equations: $x_1 = 3$ and $x_2 = -1$. So $S = \{(3, -1)\}$. ∎

Example 1.5. Solve

$$\begin{cases} x_1 + 2x_2 = 1 \\ 2x_1 + 4x_2 = 3 \end{cases}$$

Solution. Step 1.

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 3 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(The two elementary row operations used here are: add -2 multiple of row 1 to row 2, add -1 multiple of row 2 to row 1.)

Step 2. The last column of $\bar{A}$ is a pivotal column, so $S = \phi$. (The last row of $\bar{A}$ says that $0 = 1$.) ∎

Example 1.6. Solve

$$\begin{cases} x_1 + 2x_2 = 1 \\ 2x_1 + 4x_2 = 2 \end{cases}$$

Solution. Step 1.

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

(The only elementary row operations used here are: add -2 multiple of row 1 to row 2.)

Step 2. x_1 is a pivotal variable and x_2 is a non-pivotal variable. So we set $x_2 = r$. Then the 1st row correspond to equation $x_1 + 2r = 1$, so $S = \{(1 - 2r, r) \mid r \text{ is a free parameter}\}$. ∎

Example 1.7. Solve

$$\begin{cases} 0 \cdot x_1 + x_2 + 2x_3 = 1 \\ 0 \cdot x_1 + 2x_2 + 4x_3 = 2 \end{cases}$$

Solution. Step 1.

$$\begin{bmatrix} 0 & 1 & 2 & 1 \\ 0 & 2 & 4 & 2 \end{bmatrix} \sim_r \begin{bmatrix} 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

(The only elementary row operations used here are: add -2 multiple of row 1 to row 2.)

Step 2. x_2 is a pivotal variable and x_1 and x_3 are non-pivotal variables. So we set $x_1 = r_1$ and $x_3 = r_2$. Then the 1st row corresponds to equation $x_2 + 2r_2 = 1$, so

$$S = \{(r_1, 1 - 2r_2, r_2) \mid r_1 \text{ and } r_2 \text{ are free parameters}\}.$$

∎

Exercise 1.8. Consider the following linear system

$$\begin{cases} x_1 - x_2 + (h-1)x_3 + x_5 = 1 \\ x_1 + (h+2)x_3 - x_5 = 1 \\ 2x_2 + 6x_3 + x_4 = 2 \\ -x_1 + 2x_2 + (4-h)x_3 + x_4 + x_5 = h+4 \end{cases}$$

Determine the values of h for which the system has a solution and find the general solution for these cases.

1.5 Elementary Matrices

Start with I_n — the $n \times n$ identity matrix, if we apply an elementary row operation to it, we get a new $n \times n$ matrix, called the **elementary matrix**(please compare with Exercise 2.9) associated to that elementary row operation. It is clear that distinct elementary row operations will produce distinct elementary matrices. Therefore, we have an one-one correspondence:

$$\boxed{\begin{array}{c} \text{elementary row operations on } m \times n\text{-matrices} \\ \updownarrow \\ \text{elementary } m \times m\text{-matrices.} \end{array}}$$

A natural question arises here is this: suppose that E is known to be an elementary $m \times m$-matrix, what kind of elementary row operation does it correspond to? Well, if you work out a few examples, you should be able to find this answer: it is the one such that if you apply it to an $m \times n$-matrix A, the resulting matrix is precisely EA.

Exercise 1.9. Prove this claim now or after you have done Exercise 2.9.

Should we call the elementary matrices the elementary row matrices because they are related to elementary row operations? Well, we should not, that is because one realizes that any $n \times n$

elementary matrix can also be obtained by applying a unique elementary column operation to I_n. (Prove it) It is not hard to see that

$$\boxed{\begin{array}{c}\text{elementary column operations on } m \times n\text{-matrices}\\ \updownarrow \\ \text{elementary } n \times n\text{-matrices.}\end{array}}$$

Now if suppose that E is known to be an elementary $n \times n$-matrix, what kind of elementary column operation does it correspond to? Well, if you work out a few examples, you should be able to find this answer: it is the one such that if you apply it to an $m \times n$-matrix A, the resulting matrix is precisely AE.

Exercise 1.10. Prove this claim now or after you have done Exercise 2.9.

1.6 Additional Exercise Problems

The purpose here is to list a few exercise problems. These problems are helpful to your understanding of the course material, and some of them will be used later again and again. I strongly encourage you do all of them. Perhaps it is a good idea to discuss them with your classmates if you are not so confident in doing them alone.

Exercise 1.11. Show that[4] any matrix can be reduced to a unique matrix in reduced row echelon form by doing finitely many elementary row operations.

Exercise 1.12. Define elementary column operations, column equivalence, column echelon form and reduced column echelon form.

Exercise 1.13. We say a matrix is in **reduced form** if it is in both reduced row echelon and reduced column echelon form. Please describe the reduced form more directly. Show that any matrix can be reduced to a unique matrix in reduced form by a combination of elementary row and column operations.

Exercise 1.14. List all 2×3-matrices in reduced row echelon form, all 2×3-matrices in reduced column echelon form and all 2×3-matrices in reduced form.

Exercise 1.15. Let A, B be two matrices. We say $A \sim B$ if after doing finitely many elementary row or column operations we can turn A into B. Show that $\sim$ defines an equivalence relation on the set of matrices. We say A, B are **row-column equivalent** or simply **equivalent** if $A \sim B$. Compute the number of row-column equivalence classes on the set of all $m \times n$-matrices.

Exercise 1.16. We say a matrix is invertible if the corresponding linear map is invertible. Show that elementary matrices are all invertible.

Exercise 1.17. Show that $A \sim_r B \iff$ there are elementary matrices $E_1, \ldots, E_k$ such that $A = E_1 \cdots E_k B$. Can you formulate the corresponding statement in the case when $\sim_r$ is replaced by either column equivalence or row-column equivalence.

[4] You may skip the uniqueness part because it will be addressed in Ex. 2.13 later.

Exercise 1.18. Consider a linear system whose augmented coefficient matrix is $[A, \vec{b}]$. Show that

1) Its solution set is non-empty if and only if the last column of $[A, \vec{b}]$ is a non-pivotal column.

2) Its solution set has a unique element if and only if the last column of $[A, \vec{b}]$ is the only non-pivotal column.

3) Its solution set has infinitely many elements if and only if the last column of $[A, \vec{b}]$ is not the only non-pivotal column.

Exercise 1.19. Consider a linear system whose augmented coefficient matrix is $[A, \vec{0}]$. Show that

1) Its solution set has a unique element if and only if the number of pivotal columns of A is equal to the number of columns of A.

2) Its solution set has infinitely many elements if and only if the number of pivotal columns of A is less than the number of columns of A.

Exercise 1.20. Consider a linear system whose augmented coefficient matrix is $[A, \vec{b}]$. Show that this system has a solution for any $\vec{b}$ if and only if the number of pivotal columns of A is the same as the number of rows of A.

Chapter 2

Model Vector Spaces(2.5 weeks)

The basic concepts of Linear Algebra are developed in this and the next chapter. The ground field is assumed to be the field of real numbers; however, when it is replaced everywhere by the field of complex numbers, every piece developed here is just equally fine.

Major efforts are spent on model vector spaces and the linear maps between them. There are good reasons for that. Firstly, model vector spaces and the linear maps between them are very close to the computational side, and being able to compute always makes us feel sound and comfortable. Secondly, model vector spaces are very close to our geometric intuition, because they are essentially just Euclidean spaces. Lastly, they are the **models**. If we understand everything about them in non-superficial way, we actually already understand everything about general vector spaces, and the only things we need to do are translations or interpretations. This is because, although vector spaces appear in different guises in real world, as far as mathematics is concerned, each is equivalent to a model vector space.

2.1 Vectors

Vectors are geometrical or physical quantity. Many important physical quantities such as velocity, magnetic field are vectors in the space that we live in. Geometrically a vector in a flat space (official name: Euclidean space) is just an arrow, i.e., a line segment together with a direction. If P, Q are two distinct points in an n-dimensional flat space, then the direct line segment from P to Q, denoted by $\overrightarrow{PQ}$, is a vector with Q being its head and P being its tail (also called its location). By convention, $\overrightarrow{PP}$ also denotes a vector, called the zero vector at P. Note that vectors at different locations of a flat space can be identified with each other by parallel transport. With this identification in mind, people often consider any two vectors that are parallel and have the same length as the same vector.

Upon a choice a *rectangular* coordinate system, an n-dimensional flat space can be identified with $\mathbb{R}^n$. Moreover, the vectors located at the origin O of the coordinate system can be identified with column matrices with n entries: if the point P has coordinate $x = (x_1, \ldots, x_n)$, then vector

$\overrightarrow{OP}$ can be identified with column matrix

$$\vec{x} \equiv \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

and we often write

$$\overrightarrow{OP} = \vec{x}. \tag{2.1}$$

Note that Eq. (2.1) should be understood this way: the column matrix on the right is the coordinate representation of the vector on the left. Because of (2.1), we often call such a column matrix a **column vector** in $\mathbb{R}^n$, or a vector in $\mathbb{R}^n$. The set of all column matrices with n real entries is our model vector (or linear) space of dimension n and is also denoted by $\mathbb{R}^n$. Note that we have a natural one-one correspondence: $x \leftrightarrow \vec{x}$. Please keep this one-one correspondence in mind:

$$\boxed{\begin{array}{c}\text{The set of vectors in an } n \text{ dimensional flat spaces with a fixed common tail} \\ \updownarrow \\ \mathbb{R}^n = \text{the set of column vectors in } \mathbb{R}^n.\end{array}}$$

This correspondence is never unique because it depends on the choice of a rectangular coordinate system.

Note that any two vectors in the same location can be naturally added and any vector can be naturally multiplied by any number. These operations are defined geometrically; for example, addition is defined via the parallelogram rule. In terms of coordinate representation, they are defined as follows: suppose that $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ and c is a number, then

$$\vec{x} + \vec{y} := \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}, \qquad c\vec{x} := \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}.$$

These operations on vectors are called **vector addition** and **scalar multiplication** respectively, and should be viewed as maps: $\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ $((\vec{x}, \vec{y}) \mapsto \vec{x} + \vec{y})$ and $\mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ $((c, \vec{x}) \mapsto c\vec{x})$.

Exercise 2.1. Draw some pictures on a plane to figure out the geometric meanings of vector addition and scalar multiplication. How about the dimension 3 case?

Span

It is clear how to add any finite number of vectors in $\mathbb{R}^n$. It is also clear how to make sense of expression $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k$ (where c_i's are real numbers, $\vec{v}_i$'s are vectors in $\mathbb{R}^n$.). This last expression is called the **linear combination** of $\vec{v}_1, \ldots, \vec{v}_k$ weighted by $c_1, \ldots, c_k$, or simply the linear combination of $\vec{v}_1, \ldots, \vec{v}_k$. By definition, the set of all possible linear combinations of $\vec{v}_1, \ldots, \vec{v}_k$ is called the **(linear) span** of $\vec{v}_1, \ldots, \vec{v}_k$, denoted by $\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_k\}$.

By definition, the span of the empty set of vectors in $\mathbb{R}^n$ is the set consisting of only the zero vector in $\mathbb{R}^n$. In general, if S is a set of vectors in $\mathbb{R}^n$ (may not be finite), the span of S, denoted by $\operatorname{span} S$, is defined to be the collection of all possible linear combinations of finitely many vectors in S. In other words, $\operatorname{span} S$ is the union of the span of all possible finite subsets of S.

Exercise 2.2. Suppose that $v_1, \ldots, v_k$ are vectors in $\mathbb{R}^n$. Show that, if $\vec{v}_k$ is a linear combination of the other $\vec{v}_i$'s, then

$$\mathrm{span}\{\vec{v}_1, \ldots, \vec{v}_k\} = \mathrm{span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}.$$

2.2 Vector/Matrix Equations

Linear system (1.3) can be put in a compact form

$$T(x) = b \tag{2.2}$$

where $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$, $x = (x_1, \ldots, x_n)$, and $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is the linear map[1]:

$$T(x) = (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n,\ \ldots,\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n). \tag{2.3}$$

Via the natural one-one correspondence between points in $\mathbb{R}^n$ (or $\mathbb{R}^m$) and column vectors in $\mathbb{R}^n$ (or $\mathbb{R}^m$), we can view T in (2.3) as a map between model vector spaces and rewrite (2.2) as

$$\boxed{T(\vec{x}) = \vec{b}.} \tag{2.4}$$

It is then clear that the solution set of linear system (1.3) is just $T^{-1}(\vec{b})$. While linear system (1.3) is more concrete and elementary and very close to the computational side, the equivalent formulation (2.4) is more concise and conceptual and very close to the geometrical or physical thinking. From now on, we will think in terms of (2.4) and come back to (1.3) or matrix (1.4) when we need to do concrete computations.

Let $\vec{e}_i$ be the column vector in $\mathbb{R}^n$ whose i-th entry is one and all other entries are zero, $\vec{a}_i$ be i-th column of matrix (1.4). Then $T(\vec{e}_i) = \vec{a}_i$, and

$$T(\vec{x}) = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n.$$

So (2.4) can be rewritten as the following vector equation:

$$\boxed{x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \vec{b}.} \tag{2.5}$$

Let A be the matrix obtained from matrix (1.4) by deleting its last column, i.e.,

$$A = [\vec{a}_1\ , \ldots,\ \vec{a}_n].$$

Then T and A uniquely determine each other. In fact, they are related to each other in very explicit ways:

$$\boxed{\begin{array}{c} T(\vec{x}) = A\vec{x} \quad \text{for any } \vec{x} \in \mathbb{R}^n, \\[1em] A = [T(\vec{e}_1), \ldots, T(\vec{e}_n)]. \end{array}} \tag{2.6}$$

Here $A\vec{x}$ is the matrix multiplication of A with $\vec{x}$.

[1]By definition, a **linear map** from $\mathbb{R}^n$ to $\mathbb{R}^m$ is just a map whose each component is a linear polynomial without constant term.

Exercise 2.3. Verify Eq. (2.6).

Therefore,

$$\boxed{A\vec{x} = x_1\vec{a}_1 + \cdots + x_n\vec{a}_n,} \tag{2.7}$$

and Eq. (2.4) can also be rewritten as the following matrix equation:

$$\boxed{A\vec{x} = \vec{b}.} \tag{2.8}$$

One should keep in mind that (1.3), (2.4), (2.5) and (2.8) are just different forms of the same object. Note that $A\vec{e}_i = \vec{a}_i$. We also note that matrix (1.4) can be compactly written as $[A,\ \vec{b}]$, i.e.,

$$[\vec{a}_1\ , \ldots,\ \vec{a}_n,\ \vec{b}].$$

It is a good time to introduce some basic terminologies. Let T be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, A be the corresponding coordinate matrix (called the **standard coordinate matrix** or simply **standard matrix** for T). The **kernel** of T, denoted by $\ker T$, is defined to be the solution set of equation $T(\vec{x}) = \vec{0}$. The **null space** of A, denoted by $\mathrm{Nul}\, A$, is defined to be the solution set of equation $A\vec{x} = \vec{0}$. The **column space** of A, denoted by $\mathrm{Col}\, A$, is defined to be the span of the columns of A. It is not difficult to see that

1. $\mathrm{Col}\, A = \mathrm{Im}\, T$.
2. T is onto if and only if $\mathrm{Col}\, A = \mathbb{R}^m$.
3. T is one-to-one if and only if $\ker T$ is trivial, i.e., consists of only the zero vector.
4. $\ker T = \mathrm{Nul}\, A$.
5. T is one-to-one if and only if the columns of $\bar{A}$ are all pivotal.
6. T is onto if and only if the rows of $\bar{A}$ are all nonzero.

Exercise 2.4. Prove the above six points.

Exercise 2.5. Let $T : \mathbb{R}^4 \longrightarrow \mathbb{R}^3$ be a linear map defined by

$$T(x_1, x_2, x_3, x_4) = (x_1 - x_2 + x_3 + x_4, x_1 + 2x_2 - x_4, x_1 + x_2 + 3x_3 - 3x_4).$$

(a) Find the coordinate matrix of T; (b) Is T one to one? (c) Is T onto? (d) Find all x such that $T(x) = (0, 1, 2)$.

Exercise 2.6. Prove that a map T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear if and only if it respects the addition and scalar multiplication on $\mathbb{R}^n$ and $\mathbb{R}^m$:

$$\begin{aligned} T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}) \quad & \text{for all } \vec{x} \text{ and } \vec{y} \text{ in } \mathbb{R}^n, \\ T(c\vec{x}) = cT(\vec{x}) \quad & \text{for all } \vec{x} \text{ in } \mathbb{R}^n \text{ and all } c \in \mathbb{R}, \end{aligned}$$

or equivalently, for all $\vec{x}$, $\vec{y}$ in $\mathbb{R}^n$ and all real numbers c and d,

$$\boxed{T(c\vec{x} + d\vec{y}) = cT(\vec{x}) + dT(\vec{y}).}$$

Exercise 2.7. Let S: $\mathbb{R}^p \longrightarrow \mathbb{R}^n$, T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ be linear maps.

1) Show that TS: $\mathbb{R}^p \longrightarrow \mathbb{R}^m$ is also a linear map

2) Suppose that $T(\vec{x}) = A\vec{x}$, $S(\vec{y}) = B\vec{y}$, $TS(\vec{z}) = C\vec{z}$ for some unique matrix A, B and C. By definition, we have $C = AB$ — the matrix product of A with B. Show that, if we write $B = [\vec{b}_1, \ldots, \vec{b}_p]$, then

$$\boxed{AB = [A\vec{b}_1, \ldots, A\vec{b}_p]\,.} \tag{2.9}$$

Exercise 2.8. Let T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear map. Show that[2]

1) T is one-to-one implies that $n \le m$;

2) T is onto implies that $n \ge m$;

3) T is both one-to-one and onto implies that $n = m$.

Exercise 2.9. View $\mathbb{R}^n$ as the set of column matrices with n entries. Let T: $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ be the map that corresponds to an elementary row operation; for example, if the elementary row operation is the switching of row one with row two, then T maps

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} \text{ to } \begin{bmatrix} a_2 \\ a_1 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}.$$

Show that T is a linear map and the standard matrix for T is exactly the elementary matrix obtained from the $n \times n$-identity matrix by applying the given elementary row operation.

Linearly Independent

Let S be a non-empty *finite* set of vectors in $\mathbb{R}^n$. By definition, a **linear dependence relation** on S is just a way to write the zero vector $\vec{0}$ as a linear combination of *all* vectors in S. There is at least one such thing, namely, the trivial one: $\vec{0} = \sum_{a\in S} 0\,\vec{a}$. We say S is a **linearly independent set**, or the vectors in S are linearly independent, if the only linear dependence relation on S is the trivial one.

Let us use all of the vectors in S as column vectors to form a matrix A. In view of (2.7), one can see that S is a linear independent set $\iff$ equation $A\vec{x} = \vec{0}$ has only the trivial solution: $\vec{x} = \vec{0}$. Then it is easy to see from Ex. 1.19 that S is a linear independent set $\iff$ the columns of $\bar{A}$ are all pivotal. Therefore, if S is an independent set of vectors in $\mathbb{R}^n$, then $|S| \le n$ by (1.5).

Remark 2.1. 1) A single vector is linearly independent means that it is not zero.

2) Two vectors are linearly independent means that they are not parallel or antiparallel to each other

3) Three vectors are linearly independent means that they are not in a two-(or lower-)dimensional flat space.

4) In general, intuitively, n vectors are linearly independent means that they are not in a flat space of dimension less than n.

[2] Hint: use the exercises in the last chapter.

Exercise 2.10. Show that the linear dependence relation is preserved under a linear map. (I.e., if T is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, and $\vec{v}_1$, $\vec{v}_2$, ... are vectors in $\mathbb{R}^n$, c_1, c_2, ... are some numbers. If $\vec{0} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots$, then $\vec{0} = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots$.)

Exercise 2.11. Let T be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Show that T is one-to-one $\iff$ T always maps a linearly independent set of vectors in $\mathbb{R}^n$ to a linearly independent set of vectors in $\mathbb{R}^m$.

Exercise 2.12. Let A be a matrix in reduced row echelon form. Show that
1) the pivotal columns of A are linearly independent;
2) the non-pivotal columns of A are in the span of the pivotal columns of A.

Exercise 2.13. 1) Let A and B be two matrices in reduced row echelon form. Show that[3] $A = B$ $\iff$ $A \sim_r B$.

Therefore, the row reduced echelon form of a matrix is unique, and the notion of pivotal or non-pivotal columns extend to all matrices.

2) Show that, for any matrix A, the pivotal columns of A are linearly independent and the non-pivotal columns of A are in the span of the pivotal columns of A.

3) Show that $\operatorname{Col} A$ is the span of pivotal columns.

4) Show that if $A \sim_r B$, then $\operatorname{Nul} A = \operatorname{Nul} B$. So in particular $\operatorname{Nul} A = \operatorname{Nul} \bar{A}$.

Exercise 2.14. Consider vectors

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \quad \vec{a}_3 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}.$$

(1) Determine whether $\vec{b}$ is in $\operatorname{span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3\}$;
(2) Determine whether $\vec{a}_1$, $\vec{a}_2$, $\vec{a}_3$ span $\mathbb{R}^3$, i.e., whether $\mathbb{R}^3 = \operatorname{span}\{\vec{a}_1, \vec{a}_2, \vec{a}_3\}$;
(3) Does the homogeneous equation $x_1\vec{a}_1 + x_2\vec{a}_2 + x_3\vec{a}_3 = 0$ have nontrivial solutions?

Exercise 2.15. Determine which sets of vectors are linearly independent:

(1)
$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}.$$

(2)
$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 0 \\ 3 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} \quad \begin{bmatrix} -1 \\ 5 \\ 3 \\ 0 \end{bmatrix}.$$

(3)
$$\begin{bmatrix} 255 \\ -36 \\ 78 \end{bmatrix}, \quad \begin{bmatrix} 51 \\ -411 \\ 9 \end{bmatrix}, \quad \begin{bmatrix} -19 \\ 112 \\ 50 \end{bmatrix}, \quad \begin{bmatrix} 256 \\ -35 \\ 79 \end{bmatrix}.$$

[3] Hint: use the previous two exercises and work on columns from left to right.

Exercise 2.16. 1) Find the solution set of

$$\begin{bmatrix} 2 & 1 & 7 & 2 \\ -1 & 2 & -6 & -1 \end{bmatrix} \vec{x} = 0.$$

Please express your answer as a span of vectors in $\mathbb{R}^4$.

2) Find all the vectors

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \in \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\},$$

such that

$$\begin{bmatrix} 2 & 1 & 7 & 2 \\ -1 & 2 & -6 & -1 \end{bmatrix} \vec{x} = 0.$$

2.3 Linear and Affine Subspaces of $\mathbb{R}^n$

Let W be a subset of $\mathbb{R}^m$. We say W is a **(linear) subspace** of $\mathbb{R}^m$ if W is the range of some linear map[4] $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Assume $n > 0$. Let $\vec{e}_i$ be the vector in $\mathbb{R}^n$ whose i-th component is 1 and other components are zero. Then $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$. Since T is linear, we have $T(\vec{x}) = T(\sum_i x_i\vec{e}_i) = \sum_i x_i T(\vec{e}_i)$, then

$$W = \left\{ \sum_i x_i T(\vec{e}_i) \,|\, x_i\text{'s are arbitrary} \right\} = \operatorname{span}\{T(\vec{e}_1), \ldots, T(\vec{e}_n)\}.$$

In fact, one can see that W is a subspace of $\mathbb{R}^m$ if and only if it is the span of a finite number vectors in $\mathbb{R}^m$. Intuitively, a subspace of $\mathbb{R}^m$ is just a flat space sitting inside $\mathbb{R}^m$ and passing through the origin of $\mathbb{R}^m$.

Example 2.1. Let $\mathbb{R}^m \times 0$ be the set of all vectors in $\mathbb{R}^{m+n}$ whose bottom n entries are zero, $0 \times \mathbb{R}^n$ be the set of all vectors in $\mathbb{R}^{m+n}$ whose top m entries are zero. Then both $\mathbb{R}^m \times 0$ and $0 \times \mathbb{R}^n$ are subspaces of $\mathbb{R}^{m+n}$. ▮

For $\vec{y} \in \mathbb{R}^m$, $\vec{y} + W := \{\vec{y} + \vec{w} \,|\, \vec{w} \in W\}$ is called a **parallel translate** of W by $\vec{y}$. By definition, an **affine subspace** of $\mathbb{R}^m$ is a parallel translate of a subspace of $\mathbb{R}^m$. A (linear) subspace of $\mathbb{R}^m$ is then just an affine subspace of $\mathbb{R}^m$ containing the origin of $\mathbb{R}^m$. Geometrically, a k-dimensional affine subspace of $\mathbb{R}^n$ is just a flat k-dimensional space inside $\mathbb{R}^n$, and a k-dimensional subspace of $\mathbb{R}^n$ is just a flat k-dimensional space inside $\mathbb{R}^n$ and passing through the origin. So a two dimensional affine subspace in $\mathbb{R}^3$ is just a plane inside $\mathbb{R}^3$ and a one-dimensional affine subspace in $\mathbb{R}^3$ is just a straight line inside $\mathbb{R}^3$.

[4]We allow $n = 0$ and take $\mathbb{R}^0 = \{0\}$.

Exercise 2.17. Let S be a subset of $\mathbb{R}^m$. Show that

1) $\operatorname{span} S$ is a subspace of $\mathbb{R}^m$.

2) $\operatorname{span} S$ is the smallest subspace of $\mathbb{R}^m$ that contains S, i.e., $\operatorname{span} S$ is the *subspace completion* of S.

3) $\operatorname{span} S = S$ if and only if S is a subspace of $\mathbb{R}^m$.

Exercise 2.18. Let W be a non-empty subset of $\mathbb{R}^m$. Show that[5] W is a subspace of $\mathbb{R}^m$ if and only if for any u, v in W, $\operatorname{span}\{u, v\}$ is inside W. (In other words, a nonempty subset of $\mathbb{R}^m$ is a subspace of $\mathbb{R}^m$ if and only if it is invariant under scalar multiplication and addition.)

Exercise 2.19. Let W_1 and W_2 be two subspaces of $\mathbb{R}^n$. Let $W_1 + W_2$ be the following subset of $\mathbb{R}^n$:

$$W_1 + W_2 = \{u + v \mid u \in W_1, v \in W_2\}.$$

Show that $W_1 + W_2$ is a subspace of $\mathbb{R}^n$ and is the smallest subspace that contains both W_1 and W_2: if V is a subspace of $\mathbb{R}^n$ that contains both W_1 and W_2, and $V \subset W_1+W_2$, then $V = W_1+W_2$.

2.4 Minimal Spanning Sets and Dimension

Let W be a subspace of $\mathbb{R}^m$, S a finite set of vectors in $\mathbb{R}^m$. We say S is a **minimal spanning set** for W if W is the span of S but not the span of any proper subset of S. It is easy to see that minimal spanning set for W always exists. (You prove it)

Exercise 2.20. Show that a minimal spanning set is also a linearly independent set.

Claim 1. *Let S_1, S_2 be two minimal spanning sets for W, then $|S_1| = |S_2|$.*

Proof. The case $W = \{\vec{0}\}$ is trivial, where both S_1 and S_2 must be empty.

Otherwise, if we let S be either S_1 or S_2, then S is non-empty. Upon fixing an order on the spanning set S for W, we have a natural linear map from $\mathbb{R}^{|S|}$ to $\mathbb{R}^m$:

$$\boxed{T_S\colon \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix} \mapsto c_1\vec{a}_1 + c_2\vec{a}_2 + \cdots} \tag{2.10}$$

where $\vec{a}_i$'s are elements in S ordered by their subscripts.

Next we observe that T_S is one-to-one $\iff$ $\ker T_S$ is trivial $\iff$ S is a minimal spanning set for W. Therefore, linear map $T_{S_1}^{-1}T_{S_2}\colon \mathbb{R}^{|S_2|} \longrightarrow \mathbb{R}^{|S_1|}$ is both one-to-one and onto, so by Ex. 2.8 $|S_1| = |S_2|$. $\square$

By definition, the **dimension** of a subspace W, denoted by $\dim W$, is the cardinality of a minimal spanning set for W. The above claim says that this definition makes sense.

Suppose that $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear map. From the definition we know $\operatorname{Im}(T)$ is a subspace of $\mathbb{R}^m$. The dimension of this subspace is called the **rank** of T, denoted by $r(T)$.

[5] Hint: use the result of the previous exercise

Exercise 2.21. Show that $\dim W = \min |S|$, where the minimum is taken over the collection of all spanning set for W.

Exercise 2.22. Show that $\mathbb{R}^n$ is an n dimensional subspace of itself.

Exercise 2.23. Let A be an $m \times n$-matrix. Show that any two of the following three statements imply the third.
1) $m = n$.
2) The column space of A is equal to $\mathbb{R}^m$.
3) The null space of A is trivial.

Exercise 2.24. Let T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear map. Show that any two of the following three statements imply the third.
1) $m = n$.
2) T is onto.
3) T is one-to-one.

Exercise 2.25. Let S be a finite set of vectors in subspace W. Show that any two of the following three statements imply the third.
1) $\dim W = |S|$.
2) The span of S is equal to W.
3) S is a linearly independent set.

Exercise 2.26. Let A be an $m \times n$-matrix. The **rank** of A, denoted by $r(A)$, is defined to be the dimension of the column space. Of course, this is the rank of the corresponding linear map.
1) Show that[6] $\operatorname{Nul} A$ is a subspace of $\mathbb{R}^n$.
2) Show that the pivotal column of A is a minimal spanning set for $\operatorname{Col} A$.
3) Show that $r(A) \leq m$, $r(A) \leq n$, and

$$\boxed{r(A) + \dim \operatorname{Nul} A = n\,.} \tag{2.11}$$

Exercise 2.27. Let W be a subspace of $\mathbb{R}^n$, T be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Show that
1) $T(W)$ is a subspace of $\mathbb{R}^m$.
2) $\dim T(W) \leq \dim W$ and the equality holds when T is one-to-one.
3) T is one-to-one $\iff$ $\dim T(W) = \dim W$ for any subspace W of $\mathbb{R}^n$.

Exercise 2.28. Let T, C and R be three linear maps such that CTR is defined. Show that

$$r(CTR) = r(T)$$

provided that C is one-to-one and R is onto.

Exercise 2.29. Show that the rank of matrices is invariant under both the column and row operations.

[6]Hint: you can assume A is already in reduced row echelon form (why?). If we set the i-th non-pivotal variable be 1 and all other non-pivotal variables be zero, then we get a unique solution $\vec{x}_i$. Now you just need to show that the collection of such $\vec{x}_i$'s is a spanning set (in fact a minimal spanning set) for $\operatorname{Nul} A$.

Exercise 2.30. The **transpose** of a matrix is the matrix obtained from it by switching its columns with its rows. (Compare with proposition 5.2) For example, the transpose of

$$\begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

is

$$\begin{bmatrix} 1 & 7 \\ 2 & 8 \\ 3 & 9 \end{bmatrix}.$$

It is customary to use A' to denote the transpose of A. Show that
1) $(A')' = A$ and $(AB)' = B'A'$
2) $(A+B)' = A' + B'$ and $(cA)' = cA'$.
3) $r(A') = r(A)$.

Exercise 2.31. Let W_1 and W_2 be two subspaces of $\mathbb{R}^n$. Show that

$$\dim(W_1 + W_2) \le \dim W_1 + \dim W_2. \tag{2.12}$$

Exercise 2.32. Suppose that W_1 and W_2 are two subspaces of $\mathbb{R}^n$, and $W_1 \subset W_2$. Show that $\dim W_1 \le \dim W_2$ and the equality holds if and only if $W_1 = W_2$.

Exercise 2.33. Let A, B be two matrices such that AB is defined. Show that

$$\boxed{r(AB) \le r(B),\, r(A)}$$

and[7]

$$r(A+B) \le r(A) + r(B).$$

Exercise 2.34. From definition we know that the span of k vectors in $\mathbb{R}^n$ has dimension at most k. Show that k vectors in $\mathbb{R}^n$ are linearly independent $\iff$ they lie in a k-dimensional subspace of $\mathbb{R}^n$ but not in a lower dimensional subspace.

Exercise 2.35. Let W be a subspace of $\mathbb{R}^n$ with $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ being a minimal spanning set for W. Show that we can find additional vectors $\vec{v}_{k+1}, \ldots, \vec{v}_n$ in $\mathbb{R}^n$ such that $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a minimal spanning set for $\mathbb{R}^n$.

Let us end this section with

Theorem 2.1 (The Structure Theorem of Linear Systems). *Let $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear map. Then*

i) The solution set of the homogeneous linear system $T(\vec{x}) = \vec{0}$ is a linear subspace of $\mathbb{R}^n$ of dimension $n - r(T)$.

ii) The solution set of the nonhomogeneous linear system $T(\vec{x}) = \vec{b}$ ($\vec{b} \ne \vec{0}$) is either empty or an affine subspace of $\mathbb{R}^n$ of dimension $n - r(T)$. (In fact the affine subspace, if it exists, is a parallel translate of the linear subspace in i).)

Exercise 2.36. Prove this theorem.

[7] Hint: $\mathrm{Col}(A+B) \subset \mathrm{Col}(A) + \mathrm{Col}(B)$

2.5 Invertible Linear Maps and Invertible Matrices

Let $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear map. Recall that T is invertible means that T is both one-to-one and onto. By Ex. 2.8 we know that $m = n$ if T is invertible. It is clear that the inverse of linear maps are also invertible linear maps and the compositions of invertible linear maps are invertible linear maps.

On the computational side, we say a matrix A is invertible if the corresponding linear map T: $\vec{x} \mapsto A\vec{x}$ is invertible. The inverse of A, denoted by A^{-1}, is the coordinate matrix for T^{-1}. It is clear from the previous paragraph that $AA^{-1} = A^{-1}A = I$; also, invertible matrices must be square matrices and product of invertible matrices are invertible.

Given a square matrix A, how do we know it is invertible matrix? If A is invertible, how do we find A^{-1}. Once these questions are answered, the corresponding questions for linear maps are also answered.

Since A is a square matrix, based on some exercise problems we have done before, we observe that A is invertible $\iff$ $\operatorname{Nul} A$ is trivial $\iff$ $\operatorname{Nul} \bar{A}$ is trivial $\iff$ $\bar{A} = I$ $\iff$ A is a product of elementary matrices.

Exercise 2.37. Show that elementary matrices are invertible and their inverse are also elementary matrices.

Exercise 2.38. Show that $A \sim_r B \iff$ there is an invertible matrix E such that $A = EB$. Can you formulate the corresponding statement in the case when $\sim_r$ is replaced by either column equivalence or row-column equivalence?

Next, we observe that, if A^{-1} exists, then

$$\begin{aligned} [A \quad I] &\sim_r A^{-1}[A \quad I] \\ &= [A^{-1}A \quad A^{-1}I] \qquad \text{Eq. (2.9)} \\ &= [I \quad A^{-1}]. \end{aligned} \tag{2.13}$$

Since $[I \quad A^{-1}]$ is in reduced echelon row form (why?), we have the following algorithm:

Step 1. Form the $n \times 2n$-matrix $[A \quad I]$.

Step 2. Reduce $[A \quad I]$ to its reduced row echelon form $[\tilde{A} \quad \tilde{I}]$.

Step 3. If $\tilde{A} \neq I$, then A is not invertible. If $\tilde{A} = I$, then A is invertible and $A^{-1} = \tilde{I}$.

Example 2.2. Let

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

Then

$$\begin{aligned} [A \quad I] &= \begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 3 & 0 & 1 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & -1 & -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & -1 & -2 & 1 \end{bmatrix} \sim_r \begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & 2 & -1 \end{bmatrix} \end{aligned}$$

So A is invertible with

$$A^{-1} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}.$$

∎

Exercise 2.39. Let $T \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be this linear map:

$$T(x_1, x_2) = (x_1 + 2x_2, 2x_1 + 3x_2).$$

Is T invertible? If yes, find its inverse. (The computational part is already done in the example, you just need to translate the result.)

Exercise 2.40. Which of the following linear maps are invertible?

1. $T(x_1, x_2, x_3) = (x_2 + 7x_3, x_1 + 3x_2 - 2x_3)$;
2. $T(x_1, x_2, x_3) = (x_1 + 2x_3, 2x_1 - x_2 + 3x_3, 4x_1 + x_2 + 8x_3)$;
3. $T(x_1, x_2, x_3) = (x_1 + x_2 + x_3, x_1 + 2x_2, x_1 + 2x_3)$.

Chapter 3

General Vector Spaces(2.5 weeks)

3.1 Vector Spaces and Linear Maps

Intuitively we all know what a line and a plane are: they are a flat 1-dimensional space and a flat 2-dimensional space respectively. You may imagine what an n-dimensional flat space is. Suppose that E is an n-dimensional flat space, and suppose a point O in E is chosen once for all. Then the arrows in E with tail at O is in one-one correspondence with points in E. Now, addition and multiplication by real numbers are naturally defined for those arrows; therefore, addition and multiplication by real numbers are naturally defined for points in E via the one-one correspondence. Equipped with these addition and scalar multiplication, E is then something which will later be called a vector space — the **tangent space** of the flat space E at point O. Algebraically, upon a choice of a rectangular coordinate system on E with the coordinate origin at O, this corresponds to $\mathbb{R}^n$ and the addition and scalar multiplication that we have introduced early in section 2.1. In other word, even with the natural addition and scalar multiplication considered for both E and $\mathbb{R}^n$, E still looks like $\mathbb{R}^n$.

By definition, the triple $(\mathbb{R}^n, +, \cdot)$ — where $+$ and $\cdot$ mean the natural addition and scalar multiplication that we have introduced early in section 2.1 — is called the real n-dimensional **model vector space**. For simplicity, people often write $\mathbb{R}^n$ for $(\mathbb{R}^n, +, \cdot)$. What is a general n-dimensional real vector space? Well, it is just something that looks like the real n-dimensional model vector space $\mathbb{R}^n$. More precisely, we have [1]

Definition 3.1 (n-dimensional vector space). An n-dimensional real **vector space** is a triple $(V, +, \cdot)$ where V is a non-empty set, $+$ is a map from $V \times V$ to V (the addition), and $\cdot$ is a map from $\mathbb{R} \times V$ to V (the scalar multiplication); such that, as far as addition and scalar multiplication are concerned, $(V, +, \cdot)$ is just like[2] $\mathbb{R}^n$. To be precise, if we write the image of $(u, v) \in V \times V$ under $+$ as $u + v$ and the image of (c, v) under $\cdot$ as cv, this means that there is a one-one correspondence $T\colon V \longrightarrow \mathbb{R}^n$ such that

[1] In this section we only talk about finite dimensional vector spaces. The study of all interesting infinite dimensional vector spaces requires analysis and is a special area of analysis; it is not part of Linear Algebra where geometry rather than analysis is involved. The axiomatic definition of vector space (finite or infinite dimensional) is given later.

[2] We allow $n = 0$ and take $\mathbb{R}^0 = \{0\}$. The addition and scalar multiplication on $\mathbb{R}^0$ are of course unique.

1. $T(x+y) = T(x) + T(y)$ for all x, y in V,

2. $T(cx) = cT(x)$ for all $c \in \mathbb{R}$ and all x in V.

Remark 3.1. Roughly speaking, a vector space is just a set together with a **linear structure** on that set. Elements of V are called **vectors** in V. The element that corresponds to the zero vector $\vec{0}$ in $\mathbb{R}^n$ is called the **zero vector** of V, denoted by 0. It is clear that $0v = 0$ for any v in V because the corresponding thing is true when $V = \mathbb{R}^n$. We write $-v$ for $(-1)v$ for any v in V. $-v$ is called the **negative** of v. It is clear that $v + (-v) = 0$ for any v in V because the corresponding thing is true when $V = \mathbb{R}^n$. Note that the dimension of V is well-defined; otherwise, we would have a linear map $\mathbb{R}^m \longrightarrow \mathbb{R}^n$ which is one-one correspondence for some (m, n) with $m \neq n$.

Example 3.1. [3]Let W be a subspace of $\mathbb{R}^n$ and $\dim W = k$. Then we know that there is a minimal spanning set S for W. Once we fix an order on S, we can construct a linear map T_S: $\mathbb{R}^k \longrightarrow \mathbb{R}^n$(defined in (2.10)). The fact that S is minimal implies that T_S is one-to-one, so T_S is an one-one correspondence between $\mathbb{R}^k$ and W. The addition and scalar multiplication on W are that of $\mathbb{R}^n$ restricted to W, and they correspond to that of $\mathbb{R}^k$ under T_S. So W is a vector space of dimension k. In particular, $\mathbb{R}^n$ is a vector space of dimension n. In particular, if A is an $m \times n$-matrix, then $\operatorname{Col} A$ is a vector space of dimension $r(A)$ and $\operatorname{Nul} A$ is a vector space of dimension $n - r(A)$. ▮

Example 3.2. Let S be a finite set, $\mathbb{R}^S$ be the set of all real valued functions on S. Then $\mathbb{R}^S$ is a vector space of dimension $|S|$. ▮

Example 3.3. Let $\mathscr{P}_n$ be the set of all polynomials in t of degree at most n. Then $\mathscr{P}_n$ with the natural addition and scalar multiplication is an $(n+1)$-dimensional vector space. One may choose the one-one correspondence T: $\mathscr{P}_n \longrightarrow \mathbb{R}^{n+1}$ to be $T(a_0 + a_1 t + \cdots + c_n t^n) = [a_0, a_1, \ldots, a_n]'$. ▮

Example 3.4. Let $\mathcal{M}_{m\times n}$ be the set of all $m \times n$-real matrices. Then $\mathcal{M}_{m\times n}$ with the natural addition and scalar multiplication is an mn-dimensional vector space. One may choose the one-one correspondence T: $\mathcal{M}_{m\times n} \longrightarrow \mathbb{R}^{mn}$ as

$$T([a_{ij}]) = [a_{11}, a_{12}, \ldots, a_{1n},\ a_{21}, a_{22}, \ldots, a_{2n},\ \ldots,\ a_{m1}, a_{m2}, \ldots, a_{mn}]'.$$

▮

To get more examples of vector spaces, let us introduce linear maps. Suppose V and W are vector spaces of dimension n and m respectively, and T: $V \longrightarrow W$ is a map. We say T is a **linear map** if T respects the linear structures (i.e, additions and scalar multiplications) on V and W, i.e.,

1. $T(x+y) = T(x) + T(y)$ for all x, y in V,

2. $T(cx) = cT(x)$ for all $c \in \mathbb{R}$ and all x in V.

[3]Please review section 2.4 if you have difficulty with this example.

An invertible linear map is also called a **linear equivalence**. If $T\colon V \longrightarrow W$ is a linear equivalence, then we say V and W are linearly equivalent or **isomorphic** and we write this as $V \cong W$. Then it follows from the definition of vector space that $V \cong \mathbb{R}^{\dim V}$; i.e., in each dimension, there is only one vector space up to linear equivalence. With the help of Ex. 2.8, it is then easy to see that $V \cong W \iff \dim V = \dim W$.

Here comes an important remark: *the identification of a vector space with a model vector space enables us to solve problems in general vector spaces in term of matrix techniques. Statements and problems for general vector spaces are just translations of the corresponding ones for model vector spaces, hence can be proved or solved by dealing with column vectors and matrices.* From now on, if you have difficulty understanding or proving anything about general vector space and linear maps, just assume your vector spaces are the model vector spaces and the linear maps are matrices. The one-one correspondence in the definition of vector spaces allows you do the straightforward translation of anything involved.

The **kernel** of T, denoted by $\ker T$, is defined to be solution set of equation $T(x) = 0$. It is just a matter of translation of the results in example 3.1 that both the kernel and the image of T are vector spaces. The rank of T, denoted by $r(T)$, is defined to be the dimension of the image of T. It is just a matter of translation of Eq. (2.11) that

$$\boxed{\dim \ker T + r(T) = \dim V\,.} \tag{3.1}$$

Example 3.5. The set of all upper triangular 2×2-matrix is a vector space because it is $\ker T$ where $T\colon \mathcal{M}_{2\times 2} \longrightarrow \mathbb{R}$ is the linear map sending $[a_{ij}]$ to a_{21}. ▮

Example 3.6. The set of all even polynomials in t of degree at most n is a vector space. ▮

The notion of linear independence, linear combination, span, minimal spanning set and subspace, invertible map, can all be trivially translated to general vector spaces. For example, if V is a vector space, W is a non-empty set of V, then we say W is a **subspace** of V if $\operatorname{span} W = W$.

Exercise 3.1. Let V be a vector space, W be a subspace of V. Show that W is also a vector space.

Exercise 3.2. Let W be a nonempty set of V. Show that W is a subspace of V if and only if W satisfies this property: for any u, v in W, $\operatorname{span}\{u, v\} \subset W$.

Example 3.7. Let A be an $m \times n$-matrix. The rows of A are vectors in $\mathcal{M}_{1\times m}$. The **row space** of A, denoted by $\operatorname{Row} A$, is defined to be the span of rows of A. It is clear that $\operatorname{Row} A$ is a subspace of $\mathcal{M}_{1\times m}$. ▮

Example 3.8. Let V, W be two vector spaces, $\operatorname{Hom}(V, W)$ be the set of all linear maps from V to W. Then, it is just a matter of translation that $\operatorname{Hom}(V, W)$ is a vector space of dimension $\dim V \cdot \dim W$. $\operatorname{Hom}(V, \mathbb{R})$ is called the **dual space** of V, denoted by V^*. ▮

Exercise 3.3. Assume that W is a subspace of V. Show that $\dim W \le \dim V$ and the equality holds only when $W = V$.

Exercise 3.4. Let S and T be linear maps. Show that 1) $r(TS) \le r(T)$ and $r(TS) \le r(S)$ whenever TS makes sense; 2) $r(T+S) \le r(T) + r(S)$ whenever $T+S$ makes sense.

Exercise 3.5. Show that the row spaces are invariant under elementary row operations, but may change under elementary column operations. Likewise, show that the column spaces are invariant under elementary column operations, but may change under elementary row operations. How about the null space? Also, show that $\mathrm{Row}\, A \cong \mathrm{Col}\, A$.

Exercise 3.6. Let V be a vector space, W a subspace of V. Suppose that $\{v_1, v_2, \ldots, v_k\}$ is a minimal spanning set for W. Show that we can find additional vectors $v_{k+1}, \ldots, v_n$ in V such that $\{v_1, v_2, \ldots, v_n\}$ is a minimal spanning set for V.

Exercise 3.7. Let $(V, +, \cdot)$ be a vector space and $(W, +, \cdot)$ be a triple as in our definition of vector spaces. Suppose that $T\colon V \longrightarrow W$ is a set map which is one-to-one, onto and preserves $+$ and $\cdot$. Show that $(W, +, \cdot)$ is a vector space which is linearly equivalent to $(V, +, \cdot)$.

Exercise 3.8. Let V be a vector space, W a subspace of V. If $x \in V$, we use $x + W$ to denote $\{x + w \mid w \in W\}$. Let

$$V/W = \{x + W \mid x \in V\}.$$

Now the addition and scale multiplication on V induce one on V/W: $(x+W)+(y+W) := (x+y)+W$, $c(x+W) := cx+W$. Show that[4], with the addition and scale multiplication induced from V, V/W becomes a vector space of dimension equal to $(\dim V - \dim W)$, called the **quotient space** of V by W.

Exercise 3.9. Let V be a vector space and W a subspace of V. The map $p\colon V \longrightarrow V/W$ sending x to $x+W$ is called the projection map. p is a linear map which is zero when restricted to W. There are many linear maps from V to vector spaces such that when it is restricted to W it becomes zero. However, p is a universal one, that means that, if $T\colon V \longrightarrow Z$ is a linear map such that $T|_W = 0$, then there is a unique linear map $\bar{T}\colon V/W \longrightarrow Z$ such that the triangle

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "p"'] & Z \\ V/W \arrow[ur, "\bar{T}"'] & \end{tikzcd}$$

is commutative. Please prove this last statement.

Exercise 3.10. 1) Show that if Z is a subspace of W and W is a subspace of V, then Z is a subspace of V.

2) Let W_1 and W_2 be two subspaces of V. Show that the intersection $W_1 \cap W_2$ is a subspace of V, and

$$\dim(W_1 \cap W_2) \ \ge\ \dim W_1 + \dim W_2 - \dim V. \tag{3.2}$$

[4]Hint: use the previous two exercises.

Exercise 3.11. Let $S\colon V \longrightarrow W$ be a linear map, we know that there is a unique linear map $\bar{S}\colon V/\ker S \longrightarrow W$ such that the triangle

$$\begin{tikzcd} V \arrow[r, "S"] \arrow[d, "p"'] & W \\ V/\ker S \arrow[ur, "\bar{S}"'] & \end{tikzcd}$$

is commutative. Show that $\bar{S}$ has a linear left inverse — a left inverse which is also a linear map. (In general, any one-to-one linear map has a linear left inverse.)

Exercise 3.12. Let $T\colon U \longrightarrow V$ be a linear map, $f\colon V \longrightarrow Z$ be another linear map. Suppose that $fT = 0$, show that there is a unique linear map $\bar{f}\colon V/\mathrm{Im}T \longrightarrow Z$ such that the triangle

$$\begin{tikzcd} V \arrow[d, "f"'] \arrow[dr, "p"] & \\ Z & V/\mathrm{Im}T \arrow[l, "\bar{f}"] \end{tikzcd}$$

is commutative.

Exercise 3.13. Let $T\colon V \longrightarrow W$ be a linear map. If $f\colon W \longrightarrow \mathbb{R}$ is a linear map, then $T^*(f) := fT\colon V \longrightarrow \mathbb{R}$ is also a linear map. So T^* defines a map from W^* to V^*. Show that 1) T^* is a linear map; 2) $0^* = 0$, $I^* = I$, and $(TS)^* = S^*T^*$ if either side is defined.

Exercise 3.14. Let V be a vector space, V^{**} be the dual of V^*. For any $v \in V$, we denote by v^{**} the linear map from V^* to $\mathbb{R}$ which maps $f \in V^*$ to $f(v)$. Let $\iota_V\colon V \longrightarrow V^{**}$ be the map sending $v \in V$ to v^{**}. Show that ι_V is a linear equivalence.

Exercise 3.15. Suppose that $T\colon V \longrightarrow W$ is a linear map. Show that the following square

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\iota_V"'] & W \arrow[d, "\iota_W"] \\ V^{**} \arrow[r, "T^{**}"'] & W^{**} \end{tikzcd}$$

is a commutative.

Exercise 3.16. Let U, V and W be vector spaces. We say $U \xrightarrow{T} V \xrightarrow{S} W$ is **exact** at V if $\ker S = \mathrm{Im}\, T$. Show that 1) if $U \xrightarrow{T} V \xrightarrow{S} W$ is exact at V then $ST = 0$; 2) $U \xrightarrow{T} V \xrightarrow{S} W$ is exact at $V \Leftrightarrow W^* \xrightarrow{S^*} V^* \xrightarrow{T^*} U^*$ is exact at V^*.

Exercise 3.17. Let 0 be the 0-dimensional vector space. We say

$$0 \longrightarrow U \xrightarrow{T} V \xrightarrow{S} W \longrightarrow 0 \tag{3.3}$$

is a **short exact sequence** if it is exact at U, V and W, i.e., T is one-to-one, S is onto and $\ker S = \operatorname{Im} T$. Show that, 1) if (3.3) is a short exact sequence, then $\dim V = \dim U + \dim W$; 2) (3.3) is a short exact sequence if and only if

$$0 \longrightarrow W^* \xrightarrow{S^*} V^* \xrightarrow{T^*} U^* \longrightarrow 0$$

is a short exact sequence.

Exercise 3.18. Let $T\colon V \longrightarrow W$ be a linear map. The **cokernel** of T, denoted by $\operatorname{coker} T$, is the quotient space of W by the image of T. So T is onto means that the cokernel of T is trivial. Also, T is one-to-one means that the kernel of T is trivial.

Show that 1) $0 \longrightarrow \ker T \longrightarrow V \xrightarrow{T} W$ is exact at both $\ker T$ and V, and then conclude that $(\ker T)^* \cong \operatorname{coker} T^*$, 2) $V \xrightarrow{T} W \longrightarrow \operatorname{coker} T \longrightarrow 0$ is exact at both W and $\operatorname{coker} T$, and then conclude that $\ker T^* \cong (\operatorname{coker} T)^*$.

Exercise 3.19. Let $T\colon V \longrightarrow W$ be a linear map. Show that 1) T is one to one $\iff$ T^* is onto; 3) T is onto $\iff$ T^* is one-to-one; 3) $r(T) = r(T^*)$.

Exercise 3.20. Let $0 \longrightarrow W_i \xrightarrow{T_i} V_i$ be exact at W_i for $i = 1$ or 2. Suppose that

$$\begin{array}{ccc}
W_1 & \xrightarrow{T_1} & V_1 \\
{\scriptstyle\psi}\big\downarrow & & \big\downarrow{\scriptstyle\phi} \\
W_2 & \xrightarrow[T_2]{} & V_2
\end{array}$$

is commutative, then we have a map of short exact sequences:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & W_1 & \xrightarrow{T_1} & V_1 & \xrightarrow{p_1} & \operatorname{coker} T_1 & \longrightarrow & 0 \\
\big\downarrow & & {\scriptstyle\psi}\big\downarrow & & \big\downarrow{\scriptstyle\phi} & & \big\downarrow{\scriptstyle\bar\phi} & & \big\downarrow \\
0 & \longrightarrow & W_2 & \xrightarrow{T_2} & V_2 & \xrightarrow{p_2} & \operatorname{coker} T_2 & \longrightarrow & 0.
\end{array}$$

I.e., rows are exact and squares are commutative. Please prove this statement.

Exercise 3.21 (Injective Limit). Let V_1, V_2, ... be a sequence of nested subspaces of a (finite dimensional) vector space V:

$$V_1 \subset V_2 \subset \dots$$

Show that this sequence must stabilize somewhere, i.e., there is an integer $k \geq 1$ such that $V_k = V_{k+1} = \dots$. (This stabilized subspace is called the **injective limit** of this sequence, denoted by $\varinjlim V_i$.)

3.2 Axiomatic Definition of Vector Spaces

In (almost) all standard textbooks on Linear Algebra, the following definition for vector space is given.

Definition 3.2 (Axiomatic Definition of Vector Spaces). A real **vector space** is a triple $(V, +, \cdot)$ where V is a non-empty set of objects (called vectors), $+$ is an addition, and $\cdot$ is a multiplication by $\mathbb{R}$ (called scalar multiplication), and addition and scalar multiplication are subject to the following ten axioms. (The axioms must hold for all vectors u, v, and w in V and for all scalars c and d.)

1. The sum of u with v, denoted by $u + v$, is in V.
2. $u + v = v + u$.
3. $(u + v) + w = u + (v + w)$.
4. There is a **zero vector** 0 in V such that $u + 0 = u$.
5. For each u in V, there is a **negative** of u in V, denoted by $-u$, such that $u + (-u) = 0$.
6. The multiplication of u by c, denoted by cu, is in V.
7. $c(u + v) = cu + cv$.
8. $(c + d)u = cu + du$.
9. $c(du) = (cd)u$.
10. $1u = u$.

Exercise 3.22. From the axioms, show that there is a unique zero vector and the negative of any u is also unique, moreover, $0u = 0 = c0$ and $(-1)u = -u$.

From the axioms one can see that the concepts of linear combination, linear indepedence and span make sense. In the case V is the span of finitely many vectors in V, we say V is a finite dimensional real vector space; otherwise, we say V is an infinite dimensional real vector space.

If V is a vector space, W is a non-empty subset of V, then we say W is a **subspace** of V if $W = \mathrm{span} W$.

Example 3.9. The set of all real polynomials in t is an infinite dimensional real vector space.

Example 3.10. Let X be a set (could be infinite), V be a real vector space(could be infinite dimensional). Let V^X be the set of all set maps from X to V. The addition and scalar multiplication on V naturally gives rise to an addition and scalar multiplication on V^X, with which, V^X then becomes a real vector space. V^X is finite dimensional if and only if X is finite and V is finite dimensional.

Remark 3.2. Almost any interesting infinite dimensional vector space that you are going to meet in your lifetime is just a subspace of V^X for some set X and some finite dimensional vector space V. For example, the set of all continuous functions on interval $[0,1]$, the set of all smooth functions on interval $[0,1]$, the set of all continuous and piecewise smooth functions on $[0,1]$, the set of all bounded and piecewise continuous functions on $[0,1]$, etc., are all vector spaces because they are all subspaces of $\mathbb{R}^{[0,1]}$. So you don't need to check the ten axioms; instead, you just check whether a candidate for vector space is a subspace or not. By the way, the set of all real polynomials in t is a vector space because it is a subspace of $\mathbb{R}^{\mathbb{Z}_{\geq 0}}$ where $\mathbb{Z}_{\geq 0}$ is the set of nonnegative integers.

Exercise 3.23. Show that the two definitions of finite dimensional real vector spaces are equivalent.

Exercise 3.24. Let I be a set. Suppose that, for each $\alpha \in I$, we have a vector space V_α. Let $\prod_{\alpha\in I} V_\alpha$ be the Cartesian product of V_α's, i.e., the set of all I-sequences $\{x_\alpha\}_{\alpha\in I}$ with $x_\alpha \in V_\alpha$ for each $\alpha \in I$. Define the addition and scalar multiplication on $\prod_{\alpha\in I} V_\alpha$ as follows: $\{x_\alpha\} + \{y_\alpha\} = \{x_\alpha + y_\alpha\}$ and $c\{x_\alpha\} = \{cx_\alpha\}$. Show that $\prod_{\alpha\in I} V_\alpha$ then becomes a vector space — the so called **direct product** of V_α's.

Let $\oplus_{\alpha\in I} V_\alpha$ be the subset of $\prod_{\alpha\in I} V_\alpha$ consisting of all $\{x_\alpha\}_{\alpha\in I}$ with $x_\alpha = 0$ for all but finitely many $\alpha \in I$. Show that $\oplus_{\alpha\in I} V_\alpha$ is a subspace, so it is also a vector space — the so called **direct sum** of V_α's. (The direct product and the direct sum are equal if I is a finite set.)

3.3 Bases and Coordinates

Let V be an n-dimensional space, S be a minimal spanning set for V. (Such a S exists when $V = \mathbb{R}^n$, so exists for general V.) If we choose an order on S, we can have a linear map T_S: $\mathbb{R}^{|S|} \longrightarrow V$ defined by mapping $[x_1, x_2, \ldots]'$ into $x_1 v_1 + x_2 v_2 + \cdots$ where v_i's are all of the vectors in S ordered by their subindex. The fact that S is a minimal spanning set for V precisely implies that T_S is a linear equivalence, so the inverse of T_S[5] assigns a column vector to each vector in V, called the **coordinate vector** of the vector with respect to the ordered minimal spanning set. By definition, a **basis** for V is nothing but an ordered minimal spanning set.

Remark 3.3. $\mathbf{e} = (\vec{e}_1, \ldots, \vec{e}_n)$ is a basis for $\mathbb{R}^n$, called the standard basis for $\mathbb{R}^n$. If we have a linear equivalence T: $\mathbb{R}^n \longrightarrow V$, then the linear equivalence T will map $\mathbf{e}$ into $T(\mathbf{e}) = (T(\vec{e}_1), \ldots, T(\vec{e}_n))$ — a basis for V. It is clear that this assignment of a basis for V to a linear equivalence defines a one-one correspondence:

$$\boxed{\text{linear equivalences between } V \text{ and } \mathbb{R}^{\dim V} \leftrightarrow \text{bases for } V\ .}$$

And a linear equivalence $V \longrightarrow \mathbb{R}^n$ is just a coordinate map with respect to the corresponding basis.

Example 3.11. Let $\mathscr{P}_2$ be the set of polynomials in t of degree at most 2. $S = \{1, 1+t, 1+t^2\}$. Then S is a minimal spanning set for $\mathscr{P}_2$. Choose an order on S, say $p_1 = 1+t$, $p_2 = 1+t^2$, $p_3 = 1$. Then $\mathbf{p} = (p_1, p_2, p_3)$ is a basis for $\mathscr{P}_2$. Now what is the coordinate vector for $1 + 2t + 3t^2$

[5] T_S is called the parametrization map and its inverse is called the coordinate map.

with respect to $\mathbf{p}$? One way to do this problem is to use the natural equivalence of $\mathscr{P}_2$ with $\mathbb{R}^3$: $a_0+a_1t+a_2t^2 \leftrightarrow [a_0, a_1, a_2]'$ and translate the problem into a problem in $\mathbb{R}^3$: what is the coordinate vector for $[1, 2, 3]'$ with respect to basis $([1, 1, 0]', [1, 0, 1]', [1, 0, 0)]')$? But this is just this question: what is the unique solution of equation

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} ?$$

By doing computation with matrices, we get the solution $\vec{x} = [2, 3, -4]'$. You can check that indeed

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

∎

Example 3.12. Let V be a vector space with basis $\mathbf{v} = (v_1, \ldots, v_n)$. For each i, let $\hat{v}_i$ be the element in V^* such that $\hat{v}_i(v_j) = 1$ if $i = j$ and $= 0$ if $i \neq j$. Then $\hat{\mathbf{v}} = (\hat{v}_1, \ldots, \hat{v}_n)$ is a basis for V^*. Such a basis for V^* is called a **dual basis** of $\mathbf{v}$. It is then clear that $V^* \cong V$ because $\dim V^* = \dim V$. ∎

Exercise 3.25. Show that[6] any basis for V^* is a dual basis.

Exercise 3.26. Let $\vec{v}_1$, $\vec{v}_2$, $\ldots$, $\vec{v}_n$ be n elements of $\mathbb{R}^n$. Then the following statements are all equivalent.

1. $(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n)$ is a basis for $\mathbb{R}^n$.

2. $\vec{v}_1$, $\vec{v}_2$, $\ldots$, $\vec{v}_n$ are linearly independent.

3. $\vec{v}_1$, $\vec{v}_2$, $\ldots$, $\vec{v}_n$ span $\mathbb{R}^n$.

4. $[\vec{v}_1, \ldots, \vec{v}_n]$ is an invertible matrix.

Exercise 3.27. Let A be a matrix. Show that, 1) the pivotal columns of A form a basis for $\operatorname{Col} A$; 2) the nonzero rows of $\bar{A}$ form a basis for $\operatorname{Row} A$; 3) the special solutions x_i's described in the footnote attached to Ex. 2.26 form a basis for $\operatorname{Nul} A$.

Exercise 3.28. Consider the matrix

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

(a) What is the rank of A? (b) Find a basis for $\operatorname{Nul} A$, a basis for $\operatorname{Col} A$ and a basis for $\operatorname{Row} A$.

[6] Hint: you may use exercise 3.14.

Exercise 3.29. Consider polynomials

$$p_1 = 1 + t - t^3, \quad p_2 = 2t - t^2 - t^3, \quad p_3 = 1 - t + t^2, \quad p_4 = 2 + t^2 - t^3.$$

(a) Find a basis for the subspace $\operatorname{span}\{p_1, p_2, p_3, p_4\}$. (b) Find the coordinate vectors of p_1, p_2, p_3, p_4 with respect to the basis. (c) Extend the basis to a basis for $\mathscr{P}_3$ — the vector space of polynomials of degree ≤ 3.

Exercise 3.30. Let W be the span of the following four matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}.$$

Find a basis for W and compute the dimension of W.

Coordinate Matrices for Linear Maps

Let $T\colon V \longrightarrow W$ be a linear map. Suppose that we have a basis $\mathbf{v}$ chosen for V and a basis $\mathbf{w}$ chosen for W. Then there is a one-one correspondence between $\operatorname{Hom}(V, W)$ and $\operatorname{Hom}(\mathbb{R}^n, \mathbb{R}^m)$. Also, there is a one-one correspondence between $\operatorname{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ and the set of all $m \times n$-matrices. Therefore we have an assignment of an $m \times n$-matrix A to each linear map T from V to W. Of course, A is the unique matrix such that

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \\ {\scriptstyle T_{\mathbf{v}}}\downarrow & & \downarrow{\scriptstyle T_{\mathbf{w}}} \\ V & \xrightarrow[T]{} & W \end{array}$$

is commutative. (Here the vertical maps are the inverse of the coordinate maps and the bottom map is the multiplication by A.) By definition, A is called the **coordinate matrix** for T with respect to bases $\mathbf{v}$ and $\mathbf{w}$.

When $W = V$ and T is the identity map on V, A is called the **coordinate change matrix**: Given a vector in V, it has a coordinate vector $\vec{x}$ with respect to basis $\mathbf{v}$, it also has a coordinate vector $\vec{y}$ with respect to basis $\mathbf{w}$, then $\vec{y} = A\vec{x}$. We know that there is an invertible matrix $P = [p^i_j]$ (here i is the row index and j is the column index) such that $[v_1, \ldots, v_n]P = [w_1, \ldots, w_n]$, i.e., $\sum_i v_i {p^i}_j = w_j$, then from the identity $[v_1, \ldots, v_n]\vec{x} = [w_1, \ldots, w_n]\vec{y}$ we have $\vec{x} = P\vec{y}$. So $A = P^{-1}$.

Exercise 3.31. Let T be an endomorphism[7] on V, $\mathbf{v}$ and $\mathbf{w}$ be two bases for V and $[v_1, \ldots, v_n]P = [w_1, \ldots, w_n]$. Suppose that the coordinate matrix for T with respect to $\mathbf{v}$ is A and the coordinate matrix for T with respect to $\mathbf{w}$ is B. Show that $A = PBP^{-1}$. (Two matrices A and B are related to each other this way are called **similar**. It is clear that similar matrices are just different coordinate matrices for the same endomorphism.)

[7] I.e., T is a linear map from V to V.

3.4 Direct Sums and Block Matrices

Let V be a vector space, W_1 and W_2 be subspaces of V. The **sum** of W_1 with W_2, denoted by $W_1 + W_2$, is defined to be

$$\{v_1 + v_2 \mid v_1 \in W_1,\ v_2 \in W_2\}.$$

It is easy to see that $W_1 + W_2$ is a subspace of V, in fact the smallest subspace containing both W_1 and W_2. Sum of a finitely many subspaces W_1, ..., W_k are similarly defined, denoted by $W_1 + W_2 + W_3 + \cdots + W_k$.

If $W_1 \cap W_2$ is the trivial subspace of V or equivalently if

$$\{(x_1, x_2) \in W_1 \times W_2 \mid x_1 + x_2 = 0\}$$

is trivial (i.e, has $(0, 0)$ as its only element), then $W_1 + W_2$ is called the (internal) **direct sum** of W_1 with W_2, denoted by $W_1 \oplus_i W_2$. In general, if

$$\{(x_1, \ldots, x_k) \in W_1 \times \cdots \times W_k \mid x_1 + \cdots + x_k = 0\}$$

is trivial, i.e., there is only one way to expresses 0 as a sum of k vectors, one from each W_i, then $W_1 + \cdots + W_k$ is called the (internal) direct sum of W_1, ..., W_k, denoted by $W_1 \oplus_i \cdots \oplus_i W_k$.

Remark 3.4. It is convenient to introduce the direct minus sign $\ominus$: If $V = W_1 \oplus_i W_2$, we write $W_1 = V \ominus W_2$ and $W_2 = V \ominus W_1$.

Example 3.13. Take $V = \mathbb{R}^2$, $W_1 = \operatorname{span}\{\vec{e}_1\}$ and $W_2 = \operatorname{span}\{\vec{e}_1 + \vec{e}_2\}$. Then $W_1 + W_1 = W_1$, $W_1 + V = V$, $W_1 + W_2 = W_1 \oplus_i W_2 = V$. ∎

Example 3.14. Let W be a subspace of V. Then $W + W = W$, $W + V = V$ and $\{0\} + W = \{0\} \oplus_i W = W$. Also, if W_1, W_2 and W_3 are subspace of V, then $W_1 + W_2 + W_3 = (W_1 + W_2) + W_3 = W_1 + (W_2 + W_3)$. Also $W_1 + W_2 = W_2 + W_1$. ∎

Example 3.15. Suppose that v_1, ..., v_n form a basis for V. Let V_i be the span of v_i. Then $V = V_1 \oplus_i \cdots \oplus_i V_n$.

∎

Example 3.16. Let $\mathbb{R}^m \times 0$ be the subspace of $\mathbb{R}^{m+n}$ consisting of the column vectors whose last n entries are zero, and $0 \times \mathbb{R}^n$ be the subspace of $\mathbb{R}^{m+n}$ consisting of the column vectors whose first m entries are zero. Then $\mathbb{R}^m \times 0 \oplus_i 0 \times \mathbb{R}^n = \mathbb{R}^{m+n}$. ∎

Example 3.17. Let T be an invertible map from V to itself, and $V = V_1 \oplus_i V_2$. Then $V = T(V) = T(V_1) \oplus_i T(V_2)$. Also $V/V_1 \cong V_2$. ∎

Exercise 3.32. Suppose that $V = V_1 \oplus_i \cdots \oplus_i V_k$.

1) If $S_i \subset V_i$ $(1 \le i \le k)$ is a linearly independent set of vectors, show that the (disjoint) union of S_i's is a linearly independent set of vectors.

2) Pick a basis for each V_i. Show that the juxtaposition of these bases is a basis for V.

Exercise 3.33. Let W_i's be subspaces of a vector space. Show that

$$\dim(W_1 + \cdots + W_k) \leq \sum_i \dim W_i$$

with equality holds if and only if the sum is actually a direct sum.

Let V, W be vector spaces of dimension n and m respectively. The (external) direct sum of V with W, denoted by $V \oplus_e W$, is this triple: $(V \times W, +, \cdot)$, where addition rule is $(v, w) + (v', w') = (v + v', w + w')$ and the scalar multiplication rule is $c(v, w) = (cv, cw)$. It is easy to see that $\mathbb{R}^n \oplus_e \mathbb{R}^m = \mathbb{R}^{m+n}$ and $\mathbb{R}^m \times 0 \oplus_e 0 \times \mathbb{R}^n \cong \mathbb{R}^{m+n}$ (naturally). The (external) direct sum of a finite number of vector spaces is similarly defined. It is easy to see that $\mathbb{R} \oplus_e \mathbb{R}^2 \oplus_e \mathbb{R} = \mathbb{R}^4$.

Exercise 3.34. Let V and W be two vector spaces, T: $V \longrightarrow W$ a set map. Show that T is a linear map if and only if $\Gamma(T)$ is a subspace of $V \oplus_e W$.

Let $V = V_1 \oplus_i V_2$ and $W = W_1 \oplus_i W_2$, T be a linear map from V to W. With respect to the decomposition of V and W, we have a **block decomposition** of T:

$$T = \left[\begin{array}{c|c} T_{11} & T_{12} \\ \hline T_{21} & T_{22} \end{array}\right]$$

where T_{ij} is a linear map from V_j to W_i.

From Ex. 1.13 we know that if A is an $m \times n$-matrix with rank r, then the reduced form of A looks like this:

$$\left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right]$$

where I_r is the $r \times r$-identity matrix and O's are zero matrices. Therefore, by Ex. 2.38, there are invertible matrices E and F such that

$$EAF = \left[\begin{array}{c|c} I_r & O \\ \hline O & O \end{array}\right].$$

It is then just a matter of translation that, if T: $V \longrightarrow W$ is a linear map of rank r, then there are decompositions $V = V_1 \oplus_i V_2$ and $W = W_1 \oplus_i W_2$, with respect to which, T has this decomposition:

$$T = \left[\begin{array}{c|c} T_r & O \\ \hline O & O \end{array}\right] \tag{3.4}$$

where T_r is a linear equivalence and the O's are zero maps.

Exercise 3.35. Let W be a subspace of V. Show that[8] there is a subspace W' of V such that $V = W \oplus_i W'$.

Exercise 3.36. Let T: $V \longrightarrow W$ be a linear map of rank r. Show that there is a subspace V_1 of V and a subspace W_2 of W, such that $V = V_1 \oplus_i \ker T$ and $W = \mathrm{Im}(T) \oplus_i W_2$. So $V/\ker T \cong V_1 \cong \mathrm{Im}(T)$.

[8]Hint: use (3.4) with T being the inclusion map: $W \longrightarrow V$.

If X, Y and Z are three vector spaces, S is a linear map from X to Y and T is a linear map from Y to Z. Suppose that $X = X_1 \oplus_i X_2$, $Y = Y_1 \oplus_i Y_2$, $Z = Z_1 \oplus_i Z_2$. Then the block decomposition of TS is

$$TS = \left[\begin{array}{c|c} T_{11}S_{11} + T_{12}S_{21} & T_{11}S_{12} + T_{12}S_{22} \\ \hline T_{21}S_{11} + T_{22}S_{21} & T_{21}S_{12} + T_{22}S_{22} \end{array}\right]$$

and this gives the multiplication rule for block matrices:

$$\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right] = \left[\begin{array}{c|c} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ \hline A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{array}\right].$$

In general we can decompose V and W into many pieces, then T will be decomposed into many blocks. Can you derive the multiplication rule in this general case?

Let $T\colon V \longrightarrow W$ be a linear map. Suppose that V and W can each be decomposed into the direct sum of k subspaces, and with respect to which, we have

$$T = \begin{bmatrix} T_1 & & \\ & \ddots & \\ & & T_k \end{bmatrix}$$

where the off-diagonal blocks are all zero maps, then we say that T is the direct sum of T_i's and write $T = T_1 \oplus T_2 \oplus \cdots \oplus T_k$, or $T = \oplus_{i=1}^k T_i$.

3.5 Nilpotent Endomorphisms*

An **endomorphism** is just a linear map from a vector space to itself. In general it could be very complicated. However, endomorphisms of complex vector spaces are reasonably simple, they are direct sums of basic building blocks[9], each of which is of this form: a scalar multiplication + a nilpotent endomorphism.

What is a nilpotent endomorphism? Well, an endomorphism N is said to be **nilpotent** if $N^k = 0$ for an integer $k \geq 1$.

Let $V(i) = \ker N^i$ for $i \geq 0$ (by convention $N^0 = I$). Then $V(i-1) \subset V(i)$ because if $N^{i-1}(v) = 0$ then $N^i(v) = 0$. Here is our key observation:

***N* maps $V(i)$ to $V(i-1)$; passing to the quotient spaces, it actually induces a one-to-one linear map $\overline{N}$: $V(i)/V(i-1) \longrightarrow V(i-1)/V(i-2)$.**

Exercise 3.37. 1) Prove the key observation and then conclude that $N^n = 0$ if N is a nilpotent endomorphism on an n-dimensional vector space.

2) Suppose that $V = W \oplus_i V_3$ and $W = V_1 \oplus_i V_2$ then $V = V_1 \oplus_i V_2 \oplus_i V_3$.

3) Suppose that $N\colon V \longrightarrow W$ is injective and $V = V_1 \oplus_i V_2$, then $N(V) = N(V_1) \oplus_i N(V_2)$.

[9]We will prove this statement in chapter 6.

Lemma 3.1. *Let N be an endomorphism such that $N^k = 0$ and $N^{k-1} \neq 0$ for some $k \geq 1$. Then,*

1) for each $1 \leq j \leq k$, there is a nontrivial subspace V_j of $V(j)$ such that N maps V_j into V_{j-1} injectively for each $2 \leq j \leq k$ and

$$V(j) = V_1 \oplus_i V_2 \oplus_i \cdots \oplus_i V_j \;;$$

2) for each $1 \leq j \leq k$, there is a subspace W_j (could be trivial) of V_j such that

$$V_j = N^{k-j}(W_k) \oplus_i N^{k-j-1}(W_{k-1}) \oplus_i \cdots \oplus_i W_j \;.$$

Proof. 1) Since $V(j-1) \subset V(j)$ for $j \geq 1$, by Ex. 3.35, we can pick a subspace V_j of $V(j)$ such that $V(j) = V(j-1) \oplus_i V_j$ for each $j \geq 1$, so by the 2nd part of the last exercise, we have $V(j) = V_1 \oplus_i V_2 \oplus_i \cdots \oplus_i V_j$. Note that, for each $j \geq 2$, $x \in V_j$ and $x \neq 0 \iff N^j(x) = 0$ and $N^{j-1}(x) \neq 0 \iff N^{j-1}(N(x)) = 0$ and $N^{j-2}(N(x)) \neq 0 \iff N(x) \in V_{j-1}$ and $N(x) \neq 0$; therefore N maps V_j into V_{j-1} injectively for each $j \geq 2$; consequently, $\dim V_j \geq \dim V_{j+1} \geq \cdots \geq \dim V_k$ if $1 \leq j \leq k$. Since $N^{k-1} \neq 0$, V_k must be nontrivial, then $\dim V_j > 0$, i.e., V_j is nontrivial for $1 \leq j \leq k$.

2) Since $N(V_{j+1}) \subset V_j$ for $j \geq 1$, by Ex. 3.35, we can pick a subspace W_j of V_j such that $V_j = N(V_{j+1}) \oplus_i W_j$ for each $j \geq 1$, so by the 2nd and the 3rd parts of the last exercise, we have $V_j = N^{k-j}(W_k) \oplus_i N^{k-j-1}(W_{k-1}) \oplus_i \cdots \oplus_i W_j$ for each $j \geq 1$. □

Since $V = V(k)$, in view of lemma 3.1 and the 2nd part of the last exercise, we have

$$\begin{array}{rcllll}
V & = & W_k & & & = & V_k \\
 & & \oplus_i N(W_k) & \oplus_i W_{k-1} & & & \oplus_i V_{k-1} \\
 & & \oplus_i N^2(W_k) & \oplus_i N(W_{k-1}) \oplus_i W_{k-2} & & & \oplus_i V_{k-2} \\
 & & \vdots & & & & \vdots \\
 & & \oplus_i N^{k-1}(W_k) \oplus_i & \cdots & \oplus_i W_1 & & \oplus_i V_1 \\
 & = & \multicolumn{3}{l}{\boxed{e^N(W_k) \quad \oplus_i \quad \cdots \quad \oplus_i e^N(W_1).}} & &
\end{array} \tag{3.5}$$

Here, by definition, $e^N(W_i) = W_i \oplus_i N(W_i) \oplus_i N^2(W_i) \oplus_i \cdots \oplus_i N^{i-1}(W_i)$. Note that N maps each $e^N(W_i)$ into itself, so if we let N_i be N restricted to $e^N(W_i)$, then

$$\boxed{N = N_1 \oplus N_2 \oplus \cdots \oplus N_k.} \tag{3.6}$$

Corollary 3.1. *1. $N^j|_{W_i}$ is one-to-one if $j \leq i-1$. Therefore, $\boxed{\dim e^N(W_i) = i \dim W_i}$, and*

$$\boxed{\dim V = \sum_{i=1}^k i \dim W_i.} \tag{3.7}$$

2. $\dim W_k \geq 1$, therefore, $k \leq \dim V$ and

$$\boxed{N^{\dim V} = 0.} \tag{3.8}$$

3. $\dim V_1 = 1 \iff k = \dim V$.

Lemma 3.2. *Assume* $\dim W_i \geq 1$. *Then* $\dim W_i = 1 \iff N_i$ *is* **indecomposable**, *i.e., not a direct sum of endomorphisms on non-trivial subspaces.*

Proof. Suppose that $\dim W_i = m > 1$. Let $u_1, \ldots, u_m$ be a basis for W_i. Then N_i is a direct sum with respect to this decomposition:

$$e^N(W) = \operatorname{span}\{u_1, N(u_1), \ldots, \} \oplus_i \operatorname{span}\{u_2, N(u_2), \ldots, \} \oplus_i \cdots.$$

Suppose that $\dim W_i = 1$, so $\dim e^N(W_i) = i$. If $v \in W_i$ and $v \neq 0$, then $v_i := v$, $v_{i-1} := N(v)$, $\ldots$, $v_1 := N^{i-1}(v)$ form a basis for $e^N(W_i)$ such that $N(v_1) = 0$, $N(v_2) = v_1$, $\ldots$, $N(v_i) = v_{i-1}$. In particular, $N^{i-1}(v_i) = v_1 \neq 0$, so $(N_i)^{i-1} \neq 0$. But that implies that N_i is indecomposable; otherwise, say $N_i = L \oplus M$, where L and M are endomorphisms on proper subspaces; then both L and M are nilpotent, and $(N_i)^{i-1} = L^{i-1} \oplus M^{i-1} = 0$ by applying Eq. (3.8) to both L and M, a contradiction! ☐

Let $n = \dim V$, in view of Eq. (3.8) and Eq. (3.5) we have

$$\begin{aligned} N^{n-1} \neq 0 &\iff V(n-1) \subsetneq V(n) \iff \dim W_n \geq 1 \\ &\iff \dim W_i = 0 \text{ if } i < n \text{ and } \dim W_n = 1 \qquad \text{use Eq. (3.7)} \\ &\iff N \text{ is indecomposable} \qquad \text{use Lemma 3.2 and Eq. (3.6).} \end{aligned}$$

In summary, we have proved the following

Theorem 3.2. *Let N be a nilpotent endomorphism of V, and $n = \dim V$. Then*

1. $N^n = 0$;
2. *If N is indecomposable, then there is a basis $\mathbf{v} = (v_1, \ldots, v_n)$ for V such that*

$$\begin{aligned} N(v_1) &= 0 \\ N(v_2) &= v_1 \\ &\vdots \\ N(v_n) &= v_{n-1}; \end{aligned} \tag{3.9}$$

3. *In general $N = N_1 \oplus N_2 \oplus \cdots$ where each N_i is an indecomposable nilpotent endomorphism.*
4. $r(N) = n - 1 \iff N^{n-1} \neq 0 \iff N$ *is indecomposable;*

Corollary 3.3. *Let N be an indecomposable nilpotent $n \times n$-matrix, then there is an invertible matrix P such that $N = PJ_nP^{-1}$ where*

$$J_n = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

In general, if N is a nilpotent matrix, then there is an invertible matrix P and positive integers $n_1, \ldots, n_l$ (not necessarily distinct) such that $N = P(J_{n_1} \oplus \cdots \oplus J_{n_l})P^{-1}$. Moreover, if we require $n_1 \geq n_2 \geq \cdots \geq n_l$, then $(n_1, n_2, \ldots, n_l)$ is uniquely determined by N.

Remark 3.5. For any number λ, we call $\lambda I + J_n$ a **Jordan block**. In the preceding corollary, there are l Jordan blocks in N.

Proof. Suppose that N is indecomposable. Let $P = [v_1, \ldots, v_n]$ where $v_1, \ldots, v_n$ are chosen as in point 3 of the theorem. Then $NP = [Nv_1, \ldots, Nv_n] = [0, v_1, \ldots, v_{n-1}] = PJ_n$, so $N = PJ_nP^{-1}$. The general case is also easy. □

Exercise 3.38. Let

$$N = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}.$$

It is clear that $r(N) = 2$. Find an invertible matrix P such that[10]

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

How about N^2?

Exercise 3.39. Let D: $\mathscr{P}_n \longrightarrow \mathscr{P}_n$ be the linear map which maps t^k to kt^{k-1} for any $0 \le k \le n$. Show that D is nilpotent and indecomposable.

Exercise 3.40. Let N be a nilpotent endomorphism on an n-dimensional vector space. Show that the number of Jordan blocks in N is $n - r(N)$.

To introduce the next exercise, we need to introduce the notion of **Young diagrams**. Roughly speaking, a Young diagram is a rectangular array of squares where some bottom right squares might be removed. Note that, in a Young diagram, the squares in each row must be adjacent to each other and the number of squares in the rows must be monotone decreasing when the row runs from top to bottom; consequently, the squares in each column must be adjacent to each other and the number of squares in the columns must be monotone decreasing when the column runs from left to right. Here are two Young diagrams:

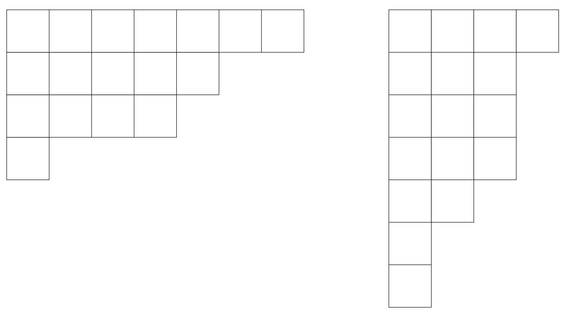

Figure 3.1: A Young diagram and its dual.

[10]Choose a $v_3 \notin \ker N^2$, and let $P = [N^2(v_3), N(v_3), v_3]$.

Note also that, a Young diagram determines a monotone decreasing finite sequence of positive integers n_1, n_2, …, n_l, where for each $1 \leq i \leq l$, n_i is the number of squares in the i-th row. In fact, this defines a one-one correspondence:

$$\boxed{\{\text{Young diagrams}\} \leftrightarrow \{\text{monotone decreasing finite sequences of positive integers}\}.} \tag{3.10}$$

Finally, each Young diagram has its dual Young diagram, this is like the fact that each matrix has its transpose. More precisely, given a Young diagram D, let m_i be the number of squares in the i-th *column*, then m_1, m_2, … is a monotone decreasing finite sequences of positive integers. Under the correspondence 3.10, $(m_1, m_2, \ldots)$ defines a Young diagram D' — the dual of D. For example, the two Young diagrams in figure 3.1 are dual of each other.

Exercise 3.41. 1) Let D be a Young diagram, show that $(D')' = D$.

2) Let N be a nilpotent endomorphism. For $i \geq 1$, we let $m_i = \dim\ker N^i - \dim\ker N^{i-1} = r(N^{i-1}) - r(N^i)$. The key observation in this section says that infinite sequence m_1, m_2, … is monotone decreasing; in fact, it is a monotone decreasing finite sequence of positive integers if we ignore the zeros. On the other hand, there is another monotone decreasing finite sequence of positive integers associated with N, it is n_1, …, n_l in corollary 3.3. Show that, under the correspondence 3.10, the Young diagrams determined by these two monotone decreasing finite sequence of positive integers are dual of each other. For example, suppose that for some nilpotent endomorphism N, we have $m_1 = 4$, $m_2 = m_3 = m_4 = 3$, $m_5 = 2$, $m_6 = 1$ and $m_7 = 1$ and $m_i = 0$ if $i > 7$, under the correspondence 3.10, we have the Young diagram on the right in figure 3.1. From its dual on the left in figure 3.1, we can read off $n_1 = 7$, $n_2 = 5$, $n_3 = 4$ and $n_4 = 1$ and $n_j = 0$ when $j > 4$. Therefore, N is similar to $J_7 \oplus J_5 \oplus J_4 \oplus J_1$.

Part II

Intermediate Part

Chapter 4

Determinant and Diagonalization(2 weeks)

This chapter contains materials that have found many applications in mathematics, science, engineering, economics, ... That is why every college student needs to learn Linear Algebra.

We start with the notion of Determinant. By emphasizing the algebraic and geometric origin of this basic notion, I hope that any mystery about its computational definition will disappear and its many properties will become transparent.

The problem of decomposing an endomorphism into a direct sum of simple pieces is a key step in simplifying and clarifying problems involving endomorphisms. In the simplest case, each simple piece is just a scalar multiplication. In general, these simple pieces are Jordan Blocks which are still manageable. This problem is completely solvable by studying eigenvalues, eigenspaces and generalized eigenspaces.

4.1 Multi-linear Antisymmetric maps

Let V be a vector space of dimension n. Consider a map $f\colon V^n \longrightarrow W$ where W is a vector space. We say f is **multi-linear** if it is linear in each variable, i.e., for each i,

$$f(\ldots, cv_i + c'v_i', \ldots) = cf(\ldots, v_i, \ldots) + c'f(\ldots, v_i', \ldots).$$

We say f is **antisymmetric** if for each pair (i, j) with $i < j$, we have

$$f(\ldots, v_i, \ldots, v_j, \ldots) = -f(\ldots, v_j, \ldots, v_i, \ldots).$$

Proposition 4.1. *Let V and W be vector spaces. Suppose that $\dim V = n$. Let $D(V, W)$ be the set of all multi-linear and antisymmetric maps from V^n to W. Then $D(V, W)$ is a vector space which is linearly equivalent to W.*

Proof. First of all, the addition and scalar multiplication is naturally defined on $D(V, W)$. Next, we observe that if $\mathbf{v} = (v_1, \ldots, v_n)$ is a basis for V, any $f \in D(V, W)$ is uniquely determined by its value at $\mathbf{v}$ (i.e., $f(v_1, \ldots, v_n)$), so the map sending $f \in D(V, W)$ to $f(\mathbf{v}) \in W$ is a one-one correspondence; moreover, this map respects the additions and scalar multiplications. The rest is clear. ☐

Remark 4.1. Let V be an n-dimensional vector space, V^* be the dual vector space of V. Then the set of all multi-linear antisymmetric maps from $(V^*)^n$ to the ground field K is a one dimensional vector space, it is called the **determinant line** of V, and is denoted by $\det V$.

To continue, a digression on **permutation groups** is needed. Let $\mathbf{n}$ be a set of n objects, to be specific, say $\mathbf{n} = \{1, 2, \ldots, n\}$ — the set of first n natural numbers. Let Σ_n be the set of all permutations of n objects, i.e., the set of all invertible maps from $\mathbf{n}$ to itself. (Officially we say Σ_n is the automorphism group of $\mathbf{n}$.) From high school mathematics we know $|\Sigma_n| = n!$. For each $1 \le i \le n-1$, we let $\sigma_i \in \Sigma_n$ be the permutation that exchanges i with $i+1$ and leaves the other elements of $\mathbf{n}$ fixed, i.e., $\sigma_i(i) = i+1$, $\sigma_i(i+1) = i$, and $\sigma_i(j) = j$ if $j \neq i$ and $\neq i+1$. It is not hard to see that

$$\boxed{\begin{array}{cl} \sigma_i\sigma_j = \sigma_j\sigma_i & \text{for each pair } (i,j) \text{ with } |i-j| > 1 \\ \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1} & \text{for each } i < n-1 \\ \sigma_i^2 = I & \text{for each } i. \end{array}} \tag{4.1}$$

Here I is the identity map and $\sigma_i\sigma_j$ is the composition of σ_i and σ_j.

Theorem 4.1. *1) $\sigma \in \Sigma_n \iff \sigma$ is a composition of finitely many $\sigma_1, \ldots, \sigma_{n-1}$.*

2) Two compositions of finitely many $\sigma_1, \ldots, \sigma_{n-1}$ are the same element of $\Sigma_n \iff$ one can be turned into the other by applying identities in Eq. (4.1) finitely many times.

For example, in Σ_3, $I = \sigma_1\sigma_1$; the map sending $(1,2,3)$ to $(3,1,2)$ is $\sigma_1\sigma_2$; the map sending $(1,2,3)$ to $(3,2,1)$ is $\sigma_1\sigma_2\sigma_1$, etc.

Remark 4.2. Officially, we state this theorem by saying that Σ_n is generated by σ_i's subject to the identities in Eq. (4.1). σ_i's are called the **generators**, and the identities in Eq. (4.1) are called **relations**.

Proof. 1) A simple induction on n will yield a proof. I leave it to you. 2) A little harder, just take it for granted[1]. ☐

Let $\mathbb{Z}_2$ be the set $\{1, -1\}$ together with the following multiplication rules: $1 \cdot 1 = (-1)\cdot(-1) = 1$, $1 \cdot (-1) = (-1) \cdot 1 = -1$. From the above theorem and Eq. (4.1), we have

Corollary 4.2. *There is a unique map* $\mathrm{sign}\colon \Sigma_n \longrightarrow \mathbb{Z}_2$ *such that 1)* $\mathrm{sign}(\sigma\sigma') = \mathrm{sign}(\sigma)\cdot\mathrm{sign}(\sigma')$*; 2)* $\mathrm{sign}(\sigma_i) = -1$ *for each i.*

Corollary 4.3. *Suppose that $f\colon V^n \longrightarrow \mathbb{R}$ is a multi-linear map. Let $A(f)$ be the anti-symmetrization of f, i.e.,*

$$A(f)(v_1, \ldots, v_n) = \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \mathrm{sign}(\sigma) f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}).$$

Then $A(f)$ is indeed anti-symmetric; moreover, $A(f) = f$ if f is antisymmetric.

[1]The proof is not important here, you may find a proof in a textbook on Group Theory. This is the only place in this book that a complete proof or a hint for it is not given.

Exercise 4.1. 1) Let $f\colon V^n \longrightarrow \mathbb{R}$ be a multi-linear map. Show that f is antisymmetric if and only if

$$f(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \operatorname{sign}(\sigma) f(v_1, \dots, v_n)$$

for any $\sigma \in \Sigma_n$.

2) Prove the above two corollaries.

Exercise 4.2. Let V be an n-dimensional vector space, $T\colon V^n \longrightarrow \det V$ be the map sending $\mathbf{u} = (u_1, \cdots, u_n) \in V^n$ to the following element of $\det V$:

$$\mathbf{g} = (g_1, \dots, g_n) \mapsto g_1(u_1) g_2(u_2) \cdots g_n(u_n).$$

1) Show that T is a multi-linear map.

2) Show that $A(T)$ is the map sending $\mathbf{u} = (u_1, \dots, u_n) \in V^n$ to the following element of $\det V$:

$$\mathbf{g} = (g_1, \dots, g_n) \mapsto \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) g_{\sigma(1)}(u_1) g_{\sigma(2)}(u_2) \cdots g_{\sigma(n)}(u_n).$$

3) Show that $A(T)$ is also the map sending $\mathbf{u} = (u_1, \dots, u_n) \in V^n$ to the following element of $\det V$:

$$\mathbf{g} = (g_1, \dots, g_n) \mapsto \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) g_1(u_{\sigma(1)}) g_2(u_{\sigma(2)}) \cdots g_n(u_{\sigma(n)}).$$

4) Let $\det\colon V^n \longrightarrow \det V$ be $n!A(T)$. Show that det is a multi-linear antisymmetric map.

4.2 Determinant

With exercise 4.2 in mind, we are ready to state the following key theorem.

Theorem 4.4 (Universal Property). *Let $D(V)$ be the collection of all multi-linear anti-symmetric maps from V^n to vector spaces. Then, for any $f\colon V^n \longrightarrow Z$ in $D(V)$, there is a* **unique** *linear map $\phi_f\colon \det V \longrightarrow Z$ such that $\phi_f \det = f$, i.e., such that diagram*

$$\begin{tikzcd} V^n \arrow[r, "\det"] \arrow[dr, "f"'] & \det V \arrow[d, dotted, "\phi_f"] \\ & Z \end{tikzcd}$$

is commutative[2]*.*

[2] This is called the **universal property** for $\det V\colon V^n \longrightarrow \det V$; and $(\det, \det V)$ is called a **universal object** in $D(V)$.

Proof. Recall that det: $V^n \longrightarrow \det V$ is this multi-linear antisymmetric map: If $\mathbf{u} = (u_1, \ldots, u_n) \in V^n$, then $\det(\mathbf{u})$ is the following multi-linear anti-symmetric maps from $(V^*)^n$ to $\mathbb{R}$:

$$\begin{aligned} \mathbf{g} = (g_1, \ldots, g_n) \quad &\mapsto \quad \sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) g_{\sigma(1)}(u_1) \cdots g_{\sigma(n)}(u_n) \\ &= \quad \sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) g_1(u_{\sigma(1)}) \cdots g_n(u_{\sigma(n)}). \end{aligned} \tag{4.2}$$

Let $\mathbf{v} = (v_1, \ldots, v_n)$ be a basis for V, and $\hat{\mathbf{v}} = (\hat{v}_1, \ldots, \hat{v}_n)$ be the corresponding dual basis for V^*. We know that $\det V$ is the set of all multi-linear and anti-symmetric maps from $(V^*)^n$ to $\mathbb{R}$. From the proof of proposition 4.1, we know the map $S_{\mathbf{v}}$: $\det V \longrightarrow \mathbb{R}$ that assigns $f(\hat{\mathbf{v}})$ to $f \in \det V$ is a linear equivalence. The fact that $\mathbf{v} = (v_1, \ldots, v_n)$ is a basis for V implies that $\det(\mathbf{v})$ is a basis for $\det V$, that is because $S_{\mathbf{v}}(\det(\mathbf{v})) = 1$.

Suppose that f: $V^n \longrightarrow Z$ ia a multi-linear and anti-symmetric map. Then f is uniquely determined by $f(\mathbf{v})$. To define ϕ_f: $\det V \longrightarrow Z$ with $\phi_f \det = f$, recall that $\det V$ is one dimensional with $\det(\mathbf{v})$ as a basis, then we are forced to define ϕ_f: $\det V \longrightarrow Z$ to be the unique linear map such that $\phi_f(\det(\mathbf{v})) = f(\mathbf{v})$. □

Let T: $V \longrightarrow W$ be a linear map and $\dim V = \dim W = n$. The property of universal objects discussed above says that there is a unique linear map from $\det V$ to $\det W$, denoted by $\det T$, which makes square diagram

$$\begin{array}{ccc} V^n & \xrightarrow{\det} & \det V \\ {\scriptstyle T^{\times n}}\downarrow & & \downarrow{\scriptstyle \det T} \\ W^n & \xrightarrow[\det]{} & \det W \end{array} \tag{4.3}$$

commutative, that is because the composition of the bottom det with $T^{\times n}$ is multi-linear and antisymmetric.

Now suppose $W = V$, i.e., T is an endomorphism, then $\det T$ is an endomorphism of $\det V$ — a one dimensional vector space, so it must be a scalar multiplication, or simply a number. This number is called the **determinant** of T, and is denoted by $\operatorname{Det} T$.

Exercise 4.3. 0) Show that $\operatorname{Det} I = 1$.

1) Let T: $V \longrightarrow W$ and S: $W \longrightarrow Z$ be linear maps. Suppose that $\dim V = \dim W = \dim Z$. Use the universal property to show that[3]

$$\boxed{\det ST = \det S \, \det T.}$$

[3] Hint: Use square (4.3) and another square similar to that to prove that

$$\begin{array}{ccc} V^n & \xrightarrow{\det} & \det V \\ {\scriptstyle (ST)^{\times n}}\downarrow & & \downarrow{\scriptstyle \det S\ \det T} \\ Z^n & \xrightarrow[\det]{} & \det Z \end{array}$$

is a commutative square. Then use the uniqueness to conclude that $\det ST = \det S \, \det T$.

2) If $W = Z = V$ in part 1), show that

$$\boxed{\operatorname{Det} ST = \operatorname{Det} S \operatorname{Det} T.}$$

3) We write $\operatorname{Det} A$ for the determinant of the linear map $\vec{x} \mapsto A\vec{x}$ where $\vec{x} \in \mathbb{R}^n$. Show that

$$\boxed{\operatorname{Det} AB = \operatorname{Det} A \operatorname{Det} B.}$$

4) Let A be a coordinate matrix for endomorphism T. Show that $\operatorname{Det} T = \operatorname{Det} A$. So similar matrices have the same determinant because they are the coordinate matrices of the same linear map.

Example 4.1. Let A be an $n \times n$-matrix. Then A determines an endomorphism of $\mathbb{R}^n$ via matrix multiplication. The determinant of this endomorphism is called the determinant of A, and is denoted by $\operatorname{Det} A$. To get a formula for $\operatorname{Det} A$, we write $A = [\vec{a}_1, \ldots, \vec{a}_n] = [a_{ij}]$, and let $\mathbf{e} = (\vec{e}_1, \ldots, \vec{e}_n)$, $\hat{\mathbf{e}} = (\hat{\vec{e}}_1, \ldots, \hat{\vec{e}}_n)$. Then Eq. (4.3) implies that

$$\operatorname{Det}[\vec{a}_1, \ldots, \vec{a}_n] \cdot \det(\mathbf{e}) = \det(\vec{a}_1, \ldots, \vec{a}_n). \tag{4.4}$$

Here $\cdot$ is the scalar multiplication. Both $\det(\mathbf{e})$ and $\det(\vec{a}_1, \ldots, \vec{a}_n)$ is a multi-linear and antisymmetric map from $(\mathbb{R}^{n*})^n$ to $\mathbb{R}$. Now, according to Eq. (4.2), $\det(\mathbf{e})$ maps $\hat{\mathbf{e}}$ to 1, so it follows from Eq. (4.4) that

$$\boxed{\operatorname{Det}[\vec{a}_1, \ldots, \vec{a}_n] = \det(\vec{a}_1, \ldots, \vec{a}_n)(\hat{\mathbf{e}}).} \tag{4.5}$$

Now $\det(\vec{a}_1, \ldots, \vec{a}_n)$ maps $\hat{\mathbf{e}}$ to

$$\sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) \hat{\vec{e}}_{\sigma(1)}(\vec{a}_1) \cdots \hat{\vec{e}}_{\sigma(n)}(\vec{a}_n) = \sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n}$$

or

$$\sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) \hat{\vec{e}}_1(\vec{a}_{\sigma(1)}) \cdots \hat{\vec{e}}_n(\vec{a}_{\sigma(n)}) = \sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

It then follows from Eq. (4.5) that

$$\boxed{\begin{aligned} \operatorname{Det}[a_{ij}] &= \textstyle\sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} \\ &= \textstyle\sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}. \end{aligned}} \tag{4.6}$$

∎

Remark 4.3. The second equality in Eq. (4.6) says that determinant is invariant under the transpose. Compare with Ex. 5.22.

Remark 4.4. Since det: $V^n \longrightarrow \det V$ is multi-linear and antisymmetric, it follows from Eq. (4.5) that

$$\boxed{\begin{aligned} &\mathrm{Det}\,[\ldots,\, c\vec{a}_i + c'\vec{a}_i',\, \ldots] = c\mathrm{Det}\,[\ldots,\, \vec{a}_i,\, \ldots] + c'\mathrm{Det}\,[\ldots,\, \vec{a}_i',\, \ldots], \\ &\mathrm{Det}\,[\ldots,\, \vec{a}_i,\, \ldots,\, \vec{a}_j,\, \ldots] = -\mathrm{Det}\,[\ldots,\, \vec{a}_j,\, \ldots,\, \vec{a}_i,\, \ldots] \text{ for each } i \neq j. \end{aligned}} \tag{4.7}$$

Consequently, if a square matrix has a zero column or has two columns that are proportional to each other, then its determinant is zero.

In view of the previous remark, similar statements are also true when columns are replaced by rows.

Exercise 4.4. Investigate how the determinant changes when we apply elementary row or column operations.

Exercise 4.5. Show that the determinant of a square matrix is zero $\iff$ the columns of the square matrix are linearly dependent. Consequently, $\mathrm{Det}\,A \neq 0 \iff A$ is invertible.

Exercise 4.6. 1) Let $\vec{v}$ be a column matrix with n entries, if we delete its i-th entry $(1 \leq i \leq n)$, we get a column matrix with $(n-1)$ entries, denoted by $\vec{v}(i)$. Use Eq. (4.6) to show that

$$\mathrm{Det}\,[\vec{a}_1, \ldots, \vec{a}_{n-1}, \vec{e}_n] = \mathrm{Det}\,[\vec{a}_1(n), \ldots, \vec{a}_{n-1}(n)]\,.$$

2) Let $A = [a_{ij}]$. Let A^{ij} be the matrix obtained from A with its i-th row and j-th column removed. Use 1) and other properties of determinant to derive the **Laplace expansion** along the j-th column:

$$\boxed{\mathrm{Det}\,A = \textstyle\sum_{1\leq i\leq n}(-1)^{i+j}a_{ij}\mathrm{Det}\,A^{ij}.}$$

How about the expansion along a row?

3) We say a square matrix is a **upper triangular** matrix if any entry below the diagonal is zero. You know what a lower triangular matrix should be. A diagonal matrix is both a lower and a upper triangular matrix. Show that the determinant of a (upper or lower) triangular matrix is just the product of the diagonal entries.

4) Let $C_{ij}(A) = (-1)^{i+j}\mathrm{Det}\,A^{ij}$ (called the (i,j)-**cofactor** of A) and $C = [C_{ij}(A)]$. Show that

$$\boxed{AC' = \mathrm{Det}\,A\,I,}$$

where on the left hand side we have the matrix multiplication, on the right hand side we have scalar multiplication, and I is the $n \times n$-identity matrix. Consequently, if A is invertible, we have

$$\boxed{A^{-1} = \tfrac{1}{\mathrm{Det}\,A}C'.}$$

5) Suppose that A is an invertible $n \times n$-matrix, $\vec{b} \in \mathbb{R}^n$. Replacing the i-th column of A by $\vec{b}$, we produce a new matrix which is denoted by A_i. Let $\Delta = \mathrm{Det}\,A$ and $\Delta_i = \mathrm{Det}\,A_i$. Show that the unique solution of equation $A\vec{x} = \vec{b}$ can be written as

$$\boxed{x_i = \tfrac{\Delta_i}{\Delta}.}$$

(This is called the **Cramer's rule**.)

Exercise 4.7. Suppose that $n \times n$-matrix $A(t) = [a_{ij}(t)]$ depends on t smoothly. Show that

$$\boxed{\frac{d}{dt}\operatorname{Det} A = \sum_{1\le i,j\le n} \frac{da_{ij}}{dt} C_{ij}(A).}$$

Example 4.2. [Vandermonde determinant] Let $\vec{a}_i = [1, x_i, x_i^2, \ldots, x_i^{n-1}]'$, $1 \le i \le n$. Then $\operatorname{Det}[\vec{a}_1, \ldots, \vec{a}_n]$ is divisible by $(x_1 - x_2)$ because it is equal to $\operatorname{Det}[\vec{a}_1 - \vec{a}_2, \vec{a}_2, \ldots, \vec{a}_n]$. Similarly, it is divisible by $(x_i - x_j)$ for any $i \neq j$. So it is divisible by $\prod_{1\le i<j\le n}(x_j - x_i)$. By degree counting we know that

$$\operatorname{Det}[\vec{a}_1, \ldots, \vec{a}_n] = c_n \prod_{1\le i<j\le n} (x_j - x_i) \tag{4.8}$$

where c_n is a constant. Now expand the left-hand side of (4.8) along the last column, we know the coefficient in front of x_n^{n-1} is $c_{n-1}\prod_{1\le i<j\le n-1}(x_j - x_i)$. We can also expand the right-hand side of (4.8) in powers of x_n, and the coefficient in front of x_n^{n-1} is $c_n \prod_{1\le i<j\le n-1}(x_j - x_i)$. So $c_n = c_{n-1}$. Since $c_2 = 1$, so $c_n = 1$. Therefore $\operatorname{Det}[\vec{a}_1, \ldots, \vec{a}_n] = \prod_{1\le i<j\le n}(x_j - x_i)$. ▮

Exercise 4.8. 1) Show that $\operatorname{Det}(I \oplus A) = \operatorname{Det} A = \operatorname{Det}(A \oplus I)$ for any square matrix A.

2) Show that $\operatorname{Det}(T_1 \oplus T_2) = \operatorname{Det} T_1 \cdot \operatorname{Det} T_2$ for any endomorphisms T_1 and T_2.

3) Suppose that A, D are endomorphisms or square matrices and A is invertible. Show that

$$\operatorname{Det}\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \operatorname{Det}\left[\begin{array}{c|c} A & B \\ \hline 0 & D - CA^{-1}B \end{array}\right] = \operatorname{Det}\left[\begin{array}{c|c} A & 0 \\ \hline C & D - CA^{-1}B \end{array}\right].$$

4) Let A, D be endomorphisms or square matrices. Show that[4]

$$\operatorname{Det}\left[\begin{array}{c|c} A & B \\ \hline 0 & D \end{array}\right] = \operatorname{Det} A \operatorname{Det} D = \operatorname{Det}\left[\begin{array}{c|c} A & 0 \\ \hline C & D \end{array}\right].$$

Consequently, the determinant of a block triangular matrix is the product of the determinant of the block diagonals.

5) Let A, D be endomorphisms or square matrices and A is invertible. Show that

$$\boxed{\operatorname{Det}\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \operatorname{Det} A \operatorname{Det}(D - CA^{-1}B).}$$

Exercise 4.9. Calculate the determinant of

$$\left[\begin{array}{c|c} I & \vec{x} \\ \hline \vec{y}' & d \end{array}\right]$$

where d is a number, $\vec{x}$ and $\vec{y}$ are vectors in $\mathbb{R}^n$ and I is the $n \times n$-identity matrix.

[4]Hint: it is true from 2) and 3) when A is invertible. So it is true when A is replaced by $A + tI$ for all sufficiently large t, so must be true for all t because both sides are polynomial in t. Then set $t = 0$.

Exercise 4.10.

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

Find $\operatorname{Det} A$, $\operatorname{Det} A^{-1}$, $\operatorname{Det}(A^T)^{-1}$, $\operatorname{Det} A^T$, $\operatorname{Det} A^5$, $\operatorname{Det} A^{-4}$.

Remark 4.5. We would like to remark that the universal property actually implies that the universal objects are unique up to linear equivalence. That is to say that, if $(\det' V, \det')$ is another such object, then there is a unique linear equivalence such that triangle

$$\begin{tikzcd} & \det V \arrow[dd, leftrightarrow, "\cong"] \\ V^n \arrow[ur, "\det"] \arrow[dr, "\det'"'] & \\ & \det' V \end{tikzcd}$$

is commutative.

Exercise 4.11. The determinant of square matrices defines a multi-linear and antisymmetric map $\operatorname{Det}\colon (\mathbb{R}^n)^n \longrightarrow \mathbb{R}$:

$$(\vec{a}_1, \ldots, \vec{a}_n) \quad \mapsto \quad \operatorname{Det}[\vec{a}_1, \ldots, \vec{a}_n].$$

Show that $(\mathbb{R}, \operatorname{Det})$ is a universal object and then conclude that $\operatorname{Det}(AB) = \operatorname{Det}(A) \cdot \operatorname{Det}(B)$ for any two $n \times n$-matrices A and B.

Everything we have developed so far in this section is valid when the ground field is replaced by $\mathbb{C}$. However, when the ground field is $\mathbb{R}$, the determinant of an $n \times n$-matrix has a geometric meaning: *it is the signed volume of the oriented bounded geometric region in $\mathbb{R}^n$ spanned by the columns of the matrix.*

To explain that, note that $\mathbb{R}^n$ carries a standard (the one that agrees with your intuition) geometrical structure, i.e., something that allows you to talk about length and angle, hence area and volume. Suppose that $\vec{a}_1, \ldots, \vec{a}_n$ are vectors in $\mathbb{R}^n$, then they span an oriented bounded geometric region in $\mathbb{R}^n$. Let me not give a precise description here; however, I hope you get the idea if you know what it is in dimension one, two and three. In dimension one, it is just the unique vector, but considered as an oriented line segment, it is positively (negatively) oriented if it points to the right (left). In dimension two, it is the oriented parallelogram with the two vectors as a pair of neighboring edges. The orientation is given this way: assume the two vectors are linearly independent and the angle from the first vector to the second vector is positive and less than π, then the orientation is called positive; otherwise, the orientation is called negative. In dimension three, the orientation is positive(negative) if $\vec{a}_1$, $\vec{a}_2$, $\vec{a}_3$ obey the right-hand (left-hand) rule.

The general claim is this: Det $[\vec{a}_1, \ldots, \vec{a}_n]$ *is equal to the volume of the bounded geometric region, but this is true modulo a sign if the volume is not zero. The sign is positive if the orientation is positive and is negative if the orientation is negative.*

Exercise 4.12. Prove this claim in the case $n = 1$, 2 and 3.

In summary, we have shown that for each vector space V of dimension n, there is a universal object det: $V^n \longrightarrow \det V$ in the collection of all multi-linear antisymmetric maps from V^n to vector spaces. It is unique up to linear equivalences. In general, if T is a linear map between vector spaces of equal dimension, then $\det T$ is a linear map between the corresponding determinant lines. In the case that T is an endomorphism, $\det T$ is just a scalar multiplication, or just a number.

4.3 Eigenvectors, Eigenvalues and Diagonalizations

In many interesting linear problems, an endomorphism T on a vector space V is encountered. If T is just a scalar multiplication, the problem gets solved quickly. In general, T is not like that. However, the problem can still be solved quickly if it admits a decomposition into a direct sum of scalar multiplications. So it is important to know whether a given endomorphism is decomposable into a direct sum of scalar multiplications, and if it is, how we do it. We will answer these questions here.

Suppose that $V = V_1 \oplus \cdots \oplus V_k$ where $\dim V_i > 0$ for each $1 \leq i \leq k$, and with respect to this decomposition, $T = \lambda_1 I \oplus \cdots \oplus \lambda_k I$ where λ_i's are some distinct numbers. Then it is clear that $\lambda = \lambda_i$ for some $i \iff T - \lambda I$ is not invertible $\iff \mathrm{Det}(T - \lambda I) = 0$. (True for matrices, so true in general) Moreover, $\ker(T - \lambda_i I) = V_i$.

Before we proceed to answer the questions mentioned above, it is a good time to introduce some terminologies. Each of these special numbers λ_i's is called an **eigenvalue** of T and each of these special subspaces V_i's is called the **eigenspace** for T, and a nonzero vector in V_i is called an **eigenvector** of T with eigenvalue λ_i. Note that if v is an eigenvector of T with eigenvalue λ_i, then $T(v) = \lambda_i v$ because T is just a scalar multiplication by λ_i on V_i.

We extend these terminologies to general endomorphisms as well. We say λ is an eigenvalue of T if $T - \lambda I$ is not invertible. If λ is an eigenvalue of T, then $\ker(T - \lambda I)$ is called an eigenspace with eigenvalue λ, denoted by E_λ. If $v \in E_\lambda$ and $v \neq 0$, then v is called an eigenvector of T with eigenvalue λ.

Exercise 4.13. Show that the dimension of an eigenspace is at least one. Show that λ is an eigenvalue of $T \iff$ there is a nonzero vector v such that $T(v) = \lambda v$.

Therefore, to solve the problem of decomposing an endomorphism into a direct sum of scalar multiplications, all we need to do are the followings:

1. Find all solutions[5] of the **characteristic equation**

$$\mathrm{Det}(T - \lambda I) = 0.$$

[5]This means all real solutions if we work with a real vector space, or all complex solutions if we work with a complex vector space.

I.e., *find all eigenvalues of* T.

2. For each solution λ_i, find $\ker(T - \lambda_i I)$. I.e., *find all eigenspaces.*

3. Write down the decomposition of V into the direct sum of eigenspaces (if possible) and write T as the corresponding decomposition into the direct sum of scalar multiplication by eigenvalues on the associated eigenspaces.

Note that, the last step may not be always possible even if we work with complex numbers and complex vector spaces, because there are situations where $\sum_i \dim V_i < \dim V$, see the example 4.3 below.

Remark 4.6. $\mathrm{Det}\,(\lambda I - T) = \lambda^n + \cdots$ is a degree n $(= \dim V)$ polynomial in λ, so counting with multiplicity it always has n complex solutions. (It is called the **characteristic polynomial** for T) Over real numbers, it may have no solutions at all. Because of that, it is more convenient to work with complex vector spaces and (complex) linear maps.

Example 4.3. Consider the two endomorphisms of $\mathbb{R}^2$ whose standard matrices are

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

respectively. The characteristic equation for A is $(1-\lambda)(3-\lambda) = 0$, its solutions are 1 and 3. Then

$$E_1 = \mathrm{Nul}\,(A - 1 \cdot I) = \mathrm{Nul} \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} = \mathrm{Nul} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \mathrm{span} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}.$$

$$E_3 = \mathrm{Nul}\,(A - 3 \cdot I) = \mathrm{Nul} \begin{bmatrix} -2 & 1 \\ 0 & 0 \end{bmatrix} = \mathrm{Nul} \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix} = \mathrm{span} \left\{ \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \right\}.$$

It is clear that $\mathbb{R}^2 = E_1 \oplus E_2$, so the endomorphism whose standard matrix is A is decomposable into a direct sum of scalar multiplications.

The characteristic equation for B is $(1 - \lambda)^2 = 0$, it has only a single solution: 1. Then

$$E_1 = \mathrm{Nul}\,(A - 1 \cdot I) = \mathrm{Nul} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \mathrm{span} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}.$$

It is clear that $\mathbb{R}^2 \neq E_1$, so the endomorphism whose standard matrix is B is not decomposable into a direct sum of scalar multiplications. This conclusion holds even if we work with complex numbers. ∎

If an endomorphism on $\mathbb{R}^n$ is decomposable into a direct sum of scalar multiplications, we say its standard matrix A is **diagonalizable**, i.e., there is an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$. That is because if $\mathbb{R}^n = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$, pick a basis for each eigenspace, juxtapose these bases together, we get a basis[6] for $\mathbb{R}^n$: $(\vec{v}_1, \ldots, \vec{v}_n)$. Take $P = [\vec{v}_1, \ldots, \vec{v}_n]$, then

$$AP \quad = \quad [A\vec{v}_1, \ldots, A\vec{v}_n] = [\lambda_1 \vec{v}_1, \ldots, \lambda_n \vec{v}_n]$$

[6] See Ex. 3.32.

$$\begin{aligned} &= [\vec{v}_1, \ldots, \vec{v}_n] \begin{bmatrix} \lambda_1 I & & \\ & \ddots & \\ & & \lambda_k I \end{bmatrix} \\ &= PD, \end{aligned}$$

so $A = PDP^{-1}$. Note that *the columns of P are linearly independent eigenvectors and their corresponding eigenvalues are the diagonals of D.*

Example 4.4. In the previous example, matrix B is not diagonalizable. Matrix A is diagonalizable, what we have discussed so far enable us quickly to write $A = PDP^{-1}$ where

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

∎

Exercise 4.14. Let T be an endomorphism on V. Show that T is decomposable into a direct sum of scalar multiplications $\iff$ there is a basis for V such that the coordinate matrix for T with respect to this basis is a diagonal matrix.

Exercise 4.15. Let A be a square matrix. Suppose that $A^2 - 3A + 2I = 0$ and λ is an eigenvalue of A. Show that λ is either 1 or 2.

Exercise 4.16. Consider the matrix

$$A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}.$$

(a) Diagonalize A; (b) Use the diagonalization to find A^{10}; (c) Determine the invertibility of A from its diagonalization; (d) Use the diagonalization to compute $A^5 - 2A^4 + A^3$.

Exercise 4.17. Let T be an endomorphism on vector space V. Show that

1) If $v_1, \ldots, v_k$ are eigenvectors of T with distinct eigenvalues, then[7] they are linearly independent.

2) The sum of eigenspaces of T with distinct eigenvalues is a (internal) direct sum.

3) The sum of the dimension of the eigenspaces is at most $\dim V$.

Exercise 4.18. Let A be an $n \times n$-matrix. Show that A is diagonalizable $\iff$ there is a basis for $\mathbb{R}^n$ consisting of the eigenvectors of A $\iff$ the sum of the dimension of the eigenspaces is equal to n. In particular, if A has n distinct eigenvalues, then A is diagonalizable.

Example 4.5. Let

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

[7]Hint: induction on k.

The characteristic equation is $\lambda^2+1=0$. Over real numbers there are no eigenvalues and no eigenspaces, so considered as endomorphism of $\mathbb{R}^2$, the linear map $\vec{x}\mapsto A\vec{x}$ is not decomposable into a direct sum of scalar multiplications. However, over complex numbers, we have eigenvalues $\pm i$, so considered as endomorphism of $\mathbb{C}^2$, the linear map $\vec{x}\mapsto A\vec{x}$ is decomposable into a direct sum of scalar multiplications. ∎

Exercise 4.19 (Cayley-Hamilton Theorem). Let A be an $n\times n$-matrix and

$$P_A(\lambda)=\mathrm{Det}\,(\lambda I-A)=\lambda^n+\cdots+\det(-A).$$

1) Show that if U is invertible then $P_A(\lambda)=P_{UAU^{-1}}(\lambda)$.
2) Show that if A is a diagonal matrix then $P_A(\lambda)|_{\lambda=A}=A^n+\cdots+\det(-A)I=0$.
3) Show that if A is a diagonalizable matrix then $P_A(\lambda)|_{\lambda=A}=0$.
4) Let D_n be the diagonal matrix whose diagonal entries are 1, 2, ..., n. For a fixed A, show that tD_n+A has n distinct eigenvalues for all sufficiently large t.
5) Show that $P_A(\lambda)|_{\lambda=A}=0$ for any square matrix or endomorphism A.

Exercise 4.20 (Trace). Let T be an endomorphism of an n-dimensional real (complex) vector space V. Consider the expansion of $\mathrm{Det}\,(I+tT)$ in t:

$$\mathrm{Det}\,(I+tT)=1+tc_1(T)+t^2c_2(T)+\cdots+t^nc_n(T).$$

1) Show that c_1 is a real (complex) linear function on $\mathrm{End}\,(V):=\mathrm{Hom}\,(V,V)$;
2) Show that $c_1(TS)=c_1(ST)$ for any T, S in $\mathrm{End}\,(V)$ under the assumption that S is invertible;
3) Show that $c_1(TS)=c_1(ST)$ for any T, S in $\mathrm{End}\,(V)$;
4) Show that if the coordinate matrix of T with respect to a basis for V is $A=[a_{ij}]$, then $c_1(T)=\sum_i a_{ii}$.
5) Any linear function f on $\mathrm{End}\,(V):=\mathrm{Hom}\,(V,V)$ which satisfies the relation $f(TS)=f(ST)$ for any T, S in $\mathrm{End}\,(V)$ is called a **trace** on $\mathrm{End}\,(V)$. Show that, if f is a trace on $\mathrm{End}\,(V)$, then f is a scalar multiple of c_1. So, c_1 is called *the trace*, often denoted by tr, so $c_1(T)=\mathrm{tr}\,(T)$.
6) Suppose that $\mathrm{Det}\,(\lambda I-T)=(\lambda-\lambda_1)\cdots(\lambda-\lambda_n)$. Show that

$$c_k(T)=\sum_{1\le i_1<\cdots<i_k\le n}\lambda_{i_1}\cdots\lambda_{i_k}$$

7) One can show that $c_i(T)$ is the trace of something called the i-th exterior power of T. Of course you can not prove this here. However, $c_i(T)$ is made of $\mathrm{tr}\,(T)$, $\mathrm{tr}\,(T^2)$, ..., $\mathrm{tr}\,(T^i)$, and that is something you can show[8]. For example,

$$c_2(T)=\frac{1}{2}\left((\mathrm{tr}T)^2-\mathrm{tr}T^2\right).$$

[8] Hint: just need to show that this is true in the case T is a diagonal matrix.

Chapter 5

Geometric Structures (4 weeks)

Every real vector space can carry a Riemannian geometric structure, i.e., a structure that allows us to measure lengths and angles in a way which is compatible with its linear structure. In the case that a real vector space carries a complex structure, it can carry an Hermitian geometric structure, i.e., a Riemannian geometric structure that is in compatible with the complex structure. It turns out that a complex vector space is just a real vector space together with a complex structure, and the Hermitian geometric structure can be rephrased in pure complex language. Moreover, any complex vector space can carry an Hermitian geometric structure. By presenting the materials this way, I hope that the mysterious definition of Hermitian geometric structures in pure complex language is not mysterious anymore.

5.1 Riemannian Structures on Real Vector Spaces

$\mathbb{R}^n$ is more than just a vector space, it has a standard geometrical structure, i.e., the one that agrees with our intuition. In general, a geometric structure on a vector space is a way for us to measure the lengths of vectors and the angles between vectors. Of course, that should be compatible with the linear structure, for instance, when a vector is multiplied by a real number c, the length should get multiplied by $|c|$. With lengths and angles defined on a vector space V, we can now define a product $\langle\, ,\, \rangle$ on V: $V \times V \longrightarrow \mathbb{R}$ as follows:

$$\boxed{\langle u,\, v\rangle = |u| \cdot |v| \cdot \cos\theta} \tag{5.1}$$

where u and v are vectors in V, $|u|$ and $|v|$ are the length of u and v respectively, and θ is the angle between u and v. The angle is not well defined when either u or v is the zero vector; however, in this case we define the product to be zero.

Exercise 5.1. Take the standard geometry on $\mathbb{R}^n$, i.e., the one such that each $|\vec{e}_i| = 1$ and $\vec{e}_i$, $\vec{e}_j$ are perpendicular if $i \neq j$. Show that $\langle \vec{x}, \vec{y}\rangle = x_1y_1 + \cdots + x_ny_n$. (by convention, this product is called the **dot product** and $\langle \vec{x},\, \vec{y}\rangle$ is written as $\vec{x} \cdot \vec{y}$. Note that $[\vec{x} \cdot \vec{y}] = \vec{x}'\vec{y}$ as matrix identity.)

From the fact that the geometric structure is compatible with the linear structure, or at least from the standard geometry on model space, we can see that Eq. (5.1) implies the following properties for $\langle , \rangle$:

1. $\langle , \rangle$ is *bilinear*, i.e., linear in each variable;
2. $\langle , \rangle$ is *symmetric*, i.e., $\langle u, v\rangle = \langle v, u\rangle$ for any u and v in V;
3. $\langle , \rangle$ is *positive definite*, i.e., $\langle u, u\rangle \geq 0$ for any u and $= 0$ if and only if $u = 0$.

Here comes a formal definition.

Definition 5.1 (Inner Product). Let V be a real vector space. An **inner product** on V is defined to be a bilinear map $\langle , \rangle\colon V \times V \longrightarrow \mathbb{R}$ which is both symmetric and positive definite.

Remark 5.1. An inner product on a real vector space V is also called a **Riemannian structure** on V. The dot product on $\mathbb{R}^n$ is called the **standard Riemannian structure** on $\mathbb{R}^n$. Under a given inner product $\langle , \rangle$ on V, the length of a vector u in V, denoted by $|u|$, is defined to be $\sqrt{\langle u, u\rangle}$.

A real vector space equipped with an inner product is called a (finite dimensional) real **Hilbert space**. $\mathbb{R}^n$ together with the standard Riemannian structure is a real Hilbert space. Since $V \cong \mathbb{R}^{\dim V}$, it is clear that inner product exists on any real vector space.

Remark 5.2. An inner product on V enables us to define lengths and angles for vectors in V via Eq. (5.1). It turns out, there is a one-one correspondence:

$$\boxed{\begin{array}{c}\text{inner products on } V\\ \updownarrow \\ \text{compatible ways of measuring lengths and angles on } V.\end{array}}$$

(Can you make the word 'compatible' precise?[1])

Exercise 5.2 (Cauchy-Schwartz inequality). Use the three properties listed in the definition of inner product to show that[2]

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

5.1.1 Orthogonal Projections

Throughout this subsection we assume V is a (finite dimensional) real Hilbert space. We say two vectors u and v in V are **orthogonal** to each other (denoted by $u \perp v$) if their inner product is zero.

Theorem 5.1 (Pythagorean Theorem and its Converse). $u \perp v \iff |u|^2 + |v|^2 = |u + v|^2$.

[1] Trust your intuition.
[2] Hint: $|u + tv|^2 \geq 0$ for any real parameter t.

Proof. The theorem follows from the following computations:

$$\begin{aligned} |u+v|^2 &= \langle u+v, u+v\rangle \qquad \text{definition of length} \\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \qquad \text{bilinear} \\ &= \langle u,u\rangle + 2\langle u,v\rangle + \langle v,v\rangle \qquad \text{symmetry} \\ &= |u|^2 + |v|^2 + 2\langle u,v\rangle \qquad \text{definition of length.} \end{aligned}$$

□

Let W be a subspace of V, u a vector in V. We say that v is orthogonal to W, denoted by $v \perp W$, if v is orthogonal to each vector in W, but that is equivalent to saying that v is orthogonal to each vector in a spanning set for W.

The **orthogonal complement** of W in V, denoted by $W^\perp$, is defined to be the set of all vectors that are orthogonal to W. $W^\perp$ is a subspace of V because it is the kernel of a linear map: Suppose that $v_1, \ldots, v_k$ form a spanning set for W; define $T\colon V \longrightarrow \mathbb{R}^k$ by $T(v) = [\langle v, v_1\rangle, \ldots, \langle v, v_k\rangle]'$, then $W^\perp = \ker T$. In fact $W^\perp$ is the largest subspace of V that is orthogonal to W.

Proposition 5.1. *Let W be a subspace of V. Then*

1) $W \cap W^\perp = \{0\}$.

2) $V = W \oplus_i W^\perp$. Consequently any $x \in V$ can be uniquely decomposed as

$$x = y + z \qquad y \in W \text{ and } z \in W^\perp.$$

3) $(W^\perp)^\perp = W$.

Proof. 1) If $v \in W \cap W^\perp$, then $v \cdot v = 0$, so $v = 0$.

2) From part 1) we know that the sum of W with $W^\perp$ is a direct sum. To show that $W \oplus_i W^\perp = V$, we just need to show that $\dim V = \dim W + \dim W^\perp$. But that can be proved by a sandwich argument:

$$\begin{aligned} \dim V &\geq \dim(W \oplus_i W^\perp) = \dim W + \dim W^\perp \\ &= \dim W + \dim \ker T \qquad \text{because } W^\perp = \ker T \\ &\geq \dim W + (\dim V - k) \qquad \text{by Eq. (3.1) and Ex. 3.3} \\ &= \dim V \qquad \text{assume } v_1, \ldots, v_k \text{ form a minimal spanning set for } W. \end{aligned}$$

3) It is clear that $W \subset (W^\perp)^\perp$. When replacing W by $W^\perp$ in part 2), we have $V = W^\perp \oplus_i (W^\perp)^\perp$. So $\dim W + \dim W^\perp = \dim V = \dim W^\perp + \dim(W^\perp)^\perp$, so $\dim W = \dim(W^\perp)^\perp$, so $W = (W^\perp)^\perp$ by Ex. 3.3. □

We have seen that any $x \in V$ can be uniquely written as $y + z$ with $y \in W$ and $z \in W^\perp$. We call such y the orthogonal projection of x onto W, written as $y = \mathrm{Proj}_W(x)$. So the proposition says that $z = \mathrm{Proj}_{W^\perp}(x)$. The assignment of y to x defines a linear map

$$\mathrm{Proj}_W : V \longrightarrow V.$$

Note that $\mathrm{Proj}_W \mathrm{Proj}_W = \mathrm{Proj}_W$, i.e., projection twice is the same as projection once. The range of Proj_W is W and its kernel is $W^\perp$, they are eigenspaces of Proj_W with eigenvalue 1 and 0 respectively. With respect to the decomposition $V = W \oplus_i W^\perp$, we have $\mathrm{Proj}_W = I \oplus 0$ and $\mathrm{Proj}_{W^\perp} = 0 \oplus I$, so $\mathrm{Proj}_W + \mathrm{Proj}_{W^\perp} = I$ and $\mathrm{Proj}_W \mathrm{Proj}_{W^\perp} = 0$.

Exercise 5.3. Suppose that W is a subspace of V. Show that[3], for any $x \in V$ and any $w \in W$, we have $|x - \mathrm{Proj}_W(x)| \le |x - w|$. I.e., among all vectors in W, $\mathrm{Proj}_W(x)$ gives the best approximation to x.

Exercise 5.4. Consider a parallelogram. Show that the sum of the square of the length of its diagonals is equal to the sum of the square of the length of its sides.

5.1.2 Orthonormal Bases

Throughout this subsection we assume V is a (finite dimensional) real Hilbert space, W is a subspace of V and $W^\perp$ is the orthogonal complement of W in V. Note that, W is also a Hilbert space with the inner product inherited from that of V.

A set of vectors in V is called an **orthogonal set** of vectors if any two in this set are orthogonal. *An orthogonal set of* **nonzero** *vectors is always a linearly independent set.* This can be proved by a simple computation: Suppose that $v_1, \ldots, v_k$ form an orthogonal set and

$$c_1v_1 + \cdots + c_kv_k = 0;$$

taking the inner product of this identity with v_1, we get

$$c_1\langle v_1, v_1\rangle + c_2\langle v_2, v_1\rangle + \cdots = \langle 0, v_1\rangle, \quad \text{or} \quad c_1|v_1|^2 = 0 \Rightarrow c_1 = 0.$$

Similarly, $c_i = 0$ for each i.

Let $(v_1, \ldots, v_n)$ be a basis for V. We say this is an **orthonormal basis** if $\langle v_i, v_j\rangle = 0$ if $i \ne j$ and $= 1$ if $i = j$. I.e., an orthonormal basis is a basis such that each vector in the basis has length one and any two of them are orthogonal. For example, $(\vec{e}_1, \ldots, \vec{e}_n)$ is an orthonormal basis for $\mathbb{R}^n$.

One advantage of having an orthonormal basis is that we can find the coordinates of any vector very quickly. Suppose that $\mathbf{v} = (v_1, \ldots, v_n)$ is an orthonormal basis for V, and $x \in V$. We know $x = x_1v_1 + \cdots + x_nv_n$ for some numbers x_i's. Now

$$\langle v_1, x\rangle = x_1\langle v_1, v_1\rangle + x_2\langle v_1, v_2\rangle + \cdots = x_1.$$

Similarly, $\langle v_i, x\rangle = x_i$. So we have

$$\boxed{x = \langle v_1, x\rangle v_1 + \cdots + \langle v_n, x\rangle v_n.} \tag{5.2}$$

Suppose that $(x_1, \ldots, x_n)$ is a basis for V. Let V_i be the span of first i vectors in this basis, then $V_1 \subset V_2 \subset \cdots \subset V_n = V$ and $\dim V_i = i$.

Claim 2. *Each V_i has an orthonormal basis.*

Proof. Induction on i. If $i = 1$, we let $v_1 = \frac{1}{|x_1|}x_1$. Assume V_i has an orthonormal basis $(v_1, \ldots, v_i)$. Let y_{i+1} be the orthogonal projection of x_{i+1} onto the orthogonal complement of V_i in V_{i+1}. Then $y_{i+1} \ne 0$, otherwise, $x_{i+1} \in V_i$, so $V_{i+1} = V_i + \mathrm{span}\,\{x_{i+1}\} = V_i$, a contradiction. Let $v_{i+1} = \frac{1}{|y_{i+1}|}y_{i+1}$. Then $v_1, \ldots, v_{i+1}$ form an orthonormal basis for V_{i+1}. $\square$

[3]Hint: draw a picture to help yourself, say W is a plane in $\mathbb{R}^3$.

Therefore, *orthonormal basis always exists for any finite dimensional Hilbert space*. Consequently, any real Hilbert space of dimension n is isometrically[4] linearly equivalent to $\mathbb{R}^n$. We call $\mathbb{R}^n$ the model real Hilbert space of dimension n.

Now suppose that $v_1, \ldots, v_k$ form an orthonormal basis for W, $v_{k+1}, \ldots, v_n$ form an orthonormal basis for $W^\perp$, then $v_1, \ldots, v_n$ form an orthonormal basis for V. Now if $x \in V$, from Eq. (5.2), we have

$$x = \underbrace{\langle v_1, x\rangle v_1 + \cdots + \langle v_k, x\rangle v_k}_{\text{in } W} + \underbrace{\langle v_{k+1}, x\rangle v_{k+1} + \cdots + \langle v_n, x\rangle v_n}_{\text{in } W^\perp}.$$

Therefore, we have the following formula for the orthogonal projection: Suppose that $v_1, \ldots, v_k$ form a orthonormal basis for W, and $x \in V$, then

$$\boxed{\mathrm{Proj}_W(x) = \langle v_1, x\rangle v_1 + \cdots + \langle v_k, x\rangle v_k.}$$

In the case that $v_1, \ldots, v_k$ form just an **orthogonal basis** for W (i.e., v_i's are not assumed to have length one, although they are mutually orthogonal to each other), the projection formula then becomes

$$\boxed{\mathrm{Proj}_W(x) = \tfrac{\langle v_1, x\rangle}{|v_1|^2} v_1 + \cdots + \tfrac{\langle v_k, x\rangle}{|v_k|^2} v_k.} \tag{5.3}$$

Gram-Schmidt Process

In the induction proof of claim 2, we have defined

$$y_{i+1} = x_{i+1} - \mathrm{Proj}_{V_i}(x_{i+1}).$$

Therefore, in view of Eq. (5.3), we have

$$\boxed{\begin{array}{rcll} y_1 & = & x_1 & \text{define } y_1 \text{ this way} \\[2ex] y_2 & = & x_2 - \frac{\langle y_1, x_2\rangle}{|y_1|^2} y_1 & \\[2ex] y_3 & = & x_3 - \frac{\langle y_1, x_3\rangle}{|y_1|^2} y_1 - \frac{\langle y_2, x_3\rangle}{|y_2|^2} y_2 & \begin{array}{l}\text{because } y_1,\ y_2 \text{ form an} \\ \text{orthogonal basis for } V_2\end{array} \\[2ex] & \vdots & & \\[2ex] y_i & = & x_i - \frac{\langle y_1, x_i\rangle}{|y_1|^2} y_1 - \cdots - \frac{\langle y_{i-1}, x_i\rangle}{|y_{i-1}|^2} y_{i-1} & \text{similar reason} \\[2ex] & \vdots & & \end{array}}$$

Let $v_i = \frac{1}{|y_i|} y_i$. Then we know $v_1, v_2, \ldots$ form an orthonormal basis. The process of producing an orthonormal basis out of a basis in this way is called the **Gram-Schmidt process**.

[4] I.e., preserving length and angle.

Exercise 5.5. Consider

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad u = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

(a) Find an orthogonal basis of $\operatorname{Col} A$; (b) Find the orthogonal projection of u onto $\operatorname{Col} A$; (c) Find the distance from u to $\operatorname{Col} A$.

Exercise 5.6. Consider vectors

$$u_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad u_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \quad y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

1) Use Gram-Schmidt process to turn u_1, u_2, u_3 into an orthogonal set;
2) Extend the orthogonal set in 1) to an orthogonal basis of $\mathbb{R}^4$;
3) Find the distance from y to $\operatorname{span}\{u_1, u_2, u_3\}$.

Exercise 5.7. Let W be a k-dimensional subspace of $\mathbb{R}^n$. Suppose that $\mathbf{x} = (\vec{x}_1, \ldots, \vec{x}_k)$ is an orthonormal basis for W. Show that the standard coordinate matrix for Proj_W is $\vec{x}_1\vec{x}_1' + \cdots + \vec{x}_k\vec{x}_k'$.

Proposition 5.2. *Let T be a linear map from real Hilbert space V to real Hilbert space W. Then there is a unique linear map $T'\colon W \longrightarrow V$ such that, for any $v \in V$ and any $w \in W$, $\langle w\,, T(v)\rangle = \langle T'(w)\,, v\rangle$. Moreover, $(TS)' = S'T'$ if either side is defined.*

Proof. Just need to work with model Hilbert spaces, so we can take T be a real $m \times n$-matrix A. The proof is then just a computation: Plugging $w = \vec{e}_i \in W$ and $v = \vec{e}_j$ into $\langle w\,, Av\rangle = \langle A'w\,, v\rangle$, we are forced to have $A_{ji} = A'_{ij}$, i.e., A' is obtained from A by taking transpose. The rest is clear. $\square$

Remark 5.3. T' is called the **real adjoint** of T. In the case that T is an endomorphism on V and $T' = T$, we say T is a **real self-adjoint operator** on V.

5.1.3 Linear Isometries and Orthogonal Transformations

Let T be a linear map from Hilbert space V to Hilbert space W. We say T is a **linear isometry** if T also respects the geometric structures, i.e.,

$$\boxed{\langle T(x), T(y)\rangle = \langle x, y\rangle \qquad \text{for any } x,\ y \text{ in } V.}$$

It is easy to see that if T is a linear isometry, then T is one-to-one, so $\dim V \leq \dim W$; also T preserves lengths and angles.

If an endomorphism on V is a linear isometry we say it is an **orthogonal transformation** on V. The set of all orthogonal transformations on V is denoted by $O(V)$. (It is a Lie group in case you know what a Lie group should be. The orthogonal transformations and the translations on V do not change angles and lengths, so they are called rigid motions on V — the transformations on V which fix lengths and angles. One can show that any rigid motion on V which fixes the zero vector in V is just an orthogonal transformation on V.)

The standard matrix for an orthogonal transformation on $\mathbb{R}^n$ is called an **orthogonal matrix**. It is just a matter of computations that O is an orthogonal matrix $\iff$ the columns are orthonormal basis for $\mathbb{R}^n$ $\iff$ $O'O = I$ $\iff$ $O^{-1} = O'$ $\iff$ $OO' = I$ $\iff$ the rows are orthonormal basis for $(\mathbb{R}^n)^*$.

Exercise 5.8. 1) Prove this last statement.

2) Let $T\colon V \longrightarrow W$ be a linear map. Show that T is a linear isometry if and only if T maps an orthonormal basis for V into an orthonormal set of vectors in W.

Exercise 5.9. Let A be an invertible matrix. Show that[5] there is an orthogonal matrix O and a upper triangular matrix T such that $A = OT$, or equivalently there is an orthogonal matrix O and a lower triangular matrix T such that $A = TO$.

Note that O is an orthogonal matrix $\Rightarrow O'O = I \Rightarrow \det O'O = 1 \Rightarrow \det O = \pm 1$. Geometrically an orthogonal transformation on $\mathbb{R}^n$ is a pure rotation if its determinant is 1 and is a reflection (about hyperplanes through the origin) plus a possible rotation if its determinant is -1. In fact any orthogonal transformation on $\mathbb{R}^n$ is a composition of finitely many reflections (about a hyperplane through the origin), that is a theorem[6] of E. Cartan. More generally, any rigid motion of $\mathbb{R}^n$ is a composition of finitely many reflections about hyperplanes not necessarily through the origin.

Exercise 5.10. Find numbers a, b, c so that following

$$O = \begin{bmatrix} 1/3 & 2/3 & a \\ 2/3 & 1/3 & b \\ 2/3 & -2/3 & c \end{bmatrix},$$

is an orthogonal matrix. Then find O^{-1}.

5.2 Hermitian Structures on Complex Vector Spaces

Throughout this section, unless said otherwise, we assume V, W, etc are complex vector spaces.

5.2.1 Complex Structures

Let V be a complex vector space. If we forget the complex structure and only do scalar multiplication by real numbers, then V is also a real vector space, denoted by $V_{\mathbb{R}}$, and is called the underlying real vector space of V. Of course, the underlying sets for V and $V_{\mathbb{R}}$ are the same. To remember the complex structure, we need to observe that the multiplication by $i \equiv \sqrt{-1}$ is an endomorphism on $V_{\mathbb{R}}$ whose square is $-I$.

[5] Hint: apply the Gram-Schmidt process to the columns of A.

[6] The proof is not difficult, just use induction on n.

Exercise 5.11. Let V be a complex vector space and $\dim_{\mathbb{R}} V := \dim V_{\mathbb{R}}$. Show that $\dim_{\mathbb{R}} V = 2\dim V$.

By definition, a **complex structure** on a real vector space is an endomorphism J on the real vector space whose square is $-I$, i.e., $J^2 = -I$.

Exercise 5.12. A (nontrivial) real vector space admits a complex structure if and only if its dimension is an even non-negative integer[7].

Exercise 5.13. 1) Show that the assignment of

$$(V_{\mathbb{R}}, \text{ multiplicaton by } i \text{ on } V)$$

to complex vector space V is a one-one correspondence between the complex vector spaces and pairs $(V_{\mathbb{R}}, J)$ where $V_{\mathbb{R}}$ is a real vector space and J is a complex structure on $V_{\mathbb{R}}$. I.e., a complex vector space is just an underlying real vector space together with a complex structure[8]. 2) Show that a map between two complex vector spaces is a linear map if and only if it is a linear map between the underlying real vector spaces and is compatible with the complex structures.

5.2.2 Hermitian Inner Products

Let V be a complex vector space of dimension n, J be the multiplication by i on V. Then $V_{\mathbb{R}}$ is a real vector space of dimension $2n$ with J being a complex structure. On $V_{\mathbb{R}}$ there exists many inner products, the ones that are compatible with J are the really interesting ones. Here comes a definition.

Definition 5.2 (Hermitian Inner Product: Real Version). We say an inner product $\langle\, ,\, \rangle_{\mathbb{R}}$ on $V_{\mathbb{R}}$ is an **Hermitian inner product** for $V = (V_{\mathbb{R}}, J)$ if it is compatible with the complex structure J in the sense that J is an isometry.

Hermitian inner products always exist : If $\langle\, ,\, \rangle_{\mathbb{R}}$ is an inner product on $V_{\mathbb{R}}$, then

$$(x, y) \mapsto \frac{1}{2}\left(\langle x, y\rangle_{\mathbb{R}} + \langle J(x), J(y)\rangle_{\mathbb{R}}\right)$$

is an Hermitian inner product.

Suppose that $\langle\, ,\, \rangle_{\mathbb{R}}$ is an Hermitian inner product on $(V_{\mathbb{R}}, J)$. For any x, y in $V_{\mathbb{R}}$, define

$$\boxed{\langle x \,|\, y\rangle := \langle x, y\rangle_{\mathbb{R}} - i\langle x, J(y)\rangle_{\mathbb{R}}.}$$

Then the following is just a matter of computations (assume x, y, y_1, y_2 are in V, c_1 and c_2 are complex numbers):

1. $\langle x \,|\, y\rangle = \overline{\langle y \,|\, x\rangle}$;
2. $\langle x \,|\, \cdot\rangle$ is complex linear, i.e., $\langle x \,|\, c_1 y_1 + c_2 y_2\rangle = c_1\langle x \,|\, y_1\rangle + c_2\langle x \,|\, y_2\rangle$;
3. $\langle x \,|\, x\rangle \geq 0$ for any x and $= 0$ if and only if $x = 0$.

[7]Hint: use determinant and work with model spaces.
[8]Hint: work on the model spaces.

On the other hand, if $\langle\,|\,\rangle\colon V \times V \longrightarrow \mathbb{C}$ is a map satisfying the above three properties, set

$$\boxed{\langle\,,\rangle_{\mathbb{R}} = \text{the real part of } \langle\,|\,\rangle,}$$

then $\langle\,,\rangle_{\mathbb{R}}\colon V_{\mathbb{R}} \times V_{\mathbb{R}} \longrightarrow \mathbb{R}$ is an Hermitian inner product. (You prove it)

Obviously, $\langle\,,\rangle_{\mathbb{R}}$ and $\langle\,|\,\rangle$ determine each other. Therefore, we have an equivalent definition of Hermitian inner product for V.

Definition 5.3 (Hermitian Inner Product: Complex Version). An Hermitian inner product on V is a map $\langle\,|\,\rangle\colon V \times V \longrightarrow \mathbb{C}$ which satisfies the three properties listed above.

Remark 5.4. An Hermitian inner product on a complex vector space V is also called an **Hermitian structure** on V. If $u \in V$, the length of u, denoted by $|u|$, is defined to be the square root of $\langle u\,|\,u\rangle = \langle u\,,\,u\rangle_{\mathbb{R}}$.

Example 5.1. Since $\mathbb{C} = \mathbb{R}^2$ plus $J_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, so $C^n = \underbrace{\mathbb{R}^2 \oplus \cdots \oplus \mathbb{R}^2}_{n} = \mathbb{R}^{2n}$ plus $J = \underbrace{J_2 \oplus \cdots \oplus J_2}_{n}$. The dot product is an Hermitian product and it is equivalent to the map sending $(\vec{z}, \vec{w})$ in $\mathbb{C}^n \times \mathbb{C}^n$ to

$$\langle \vec{z}\,|\,\vec{w}\rangle = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n,$$

which is called the **standard Hermitian structure** on $\mathbb{C}^n$. ∎

Exercise 5.14 (Cauchy-Schwartz inequality). Use the three properties listed in the definition of Hermitian inner product (the complex version) to show that

$$|\langle u\,|\,v\rangle| \le |u| \cdot |v|.$$

5.2.3 Orthogonal Projections and Orthonormal Bases

By definition, a (finite dimensional) complex vector space together with an Hermitian inner product is called a (finite dimensional) complex Hilbert space. Equipped with the standard Hermitian structure, $\mathbb{C}^n$ becomes a Hilbert space of dimension n.

Let V be a complex Hilbert space of dimension n, u, v be two vectors in V. We say u, v are **orthogonal** to each other if $\langle u\,|\,v\rangle = 0$.

Exercise 5.15. Let u, v be two vectors in a complex Hilbert space V. Show that if u, v are orthogonal, then, as vectors in $V_{\mathbb{R}}$, u, v are also orthogonal. Also show that $\langle u\,|\,u\rangle = \langle u\,,\,u\rangle_{\mathbb{R}}$ for any $u \in V$.

Let W be a (complex) subspace of V. We say $u \in V$ is orthogonal to W if it is orthogonal to each vector in W, or equivalently if it is orthogonal to each vector in a spanning set for W. Let $W^{\perp}$ be the set of all vectors in V that are orthogonal to W. Similar to the real case, $W^{\perp}$ is a subspace and is called the orthogonal complement of W; and we have

Proposition 5.3. *Let W be a subspace of V. Then*
1) $W \cap W^{\perp} = \{0\}$.
2) $V = W \oplus_i W^{\perp}$. Consequently any $x \in V$ can be uniquely decomposed as

$$x = y + z \qquad y \in W \text{ and } z \in W^{\perp}.$$

3) $(W^{\perp})^{\perp} = W$.

The assignment of y to x is called the **orthogonal projection** of x onto W, denoted by

$$y = \mathrm{Proj}_W(x).$$

An **orthonormal basis** for V is a basis such that any two basis vectors are orthogonal and any basis vector has length equal to 1. Similar to the real case, orthonormal basis always exists. Also, the projection formula and the formula for the Gram-Schmidt process are the same except that we need to replace $\langle\, , \rangle$ by $\langle\, | \,\rangle$ everywhere. Properties for the projection map in the real case all hold in the complex case.

The existence of orthonormal basis for V implies that there is a linear equivalence between V and $\mathbb{C}^n$ which respects the Hermitian inner products. So $\mathbb{C}^n$ is called the model complex Hilbert space of dimension n.

Exercise 5.16. 1) Let V be a complex Hilbert space, $\mathbf{v} = (v_1, \ldots, v_n)$ an orthonormal basis for V. As sets, $V = V_{\mathbb{R}}$. Show that $v_1, \ldots, v_n, iv_1, \ldots, iv_n$ form an orthonormal basis for $V_{\mathbb{R}}$.
2) Let W be a subspace of V. Show that $\mathrm{Proj}_W(x) = \mathrm{Proj}_{W_{\mathbb{R}}}(x)$ for any $x \in V = V_{\mathbb{R}}$.

Exercise 5.17. Let A be an invertible complex matrix. Show that there is a unitary matrix U and a upper triangular matrix T such that $A = UT$, or equivalently there is a unitary matrix U and a lower triangular matrix T such that $A = TU$.

5.2.4 Adjoint Maps and Self-adjoint Operators

Let T be a linear map from (complex) Hilbert space V to (complex) Hilbert space W.

Proposition 5.4. *There is a unique (complex) linear map $T^{\dagger}\colon W \longrightarrow V$ such that, for any $v \in V$ and any $w \in W$, $\langle w \,|\, T(v)\rangle = \langle T^{\dagger}(w) \,|\, v\rangle$. Moreover, $(TS)^{\dagger} = S^{\dagger}T^{\dagger}$ if either side is defined.*

Proof. Just need to work with model Hilbert spaces, so we can take T be a complex $m \times n$- matrix A. The proof is then just a computation: Plugging $w = \vec{e}_i \in W$ and $v = \vec{e}_j$ into $\langle w \,|\, Av\rangle = \langle A^{\dagger}w \,|\, v\rangle$, we are forced to have $A_{ji} = \bar{A}^{\dagger}_{ij}$, i.e., $A^{\dagger}$ is obtained from A by taking transpose and complex conjugation. The rest is clear. □

Here are some terminologies. $T^{\dagger}$ is called the **adjoint** of T. When T is an endomorphism of V and $T^{\dagger} = T$, we say T is **self-adjoint** or T is a self-adjoint operator (or **Hermitian operator**) on V. The standard matrix for an Hermitian operator on $\mathbb{C}^n$ is called an **Hermitian matrix**.

Similar to the real case, we say a linear map is an isometry if it preserves the Hermitian inner products, and such a map must be one-to-one. A **unitary transformation** on V is just an isometric endomorphism, the collection of all such things is called the group of unitary transformations on

V, denoted by $U(V)$. The standard matrix for a unitary transformation on model Hilbert spaces is called a **unitary matrix**. It is clear that it is a complex square matrix whose columns are orthonormal. It is just a matter of computations that U is a unitary matrix $\iff$ the columns are orthonormal basis for $\mathbb{C}^n \iff U^\dagger U = I \iff U^{-1} = U^\dagger \iff UU^\dagger = I \iff$ the rows are orthonormal basis for $(\mathbb{C}^n)^*$.

Exercise 5.18. 1) Prove this last statement.

2) Let $T\colon V \longrightarrow W$ be a linear map. Show that T is a linear isometry if and only if T maps an orthonormal basis for V into an orthonormal set of vectors in W.

Exercise 5.19. Let $U(t)$ be a unitary matrix for each real parameter t. Suppose that $U(t)$ depends on t smoothly and its derivative with respect to t is $\dot{U}$, then $iU^{-1}(t)\dot{U}(t)$ is an Hermitian matrix.

Exercise 5.20. Let W be a subspace of $\mathbb{C}^n$, $\mathbf{x} = (\vec{x}_1, \ldots, \vec{x}_k)$ be an orthonormal basis for W. Then the standard coordinate matrix for Proj_W is $\vec{x}_1\vec{x}_1^\dagger + \cdots + \vec{x}_k\vec{x}_k^\dagger$. (here $\dagger$ means transpose plus complex conjugation)

Exercise 5.21. Let $T\colon V \longrightarrow W$ be a linear map of Hilbert spaces.

1) Show that there is a linear equivalence $T_r\colon (\ker T)^\perp \longrightarrow \mathrm{Im}\, T$ such that

$$T = \left[\begin{array}{c|c} T_r & O \\ \hline O & O \end{array}\right]$$

with respect to decomposition $V = (\ker T)^\perp \oplus_i \ker T$ and $W = \mathrm{Im}\, T \oplus_i (\mathrm{Im}\, T)^\perp$.

2) Show that in the real case $\ker T$ and $\mathrm{Im}\, T'$ are orthogonal complement of each other in V and $\ker T'$ and $\mathrm{Im}\, T$ are orthogonal complement of each other in W, and in the complex case $\ker T$ and $\mathrm{Im}\, T^\dagger$ are orthogonal complement of each other in V and $\ker T^\dagger$ and $\mathrm{Im}\, T$ are orthogonal complement of each other in W.

3) Show that

$$\boxed{\begin{array}{rcll} r(T') & = & r(T) & \text{in the real case;} \\ r(T^\dagger) & = & r(T) & \text{in the complex case.} \end{array}}$$

Exercise 5.22. Let V be a real (complex) Hilbert space of dimension n. Then $\det V$ is a Hilbert space whose inner product (Hermitian inner product) is the one such that the length of $\det(v_1, \ldots, v_n)$ is 1 for an (hence any, why?) orthonormal basis $v_1, \ldots, v_n$.

1) Show that for any $\mathbf{v} = (v_1, \ldots, v_n)$ and $\mathbf{w} = (w_1, \ldots, w_n)$, we have

$$\begin{array}{rcll} \langle \det(\mathbf{v}) \,,\, \det(\mathbf{w}) \rangle & = & \mathrm{Det}\,[\langle v_i \,,\, w_j \rangle] & \text{in the real case;} \\ \langle \det(\mathbf{v}) \,|\, \det(\mathbf{w}) \rangle & = & \mathrm{Det}\,[\langle v_i \,|\, w_j \rangle] & \text{in the complex case.} \end{array}$$

2) Let $T\colon V \longrightarrow W$ be a linear map and $\dim V = \dim W$. Then $\det T\colon \det V \longrightarrow \det W$ is a linear map between one-dimensional Hilbert spaces. Show that

$$\begin{array}{rcll} \det T' & = & (\det T)' & \text{in the real case;} \\ \det T^\dagger & = & (\det T)^\dagger & \text{in the complex case.} \end{array}$$

3) Let T be an endomorphism on a Hilbert space. Then $\det T$ is just the scalar multiplication by $\mathrm{Det}T$. Show that

$$\boxed{\begin{array}{lll} \mathrm{Det}T' &=& \mathrm{Det}T \quad \text{in the real case;} \\ \mathrm{Det}T^\dagger &=& \overline{\mathrm{Det}T} \quad \text{in the complex case.} \end{array}}$$

Exercise 5.23. Let T be an endomorphism on vector space V. Suppose that V admits a decomposition into the direct sum of two subspaces of V, with respect to which, $T = T_1 \oplus T_2$. Show that

1) T_1 and T_2 are self-adjoint if T is self-adjoint.

2) T is self-adjoint if the two subspaces in the decomposition are orthogonal to each other and T_1, T_2 are self-adjoint.

Here comes a theorem that has many applications in geometry and physics.

Theorem 5.2 (Spectral Decomposition Theorem for hermitian operators). *Let T be a self-adjoint operator on V. Then 1) the eigenvalues of T are all real numbers; 2) eigenspaces of T with distinct eigenvalues are orthogonal; 3) V is the direct sum of all eigenspaces of T, consequently T is decomposable into a direct sum of scalar multiplications.*

Proof. 1) Suppose that λ is an eigenvalue and $T(v) = \lambda v$ for some $v \neq 0$. Plugging $T(v) = \lambda v$ into $\langle v \,|\, T(v)\rangle = \langle T(v) \,|\, v\rangle$, we have $\lambda|v|^2 = \bar\lambda|v|^2$, so $v \neq 0$ implies that $\lambda = \bar\lambda$, i.e., λ is a real number.

2) Suppose that for $i = 1$ or 2, we have $T(v_i) = \lambda_i v_i$ for some real numbers λ_i's and nonzero vectors v_i's. Plugging these relations into $\langle v_1 \,|\, T(v_2)\rangle = \langle T(v_1) \,|\, v_2\rangle$, we get $(\lambda_2 - \lambda_1)\langle v_1 \,|\, v_2\rangle = 0$, so $\lambda_2 \neq \lambda_1$ implies that $\langle v_1 \,|\, v_2\rangle = 0$.

3) Let W be the (internal) direct sum of all eigenspace of T, then W is also the span of the set of all eigenvectors of T. Claim: $V = W$. Otherwise, $W^\perp$ is nontrivial. By simple computations we know that $T(W) \subset W$ and $T(W^\perp) \subset W^\perp$, so $T = T_1 \oplus T_2$ with respect to the decomposition $V = W \oplus_i W^\perp$. Now T_2 is an endomorphism on $W^\perp$ (non-trivial), so it has at least one eigenvalue(since we are working with complex numbers), so it has an eigenvector v. But this v is also an eigenvector of T, and is not in W, a contradiction. □

Corollary 5.3. *Let A be an Hermitian matrix, i.e., $A^\dagger = A$. Then there is a unitary matrix U whose columns consist of eigenvectors of A and a real diagonal matrix D whose diagonals are the corresponding eigenvalues, such that $A = UDU^\dagger$.*

(The slightly nontrivial part is that U can be chosen to be unitary. That is because that for each eigenspaces we can choose an orthonormal basis, so the juxtaposition of these bases form an orthonormal basis for V (because distinct eigenspaces are orthogonal).)

Corollary 5.4. *Let A be a real symmetric matrix. Then there is an orthogonal matrix O whose columns consist of eigenvectors of A and a real diagonal matrix D whose diagonals are the corresponding eigenvalues, such that $A = ODO'$.*

Exercise 5.24. 1) Prove this corollary[9]. 2) State and prove a theorem similar to theorem 5.2 for real self-adjoint operator. 3) We say a square matrix A is antisymmetric if $A' = -A$. Show that

[9] Hint: if a complex column matrix is an eigenvector of A, then either its real part or its imaginary part is nonzero and must be an eigenvector, too.

a real antisymmetric matrix is always diagonalizable over the complex numbers. Use examples to demonstrate that this is not true over the real numbers.

Exercise 5.25 (A problem in quantum mechanics or calculus). 1) Let H be an Hermitian operator. Show that the maximum (minimum) value of $\langle x \mid H(x) \rangle$ subject to constraint $|x| = 1$ is the maximum (minimum) eigenvalue of H. 2) Let A be a real symmetric matrix. Show that the maximum (minimum) value of $\langle \vec{x}\,,\, A\vec{x} \rangle$ subject to constraint $|\vec{x}| = 1$ is the maximum (minimum) eigenvalue of A.

Exercise 5.26 (Characterization of orthogonal projections). Let P be an endomorphism of real (complex) Hilbert space V. Show that P is an orthogonal projection $\iff$ $P' = P$ ($P^\dagger = P$) and $P^2 = P$.

Exercise 5.27. Let V be a complex Hilbert space, and T be an endomorphism on $V_{\mathbb{R}}$. Show that T is an Hermitian operator on V if and only if T is a real self-adjoint operator on $V_{\mathbb{R}}$ and $TJ = JT$ where J is the scalar multiplication by $i = \sqrt{-1}$.

Exercise 5.28 (Lie Algebra). Let V be a complex vector space. We say A is an **anti-hermitian operator** on V if $A^\dagger = -A$. Let $su(V)$ be the set of all anti-hermitian traceless operators on V.

1) Show that[10] $su(V)$ is a real vector space. Express the dimension of $su(V)$ in terms of the dimension of V.

2) Let A, B be two anti-hermitian traceless operators. The bracket of A with B, is defined to be $AB - BA$, and is denoted by $[A, B]$. Show that $[A, B]$ is also an anti-hermitian traceless operator on V.

3) Show that the bracket defines a map from $su(V) \times su(V)$ to $su(V)$ which is bilinear, anti-symmetric and satisfies the jacobian identity:

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0.$$

I.e, show that $su(V)$ is a Lie Algebra.

4) Show that $(A, B) \mapsto \operatorname{tr} A^\dagger B$ defines a Riemannian structure on $su(V)$.

5) Repeat everything above with the assumption that V is a real vector space, you will get something which is commonly denoted by $so(V)$.

5.2.5 Simultaneous Diagonalizations

We say two endomorphisms T and S are commuting if $TS = ST$. Here comes a theorem that is important in both quantum mechanics and Lie algebras.

Theorem 5.5. *Suppose that $H_1, \ldots, H_k$ are mutually commuting Hermitian operators on Hilbert space V. Then V is the direct sum of their mutually orthogonal common eigenspaces, consequently $H_1, \ldots, H_k$ can be simultaneously decomposed into a direct sum of scalar multiplications by real numbers.*

[10] Hint: show that $su(V)$ is a subspace of $\mathrm{Hom}(V, V)$ by using Ex. 3.2

Proof. Induction on k. $k = 1$ is just part of theorem 5.2. Assume it is true for $k = i$, i.e., V is an orthogonal direct sum of the common eigenspaces of $H_1, \ldots, H_i$: $V = W_1 \oplus W_2 \oplus \cdots$. Suppose that W is one of these common eigenspaces with eigenvalues $\lambda_1, \ldots, \lambda_i$ respectively, so $v \in W \iff H_j(v) = \lambda_j v$ for $1 \le j \le i$. Then I claim $H_{i+1}(W) \subset W$, that is because, if $v \in W$, then $H_j(H_{i+1}(v)) = H_{i+1}(H_j(v)) = H_{i+1}(\lambda_j v) = \lambda_j H_{i+1}(v)$. Therefore, H_{i+1} is decomposable into a direct sum of endomorphisms on $W_1, W_2, \ldots$, in fact a direct sum of Hermitian operators on $W_1, W_2, \ldots$. Now apply theorem 5.2 to each of this Hermitian operators, we get a decomposition of each W_j into a direct sum of the common eigenspaces of $H_1, \ldots, H_i, H_{i+1}$. Therefore, V has a decomposition into a direct sum of the common eigenspaces of $H_1, \ldots, H_i, H_{i+1}$. □

Remark 5.5. In application to quantum mechanics, these H_i's represent physics observables, one of which is the Hamiltonian of the system, and others are symmetry generators of the system; and the eigenvalues are called **quantum numbers**.

Corollary 5.6. *Suppose that $A_1, \ldots, A_k$ are mutually commuting Hermitian $n \times n$-matrices. Then there are real diagonal $n \times n$-matrices $D_1, \ldots, D_k$ and a unitary $n \times n$-matrix U whose columns are the common eigenvectors, such that*

$$A_i = U D_i U^\dagger.$$

Exercise 5.29. Let A be a complex square matrix. Show that[11] $A^\dagger A = A A^\dagger \iff A$ is diagonalizable by a unitary matrix: $A = UDU^\dagger$ where D is diagonal and U is unitary.

This exercise implies that *every unitary matrix is diagonalizable by a unitary matrix.* Now if $U = PDP^\dagger$ where U, P are unitary matrices and D is diagonal, then D is also unitary matrix, and that implies that the *eigenvalues of a unitary matrix must all have absolute value equal to* 1.

Since Hermitian matrices are the infinitesimal version of unitary matrices, the simultaneous diagonalization theorem also hold for unitary matrices.

Theorem 5.7. *Suppose that $U_1, \ldots, U_k$ are mutually commuting unitary transformations on Hilbert space V. Then V is the direct sum of their mutually orthogonal common eigenspaces, i.e., $U_1, \ldots, U_k$ can be simultaneously decomposed into a direct sum of scalar multiplications by complex numbers of modulus one.*

Corollary 5.8. *Suppose that $U_1, \ldots, U_k$ are mutually commuting unitary $n \times n$-matrices. Then there are unitary diagonal $n \times n$-matrices $D_1, \ldots, D_k$ and a unitary $n \times n$-matrix U whose columns are the common eigenvectors, such that*

$$U_i = U D_i U^\dagger.$$

Exercise 5.30. Prove the last theorem and its corollary.

[11] Hint: write $A = X + iY$ where both X and Y are Hermitian matrices.

5.3 Real Quadratic Forms

A **real quadratic form** is just a degree two homogeneous real polynomial in real variables. Since it is quadratic, does it really belong to linear algebra? Yes, it does.

Let A be a real symmetric matrix, $\vec{x} = [x_1, \ldots, x_n]'$. Then

$$Q_A(\vec{x}) := \langle \vec{x},\, A\vec{x} \rangle$$

is a real quadratic form on $\mathbb{R}^n$ (here the inner product is the dot product). In fact, this assignment of a quadratic form to a real symmetric matrix defines a one-one correspondence:

$$\boxed{\begin{array}{c} \text{real symmetric } n \times n\text{-matrices} \\ \updownarrow \\ \text{real quadratic forms in } n \text{ variables} \end{array}}$$

Exercise 5.31. Verify the last statement.

Now if A is a real symmetric matrix, then corollary 5.4 says that there is a real diagonal matrix $D = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ and an orthogonal matrix O such that $A = ODO'$. Therefore,

$$\begin{aligned} Q_A(\vec{x}) &= \langle \vec{x},\, ODO'\vec{x} \rangle \\ &= \langle O'\vec{x},\, DO'\vec{x} \rangle \qquad \text{by proposition 5.2} \\ &= Q_D(\vec{y}) = \lambda_1 {y_1}^2 + \cdots + \lambda_n {y_n}^2 \end{aligned} \tag{5.4}$$

where $\vec{y} = O'\vec{x}$. I.e., up to a rotation (yes, you can choose O such that $\det O = +1$), any quadratic form is just a one that has no mixed terms.

Many applications involving quadratic form is precisely due to this fact. For example, if Q_A is **positive-definite** (i.e., $Q_A(\vec{x}) \geq 0$ for any $\vec{x}$ and $= 0$ only if $\vec{x} = \vec{0}$), then to prove this oft quoted equality

$$\int_{\mathbb{R}^n} (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}Q_A(\vec{x})}\, dx_1 \cdots dx_n = \frac{1}{\sqrt{\mathrm{Det}\, A}}.$$

one just need to prove it in the special case when A is diagonal — a much simpler case.

Here is an example in geometry.

Example 5.2. [Quadric Surfaces] A quadric surface is a surface in $\mathbb{R}^3$ that can be represented as the solution set of a quadratic polynomial equation in real variables x, y and z. For example, the unit sphere centered at the origin of $\mathbb{R}^3$ is a quadric surface because it is the solution set of equation $x^2 + y^2 + z^2 - 1 = 0$. The interesting question here is to classify the quadric surfaces up to topological shape, i.e., up to translations, invertible linear transformations and scalings along certain directions. Of course, we are only interested in the non-degenerate one, i.e., something that is not obtained by sweeping a curve along a fixed direction. It turns out there are only six of them, all but one are obtained by rotating a curve around a fixed axis, so it is very easy to sketch them. The remaining one, the only one that is difficult to sketch, is a saddle; it is called a saddle because it looks like a saddle that you would sit on when you ride a horse. The model equations for these six quadric surfaces are

1. $x^2 + y^2 + z^2 = 1$;
2. $x^2 + y^2 - z^2 = 1$;
3. $x^2 + y^2 - z^2 = -1$;
4. $x^2 + y^2 - z^2 = 0$;
5. $z = x^2 + y^2$;
6. $z = x^2 - y^2$.

I.e., start with any quadratic polynomial equation in x, y and z, after some coordinate changes involving translations, invertible linear transformations, scalings, the equation becomes one of the six listed above provided that the surface represented by the equation is non-degenerate. ∎

Exercise 5.32. Verify the last statement.

For application to finding maxima and minima in Calculus course, we need an efficient way to decide when a real quadratic form Q_A is positive definite. In dimension two, that is easy. For example, if $Q(x, y) = ax^2 + bxy + cy^2$, then from high school algebra we know Q is

1. indefinite if $b^2 - 4ac > 0$;
2. positive definite if $a > 0$ and $b^2 - 4ac < 0$;
3. negative definite if $a < 0$ and $b^2 - 4ac < 0$.

However, in higher dimensions, you need this theorem: (For simplicity, we write $Q_A > 0$ or $A > 0$ to means that Q_A is positive definite.)

Theorem 5.9. *Let A be a real symmetric $n \times n$-matrix. Let A_i be the matrix obtained from A by deleting its last $(n - i)$ rows and last $(n - i)$ columns. Let $D_i = \operatorname{Det} A_i$ (so $D_n = \operatorname{Det} A$). Then*

$$\boxed{Q_A > 0 \iff D_1,\, D_2,\, \ldots,\, D_n \text{ are all positive.}}$$

Proof. $\Rightarrow$: First we observe that Eq. (5.4) implies that $Q_B > 0 \iff$ the eigenvalues of B are all positive. So $\operatorname{Det} B = \operatorname{Det}(ODO^{-1}) = \operatorname{Det} D > 0$ if $Q_B > 0$. Now if we restrict Q_A to subspace $\mathbb{R}^i \times 0$, we get $Q_{A_i} > 0$, so $\operatorname{Det} A_i > 0$, i.e., $D_i > 0$.

$\Leftarrow$: Induction on n. It is clearly true when $n = 1$. Assume it is true when the dimension is $n-1$; so $D_1, \ldots, D_{n-1}$ are positive implies that $Q_{A_{n-1}} > 0$, so $A_{n-1} = ODO^{-1}$ for some orthogonal $(n-1) \times (n-1)$-matrix O and some diagonal matrix $D = \operatorname{diag}\{\lambda_1, \ldots, \lambda_{n-1}\}$ with λ_i's > 0. Then

$$\begin{aligned} A &= \left[\begin{array}{c|c} O & 0 \\ \hline 0 & 1 \end{array}\right] \left[\begin{array}{c|c} D & B \\ \hline B' & d \end{array}\right] \left[\begin{array}{c|c} O' & 0 \\ \hline 0 & 1 \end{array}\right] \\ &= \left[\begin{array}{c|c} O\sqrt{D} & 0 \\ \hline 0 & 1 \end{array}\right] \left[\begin{array}{c|c} I & C \\ \hline C' & d \end{array}\right] \left[\begin{array}{c|c} \sqrt{D}O' & 0 \\ \hline 0 & 1 \end{array}\right] \end{aligned}$$

where d is a real number, $\sqrt{D} = \mathrm{diag}\{\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_{n-1}}\}$ and B, C are column vector in $\mathbb{R}^{n-1}$ with $B = \sqrt{D}C$. Since

$$\begin{aligned} 0 < \mathrm{Det}\, A &= \mathrm{Det}\, D\, \mathrm{Det} \left[\begin{array}{c|c} I & C \\ \hline C' & d \end{array}\right] \\ &= \mathrm{Det}\, D \cdot (d - |C|^2) \qquad \text{by Ex. 4.9}\,, \end{aligned}$$

we have $d - |C|^2 > 0$. Now

$$\begin{aligned} A > 0 \quad \iff \quad & \left[\begin{array}{c|c} I & C \\ \hline C' & d \end{array}\right] > 0 \iff \left[\begin{array}{c|c} \vec{y}' & y_n \end{array}\right] \left[\begin{array}{c|c} I & C \\ \hline C' & d \end{array}\right] \left[\begin{array}{c} \vec{y} \\ \hline y_n \end{array}\right] \\ = \quad & |\vec{y}|^2 + 2y_n C \cdot \vec{y} + y_n^2 d \\ = \quad & |\vec{y} + y_n C|^2 + y_n^2(d - |C|^2) \quad \text{is positive for all nonzero } [\vec{y}', y_n]; \end{aligned}$$

but that is guaranteed by the positivity of $d - |C|^2$.

□

Exercise 5.33. Let A be a real symmetric $n \times n$-matrix. We say $A \geq 0$ if $Q_A(\vec{x}) \geq 0$ for all vector $\vec{x}$ in $\mathbb{R}^n$. It is clear that $A \geq 0 \iff A + tI > 0$ for any $t > 0$. Based on this observation and the theorem above to derive a sufficient and necessary condition for $A \geq 0$.

Exercise 5.34. 1) Define real quadratic forms on a general real vector space and then translate theorem 5.9.

2) Define[12] Hermitian quadratic forms on a general complex vector space and then state and prove a theorem similar to the translated version of theorem 5.9 that you have just obtained in part 1).

Exercise 5.35 (Polar Decomposition). Let A be an invertible complex matrix. Show that there is a matrix $R > 0$[13] and a unitary matrix U such that $A = RU$.

[12]Hint: they correspond to real self-adjoint operators on $V_{\mathbb{R}}$ which are compatible with the complex structure on V, i.e., Hermitian operators on V.

[13]It means that R is Hermitian and $\langle \vec{z} | R\vec{z} \rangle \geq 0$ for any $\vec{z}$ and $= 0$ only when $\vec{z} = 0$.

Chapter 6

Jordan Canonical Forms*(1 week)

Let T be an endomorphism on a (finite dimensional) complex vector space V. We wish to decompose T into a direct sum of scalar multiplications. It turns out that this is equivalent to decomposing V into the direct sum of all eigenspaces. However, this is not always possible because in general the direct sum of all eigenspaces is just a proper subspace of V, and we have seen that in example 4.3. In that example, the only eigenspace $E_1 = \mathrm{Nul}\,(B - I)$ is a proper subspace of $\mathbb{R}^2$. However, $\mathrm{Nul}\,(B - I)^2 = \mathbb{R}^2$, and this is a big hint.

The general picture is this: *V is always equal to the internal direct sum of generalized eigenspaces; with respect to this decomposition, T is a direct sum of endomorphisms of this form: a scalar multiplication + a nilpotent endomorphism.*

The details are given in the rest of this chapter.

6.1 Generalized Eigenspaces

Let λ be an eigenvalue of an endomorphism T and $E_\lambda(i) = \ker(T - \lambda I)^i$. Then we have a filtration

$$\{0\} \subsetneq E_\lambda(1) \subset E_\lambda(2) \subset \cdots .$$

Since the filtration is bounded above by V, it must stabilize from somewhere on, i.e., there is a positive integer $k > 0$ such that

$$E_\lambda(k) = E_\lambda(k+1) = \cdots .$$

Let us call this stabilized subspace in the filtration the **generalized eigenspace** with eigenvalue λ, denoted by $\mathscr{E}_\lambda$. I.e., $\mathscr{E}_\lambda = \varinjlim \ker(T - \lambda I)^i$. Note that $E_\lambda = E_\lambda(1) \subset \mathscr{E}_\lambda$. By definition, any nonzero vector in $\mathscr{E}_\lambda$ is called a **generalized eigenvector**. Note that, for any integer $n \geq 0$ and any complex number μ, $(T - \mu I)^n$ maps $\mathscr{E}_\lambda$ into $\mathscr{E}_\lambda$; and we will see shortly that it actually maps $\mathscr{E}_\lambda$ isomorphically onto $\mathscr{E}_\lambda$ if $\mu \neq \lambda$.

Proposition 6.1. *Let v be a generalized eigenvector of T with eigenvalue λ. Then*

1. There is an integer $m \geq 0$ such that $v_m := (T - \lambda I)^m(v)$ is an eigenvector of T with eigenvalue λ.

2. v is never a generalized eigenvector of T with eigenvalue $\mu \neq \lambda$.

3. For any $n \geq 0$ and complex number $\mu \neq \lambda$, $(T-\mu I)^n(v)$ is always a generalized vector of T with eigenvalue λ, in fact an eigenvector of T with eigenvalue λ if v is. Consequently, $(T-\mu I)^n$ maps $\mathscr{E}_\lambda$ isomorphically onto $\mathscr{E}_\lambda$.

Proof. 1. That is because there is an integer $m \geq 0$ such that $(T-\lambda I)^m(v) \neq 0$ but $(T-\lambda I)^{m+1}(v) = 0$, i.e., $T((T-\lambda I)^m(v)) = \lambda(T-\lambda I)^m(v)$.

2. Otherwise, by point 1, there is an integer $l \geq 0$ such that $v_l := (T-\mu I)^l(v)$ is an eigenvector of T with eigenvalue μ. Since

$$(T-\lambda I)^k(v_l) = (T-\mu I)^l((T-\lambda I)^k(v)) = (T-\mu I)^l(0) = 0$$

if k is sufficiently large, v_l is also a generalized eigenvector of T with eigenvalue λ, so by point 1, there is an integer $m \geq 0$ such that $(T-\lambda I)^m(v_l)$ is an eigenvector of T with eigenvalue λ. Now

$$T((T-\lambda I)^m(v_l)) = (T-\lambda I)^m(T(v_l)) = (T-\lambda I)^m(\mu v_l) = \mu(T-\lambda I)^m(v_l),$$

so $(T-\lambda I)^m(v_l)$ is an eigenvector of T with both eigenvalue λ and μ, a contradiction according to Ex. 4.17.

3. Simple computation shows that $(T-\mu I)^n(v) \in \mathscr{E}_\lambda$ and $\in E_\lambda$ if $v \in E_\lambda$. So we just need to show that $(T-\mu I)^n(v) \neq 0$ for any $n \geq 0$. Suppose there is an $n \geq 0$ such that $(T-\mu I)^n(v) = 0$, then $v \in \mathscr{E}_\mu$, but that is impossible according to point 2.

□

Corollary 6.1. *The sum of all generalized eigenspaces is a direct sum.*

Proof. Prove by contradiction. Suppose that $v_1, \ldots, v_k$ are generalized eigenvectors of T with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ such that

$$v_1 + v_2 + \cdots + v_k = 0.$$

According to proposition 6.1, there are integers $n_1, \ldots, n_k$ such that when we apply $(T-\lambda_1 I)^{n_1} \cdots (T-\lambda_k I)^{n_k}$ to the above equality we get

$$u_1 + u_2 + \cdots + u_k = 0$$

where $u_1, \ldots, u_k$ are eigenvectors of T with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. But that is impossible according to Ex. 4.17. □

6.2 Jordan Canonical Forms

The hope now is, while V is not always the internal direct sum of all eigenspaces of T, it is always the internal direct sum of all generalized eigenspaces of T. And this is indeed true.

Theorem 6.2. *Let T be an endomorphism on complex vector space V. Then V is the internal direct sum of all generalized eigenspaces of T.*

Proof. Let $\mathscr{E}$ be the internal direct sum of all generalized eigenspaces of T. Try to show that $x \in V \Rightarrow x \in \mathscr{E}$.

Write the characteristic polynomial of T as $(\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_k)^{n_k}$, where λ_i's are the distinct eigenvalues of T, and n_i's are their corresponding multiplicity. The Cayley-Hamilton Theorem (see Ex. 4.19) says that $(T - \lambda_1 I)^{n_1} \cdots (T - \lambda_k I)^{n_k} = 0$, so

$$(T - \lambda_1 I)^{n_1} \cdots (T - \lambda_k I)^{n_k}(x) = 0,$$

therefore, $(T - \lambda_2 I)^{n_2} \cdots (T - \lambda_k I)^{n_k}(x) \in \mathscr{E}_{\lambda_1}$. Now point 3 of proposition 6.1 says that $(T - \lambda_2 I)^{n_2} \cdots (T - \lambda_k I)^{n_k}$ is an invertible endomorphism on $\mathscr{E}_{\lambda_1}$, therefore, there is $w_1 \in \mathscr{E}_{\lambda_1}$ such that

$$(T - \lambda_2 I)^{n_2} \cdots (T - \lambda_k I)^{n_k}(x) = (T - \lambda_2 I)^{n_2} \cdots (T - \lambda_k I)^{n_k}(w_1),$$

i.e.,

$$(T - \lambda_2 I)^{n_2} \cdots (T - \lambda_k I)^{n_k}(x - w_1) = 0.$$

Continuing the above argument, it is clear that there are $w_2 \in \mathscr{E}_{\lambda_2}, \ldots, w_k \in \mathscr{E}_{\lambda_k}$ such that

$$\begin{aligned} (T - \lambda_2 I)^{n_2} \cdots (T - \lambda_k I)^{n_k}(x - w_1) &= 0, \\ (T - \lambda_3 I)^{n_3} \cdots (T - \lambda_k I)^{n_k}(x - w_1 - w_2) &= 0, \\ &\vdots \\ x - w_1 - w_2 - \cdots - w_k &= 0. \end{aligned}$$

Therefore $x = w_1 + w_2 + \cdots + w_k \in \mathscr{E}$. $\square$

Corollary 6.3. *Let T be an endomorphism on complex vector space V, $\mathscr{E}_{\lambda_i}$'s $(1 \le i \le k)$ the distinct generalized eigenspaces of T. Then there are nilpotent endomorphism N_i on $\mathscr{E}_{\lambda_i}$ for each i, such that, with respect to the decomposition $V = \mathscr{E}_{\lambda_1} \oplus \cdots \oplus \mathscr{E}_{\lambda_k}$, we have*

$$T = (\lambda_1 I + N_1) \oplus \cdots \oplus (\lambda_k I + N_k).$$

In view of theorem 3.2, we have

Corollary 6.4. *Let A be a complex square matrix. Then there is an invertible matrix P such that $A = PDP^{-1}$ where D is a direct sum of Jordan Blocks. The columns of P are generalized eigenvectors and the diagonals of D are the corresponding eigenvalues. Moreover, D is unique up to permutations of the Jordan blocks contained inside D.*

Remark 6.1. D is called a **Jordan canonical form** of A.

Exercise 6.1. Show that the dimension of the generalized eigenspace with eigenvalue λ is equal to the algebraic multiplicity of λ as a root of the characteristic polynomial. Therefore, the dimension of an eigenspace is at most the algebraic multiplicity of the corresponding eigenvalue as a root of the characteristic polynomial.

Exercise 6.2. Show that an endomorphism is a direct sum of scalar multiplication if and only if its every eigenspace is also a generalized eigenspace.

Appendix A

Advanced Part

The materials in this appendix are not included in a traditional course of linear algebra. However, we are compelled to add them here due to their natural appearances and applications in modern geometry, quantum many-body theory and quantum field theory.

In our view, it is worth the effort to introduce some efficient modern mathematical languages at an early stage. Therefore, we devote the first section to the introduction of categories and functors even though they are not part of the advanced linear algebra.

A.1 Categories and Functors

A **category** consists of a class[1] of objects and a set $\mathrm{Mor}(X, Y)$ for each pair of objects (X, Y); moreover, to any $f \in \mathrm{Mor}(X, Y)$ and $g \in \mathrm{Mor}(Y, Z)$, a unique element in $\mathrm{Mor}(X, Z)$ (called the composition of g with f, denoted by gf) is assigned, such that 1) the composition is associative, 2) there exists a composition identity 1_X in $\mathrm{Mor}(X, X)$ for each X. I.e., $h(gf) = (hg)f$ whenever both sides make sense, and $1_Y f = f = f 1_X$ for any $f \in \mathrm{Mor}(X, Y)$.

Categories are often denoted by $\mathscr{C}$, $\mathscr{D}$, etc. If $\mathscr{C}$ is a category, its class of objects is denoted by $\mathrm{Ob}(\mathscr{C})$. The elements in $\mathrm{Mor}(X, Y)$ are called **morphisms**. In case $\mathrm{Ob}(\mathscr{C})$ is a set, we say $\mathscr{C}$ is a **small category**.

Example A.1. $\mathscr{S}$ — the category of sets and set maps. Here $\mathrm{Ob}(\mathscr{S})$ is the class of all sets, $\mathrm{Mor}(X, Y)$ is the set of set maps from set X to set Y. Composition is the composition of set maps.

Example A.2. $\mathscr{S}'$ — the category of finite sets and set maps. Here $\mathrm{Ob}(\mathscr{S}')$ is the class of all finite sets, $\mathrm{Mor}(X, Y)$ is the set of set maps from set X to set Y. Composition is the composition of set maps.

Example A.3. $\mathscr{Z}$ — the category of integers. The objects are integers. Let m, n be integers. If $m \le n$, $\mathrm{Mor}(m, n) = \{m \le n\}$, otherwise $\mathrm{Mor}(m, n)$ is the empty set. This is a small category.

Example A.4. $\mathscr{V}$—the category of real vector spaces and linear maps. The objects are real vector spaces and the morphisms are linear maps.

[1] We use the word 'class', not 'set', in order to avoid the logical contradiction.

Example A.5. $\mathscr{V}'$—the category of real vector spaces and injective linear maps. The objects are real vector spaces and the morphisms are injective linear maps.

Example A.6. $\mathscr{V}''$—the category of real vector spaces and surjective linear maps. The objects are real vector spaces and the morphisms are surjective linear maps.

A.1.1 Functors

Let $\mathscr{C}$, $\mathscr{D}$ be two categories. We say $F\colon \mathscr{C} \longrightarrow \mathscr{D}$ is a **covariant functor** if to each object X in $\mathscr{C}$, a unique object $F(X)$ in $\mathscr{D}$ is assigned, and to each morphism $f \in \mathrm{Mor}(X, Y)$, a unique morphism $F(f) \in \mathrm{Mor}(F(X), F(Y))$ is assigned, such that $F(gf) = F(g)F(f)$ if g, f are two composable morphisms in $\mathscr{C}$, and $F(1_X) = 1_{F(X)}$ for any object X in $\mathscr{C}$.

Example A.7. The forgetful functor $F\colon \mathscr{V} \longrightarrow \mathscr{S}$ is a covariant functor. F maps a vector space to its underlying set and a linear map to its underlying set map. I.e., F simply forgets the linear structures.

Example A.8. Dimension dim: $\mathscr{V}' \longrightarrow \mathscr{Z}$ is a covariant functor.

Example A.9. Let V be a real vector space. Then $\mathrm{Hom}\,(V, \ )\colon \mathscr{V} \longrightarrow \mathscr{V}$ is a covariant functor. Here, $\mathrm{Hom}\,(V, \ )$ maps vector space W to vector space $\mathrm{Hom}\,(V, W)$.

Let $\mathscr{C}$, $\mathscr{D}$ be two categories. We say $F\colon \mathscr{C} \longrightarrow \mathscr{D}$ is a **contra-variant functor** if to each object X in $\mathscr{C}$, a unique object $F(X)$ in $\mathscr{D}$ is assigned, and to each morphism $f \in \mathrm{Mor}(X, Y)$, a unique morphism $F(f) \in \mathrm{Mor}(F(Y), F(X))$ is assigned, such that $F(gf) = F(f)F(g)$ if g, f are two composable morphisms in $\mathscr{C}$, and $F(1_X) = 1_{F(X)}$ for any object X in $\mathscr{C}$.

Example A.10. Fun: $\mathscr{S}' \longrightarrow \mathscr{V}$ is a contra-variant functor. Here, for a finite set X, $\mathrm{Fun}(X)$ is the vector space of real functions on X; and for a set map $f\colon X \longrightarrow Y$, $\mathrm{Fun}(f)$ is the map that sends $\phi \in \mathrm{Fun}(Y)$ to ϕf.

Example A.11. Dimension dim: $\mathscr{V}'' \longrightarrow \mathscr{Z}$ is a contra-variant functor.

Example A.12. Let V be a real vector space. Then $\mathrm{Hom}\,(, V)\colon \mathscr{V} \longrightarrow \mathscr{V}$ is a contra-variant functor. Here, $\mathrm{Hom}\,(, V)$ maps vector space W to vector space $\mathrm{Hom}\,(W, V)$.

A.1.2 Natural Transformations

Let F, $G\colon \mathscr{C} \longrightarrow \mathscr{D}$ be two covariant functors. A **natural transformation** ϕ from F to G is an assignment that assigns a morphism $\phi(X) \in \mathrm{Mor}(F(X), G(X))$ to each $X \in \mathscr{C}$, such that, for any $f \in \mathrm{Mor}(X, Y)$, the following square

$$\begin{tikzcd} F(X) \arrow[r, "F(f)"] \arrow[d, "\phi(X)"'] & F(Y) \arrow[d, "\phi(Y)"] \\ G(X) \arrow[r, "G(f)"'] & G(Y) \end{tikzcd}$$

is commutative.

Example A.13. Let V_1 and V_2 be two real vector spaces and f: $V_2 \longrightarrow V_1$ be a linear map. Then f induces a natural transformation f^* from $\mathrm{Hom}\,(V_1,\ )$ to $\mathrm{Hom}\,(V_2,\ )$: For any vector space W, $f^*(W)$: $\mathrm{Hom}\,(V_1, W) \longrightarrow \mathrm{Hom}\,(V_2, W)$ is the linear map sending ϕ: $V_1 \longrightarrow W$ to ϕf: $V_2 \longrightarrow W$. It is clear that if f, g are two linear maps such that fg is well-defined, then $(fg)^* = g^* f^*$.

Let F, G: $\mathscr{C} \longrightarrow \mathscr{D}$ be two contra-variant functors. A natural transformation ϕ from F to G is an assignment that assigns a morphism $\phi(X) \in \mathrm{Mor}(F(X), G(X))$ to each $X \in \mathscr{C}$, such that, for any $f \in \mathrm{Mor}(X, Y)$, the following square

$$\begin{tikzcd} F(Y) \arrow[r, "F(f)"] \arrow[d, "\phi(Y)"'] & F(X) \arrow[d, "\phi(X)"] \\ G(Y) \arrow[r, "G(f)"'] & G(X) \end{tikzcd}$$

is commutative.

Example A.14. Let V_1 and V_2 be two real vector spaces and f: $V_1 \longrightarrow V_2$ be a linear map. Then f induces a natural transformation f_* from $\mathrm{Hom}\,(, V_1)$ to $\mathrm{Hom}\,(, V_2)$: For any vector space W, $f_*(W)$: $\mathrm{Hom}\,(W, V_1) \longrightarrow \mathrm{Hom}\,(W, V_2)$ is the linear map sending ϕ: $W \longrightarrow V_1$ to $f\phi$: $W \longrightarrow V_2$. It is clear that if f g are two linear maps such that fg is well-defined, then $(fg)_* = f_* g_*$.

A.2 Algebras

By definition, an **algebra** over $\mathbb{R}$ is a pair (V, m) where V is a real vector space and m: $V \times V \longrightarrow V$ is a bilinear map, called the algebra multiplication. It is customary to write $m(a, b)$ as ab for any a, b in V. We say the algebra is **associative** if $(ab)c = a(bc)$ for any a, b and c in V. We say the algebra is **commutative** if $ab = ba$ for any a, b in V. We say the algebra has a unit if there is an element e in V such that $ea = ae = a$ for any a in V.

Let A, B be two algebras over $\mathbb{R}$, T: $A \longrightarrow B$ be a set map. We say T is an **algebra homomorphism** if T respects the algebra structures, i.e., T is a linear map and $T(ab) = T(a)T(b)$ for any a, b in A. If both A and B have unit, we further require T map the unit of A to the unit of B.

Example A.15 (Endomorphism Algebra). Let V be a vector space, then $\mathrm{End}(V)$ — the set of all endomorphisms on V is also a vector space. If T, S are two endomorphisms on V, we define TS to be the composition of T with S. Then $\mathrm{End}(V)$ becomes an associative algebra with unit, called the endomorphism algebra on V.

Example A.16 (Polynomial Algebra). Let $\mathscr{P}$ be the set of all real polynomials in t, then $\mathscr{P}$ is an algebra. The algebra multiplication is just the multiplication of polynomials. It is an associative and commutative algebra with unit over $\mathbb{R}$. It is a **graded algebra** indexed by $\mathbb{Z}_{\geq 0}$ because $\mathscr{P} = \mathscr{P}(0) \oplus \mathscr{P}(1) \oplus \cdots$ where $\mathscr{P}(i)$ is the the set of degree i homogeneous polynomials in t and the multiplication maps $\mathscr{P}(i) \times \mathscr{P}(j)$ to $\mathscr{P}(i+j)$. More generally, let $\mathbb{R}[x_1, x_2, \ldots, x_n]$ be the set of real polynomials in $x_1, \ldots, x_n$, then $\mathbb{R}[x_1, x_2, \ldots, x_n]$ is a graded associative and commutative algebra with unit over $\mathbb{R}$.

Example A.17. Let V be a vector space, $\mathrm{gl}(V)$ be the set of all endomorphisms on V with this multiplication: $m(S,T) = ST - TS$. Then $\mathrm{gl}(V)$ is a non-associative, non-commutative algebra without unit. It is a **Lie algebra** — an algebra whose product satisfies the following Lie conditions:

1) Skew Symmetric: $m(a,b) = -m(a,b)$ for any a, b in the algebra.

2) Jacobi Identity: $m(a, m(b,c)) + m(b, m(c,a)) + m(c, m(a,b)) = 0$ for any a, b and c in the algebra.

Let $sl(V)$ be the subset of $gl(V)$ consisting of traceless endomorphisms on V. Then $sl(V)$ is also a Lie algebra.

Example A.18. $\mathbb{R}$ is a commutative, associative algebra with unit over $\mathbb{R}$. $\mathbb{C}$ is a commutative, associative algebra with unit over $\mathbb{R}$.

Example A.19 (The quaternion algebra). $\mathbb{Q}$ be $\mathbb{R}^4$ equipped with this product: $m(\vec{e}_1, \vec{e}_i) = m(\vec{e}_i, \vec{e}_1) = \vec{e}_i$ for $1 \le i \le 4$, $m(\vec{e}_2, \vec{e}_3) = -m(\vec{e}_3, \vec{e}_2) = \vec{e}_4$, $m(\vec{e}_3, \vec{e}_4) = -m(\vec{e}_4, \vec{e}_3) = \vec{e}_2$, $m(\vec{e}_4, \vec{e}_2) = -m(\vec{e}_2, \vec{e}_4) = \vec{e}_3$, and $m(\vec{e}_2, \vec{e}_2) = m(\vec{e}_2, \vec{e}_2) = m(\vec{e}_2, \vec{e}_2) = m(\vec{e}_2, \vec{e}_2) = -\vec{e}_1$. $\mathbb{Q}$ is an associative and non-commutative algebra with unit. It is a real **division algebra**, i.e., an algebra with unit over $\mathbb{R}$ whose every non-zero element has a multiplicative inverse. Both $\mathbb{R}$ and $\mathbb{C}$ are real division algebra. There is one more real division algebra, the Cayley algebra, it is a 8-dimensional real vector space with a non-associative and non-commutative multiplication.

A.3 Multi-linear Maps and Tensor Products

Let X, Y and Z be *finite* dimensional vector spaces. Suppose that f: $X \times Y \longrightarrow Z$ is a bilinear map. The family of all possible bilinear maps with $X \times Y$ as their domain is huge; however, there are universal objects in this family. By definition, a **universal object** in this family is just a bilinear map F: $X \times Y \longrightarrow U$ such that, for any bilinear map f: $X \times Y \longrightarrow Z$, there is a **unique** linear map f_U from U to Z which makes this triangle

$$\begin{tikzcd} X \times Y \arrow[r, "F"] \arrow[dr, "f"'] & U \arrow[d, dotted, "f_U"] \\ & Z \end{tikzcd} \tag{A.1}$$

commutative: $f = f_U F$. If such universal objects do exist, then they must be unique up to linear equivalences: suppose that F': $X \times Y \longrightarrow U'$ is another universal object, then from commutative diagram

$$\begin{tikzcd} X \times Y \arrow[r, "F"] \arrow[dr, "F'"'] & U \arrow[d, shift left, "F'_U"] \\ & U' \arrow[u, shift left, "F_{U'}"] \end{tikzcd} \tag{A.2}$$

we must have both $F'_U F_{U'}$ and $F_{U'} F'_U$ equal to identity. So $U \cong U'$. A universal object F: $X \times Y \longrightarrow U$, or simply U, is called the **tensor product** of X with Y, denoted by $X \otimes Y$, and is unique up to linear equivalence. It is customary to write $F(u,v)$ as $u\otimes v$. Note that the image of F can be seen to be a spanning set for U and F is bilinear means that $(c_1u_1+c_2u_2)\otimes v = c_1(u_1\otimes v)+c_2(u_2\otimes v)$ and $u\otimes(c_1v_1+c_2v_2) = c_1(u\otimes v_1)+c_2(u\otimes v_2)$.

Claim 3. *There does exist a universal object, i.e., the tensor product does exist.*

There are many proofs, the next exercise gives a proof.

Exercise A.1. Let $F\colon \mathbb{R}^m\times\mathbb{R}^n \longrightarrow \mathcal{M}_{m\times n}$ be given by

$$(\vec{x},\vec{y})\mapsto \vec{x}\vec{y}',$$

here $\vec{x}\vec{y}'$ is the matrix product of column matrix $\vec{x}$ with row matrix $\vec{y}'$. Show that
1) F is a bilinear map.
2) $\{F(\vec{e}_i,\vec{e}_j)\,|\,1\le i\le n, 1\le j\le m\}$ is a minimal spanning set for $\mathcal{M}_{m\times n}$.
3) $(\mathcal{M}_{m\times n}, F)$ defines the tensor product of $\mathbb{R}^m$ with $\mathbb{R}^n$.
4) Claim 3 is true.

Exercise A.2. Let X, Y and U be vector spaces and $F\colon X\times Y\longrightarrow U$ be a bilinear map. Show that (U,F) defines the tensor product of X with Y $\iff$ 1) $\dim U = \dim X\,\dim Y$, 2) F is "onto", i.e., U is the span of the range of F.

Exercise A.3. Let X, Y be vector spaces, and $\mathrm{Map}^{BL}(X\times Y,\mathbb{R})$ be the set of bilinear maps from $X\times Y$ to $\mathbb{R}$. Let $F\colon X^*\times Y^*\longrightarrow \mathrm{Map}^{BL}(X\times Y,\mathbb{R})$ be the map

$$(f,g)\mapsto \text{the bilinear map: } (x,y)\mapsto f(x)\cdot g(y).$$

Show that
1) $\mathrm{Map}^{BL}(X\times Y,\mathbb{R})$ is a vector space.
2) F is bilinear.
3) $(\mathrm{Map}^{BL}(X\times Y,\mathbb{R}),F)$ defines the tensor product of X^* with Y^*.

Exercise A.4. Let X, Y be vector spaces. Show that there is a natural bilinear map $F\colon X\times Y \longrightarrow \mathrm{Map}^{BL}(X^*\times Y^*,\mathbb{R})$ such that $(\mathrm{Map}^{BL}(X^*\times Y^*,\mathbb{R}),F)$ defines the tensor product of X with Y.

Exercise A.5. Let X, Y be vector spaces. Show that there is a natural bilinear map $F\colon X\times Y \longrightarrow \mathrm{Hom}(X^*,Y)$ such that $(\mathrm{Hom}(X^*,Y),F)$ defines the tensor product of X with Y.

Exercise A.6. Let X, Y be two vector spaces over K. Let $\mathrm{span}(X\times Y)$ be the set of all finitely supported K-valued functions on $X\times Y$. It is more convenient to write a function f in $\mathrm{span}(X\times Y)$ as a formal finite sum:

$$\sum_{z\in X\times Y,\, f(z)\neq 0} f(z)z.$$

Let R be the subset of $\mathrm{span}(X\times Y)$ consisting of elements of either the form

$$c_1(x_1,y)+c_2(x_2,y)-(c_1x_1+c_2x_2,y)$$

or the form

$$c_1(x, y_1) + c_2(x, y_2) - (x, c_1 y_1 + c_2 y_2),$$

where c_1, c_2 are in K, x, y, x_1, x_2, y_1 and y_2 are in V. Let $T(X, Y) = \mathrm{span}(X \times Y)/\mathrm{span}R$, F: $X \times Y \longrightarrow T(X, Y)$ be the map sending (x, y) to $(x, y) + \mathrm{span}R$. Show that $(F, T(X, Y))$ is the tensor product of X with Y.

Exercise A.7. Show that all the linear equivalences below are natural:

1. ι_X: $X \cong X^{**}$,
2. $\sigma(X, Y)$: $X \otimes Y \cong Y \otimes X$,
3. $\psi(X, Y)$: $\mathrm{Hom}(X, Y) \cong X^* \otimes Y$,
4. $D(X, Y, Z)$: $X \otimes (Y \oplus_e Z) \cong X \otimes Y \oplus_e X \otimes Z$.

$\vdots$

Here natural means natural in the categorical sense, i.e., certain appropriate square diagram is commutative. For example, If T: $X \longrightarrow Y$ is a linear map, then this square

$$\begin{array}{ccc} X & \xrightarrow{\iota_X} & X^{**} \\ {\scriptstyle T}\downarrow & & \downarrow{\scriptstyle T^{**}} \\ Y & \xrightarrow[\iota_Y]{} & Y^{**} \end{array} \tag{A.3}$$

is commutative.

More generally, if $V_1, \ldots, V_N$ and U are vector spaces, F: $V_1 \times \cdots \times V_N \longrightarrow U$ is a multi-linear map, we say (U, F), or simply U is the tensor product of $V_1, \ldots, V_N$, written as $U = V_1 \otimes \cdots \otimes V_N$, if for any multi-linear map f: $V_1 \times \cdots \times V_N \longrightarrow W$ there is a unique linear map f_U: $U \longrightarrow W$ such that $f = f_U F$.

Exercise A.8. Show that all the linear equivalences below are natural:
1) $(X \otimes Y) \otimes Z \cong X \otimes Y \otimes Z$;
2) $X \otimes (Y \otimes Z) \cong X \otimes Y \otimes Z$.

A.4 Tensor Algebras

Let V be a vector space, V^* be the dual space of V. Let

$$T^r_s V = \overbrace{V \otimes \cdots \otimes V}^{r} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{s}. \tag{A.4}$$

(In geometry, when V is a tangent space, elements of V, V^* and $T^r_s V$ are called **tangent vectors**, **cotangent vectors** and **tensors** of type (r, s) respectively.)

We write T^rV for T^r_0V. By convention T^0V is the ground field K for V. We write T^*V for the (external) direct sum of all T^nV where n runs over $\mathbb{Z}_+$—the set of non-negative integers. T^*V is then a $\mathbb{Z}_+$-graded vector space, in fact a $\mathbb{Z}_+$-graded associative algebra with unit over K: the product (called tensor product, denoted by $\otimes$) is uniquely determined by the following natural multiplication rule: Let $u_1, \ldots, u_m, v_1, \ldots, v_n$ be in V, then

$$(u_1 \otimes u_2 \otimes \cdots \otimes u_m) \otimes (v_1 \otimes u_2 \otimes \cdots \otimes v_n) = u_1 \otimes u_2 \otimes \cdots \otimes u_m \otimes v_1 \otimes v_2 \otimes \cdots \otimes v_n.$$

Note that the multiplication is really the natural identification map: $T^rV \otimes T^sV \cong T^{r+s}V$.

Let Σ_n be the permutation group of n objects. By the universal property of tensor product, it is easy to see that Σ_n acts on T^nV. Let

$$S^nV := \left\{ t \in \overbrace{V \otimes \cdots \otimes V}^{n} \mid \sigma(t) = t \text{ for any } \sigma \in \Sigma_n \right\}, \tag{A.5}$$

it is called the **totally symmetrical tensor product** of n copies of V. Let

$$\wedge^n V = \left\{ t \in \overbrace{V \otimes \cdots \otimes V}^{n} \mid \sigma(t) = \operatorname{sign}(\sigma)t \text{ for any } \sigma \in \Sigma_n \right\}, \tag{A.6}$$

it is called the **totally anti-symmetrical tensor product** of n copies of V.

Let S be the **symmetrization operator**, i.e., the following endomorphism of T^nV: for $t \in T^nV$,

$$S(t) = \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \sigma(t). \tag{A.7}$$

It is clear that $S^2 = S$, and S^nV is both the image of T^nV under S and the kernel of $(S - I)$.

Let A be the **anti-symmetrization operator**, i.e., the following endomorphism of T^nV: for $t \in T^nV$,

$$A(t) = \frac{1}{n!} \sum_{\sigma \in \Sigma_n} \operatorname{sign}(\sigma)\sigma(t). \tag{A.8}$$

It is clear that $A^2 = A$, and $\wedge^nV$ is both the image of T^nV under A and the kernel of $(A - I)$.

Let $u_1, \ldots, u_m$ be in V, it is customary to write $u_1u_2 \cdots u_m$ for $S(u_1 \otimes u_2 \otimes \cdots \otimes u_m)$ and $u_1 \wedge u_2 \wedge \cdots \wedge u_m$ for $A(u_1 \otimes u_2 \otimes \cdots \otimes u_m)$. So, for u, v in V, we have

$$uv = \frac{1}{2}(u \otimes v + v \otimes u), \quad u \wedge v = \frac{1}{2}(u \otimes v - v \otimes u).$$

Note that $T^r_sV = T^rV \otimes T^sV^*$. It is customary to call an element of T^nV^* an n-form on V, an element of S^nV^* a symmetric n-form on V, and an element of $\wedge^nV^*$ an anti-symmetric n-form on V.

We write S^*V for the direct sum of all S^nV. It is a $\mathbb{Z}_+$-graded associative and commutative algebra with unit over K. We write $\wedge^*V$ for the direct sum of all $\wedge^nV$. It is a $\mathbb{Z}_+$-graded associative algebra with unit over K—the exterior algebra on V. The products are defined naturally: if $u \in S^mV$, $v \in S^nV$, then $uv = S(u \otimes v)$; if $u \in \wedge^mV$, $v \in \wedge^nV$, then $u \wedge v = A(u \otimes v)$.

Exercise A.9. Show that

1) Let $u_1, \dots, u_m, v_1, \dots, v_n$ be in V, then

$$(u_1 u_2 \cdots u_m)(v_1 u_2 \cdots v_n) = u_1 u_2 \cdots u_m v_1 v_2 \cdots v_n,$$

and

$$(u_1 \wedge u_2 \wedge \cdots \wedge u_m) \wedge (v_1 \wedge u_2 \wedge \cdots \wedge v_n) = u_1 \wedge u_2 \wedge \cdots \wedge u_m \wedge v_1 \wedge v_2 \wedge \cdots \wedge v_n,$$

2) The products on S^*V and $\wedge^* V$ are all associative.

3) $uv = vu$ where u, v are elements of S^*V.

4) $u \wedge v = (-1)^{|u|\,|v|} v \wedge u$, here u, v are homogeneous elements of $\wedge^* V$ and $|u|$ and $|v|$ are the degree of u and v respectively.

Exercise A.10. 1) Let V be a vector space. Fixing a basis on V, then vectors in V can be represented by its coordinates $x_1, x_2, \dots$. Show that $S^n V^*$ is naturally identified with the vector space of homogeneous polynomials of degree n in x_i's. (So $S^n V^*$ is often called the vector space of homogeneous polynomials on V of degree n.)

2) Let U, V be vector spaces. Show that $S^n(U \oplus V) \cong \oplus_{i+j=n} S^i U \otimes S^j V$ naturally and $\wedge^n (U \oplus V) \cong \oplus_{i+j=n} \wedge^i U \otimes \wedge^j V$ naturally.

3) If V is a vector space, we use $\dim V$ to denote the dimension of V. Show that, if t is a formal variable, then

$$\sum_{n \geq 0} t^n \dim S^n V = \left(\frac{1}{1-t}\right)^{\dim V}, \quad \sum_{n \geq 0} t^n \dim \wedge^n V = (1+t)^{\dim V}. \tag{A.9}$$

Exercise A.11. 1) Assume r is a non-negative integer. Show that T^r, S^r and $\wedge^r$ are covariant functors from the category of vector spaces and linear maps to itself.

2) Show that T^*, S^* and $\wedge^*$ are covariant functors from the category of vector spaces and linear maps to the category of $\mathbb{Z}_+$-graded algebras and algebra homomorphisms.

3) Let T be an endomorphism on V, $\wedge^k T$ be the induced endomorphism on $\wedge^k V$. Show that

$$\det(I + tT) = \sum_k t^k \operatorname{tr} \wedge^k T. \tag{A.10}$$

4) Show that, for any positive integer m, $\operatorname{tr} T^m$ can be expressed in terms of $\operatorname{tr} T$, $\operatorname{tr} \wedge^2 T, \dots,$ $\operatorname{tr} \wedge^{\dim V} T$.

5) Show that, for any positive integer $m \leq \dim V$, $\operatorname{tr} \wedge^m T$ can be expressed in terms of $\operatorname{tr} T$, $\operatorname{tr} T^2, \dots, \operatorname{tr} T^m$.

Index