LINEAR ALGEBRA II

DE-QI ZHANG (OWNER OF COPYRIGHT)

(Professor of Mathematics; Office: S17-0608)

9th August 2022

Contents

1 Assessment: HWs, Mid-term test, Final Exam; References; Syllabus

Assessment

(1) There are 10 tutorials. Tutorial (1) is conducted during the Academic week 3 (22 - 26 Aug 2022). Let $X = 1, 2, 3, 4, 5$. The assignments in Tutorial $(2X - 1)$ includes two questions (50 marks each) as Homework X.

Pls submit Homework X by 11:59pm on (mostly) the Monday of (mostly) the week for Tutorial $(2X - 1)$. E.g. pls submit Homework 1 by 11:59pm of Monday, 22nd Aug 2022. See file ma2101sum.doc or Schedule-Lect-HW-LinAlgII.doc for more details.

The 5 HWs will contribute **[10%].**

(2) Mid-term Test on Wed 4.10pm – 5.10pm, 19th

Oct (tentative), covering Ch 1- 6 and Tut 1- 5, at LT26, during the lecture time and at the lecture venue. **[30%].**

(3) Final examination in November. **[60%].**

References

(1) (**Main Textbook**) Stephen H. Friedberg, Arnold J. Insel and Lawrence E. Spence, Linear Algebra, Prentice Hall; for instance: 4th edition, year 2003.

Make sure it contains Ch7. Canonical form.

http://www.amazon.com/Linear-Algebra-Edition-Stephen-Friedberg/dp/0130084514

(2) (**Reference only**) Howard Anton (or with Chris Rorres), Elementary Linear Algebra (or Applications version), John Wiley & Sons.

Prerequisites:

MA2001 Linear Algebra I, or MA1101R or MA2001 or MA1506 or MA1508 or MA1508E or MA1513, is fully assumed.

MA2001 Linear Algebra I, Module description (from NUSMODS):

This module is a first course in linear algebra. Fundamental concepts of linear algebra will be introduced and investigated in the context of the Euclidean spaces $\mathbb{R}^n$. Proofs of results will be presented in the concrete setting. Students are expected to acquire computational facilities and geometric intuition with regard to vectors and matrices. Some applications will be presented. Major topics: Systems of linear equations, matrices, determinants, Euclidean spaces, linear combinations and linear span, subspaces, linear independence, bases and dimension, rank of a matrix, inner products, eigenvalues and eigenvectors, diagonalization, linear transformations between Euclidean spaces, applications.

MA2101 Linear Algebra II, Module description (from NUSMODS):

This module is a continuation of MA1101 Linear Algebra I intended for second year students. The student will learn more advanced topics and concepts in linear algebra. A key difference from MA1101 is that there is a greater emphasis on conceptual understanding and proof techniques than on computations. Major topics: Matrices over a field. Determinant. Vector spaces. Subspaces. Linear independence. Basis and dimension. Linear transformations. Range and kernel. Isomorphism. Coordinates. Representation of linear transformations by matrices. Change of basis. Eigenvalues and eigenvectors. Diagonalizable linear operators. Cayley-Hamilton Theorem. Minimal polynomial. Jordan canonical form. Inner product spaces. Cauchy-Schwartz inequality. Orthonormal basis. Gram-Schmidt Process. Orthogonal complement. Orthogonal projections. Best approximation. The adjoint of a linear operator. Normal and self-adjoint operators. Orthogonal and unitary operators.

2 Vector spaces over a field (T1)

2.1. In this section we define **rings**, **fields** and **vector spaces** over a field F and give many examples, including: Euclidean n-space $\mathbb{R}^n$, row vector n-space F^n, column vector n-space F_c^n, matrix space $M_{m\times n}(F)$ and polynomial ring $F[x]$ (as a vector space).

2.2. (**The notation for the standard sets**) In these notes, we will use the following notation:

(1) The set of **integers**:

$$\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, 3, \cdots\}$$

(2) The set of **natural numbers** (or positive integers):

$$\mathbb{N} = \{1, 2, 3, \cdots\}$$

(3) The set of nonnegative integers:

$$\mathbb{Z}_{\geq 0}$$

(4) The set of **rational numbers** (or the field of rational numbers):

$$\mathbb{Q} = \{\frac{m}{n};\, m, n \text{ are integers, } n \neq 0\}$$

(5) The set of positive rational numbers:

$$\mathbb{Q}_{>0}.$$

(6) The set of real numbers (or the filed of real numbers):

$$\mathbb{R}$$

(7) The set of positive real numbers:

$$\mathbb{R}_{>0}.$$

(8) The set of **complex numbers** (or the field of complex numbers):

$$\mathbb{C} = \{x + \sqrt{-1}\,y \,|\, x, y \in \mathbb{R}\}$$

Definition 2.3. (**Fields, rings and groups**)

Let F be a set containing at least two elements and equipped with the following two binary operations $+$ (the addition, or plus) and $\times$ (the multiplication, or times) where $F \times F := \{(x, y) \mid x, y \in F\}$ is the **product set** of F with itself:

$$\begin{aligned} + : F \times F &\to F \\ (x, y) &\mapsto x + y; \\ \times : F \times F &\to F \\ (x, y) &\mapsto x \times y \text{ (or simply } xy). \end{aligned}$$

Axiom(0). It is very important to note that we assume here that the two operations are well defined on F in the sense that:

$$\forall x \in F, \forall y \in F \Rightarrow x + y \in F$$

$$\forall x \in F, \forall y \in F \Rightarrow xy \in F.$$

Namely, the operation map $+$ (resp. $\times$) takes element (x, y) in the domain $F \times F$ to some element

$x + y$ (resp. xy) in the codomain F.

The set F or a bit precisely the triplet $(F, +, \times)$ or more precisely the quintuple $(F, +, 0; \times, 1)$ with two distinguished elements 0 (the additive identity) and 1 (the multiplicative identity), is called a **field** if the following **Eight Axioms** (and also **Axiom(0)**) are satisfied.

(1) Existence of an **additive identity** 0_F or simply 0:

$$x + 0 = x = 0 + x, \ \ \forall \ x \in \ F.$$

(2) (Additive) Associativity:

$$(x + y) + z = x + (y + z), \ \ \forall \ x, y, z \ \in F.$$

(3) Additive inverse:

for every $x \in F$, there is an **additive inverse** $-x \in F$ **of** x (or the **negative of** x) such that

$$x + (-x) = 0 = (-x) + x.$$

(4) Existence of a **multiplicative identity** 1_F or simply 1:

$$x1 = x = 1x, \quad \forall\, x \in F.$$

(5) (Multiplicative) Associativity:

$$(xy)z = x(yz), \quad \forall\, x, y, z \in F.$$

(6) Multiplicative inverse for nonzero element:

for every $0 \neq x \in F$, there is a **multiplicative inverse** $x^{-1} \in F$ **of** x such that

$$xx^{-1} = 1 = x^{-1}x.$$

(7) Distributive law:

$$(x + y)z = xz + yz, \quad \forall\, x, y, z \in F,$$

$$z(x + y) = zx + zy, \quad \forall\, x, y, z \in F.$$

(8) Commutativity for addition and multiplication:

$$x + y = y + x, \quad \forall\, x, y \in F,$$

$$xy = yx, \qquad \forall\, x, y \in F.$$

The triplet $(F, +, \times)$ with only the Axioms (1) - (5) and (7) - (8) satisfied is called a (commutative) **ring**.

The pair $(F, +)$ with only Axioms (1) - (3) satisfied by its binary operation $+$, is called an (additive) **group**. (The existence of another operation $\times$ is not needed here).

Notation 2.4. (**about $F^\times$**) For a field $(F, +, 0; \times, 1)$, we use $F^\times$ to denote the set of nonzero elements in F:

$$F^\times := F \setminus \{0\}.$$

Example 2.5. (**Fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$**)

(1) Let $F = \mathbb{Q}$ (the set of rational numbers), $\mathbb{R}$ (the set of real numbers), or $\mathbb{C}$ (the set of complex numbers).

Then F together with the usual arithmetic operations $+$, $\times$ and usual elements 0, 1, becomes a field.

They are respectively called the field of rational numbers (when $F = \mathbb{Q}$), the field of real numbers (when $F = \mathbb{R}$), and the field of complex numbers (when $F = \mathbb{C}$).

(2) Let $F = \mathbb{Z}$ (the set of integers).

Then $(F, +, 0; \times, 1)$ is a ring, but not a field (why?)

The pair $(F, +)$ is an (additive) group.

Example 2.6. (**Polynomial ring $F[x]$**) Let $(F, +, 0; \times, 1)$ be a field (or a ring), e.g., $F = \mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$.

$$g(x) = \sum_{i=0}^{n} a_i x^i = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

with the **leading coefficient** $a_n \neq 0$, is called a **polynomial of degree $n \geq 0$, in one variable**

x **and with coefficients** $a_i \in F$. Let

$$F[x] := \{\sum_{j=0}^{d} b_j x^j \,|\, d \geq 0, b_j \in F\}$$

be the set of all polynomials in one variable x and with coefficients in F.

There are natural addition and multiplication operations for polynomials

$$g(x) = \sum_{i=0}^{r} a_i x^i, \quad h(x) = \sum_{i=0}^{s} b_i x^i$$

defined as

$$g(x) + h(x) = \sum_{i \geq 0} (a_i + b_i) x^i$$
$$g(x)h(x) = \sum_{k \geq 0} c_k x^k$$

where

$$c_k = \sum_{i+j=k} a_i b_j = a_k b_0 + a_{k-1} b_1 + \cdots + a_1 b_{k-1} + a_0 b_k,$$

such that

$$(F[x], +, 0; \times, 1)$$

is a ring, called the **polynomial ring with coefficients in F**. Here if we set $R = F[x]$, then

$$0_R = 0_F, \quad 1_R = 1_F.$$

As an illustration, if

$$g(x) = 3x^2 + 2x - 5, \quad h(x) = 2x^3 + 6$$

are polynomials of degree 2 and 3, respectively, then

$$\begin{aligned} g(x) + h(x) &= (0+2)x^3 + (3+0)x^2 + (2+0)x + 6 - 5 \\ &= 2x^3 + 3x^2 + 2x + 1, \\ g(x)h(x) &= (3x^2 + 2x - 5) \times (2x^3) + (3x^2 + 2x - 5) \times (6) \\ &= (6x^5 + 4x^4 - 10x^3) + (18x^2 + 12x - 30) \\ &= 6x^5 + 4x^4 - 10x^3 + 18x^2 + 12x - 30. \end{aligned}$$

Extra exercise 2.7. (**Polynomial ring in m variables**) Can you define polynomial rings in two variables $F[x, y]$ or even polynomial rings in m variables

$$F[x_1, x_2, \ldots, x_m]?$$

Extra exercise 2.8. (**Multiplicative groups**)

(1) Let $F = \mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. Then $(F^{\times}, \times)$ is a (multiplicative) group (i.e., a group with the multiplication as its operation).

(2) Let $F = \mathbb{Z}$ (the set of integers). Is $(F, \times)$ or $(F^{\times}, \times)$ a (multiplicative) group?

(3) Let $F = \mathrm{N}$ (the set of natural numbers). Then $(F, +)$ is not an (additive) group (why?)

Extra exercise 2.9. (= Tutorial question) (**A field with finitely many elements**) Let p be a prime number and denote

$$F_p := \{0, 1, \ldots, p - 1\}.$$

Can you define two operations $+$ and $\times$ such that $(F, +, \times)$ becomes a field (called the **finite field with p elements**)?

Exercise 2.10. (**Uniqueness of identity and inverse**) Let F or more precisely $(F, +, 0; \times, 1)$ be a field.

(1) Show that the field F has only one additive identity: 0. Namely, if $c \in F$ such that

$$x + c = x = c + x, \ \ \forall\, x \in \ F$$

then $c = 0$.

(2) Show that the field F has only one multiplicative identity: 1. Namely, if $d \in F$ such that

$$xd = x = dx, \ \ \forall\, x \ \in \ F$$

then $d = 1$.

(3) Show that every $x \in F$ has only one additive inverse: $-x$. Namely, if $x' \in F$ such that

$$x + x' = 0 = x' + x$$

then $x' = -x$.

(4) Show that every $x \in F^\times = F \setminus \{0\}$ has only one multiplicative inverse: x^{-1}. Namely, if $x'' \in F$ such that

$$xx'' = 1 = x''x$$

then $x'' = x^{-1}$.

Extra exercise 2.11. (**To be inverse to each other**) Let F or more precisely $(F, +, 0; \times, 1)$ be a field (with at least two elements as required in Definition 2.3). Prove:

(1) (Cancelation law) Let $b, x, y \in F$. Then

$$b + x = b + y \ \Rightarrow \ x = y;$$

$$b \neq 0, \ bx = by \ \Rightarrow \ x = y.$$

(2) (Killing power of 0)

$$0x = 0 = x0, \quad \text{for all}\, x \in F.$$

(3) In F (with $|F| \geq 2$ as assumed in the definition),

we have:

$$0_F \neq 1_F$$

which was not so clearly mentioned in Definition 2.3.

(4) If $x \in F^\times = F \setminus \{0\}$, then its multiplicative inverse (it exists by axiom)

$$x^{-1} \in F^\times.$$

(5) If

$$x + x' = 0$$

then x and x' are additive inverse to each other. Namely,

$$x' = -x, \quad x = -x'.$$

(6) If

$$xx'' = 1$$

then x and x'' are multiplicative inverse to each other. Namely,

$$x'' = x^{-1}, \quad x = (x'')^{-1}.$$

Definition 2.12. (**Vector space**) Let F or more precisely $(F, +, 0; \times, 1)$ be a field (you may assume $F = \mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$) and V a non-empty set, with a binary **vector addition operation**

$$+ : V \times V \to V$$
$$(\mathbf{v}_1, \mathbf{v}_2) \mapsto \mathbf{v}_1 + \mathbf{v}_2$$

and **scalar multiplication** operation

$$\times : F \times V \to V$$
$$(c, \mathbf{v}) \mapsto c\mathbf{v}.$$

In the above $V \times V := \{(\mathbf{u}, \mathbf{v}) \,|\, \mathbf{u}, \mathbf{v} \in V\}$ and $F \times V := \{(c, \mathbf{v}) \,|\, c \in F, \mathbf{v} \in V\}$ are the product sets.

Axiom(0). It is very important to note that we assume here that the two operations are well defined in the sense that:

$$\forall\, \mathbf{v}_i \in V, \;\Rightarrow \mathbf{v}_1 + \mathbf{v}_2 \in V$$

$$\forall\, c \in F, \forall\, \mathbf{v} \in V \Rightarrow c\mathbf{v} \in V.$$

Namely, the operation map $+$ takes element $(\mathbf{v}_1, \mathbf{v}_2)$ in the domain $V \times V$ to some element $\mathbf{v}_1 + \mathbf{v}_2$ in the codomain V, and the operation map $\times$ takes element $(c, \mathbf{v})$ in the domain $F \times V$ to some element $c\mathbf{v}$ in the codomain V.

V or more precisely $(V, +)$ is called a **vector space over the field F** if the following **Seven Axioms** (and also Axiom(0)) are satisfied.

(1) Existence of **zero vector** $\mathbf{0}_V$ or simply $\mathbf{0}$ (which is typed in boldface):

$$\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v}, \;\; \forall\, \mathbf{v} \in \; V.$$

(2) (Additive) Associativity:

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}), \;\; \forall\, \mathbf{u}, \mathbf{v}, \mathbf{w} \; \in V.$$

(3) Additive inverse:

for every $\mathbf{v} \in V$, there is an additive inverse $-\mathbf{v}$ of $\mathbf{v}$ (or the negative of $\mathbf{v}$) such that

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0} = (-\mathbf{v}) + \mathbf{v}.$$

(4) The effect of $1 \in F$ on V:

$$1\mathbf{v} = \mathbf{v}, \ \forall \ \mathbf{v} \in V.$$

(5) (Multiplicative) Associativity:

$$(ab)\mathbf{v} = a(b\mathbf{v}), \ \forall \ a, b \in F, \ \mathbf{v} \in V.$$

(6) Distributive law:

$$(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}, \ \forall \ a, b \in F, \ \mathbf{v} \in V$$

$$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}, \ \forall \ a \in F, \ \mathbf{u}, \mathbf{v} \in V.$$

(7) Commutativity for the vector addition:

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}, \ \forall \ \mathbf{u}, \mathbf{v} \in V.$$

One sees that the seven axioms in the definition of a vector space are similar to those in the definition of a ring in Definition 2.3. In fact, the field F with its addition and multiplication operations, is a vector space over F itself.

Remark 2.13. (**Trivial, but useful fact**) Every vector space V contains the zero vector $\mathbf{0}_V$ (or simply $\mathbf{0}$).

Exercise 2.14. (**The killing power of zero**) Let V be a vector space (with zero vector $\mathbf{0}$) over a field F (with $0 \in F$). Then

$$0\,\mathbf{v} = \mathbf{0}, \qquad a\,\mathbf{0} = \mathbf{0}$$

for any vector $\mathbf{v} \in V$ and scalar $a \in F$.

Exercise 2.15. (**Negative of an element**) Let V be a vector space over a field F. Then for any vector $\mathbf{v} \in V$, we have

$$(-1_F)\mathbf{v} = -\mathbf{v}.$$

Similarly, for any $a \in F$, we have

$$(-1_F)a = -a.$$

Definition 2.16. (**Subtractions of vectors and scalars**) Let V be a vector space over a field F. For $\mathbf{u}, \mathbf{v} \in V$, we define

$$\mathbf{u} - \mathbf{v} := \mathbf{u} + (-\mathbf{v}).$$

Similarly, for $a, b \in F$, we define $a - b := a + (-b)$.

Example 2.17. (**Euclidean plane $\mathbb{R}^2$ and 3-space $\mathbb{R}^3$ as vector spaces**) Let

$$\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$$

be the 3-dimensional Euclidean xyz-space with the natural addition

$$\begin{aligned} \mathbf{v}_1 + \mathbf{v}_2 &= (x_1, y_1, z_1) + (x_2, y_2, z_2) \\ &: = (x_1 + x_2, y_1 + y_2, z_1 + z_2) \end{aligned}$$

for $\mathbf{v}_i = (x_i, y_i, z_i) \in V$, and scalar multiplication

$$a\mathbf{v} = a(x, y, z) := (ax, ay, az)$$

for $a \in \mathbb{R}$ and $\mathbf{v} = (x, y, z) \in V$. Then $\mathbb{R}^3$ is a vector space over $\mathbb{R}$.

Similarly, let

$$\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$$

be the 2-dimensional Euclidean xy-plane with the natural addition and scalar multiplication. Then $\mathbb{R}^2$ is a vector space over $\mathbb{R}$.

Example 2.18. (**Non-vector spaces**)

(1) Let

$$L = \{(x, y) \mid x,\ y \ \in \ \mathbf{R},\ x + 2y = 3\}$$

be a line in the xy-plane $\mathbb{R}^2$. Then L is not a vector space over **R** under the usual addition and scalar multiplication for $\mathbf{R}^2$ as given in Example 2.17, because the vector addition operation

$$+ : L \times L \to L$$

$$((x_1, y_1), (x_2, y_2)) \mapsto (x_1 + x_2, y_1 + y_2)$$

is not a well defined binary operation on L (i.e., Axiom(0) is not satisfied). Indeed, given $\mathbf{v}_i = (x_i, y_i) \in L$, the addition $\mathbf{v}_1 + \mathbf{v}_2$ may not be in L, i.e., the addition map $+$ may not take $(\mathbf{v}_1, \mathbf{v}_2)$ in the domain $L \times L$ to the codmain L. To be precise, let

$$\mathbf{v}_1 = (1, 1), \quad \mathbf{v}_2 = (5, -1).$$

Then both $\mathbf{v}_i \in L$. But

$$\mathbf{v}_1 + \mathbf{v}_2 = (6, 0) \ \notin \ L.$$

(2) Let

$$V = \{(x, y) \,|\, x, y \in \mathbf{R}\}$$

and define vector addition and scalar multiplication as follows:

$$(x_1, y_1) \oplus (x_2, y_2) := (x_1 + x_2, y_1 - y_2),$$

$$c(x_1, y_1) := (cx_1, cy_1).$$

Then V is not a vector space over $\mathbf{R}$, because the commutativity axiom for the vector addition $\oplus$ is not satisfied. Indeed, let

$$\mathbf{v}_1 = (1, 2), \quad \mathbf{v}_2 = (2, 3).$$

Then

$$\mathbf{v}_1 \oplus \mathbf{v}_2 = (3, -1) \neq (3, 1) = \mathbf{v}_2 \oplus \mathbf{v}_1.$$

Alternatively, you may give a concrete example to show that the associativity axiom for vector addition is not satisfied.

(3) Let $V = \{(x, y) \,|\, x, y \in \mathbf{R}\}$ and define the vector addition and scalar multiplication as follows:

$$(x_1, y_1) + (x_2, y_2) := (x_1 + y_1, 0)$$
$$c(x_1, y_1) := (cx_1, 0).$$

Then V is not a vector space over $\mathbf{R}$ because the axiom about the effect of $1 \in \mathbb{R}$ is not satisfied.

Indeed, for $\mathbf{v} = (2, 3) \in V$, we have

$$1\mathbf{v} = (2, 0) \neq \mathbf{v}.$$

Example 2.19. (**Row vector n-space F^n; Euclidean row vector n-space $\mathbb{R}^n$**) Let F be a field and

$$V := F^n = \{(x_1, \ldots, x_n) \,|\, x_i \in F\}$$

the set of n-tuples. As in Example 2.17, define the natural addition

$$\begin{aligned}\mathbf{v}_1 + \mathbf{v}_2 &= (x_1, \ldots, x_n) + (y_1, \ldots, y_n)\\ &:= (x_1 + y_1, \ldots, x_n + y_n)\end{aligned}$$

for $\mathbf{v}_1 = (x_1, \ldots, x_n)$, $\mathbf{v}_2 = (y_1, \ldots, y_n) \in V$, and scalar multiplication

$$a\mathbf{v} = a(x_1, \ldots, x_n) := (ax_1, \ldots, ax_n)$$

for $a \in F$ and $\mathbf{v} = (x_1, \ldots, x_n) \in V$. Then V is a vector space over F.

$V = F^n$ is called the **row vector n-space over the field F** or simply the **row n-space over F**.

When $F = \mathbb{R}$, we get the usual **Euclidean row vector n-space $\mathbb{R}^n$**.

Example 2.20. (**Matrix spaces $M_n(F)$, $M_{m\times n}(F)$**) Let F be a field (say $F = \mathbb{Q}, \mathbb{R}$, or $\mathbb{C}$) and let $n \geq 1$. Let

$$M_n(F) := \{A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} ; a_{ij} \in F\}$$

be the set of all $n \times n$ square matrices with entries a_{ij} in F. For matrices $A = (a_{ij}) \in M_n(F)$, $B = (b_{ij}) \in M_n(F)$ and scalar $c \in F$, define the natural matrix addition and scalar multiplication as:

$$A + B := (a_{ij} + b_{ij}) = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & \cdots & a_{1n} + b_{2n} \\ \vdots & \vdots & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{nn} + b_{nn} \end{pmatrix}$$

$$cA := (ca_{ij}) = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nn} \end{pmatrix}.$$

Then $V := M_n(F)$ together with the vector (= matrix) addition and scalar multiplication, becomes a vector space over the field F. We call $M_n(F)$ the **matrix space of $n \times n$ square matrices with entries in the field F.**

When $F = \mathbb{R}$, we get the usual matrix space $M_n(\mathbb{R})$ of all $n \times n$ real matrices which has been studied in Linear algebra I.

Similarly, for given $m \geq 1$ and $n \geq 1$, the set

$$M_{m\times n}(F)$$

of all $m \times n$ matrices $A = (a_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ with entries $a_{ij} \in F$, is equipped with the natural matrix addition and scalar multiplication as above, so that $V := M_{m\times n}(F)$ becomes a vector space over the field F. We call $M_{m\times n}(F)$ the **matrix space of $m \times n$ matrices with entries in the field F**.

Example 2.21. (**Column m-spaces F_c^m, and row n-space F^n again**) For the matrix space $M_{m\times n}(F)$ in Example 2.20, if we set $n = 1$, then we get the usual column vector space denoted as

$$F_c^m := M_{m\times 1}(F) = \{\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} ; a_i \in F\}.$$

We call F_c^m the **column vector m-space over the field F** or simply the **column m-space over F**.

For the matrix space $M_{m\times n}(F)$ in Example 2.20, if we set $m = 1$, then we recover the usual row vector space (as defined in Example 2.19):

$$F^n = M_{1\times n}(F) = \{(a_1, a_2, \ldots, a_n)\,;\, a_i \in F\}.$$

Here, the row vector is denoted as

$$(a_1, a_2, \ldots, a_n)$$

with commas added in between the letters, instead of $(a_1\, a_2\, \ldots\, a_n)$.

Example 2.22. (**Polynomial ring $F[x]$ as a vector space**) Let F be a field and

$$F[x] = \{g(x) = \sum_{i=0}^{s} a_i x^i \mid s \geq 0, a_i \in F\}$$

the polynomial ring over F as defined in Exercise 2.6. Then with the usual scalar multiplication

$$c\sum_{i=0}^{r} b_i x^i := \sum_{i=0}^{r} cb_i x^i$$

and polynomial addition $g(x)+h(x)$ as defined there, $V := F[x]$ becomes a vector space over the field F.

Extra exercise 2.23. (**Polynomial ring in m variable as a vector space**) Let $F[x_1, \ldots, x_m]$ be the polynomial ring in m variables over a field F as in Example 2.7. Show that it is a vector space over the field F, with the usual polynomial addition and scalar multiplication.

Example 2.24. (**Continuous functions**) Let

$$C^0[x]$$

be the set of all real-valued continuous functions. Then it is a vector space over the field $\mathbb{R}$, with the usual addition $f_1(x) + f_2(x)$ and scalar multiplication $cf_1(x)$ for functions $f_i(x)$ and scalar $c \in \mathbb{R}$.

Extra exercise 2.25. (**Product of vector spaces**) Let V_i be two vector spaces over the same field F. Then the product

$$V := V_1 \times V_2 = \{(\mathbf{x}_1, \mathbf{x}_2) \,|\, \mathbf{x}_i \in V_i\}$$

has a natural structure of vector space, using the natural addition:

$$(\mathbf{x}_1, \mathbf{x}_2) + (\mathbf{y}_1, \mathbf{y}_2) := (\mathbf{x}_1 + \mathbf{y}_1, \mathbf{x}_2 + \mathbf{y}_2)$$

and scalar multiplication:

$$\alpha(\mathbf{x}, \mathbf{y}) := (\alpha\mathbf{x}, \alpha\mathbf{y}).$$

3 Vector subspaces

In this section, we introduce **subspaces** of a vector space, and give many examples of them.

Definition 3.1. (**Subspace**) Let V be a vector space over a field F. A non-empty subset $W \subseteq V$ is called a **vector subspace of V** or simply a **subspace of V** if the following two conditions are satisfied:

(CA) (**C**losed under vector **A**ddition)

$$\forall\, \mathbf{w}_i \in W \;\Rightarrow\; \mathbf{w}_1 + \mathbf{w}_2 \;\in\; W,$$

(CS) (**C**losed under **S**calar multiplication)

$$\forall\, a \in F,\ \forall\, \mathbf{w} \in W \;\Rightarrow\; a\mathbf{w} \;\in\; W.$$

Example 3.2. (**Obvious subspaces**) Let V be a vector space over a field F. Then V always has two subspaces

$$V, \quad \{\mathbf{0}\}.$$

We call $\{\mathbf{0}\}$ the **zero subspace of V**.

Remark 3.3. (**Trivial, but useful fact**) Every vector subspace W (indeed being a vector space in its own right, cf. Theorem 3.8) of V contains the zero vector $\mathbf{0}_W$ (cf. Exercise 2.14).

Can you show that $\mathbf{0}_V = \mathbf{0}_W$ (so that V and its subspaces share the same zero vector)?

Example 3.4. (**Lines and planes in $\mathbb{R}^3$ may or may not be subspaces**)

(1) The plane

$$P_1 := \{(x, y, z) \in \mathbb{R}^3 | 2x + y + 3z = 0\}$$

in the Euclidean 3-space $\mathbb{R}^3$ passes through the origin $\mathbf{0} = (0, 0, 0)$. It is a vector subspace of $\mathbb{R}^3$.

In general, a plane P in $\mathbb{R}^3$ is a vector subspace of $\mathbb{R}^3$ if and only if P passes through the origin

$\mathbf{0} = (0,0,0)$; cf. Remark 3.3 and Exercise 3.5 below; see also tutorials.

(2) The line

$L_1 := \{(x,y,z) \in \mathbb{R}^3 \,|\, 2x+y+3z = 0,\ x+3y+z = 0\}$

in the Euclidean 3-space $\mathbb{R}^3$ passes through the origin $\mathbf{0} = (0,0,0)$ (and is the intersection of the plane $2x+y+3z = 0$ and the plane $x+3y+z = 0$). It is a vector subspace of $\mathbb{R}^3$.

In general, a line L in $\mathbb{R}^3$ is a vector subspace of $\mathbb{R}^3$ if and only if L passes through the origin $\mathbf{0} = (0,0,0)$; cf. Remark 3.3 and Exercise 3.5 below; see also tutorials.

(3) The line

$$L = \{(x,y) \in \mathbb{R}^2 | x, y \in \mathbf{R}, x + 2y = 3\}$$

in the Euclidean plane $\mathbb{R}^2$ is not a vector subspace of $\mathbb{R}^2$, because L is not closed under the vector addition. Indeed, let

$$\mathbf{w}_1 := (1, 1), \quad \mathbf{w}_2 := (5, -1)$$

then both $\mathbf{w}_i \in L$, but

$$\mathbf{w}_1 + \mathbf{w}_2 = (6, 0) \notin L.$$

You may also give an example to show that L is not closed under scalar multiplication and hence L is not a vector subspace.

Extra exercise 3.5. (**Vector subspaces of $\mathbb{R}^3$**)
Show that the following are vector subspaces of the Euclidean 3-space $\mathbb{R}^3$:

(1) The origin $\{\mathbf{0} = (0, 0, 0)\}$.

(2) A line L passing through the origin.

Hint. Let $\mathbf{v}$ be the direction vector of the line L.

Then

$$L = \{t\mathbf{v} \mid t \in \mathbb{R}\}.$$

(3) A plane P passing through the origin.

Hint. Let (a, b, c) be the normal direction of P. Then

$$P = \{(x, y, z) \in \mathbb{R}^3 \mid ax + by + cz = 0\}.$$

(4) $\mathbb{R}^3$.

In fact, every vector subspace of $\mathbb{R}^3$ is one of the 4 types above.

Extra exercise 3.6. Formulate and prove the $\mathbb{R}^2$-version of Exercise 3.5.

Definition 3.7. (**Linear combination**) Let V be a vector space over a field F. A vector $\mathbf{w}$ is a **linear combination of vectors** $\mathbf{v}_1, \ldots, \mathbf{v}_m \in V$, if

$$\mathbf{w} = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m$$

for some scalars $a_i \in F$.

Below is the justification to call the W in Definition 3.1 a vector subspace of V, because W itself is indeed a vector space!

Theorem 3.8. *(**Equivalent subspace definition**) Let V be a vector space over a field F and $W \subseteq V$ a non-empty subset. Then the following are equivalent.*

(1) W is a vector subspace of V, i.e., W is closed under vector addition and scalar multiplication, in the sense of Definition 3.1.

(2) W is closed under linear combination:

$$\forall\, a_i \in F, \forall\, \mathbf{w}_i \in W \;\Rightarrow\; a_1\mathbf{w}_1 + a_2\mathbf{w}_2 \in W.$$

(3) W together with the vector addition $+$ and scalar multiplication $\times$ (borrowed from those on V), becomes a vector space.

Proof. For the direction '(2) $\Rightarrow$ (1)', just take $(a_1, a_2) = (1, 1)$ or $(a, 0)$ to conclude the two closedness conditions in Definition 3.1.

For the direction '(1) $\Rightarrow$ (2)', assuming (1) and letting $a_i \in F$ and $\mathbf{w}_i \in W$, we need to show that $a_1\mathbf{w}_1 + a_2\mathbf{w}_2 \in W$. (1) implies that W is closed under scalar multiplication, and hence $a_i\mathbf{w}_i \in W$. Thus $a_1\mathbf{w}_1 + a_2\mathbf{w}_2 \in W$ since W is also closed under vector addition by (1). This proves (2).

For the direction '(1) $\Rightarrow$ (3)', the two closedness conditions in Definition 3.1 guarantee that the two operations $+$ and $\times$ are well defined on W, i.e., Axiom(0) of Definition 2.12 is satisfied. The Axioms (1) - (7) are true for all vectors in W since they are true for all vectors in a larger set V. Thus, W, together with $+$ and $\times$, becomes a vector space.

We leave the direction '(3) $\Rightarrow$ (1)' as a logical exercise for students. $\square$

Example 3.9. (**Homogeneous linear solution sets as subspaces**) Let $A = (a_{ij}) \in M_{m \times n}(F)$ be an $m \times n$ matrix with entries in a field F.

Then the following **null space of A**

$$\mathrm{Null}(A) := \{X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in F_c^n \, ; \, AX = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\}$$

is a vector subspace of the column n-space F_c^n.

Similarly, the set of solutions

$$(x_1, x_2, \ldots, x_n) \in F^n$$

of the system of m homogeneous linear equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

i.e., the set

$$\mathrm{Null}(A)^t := \{(x_1, \ldots, x_n) \in F^n \, ; \; A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \}$$

is a vector subspace of the row vector n-space F^n. Here we denote by $\mathrm{Null}(A)^t$ the transpose of $\mathrm{Null}(A)$.

Use this result to deduce that the line L_1 and plane P_1 in Example 3.4 are subspaces of $\mathbb{R}^3$.

As a concrete example, we apply Gaussian elimination to a particular matrix A below and get the solution set of the linear system (corresponding to

$AX = \mathbf{0}$):

$$A := \begin{pmatrix} 1 & -2 & 3 \\ 2 & -4 & 6 \\ 3 & -6 & 9 \end{pmatrix} \xrightarrow{\text{row op}} \begin{pmatrix} 1 & -2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{aligned} \mathrm{Null}(A)^t &= \{(x_1, x_2, x_3) = (2s - 3t, s, t) \mid s,\, t \in F\} \\ &= \{s(2,1,0) + t(-3,0,1) \mid s,\, t \in F\}. \end{aligned}$$

One can also directly verify that this solution space is indeed a vector subspace of the row vector 3-space F^3.

Exercise 3.10. (**Polynomial ring $\mathbb{R}[x]$ as a subspace of $C^0[x]$**) Show that the polynomial ring $\mathbb{R}[x]$ over the field $\mathbb{R}$ (as in Example 2.6) is a vector subspace of the vector space $C^0[x]$ (as in Example 2.24) of all real-valued continuous functions.

Example 3.11. Let $r \geq 1$. Let

$$C^r[x]$$

be the set of all real-valued functions $f(x)$ which can be differentiated r-times. Show that $C^r[x]$ is a subspace of $C^{r-1}[x]$, and hence a subspace of $C^0[x]$.

Extra exercise 3.12. (**Inhomogeneous solution set may not be a subspace**) Let $A = (a_{ij}) \in M_{m\times n}(F)$ be an $m \times n$ matrix with entries in a field F. Show that for a given column vector $\mathbf{b}$ (of size $m \times 1$) the solution set

$$S := \{X \in F_c^n \,;\, AX = \mathbf{b}\}$$

is a vector subspace of the column vector n-space F_c^n if and only if $\mathbf{b}$ is the zero column vector.

Example 3.13. (**Subspaces of matrix spaces $M_n(F)$ and $M_{m\times n}(F)$**) We fix some notations:

A **diagonal matrix** in $M_n(F)$ is denoted as

$$\mathrm{diag}[a_1, .., a_n] = \begin{pmatrix} a_1 & 0 & \cdots & 0 & 0 \\ 0 & a_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & a_n \end{pmatrix}.$$

$$I_n = \mathrm{Diag}[1, \ldots, 1] = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{n\times n},$$

is the **multiplicative identity matrix** (or simply the **identity matrix**) of $M_n(F)$.

$$\alpha I_n = \begin{pmatrix} \alpha & 0 & \cdots & 0 & 0 \\ 0 & \alpha & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha & 0 \\ 0 & 0 & \cdots & 0 & \alpha \end{pmatrix}_{n\times n} .$$

is a **scalar matrix** in $M_n(F)$.

(1) The set

$$D_n(F) = \{\mathrm{diag}[a_1, \ldots, a_n]\,;\, a_i \in F\}$$

of all diagonal matrices of **order** n over the field F is a vector subspace of $M_n(F)$.

(2) The set

$$\mathrm{UT}_n(F) = \{ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1,n} \\ 0 & a_{22} & \cdots & a_{2,n-1} & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \cdots & 0 & a_{n,n} \end{pmatrix} ; a_{ij} \in F \}$$

of all upper triangular matrices of order n over the field F is a vector subspace of $M_n(F)$.

(3) The set

$$\mathrm{LT}_n(F)$$

of all lower triangular matrices of order n over the field F is a vector subspace of $M_n(F)$.

(4) The set

$$\mathrm{TF}_n(F) = \{A = (a_{ij}) \in M_n(F) \,;\, \mathrm{Tr}(A) = 0\}$$

of all trace free (or trace vanishing) matrices of order n over the field F is a vector subspace of

$M_n(F)$.

Recall that the **trace** of a square matrix $A = (a_{ij}) \in M_n(F)$ is defined as

$$\mathrm{Tr}(A) := a_{11} + a_{22} + \cdots + a_{nn}.$$

(5) The set

$$\mathrm{FC}_n(F) = \{ \begin{pmatrix} 0 & a_{12} & \cdots & a_{1,n} \\ 0 & a_{22} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & a_{m-1,2} & \cdots & a_{m-1,n} \\ 0 & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} ; a_{ij} \in F\}$$

of all first column vanishing matrices in $M_{m\times n}(F)$ is a vector subspace of $M_{m\times n}(F)$.

Example 3.14. (**Polynomials of degree** $< n$) Fix some $n \geq 1$. Let

$$P_n[x] := \{f(x) \in F[x] \, ; \, \deg(f(x)) < n\}$$

be the set all polynomials in $F[x]$ of degree $< n$. Thus

$$P_n[x] = \{\sum_{i=0}^{n-1} a_i x^i \,;\, a_i \in F\}$$
$$= \{a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} \,;\, a_i \in F\}.$$

and it is a vector subspace of $F[x]$.

Theorem 3.15. *(= Tutorial question) (**Intersection of subspaces being a subspace**) Let V be a vector space over a field F and let $W_\alpha \subseteq V$ $(\alpha \in I)$ be vector subspaces of V. Then the intersection*

$$\cap_{\alpha \in I} W_\alpha$$

is again a vector subspace of V.

Remark 3.16. (**Null space** $\mathrm{Null}(A)$ **as the intersection**) Use 'Intersection of subspaces being a subspace' theorem 3.15 to give another approach towards Example 3.9:

First show that

$$W_i := \{(x_1, \ldots, x_n) \in F^n \,;\, a_{i1}x_1 + \cdots + a_{in}x_n = 0\}$$

is a vector subspace of the row vector n-space F^n. Then note that

$$\mathrm{Null}(A)^t = \cap_{i=1}^m W_i.$$

Hence it is a vector subspace of F^n by Theorem 3.15.

Remark 3.17. (**Union of subspaces may not be a subspace**) If W_i are vector subspaces of a vector space V over a field F, then the union

$$\cup_{i \in I} W_i$$

may not be a vector subspace of V. See also Exercise 4.15 below.

Indeed, consider

$$W_1 = \{(x, y) \in \mathbb{R}^2 \,|\, x + 2y = 0\}$$
$$W_2 = \{(x, y) \in \mathbb{R}^2 \,|\, 2x + y = 0\}.$$

Then both W_i are vector subspaces of the Euclidean plane $\mathbb{R}^2$ (as deduced from Example 3.9), but $W_1 \cup W_2$ is not a vector subspace of $\mathbb{R}^2$. To be precise, both

$$\mathbf{w}_1 := (2, -1), \quad \mathbf{w}_2 := (1, -2)$$

are contained in the union $W_1 \cup W_2$, but the vector addition

$$\mathbf{w}_1 + \mathbf{w}_2 = (3, -3)$$

is not contained in the union $W_1 \cup W_2$; thus the union $W_1 \cup W_2$ is not closed under the vector addition and hence it is not a vector subspace of $\mathbb{R}^2$ by the very Definition 3.1.

However, check that $W_1 \cup W_2$ is closed under scalar multiplication.

Similarly, one can show that neither the union of the three axes of $\mathbb{R}^3$ (i.e., the x-axis, y-axis and z-axis), nor the union of the three coordinate planes (i.e., the xy-plane, yz-plane and zx-plane) is a vector subspace of the Euclidean 3-space $\mathbb{R}^3$.

4 Linear spans and Direct sums of subspaces (T2)

In this section, we introduce linear spans, sums and direct sums of subspaces. Since the union of two vector subspaces may not be a vector space any more, as seen in §3, sum of subspaces is the right notion.

Definition 4.1. (**Linear combination and linear span**) Let V be a vector space over a field F. A vector $\mathbf{v} \in V$ is called a **linear combination of some vectors $\mathbf{v}_i \in V$** $(1 \le i \le s)$ if

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_s\mathbf{v}_s$$

for some scalars $a_i \in F$.

Let $S \subseteq V$ be a non-empty subset. The subset $\text{Span}(S) :=$

$\{\mathbf{v} \in V; \mathbf{v}$ is a linear combination of some vectors in $S\}$

of V is called the **vector subspace of V spanned by the subset S** (cf. Theorem 4.4).

In other words, if $\mathbf{u}_1, \ldots, \mathbf{u}_r$ are vectors in S, then every linear combination (of them)

$$b_1\mathbf{u}_1 + \cdots + b_r\mathbf{u}_r$$

(with $b_j \in F$) belongs to $\mathrm{Span}(S)$; conversely, if $\mathbf{u}$ is a vector in $\mathrm{Span}(S)$, then we can write

$$\mathbf{u} = c_1\mathbf{w}_1 + \cdots + c_\ell\mathbf{w}_\ell$$

for some vectors $\mathbf{w}_1, \ldots, \mathbf{w}_\ell$ in S and some scalars $c_k \in F$.

When $S = \{\mathbf{w}_1, \ldots, \mathbf{w}_t\}$ is a finite subset of V, we denote

$$\mathrm{Span}(S) = \mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_t\}$$

and it follows from the definition that

$$\mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_t\} = \{\sum_{j=1}^{t} \beta_j\mathbf{w}_j \mid \beta_j \in F\}.$$

Namely, $\mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_t\}$ consists of all linear combinations of the fixed vectors $\mathbf{w}_1, \ldots, \mathbf{w}_t$.

Let W be a vector subspace of V. If $W = \mathrm{Span}(S)$, we say that S is a **spanning set** (or **generating set**) of W, and W is **spanned** (or **generated**) by S.

Note 1. It may happen that $W = \mathrm{Span}(S) = \mathrm{Span}(T)$ for two different sets $S \neq T$. For instance,

$$\begin{aligned}\mathbb{R}^2 &= \mathrm{Span}(\{(1,0),(0,1)\})\\ &= \mathrm{Span}(\{(1,1),(1,2)\}).\end{aligned}$$

Discussion 4.2. *Discuss why it makes sense to write*

$$\mathrm{Span}\{\mathbf{w}_1, \ldots, \mathbf{w}_t\} = F\mathbf{w}_1 + \cdots + F\mathbf{w}_t.$$

Exercise 4.3. Show that if $S \subseteq T$ then $\mathrm{Span}(S) \subseteq \mathrm{Span}(T)$. Is the converse true?

The result below justifies the notion in Definition

4.1.

Theorem 4.4. *(***Span being a subspace***)*

(1) The subset $\mathrm{Span}(S)$ *of* V *in Definition* 4.1 *is indeed a vector subspace of* V*.*

(2) $\mathrm{Span}(S)$ *is the smallest vector subspace of* V *containing the set* S*: first* $\mathrm{Span}(S)$ *is a vector subspace of* V *containing* S*, and secondly, if* W *is another vector subspace of* V *containing* S *then* $W \supseteq \mathrm{Span}(S)$*.*

Proof. (1) By Theorem 3.8, we have only to show that

$$\forall c_i \in F, \forall \mathbf{v}_i \in \mathrm{Span}(S) \Rightarrow c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \in \mathrm{Span}(S).$$

For (2), we use also Theorem 3.8. □

Example 4.5. (**Standard spanning set $\{\mathbf{e}_i\}$**

for F_c^n and $\{\mathbf{e}_i^t\}$ for F^n) Consider the vectors

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

in the column vector 3-space F_c^3. Then

$$F_c^3 = \mathrm{Span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}.$$

Namely, F^3 is spanned by $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$.

In general, consider the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \dots, \ \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

in the column vector n-space F_c^n. Then

$$F_c^n = \mathrm{Span}\{\mathbf{e}_1, \dots, \mathbf{e}_n\}.$$

Namely, F_c^n is spanned by $\{e_1, \ldots, \mathbf{e}_n\}$. We call $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ the **standard spanning set of F_c^n**.

Similarly, the row vector n-space F^n is spanned by the following **standard spanning set of F^n**:

$$\{\mathbf{e}_1^t = (1, 0, \ldots, 0), \ \ldots, \ \mathbf{e}_n^t = (0, 0, \ldots, 1)\}.$$

Example 4.6. (**Standard spanning set $\{E_{ij}\}$ of $M_{m\times n}(F)$**) Consider the matrices

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

in the matrix space $M_2(F)$. Then

$$M_2(F) = \mathrm{Span}\{E_{11}, E_{12}, E_{21}, E_{22}\}.$$

Namely, $M_2(F)$ is spanned by $\{E_{11}, E_{12}, E_{21}, E_{22}\}$.

In general, consider matrix $E_{k\ell} = (e_{ij})$ in $M_{m\times n}(F)$ where $e_{k\ell} = 1$ and $e_{ij} = 0$ for all (i, j) with $(i, j) \neq$

(k, ℓ). To be precise,

$$E_{11} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad E_{12} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \ldots,$$

$$E_{1n} = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}, \quad E_{21} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \ldots,$$

$$E_{2n} = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}, \ldots, \quad E_{m,n} = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Then

$$M_{m\times n}(F) = \mathrm{Span}\{E_{11}, \dots, E_{1n}, E_{21}, \dots, E_{mn}\}.$$

Namely, $M_{m\times n}(F)$ is spanned by the so called **standard spanning set**:

$$\{E_{11}, \dots, E_{1n}, E_{21}, \dots, E_{2n}, \dots, E_{mn}\}.$$

Example 4.7. In Example 3.9, the solution space of the linear system (corresponding to $AX = \mathbf{0}$ for a particular matrix A there) is given as

$$\mathrm{Null}(A)^t = \mathrm{Span}\{(2, 1, 0), (-3, 0, 1)\}.$$

So the solution space is spanned by two vectors $(2, 1, 0)$ and $(-3, 0, 1)$ in the row vector 3-space.

Exercise 4.8. (**Standard spanning set $\{x^i\}$ of $F[x]$**) Show that the polynomial ring $F[x]$ is spanned by the following so called **standard spanning set (of $F[x]$)**:

$$\{x^i\,;\, i \geq 0\} = \{1,\ x,\ x^2,\ x^3,\ \dots\}.$$

Show that the vector subspace $P_n[x]$ (consisting of all polynomials of degree $< n$) of $F[x]$ in Example 3.14 is spanned by the following so called **standard spanning set (of $P_n[x]$)**:

$$\{1,\ x,\ x^2,\ \ldots,\ x^{n-1}\}.$$

Definition 4.9. (**Sum of subspaces**) Let V be a vector space over a field F, and let U and W be vector subspaces of V. The subset

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

is called the **sum of the subspaces U and W**. So $U+W$ consists of exactly the vectors of the form: (a vector from U) + (another vector from W).

Exercise 4.10. (= Tutorial question) (**Inclusion and sum**) Let U_i, W_i, U and W be vector subspaces of a vector space V over a field F. Show the following:

(1) If $U_1 \subseteq U_2$ and $W_1 \subseteq W_2$ then

$$U_1 + W_1 \subseteq U_2 + W_2.$$

Is the converse true?

(2)

$$W + \{\mathbf{0}\} = W, \quad W + W = W.$$

(3)

$$U + W = W \iff U \subseteq W.$$

Theorem 4.11. *(**Sum being a subspace**) Let U and W be vector subspaces of a vector space V over a field F. For the sum $U + W$ in Definition* 4.9*, we have:*

(1)

$$U + W = \mathrm{Span}(U \cup W).$$

(2) $U + W$ is indeed a vector subspace of V.

(3) $U + W$ is the smallest vector subspace of V containing both U and W: first $U + W$ is a

vector subspace of V containing both U and W, and secondly, if T is another vector subspace of V containing both U and W then $T \supseteq U + W$.

Proof. Since (2) and (3) follow from (1) and Theorem 4.4, we only need to show (1).

(1) For the inclusion $U + W \subseteq \mathrm{Span}(U \cup W)$, take any vector $\mathbf{v}$ from $U + W$. Then $\mathbf{v}$ is of the form $\mathbf{v} = \mathbf{u} + \mathbf{w}$ for some $\mathbf{u} \in U$, $\mathbf{w} \in W$. This $\mathbf{v}$ is the linear combination of vectors $\mathbf{u}$ and $\mathbf{w}$ in $U \cup W$ and hence $\mathbf{v}$ is in $\mathrm{Span}(U \cup W)$ by the very Definition 4.1. Thus we have proved the inclusion

$$(*) \quad U + W \subseteq \mathrm{Span}(U \cup W).$$

For the inclusion $\mathrm{Span}(U \cup W) \subseteq U + W$, take any vector $\mathbf{v}$ from $\mathrm{Span}(U \cup W)$. Then $\mathbf{v}$ is a linear combination

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m$$

of vectors $\mathbf{v}_i \in U \cup W$ (and with $a_i \in F$). After relabeling, we may assume that

$$\mathbf{v}_1 \in U, \ldots, \mathbf{v}_r \in U \qquad (*1)$$

$$\mathbf{v}_{r+1} \in W, \ldots, \mathbf{v}_m \in W \qquad (*2).$$

Since U is a vector subspace of V, it is closed under linear combination (cf. Theorem 3.8) and hence

$$a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r \in U \quad (*3)$$

by using condition(*1). By the same reasoning, (*2) implies

$$a_{r+1}\mathbf{v}_{r+1} + \cdots + a_m\mathbf{v}_m \in W \quad (*4).$$

Now (*3) and (*4) imply that

$$\mathbf{v} = (a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r) + (a_{r+1}\mathbf{v}_{r+1} + \cdots + a_m\mathbf{v}_m)$$
$$\in U + W.$$

Thus we have proved the inclusion

$$(**) \quad \mathrm{Span}(U \cup W) \subseteq U + W.$$

(*) and (**) imply (1) (and also the theorem). $\square$

Remark 4.12. Since the union of vector subspaces may not be a vector subspace (cf. Remark 3.17 and also Exercise 4.15 below), in view of Theorem 4.11, the sum $U+W$ of vector subspaces U and W of V, is the right notion as the smallest vector subspace of V containing both U and W.

Example 4.13. (**Sum and intersection of matrix subspaces**). Let

$$U = \{ \begin{pmatrix} a_{11} & a_{12} \\ 0 & 0 \end{pmatrix} ;\ a_{11},\ a_{12} \in F\}$$

be the subset of all matrices in the matrix space $M_2(F)$ with vanishing second row, and let

$$W = \{ \begin{pmatrix} 0 & b_{12} \\ 0 & b_{22} \end{pmatrix} ;\ b_{12},\ b_{22} \in F\}$$

be the subset of all matrices in $M_2(F)$ with vanishing first column.

One can verify that both U and W are vector subspaces of $M_2(F)$. Further, one can prove that the sum is:

$$U + W = \{ \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix} ; c_{ij} \in F\}$$

(the vector subspace of all upper triangular matrices in $M_2(F)$), and the intersection is:

$$U \cap W = \{ \begin{pmatrix} 0 & d_{12} \\ 0 & 0 \end{pmatrix} ; d_{12} \in F\}.$$

One can also directly check that both $U+W$ and $U \cap W$ are vector subspaces of $M_2(F)$. Further, verify that $U + W \not\subseteq U \cup W$ (or equivalently $U + W \neq U \cup W$) and (hence) $U \cup W$ is not a vector subspace of $M_2(F)$ (cf. Exercise 4.15).

Exercise 4.14. (**Sum and intersections of subspaces of $\mathbb{R}^3$**)

(1) Let L_1 and L_2 be the two axes (i.e. the x-axis and y-axis) in the Euclidean plane $\mathbb{R}^2$, both of

which are vector subspaces of $\mathbb{R}^2$ (cf. Exercise 3.5). Then the sum

$$L_1 + L_2 = \mathbb{R}^2.$$

The inclusion $L_1 + L_2 \subseteq \mathbb{R}^2$ is clear. For the other inclusion, to proceed, observe that

$$L_1 = \{(x, 0) \,|\, x \in \mathbb{R}\}$$
$$L_2 = \{(0, y) \,|\, y \in \mathbb{R}\}.$$

(2) In the Euclidean 3-space $\mathbb{R}^3$, let L_1 be the x-axis and P_1 the yz-plane both of which are vector subspaces of $\mathbb{R}^3$ (cf. Exercise 3.5). Then the sum

$$L_1 + P_1 = \mathbb{R}^3.$$

The inclusion $L_1 + P_1 \subseteq \mathbb{R}^3$ is clear. For the other inclusion, to proceed, observe that

$$L_1 = \{(x, 0, 0) \,|\, x \in \mathbb{R}\}$$
$$P_1 = \{(0, y, z) \,|\, y \in \mathbb{R},\ z \in \mathbb{R}\}.$$

If L_2 is the y-axis in $\mathbb{R}^3$, then the sum $L_1 + L_2$ equals the xy-plane in $\mathbb{R}^3$.

(3) More generally, in $\mathbb{R}^3$, if P is a plane passing through the origin and L is a line which passes through the origin and is not entirely contained in P, then

$$L + P = \mathbb{R}^3.$$

(4) If L_4 and L_5 are two non-identical lines in $\mathbb{R}^3$ each of which passes through the origin, then the sum $L_4 + L_5$ equals the unique plane containing both L_4 and L_5.

(5) If P_4 and P_5 are two non-identical planes in $\mathbb{R}^3$ each of which passes through the origin, then the sum

$$P_4 + P_5 = \mathbb{R}^3$$

and the intersection $P_4 \cap P_5$ is the unique line which is contained in both P_4 and P_5.

Extra exercise 4.15. (**for a Union to Be a Subspace**) Let U and W be two vector subspaces of a vector space V over a field F. Then the following are equivalent.

(1) The union $U \cup W$ is a vector subspace of V.

(2) Either $U \subseteq W$ (and hence $U \cup W = W = U + W$) or $W \subseteq U$ (and hence $U \cup W = U = U + W$); cf. Exercise 4.10.

Hint. Assume (1). If (2) is not true, then one can find some $\mathbf{u} \in U \setminus W$ and $\mathbf{w} \in W \setminus U$ and consider $\mathbf{u} + \mathbf{w}$.

Exercise 4.16. (= Tutorial question) (**Span of union = Sum of Span**) Let S, T be subsets of a vector space V over a field F. Then

$$\mathrm{Span}(S \cup T) = \mathrm{Span}(S) + \mathrm{Span}(T).$$

Definition 4.17. (**Sum of many subspaces**) Let V be a vector space over a field F, and let W_i $(1 \leq i \leq s)$ be vector subspaces of V. The subset

$$\begin{aligned}\sum_{i=1}^{s} W_i &= W_1 + \cdots + W_s \\ &= \{\sum_{i=1}^{s} \mathbf{w}_i \,;\, \mathbf{w}_i \in W\} \\ &= \{\mathbf{w}_1 + \cdots + \mathbf{w}_s \,;\, \mathbf{w}_i \in W_i\}\end{aligned}$$

is called the **sum of the subspaces W_i**.

The proof of the theorem below is similar to that of Theorem 4.11 and is left as an exercise for students. A direct proof seems to be better than the induction on s.

Theorem 4.18. *(**Sum of many being a subspace**) Let W_i $(1 \leq i \leq s)$ be vector subspaces of a vector space V over a field F. For the sum $\sum_{i=1}^{s} W_i$ in Definition 4.17, we have:*

(1)

$$\sum_{i=1}^{s} W_i = \mathrm{Span}(\cup_{i=1}^{s} W_i).$$

(2) $\sum_{i=1}^{s} W_i$ *is indeed a vector subspace of* V*.*

(3) $\sum_{i=1}^{s} W_i$ *is the smallest vector subspace of* V *containing all* W_i*: first* $\sum_{i=1}^{s} W_i$ *is a vector subspace of* V *containing all* W_i*, and secondly, if* T *is another vector subspace of* V *containing all* W_i *then* $T \supseteq \sum_{i=1}^{s} W_i$*.*

Notation 4.19. (**Span as sum of $F\mathbf{v}_i$**) Let $\mathbf{v}, \mathbf{v}_1, \ldots, \mathbf{v}_s$ be vectors in a vector space V over a field F. Denote

$$F\mathbf{v} := \{a\mathbf{v} \mid a \in F\} = \mathrm{Span}\{\mathbf{v}\}.$$

By Exercise 4.16,

$$\begin{aligned}\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_s\} &= \mathrm{Span}\{\mathbf{v}_1\} + \cdots + \mathrm{Span}\{\mathbf{v}_s\} \\ &= F\mathbf{v}_1 + \cdots + F\mathbf{v}_s.\end{aligned}$$

Definition 4.20. (**Direct sum of subspaces**) Let V be a vector space over a field F and let

W_1, W_2 be vector subspaces of V. We say that the sum $W_1 + W_2$ is a **direct sum of two vector subspaces W_1, W_2**, if the intersection

$$W_1 \cap W_2 = \{\mathbf{0}\}.$$

In this case, we denote $W_1 + W_2$ as $W_1 \oplus W_2$, i.e.,

$$W_1 + W_2 = W_1 \oplus W_2.$$

We write

$$W = W_1 \oplus W_2$$

if W is a direct sum of W_1 and W_2, i.e., if

$$W = W_1 + W_2, \quad W_1 \cap W_2 = \{\mathbf{0}\}.$$

Example 4.21. (**F^2 as direct sum of its axes**)
For the standard spanning set

$$\{\mathbf{e}_1^t = (1,\ 0), \quad \mathbf{e}_2^t = (0,\ 1)\}$$

of the row vector 2-space F^2 (cf. Example 4.5) we have (cf. Notation 4.19):

$$\begin{aligned} F^2 &= \mathrm{Span}\{\mathbf{e}_1^t, \mathbf{e}_2^t\} \\ &= \mathrm{Span}\{\mathbf{e}_1^t\} \oplus \mathrm{Span}\{\mathbf{e}_2^t\} \\ &= F\mathbf{e}_1^t \oplus F\mathbf{e}_2^t. \end{aligned}$$

Here

$$F\mathbf{e}_1^t = \{x\mathbf{e}_1^t \,|\, x \in F\} = \{(x, 0) \,|\, x \in F\}$$

is the x-axis of the plane F^2 while

$$F\mathbf{e}_2^t = \{y\mathbf{e}_2^t \,|\, y \in F\} = \{(0, y) \,|\, y \in F\}$$

is the y-axis of F^2.

Example 4.22. (**Direct sum of matrix subspaces**)

(1) Consider subsets

$$V_1 := \{ \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} ; a_{ij} \in F\},$$

$$V_2 := \{ \begin{pmatrix} 0 & 0 \\ a_{21} & 0 \end{pmatrix} ; a_{21} \in F\}.$$

of the matrix space $M_2(F)$. One can verify that both V_i are vector subspaces of $M_2(F)$ such that

$$M_2(F) = V_1 \oplus V_2.$$

Namely, $M_2(F)$ is a direct sum of V_i.

(2) Consider subsets

$$W_1 := \{ \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} ; a_{ij} \in F\},$$

$$W_2 := \{ \begin{pmatrix} 2a & 3a \\ a & 4a \end{pmatrix} ; a \in F\}.$$

of the matrix space $M_2(F)$. One can verify that both W_i are vector subspaces of $M_2(F)$ such that

$$M_2(F) = W_1 \oplus W_2.$$

(3) Consider subsets

$$U_1 := \{ \begin{pmatrix} a_{11} & 0 \\ a_{21} & 0 \end{pmatrix} ; a_{i1} \in F\},$$

$$U_2 := \{ \begin{pmatrix} 0 & a_{12} \\ 0 & a_{22} \end{pmatrix} ; a_{i2} \in F\}.$$

of the matrix space $M_2(F)$. One can verify that both U_i are vector subspaces of $M_2(F)$ such that

$$M_2(F) = U_1 \oplus U_2.$$

(4) The sum $U + W$ in Example 4.13 is not a direct sum because $U \cap W \neq \{\mathbf{0}\}$.

Example 4.23. (**Direct sum of subspaces of F^3**)

(1) Consider subsets

$$V_1 := \{(x, 2x, 3x) | \, x \in F\},$$

$$V_2 := \{(0, y, z) | \, y, z \in F\}$$

of the row vector 3-space F^3. One can verify both V_i are vector subspaces of F^3 such that

$$F^3 = V_1 + V_2 = V_1 \oplus V_2.$$

Namely, F^3 is a direct sum of V_i.

(2) Consider the sums in Exercise 4.14. The sums

$$\begin{aligned} L_1 + L_2 &\text{ in } 4.14(1), \\ L_1 + P_1 &\text{ in } 4.14(2), \\ L_1 + L_2 &\text{ in } 4.14(2), \\ L + P &\text{ in } 4.14(3), \\ L_4 + L_5 &\text{ in } 4.14(4) \end{aligned}$$

are all direct sums. But the sum $P_4 + P_5$ in 4.14(5) is not a direct sum because $P_4 \cap P_5 \neq \{\mathbf{0}\}$.

Example 4.24. (**Splitting the standard spanning set to construct direct sums**)

(1) For the standard spanning set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of the column vector n-space F_c^n (cf. Example 4.5), take two disjoint subsets

$$\{i_1, \ldots, i_r\}, \quad \{j_1, \ldots, j_s\}$$

of $\{1, 2, \ldots, n\}$ and set

$$W_1 := \mathrm{Span}\{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_r}\}$$
$$W_2 := \mathrm{Span}\{\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_s}\}.$$

Then the sum $W_1 + W_2$ is a direct sum, i.e.,

$$W_1 + W_2 = W_1 \oplus W_2.$$

If

$$\{i_1, \ldots, i_r\} \coprod \{j_1, \ldots, j_s\} = \{1, 2, \ldots, n\}$$

is further assumed, then

$$F_c^n = W_1 + W_2 = W_1 \oplus W_2.$$

Recall that two subsets S, T of some ambient set, are **disjoint** if the intersection $S \cap T = \emptyset$;

the union $S \cup T$ is disjoint, denoted as

$$S \coprod T,$$

and said to be a disjoint union of S and T, if S and T are disjoint, i.e., if $S \cap T = \emptyset$.

(2) For the standard spanning set

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

of the matrix space $M_{m\times n}(F)$ (cf. Example 4.6), take two disjoint subsets

$$\{E_{ij} \,|\, (i,j) \in I\}, \quad \{E_{k\ell} \,|\, (k,\ell) \in J\}$$

of $\{E_{ij} \,|\, 1 \leq i \leq m,\ 1 \leq j \leq n\}$ and set

$$W_1 := \mathrm{Span}\{E_{ij} \,|\, (i,j) \in I\}$$
$$W_2 := \mathrm{Span}\{E_{k\ell} \,|\, (k,\ell) \in J\}.$$

Then the sum $W_1 + W_2$ is a direct sum, i.e.,

$$W_1 + W_2 = W_1 \oplus W_2.$$

If

$$\{E_{ij} \,|\, (i,j) \in I\} \coprod \{E_{k\ell} \,|\, (k,\ell) \in J\}$$
$$= \{E_{ij} \,|\, 1 \le i \le m,\ 1 \le j \le n\}$$

is further assumed, then

$$M_{m\times n}(F) = W_1 + W_2 = W_1 \oplus W_2.$$

Theorem 4.25. *(***Equivalent direct sum definition***) Let W_1 and W_2 be two vector subspaces of a vector space V over a field F. Set $W := W_1 + W_2$. Then the following are equivalent.*

(1) We have

$$W_1 + W_2 = W_1 \oplus W_2$$

i.e., $W_1 + W_2$ is a direct sum of W_1, W_2, i.e., $W_1 \cap W_2 = \{\mathbf{0}\}$.

(2) (Unique expression condition) Every vector $\mathbf{w} \in W$ can be expressed as

$$\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$$

for some $\mathbf{w}_i \in W_i$ *and such expression of* $\mathbf{w}$ *is unique: whenever* $\mathbf{w} = \mathbf{w}'_1 + \mathbf{w}'_2$ *for some* $\mathbf{w}'_i \in W_i$ *we have* $\mathbf{w}'_i = \mathbf{w}_i$.

Proof. Equivalently, one shows the contraposition of the theorem below:
'Not(1): $W_1 \cap W_2 \neq \{\mathbf{0}\}$' $\Longleftrightarrow$ 'Not(2): some $\mathbf{w} \in W$ has two different expressions $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$ and $\mathbf{w} = \mathbf{w}'_1 + \mathbf{w}'_2$ with $(\mathbf{w}_1, \mathbf{w}_2) \neq (\mathbf{w}'_1, \mathbf{w}'_2)$.' □

The concept of direct sum can be generalized to a direct sum of many subspaces.

Definition 4.26. (**Direct sum of many subspaces**) Let V be a vector space over a field F and let

$$W_i \ \ (1 \le i \le s;\ s \ge 2)$$

be vector subspaces of V. We say that the sum $\sum_{i=1}^{s} W_i$ is a **direct sum of vector subspaces**

W_i, if the intersection

$$(*) \quad (\sum_{i=1}^{k-1} W_i) \cap W_k = \{\mathbf{0}\} \quad (2 \leq \forall\, k \leq s).$$

In this case, we denote

$$\sum_{i=1}^{s} W_i = W_1 + \cdots + W_s$$

as

$$\oplus_{i=1}^{s} W_i = W_1 \oplus \cdots \oplus W_s$$

i.e.,

$$W_1 + \cdots + W_s = W_1 \oplus \cdots \oplus W_s.$$

We write

$$W = \oplus_{i=1}^{s} W_i = W_1 \oplus \cdots \oplus W_s$$

if W is a direct sum of $W_1, \ldots, W_s$, i.e., if $W = \sum_{i=1}^{s} W_i$ and the condition (*) above is satisfied.

Example 4.27. (**F^n as direct sum of its axes**)

For the standard spanning set

$$\{\mathbf{e}_1^t = (1, 0, 0), \mathbf{e}_2^t = (0, 1, 0), \mathbf{e}_3^t = (0, 0, 1)\}$$

of the row vector 3-space F^3 (cf. Example 4.5) we have (cf. Notation 4.19):

$$\begin{aligned} F^3 &= \mathrm{Span}\{\mathbf{e}_1^t,\, \mathbf{e}_2^t,\, \mathbf{e}_3^t\} \\ &= \mathrm{Span}\{\mathbf{e}_1^t\} \oplus \mathrm{Span}\{\mathbf{e}_2^t\} \oplus \mathrm{Span}\{\mathbf{e}_3^t\} \\ &= F\mathbf{e}_1^t \oplus F\mathbf{e}_2^t + \oplus F\mathbf{e}_3^t. \end{aligned}$$

Here

$$F\mathbf{e}_1^t = \{x\mathbf{e}_1^t \,|\, x \in F\} = \{(x, 0, 0) \,|\, x \in F\}$$

is the x-axis of F^3,

$$F\mathbf{e}_2^t = \{y\mathbf{e}_2^t \,|\, y \in F\} = \{(0, y, 0) \,|\, y \in F\}$$

the y-axis of F^3, and

$$F\mathbf{e}_3^t = \{z\mathbf{e}_3^t \,|\, z \in F\} = \{(0, 0, z) \,|\, z \in F\}$$

the z-axis of F^3.

In general, for the standard spanning set

$$\{\mathbf{e}_1^t = (1, 0, \ldots, 0),\ \ldots,\ \mathbf{e}_n^t = (0, \ldots, 0, 1)\}$$

of the row vector n-space F^n (cf. Example 4.5), we have (cf. Notation 4.19):

$$\begin{aligned} F^n &= \mathrm{Span}\{\mathbf{e}_1^t, \ldots, \mathbf{e}_n^t\} \\ &= \mathrm{Span}\{\mathbf{e}_1^t\} \oplus \cdots \oplus \mathrm{Span}\{\mathbf{e}_n^t\} \\ &= F\mathbf{e}_1^t \oplus + \cdots \oplus F\mathbf{e}_n^t. \end{aligned}$$

Here

$$\begin{aligned} F\mathbf{e}_i^t &= \{x_i \mathbf{e}_i^t \,|\, x_i \in F\} \\ &= \{(0, \ldots, 0, x_i, 0, \ldots, 0) \,|\, x_i \in F\} \end{aligned}$$

is the x_i-axis of F^n.

Example 4.28. ($M_{m\times n}(F)$ **as direct sum of** FE_{ij}) For the standard spanning set

$$\{E_{ij} \,|\, 1 \le i \le m,\ 1 \le j \le n\}$$

of the matrix space $M_{m\times n}(F)$ (cf. Example 4.6), we have (cf. Notation 4.19):

$$\begin{aligned} M_{m\times n}(F) &= \mathrm{Span}\{E_{ij} \,|\, 1 \le i \le m,\ 1 \le j \le n\} \\ &= \oplus_{1\le i\le m, 1\le j\le n} \mathrm{Span}\{E_{ij}\} \\ &= \oplus_{1\le i\le m, 1\le j\le n} FE_{ij}. \end{aligned}$$

Here FE_{ij} consists of all matrices in $M_{m\times n}(F)$ with entries all vanishing except the (i,j)-entry.

Example 4.29. (**Direct sum of subspaces of F^6**)

(1) Consider subsets

$$V_1 := \{(x_1, x_2, 0, 0, 0, 0) \mid x_1, x_2 \in F\},$$

$$V_2 := \{(0, 0, x_3, 0, 0, 0) \mid x_3 \in F\},$$

$$V_3 := \{(0, 0, 0, x_4, x_5, x_6) \mid x_4, x_5, x_6 \in F\}$$

of the row vector 6-space F^6. One can verify that all of V_i are vector subspaces of F^6 such that

$$F^6 = V_1 + V_2 + V_3 = V_1 \oplus V_2 \oplus V_3.$$

Namely, F^6 is a direct sum of V_i.

(2) Consider subsets

$$W_1 := \{(x_1, x_2, 2x_1, 3x_2, 4x_1) \mid x_1, x_2 \in F\},$$

$$W_2 := \{(0, 0, x_3, 3x_3, 5x_3, 7x_3) \mid x_3 \in F\},$$

$$W_3 := \{(0, 0, 0, x_4, x_5, x_6) \mid x_4, x_5, x_6 \in F\}$$

of the row vector 6-space F^6. One can verify that all of W_i are vector subspaces of F^6 such that

$$F^6 = W_1 + W_2 + W_3 = W_1 \oplus W_2 \oplus W_3.$$

Extra exercise 4.30. (**Splitting the standard spanning set to construct multiple direct sums**)

(1) For the standard spanning set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of the row vector n-space F^n (cf. Example 4.5), take pairwise disjoint subsets

$$\{\mathbf{e}_i \,|\, i \in I_1\}, \; \ldots, \; \{\mathbf{e}_i \,|\, i \in I_s\}$$

of $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and set

$$W_k = \mathrm{Span}\{\mathbf{e}_i \,|\, i \in I_k\}.$$

Then the sum $W_1 + \cdots + W_s$ is a direct sum, i.e.,

$$W_1 + \cdots + W_s = W_1 \oplus \cdots + \oplus W_s.$$

If

$$\{\mathbf{e}_i \,|\, i \in I_1\} \coprod \cdots \coprod \{\mathbf{e}_i \,|\, i \in I_s\} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$$

is further assumed, then

$$F^n = W_1 + \cdots + W_s = W_1 \oplus \cdots + \oplus W_s.$$

(2) For the standard spanning set

$$\{E_{ij} \,|\, 1 \le i \le m,\ 1 \le j \le n\}$$

of the matrix space $M_{m\times n}(F)$ (cf. Example 4.6), take pairwise disjoint subsets

$$\{E_{k\ell} \,|\, (k, \ell) \in I_1\},\ \ldots,\ \{E_{k\ell} \,|\, (k, \ell) \in I_s\}$$

of $\{E_{ij} \,|\, 1 \le i \le m,\ 1 \le j \le n\}$ and set

$$W_k := \mathrm{Span}\{E_{ij} \,|\, (i, j) \in I_k\}.$$

Then the sum $W_1 + \cdots + W_s$ is a direct sum, i.e.,

$$W_1 + \cdots + W_s = W_1 \oplus \cdots + \oplus W_s.$$

If

$$\{E_{k\ell} \,|\, (k,\ell) \in I_1\} \coprod \cdots \coprod \{E_{k\ell} \,|\, (k,\ell) \in I_s\}$$
$$=\{E_{ij} \,|\, 1 \le i \le m,\ 1 \le j \le n\}$$

is further assumed, then

$$M_{m\times n}(F) = W_1 + \cdots + W_s = W_1 \oplus \cdots + \oplus W_s.$$

The proof of the theorem below is similar to that of Theorem 4.25 and is left as an exercise for students.

Theorem 4.31. *(= Tutorial question) (***Equivalent direct multiple sum definition***) Let W_i $(1 \le i \le s, s \ge 2)$ be vector subspaces of a vector space V over a field F. Set $W := \sum_{i=1}^{s} W_i$. Then the following are equivalent.*

(1) We have

$$W_1 + \cdots + W_s = W_1 \oplus \cdots \oplus W_s$$

i.e., $\sum_{i=1}^{s} W_i$ is a direct sum of W_i.

(2)

$$(\sum_{i\neq\ell} W_i) \cap W_\ell = \{\mathbf{0}\} \quad (\forall\ 1 \le \ell \le s).$$

Here

$$\sum_{i\neq\ell} W_i := W_1 + \cdots + W_{\ell-1} + W_{\ell+1} + \cdots + W_s.$$

(3) (Unique expression condition) Every vector $\mathbf{w} \in W$ *can be expressed as*

$$\mathbf{w} = \mathbf{w}_1 + \cdots + \mathbf{w}_s$$

for some $\mathbf{w}_i \in W_i$ *and such expression of* $\mathbf{w}$ *is unique: whenever* $\mathbf{w} = \mathbf{w}'_1 + \cdots + \mathbf{w}'_s$ *for some* $\mathbf{w}'_i \in W_i$*, we have* $\mathbf{w}'_i = \mathbf{w}_i$*.*

Remark 4.32. (**Fake example of direct sum**) Find a vector space V over a field F, with three distinct nonzero subspaces W_1, W_2, W_3 such that

$$W_i \cap W_j = \{\mathbf{0}\}$$

for all $i \neq j$, but the sum $W_1 + W_2 + W_3$ is not a direct sum.

Extra exercise 4.33. (**Product v.s. Direct sum**) Consider the product $V := V_1 \times V_2 = \{(\mathbf{v}_1, \mathbf{v}_2) \mid \mathbf{v}_i \in V_i\}$ of two vector spaces V_i over the same field F (cf. Exercise 2.25) If we do the identification

$$V_1 = V_1 \times \{\mathbf{0}_{V_2}\} = \{(\mathbf{v}_1, \mathbf{0}_{V_2}) \mid \mathbf{v}_1 \in V_1\},$$
$$V_2 = \{\mathbf{0}_{V_1}\} \times V_2 = \{(\mathbf{0}_{V_1}, \mathbf{v}_2) \mid \mathbf{v}_2 \in V_2\}$$

show that $V = V_1 \oplus V_2$.

5 Linear Independence, Basis and Dimension (T3)

In this section, we cover **linear independence** of vectors, **basis** B of a vector space V and criteria to be a basis, as well as the **dimension** of a vector space: $\dim V = |B|$ (well defined).

Definition 5.1. (**Linear (in)dependence**) Let V be a vector space over a field F. Let T be a (not necessarily finite) subset of V and let

$$S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$$

be a finite subset of V.

(1) We call S **a linear independent set** (or **the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent**) or simply **L.I.**, if the vector equation below

$$x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m = \mathbf{0} \quad (\#1)$$

has only the so called **trivial solution**

$$(x_1, \ldots, x_m) = (0, \ldots, 0),$$

or equivalently if

$$x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m = \mathbf{0} \implies x_1 = \cdots = x_m = 0.$$

In this case, we say that there is only the **trivial relation**

$$\textbf{(i.e., } 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_m = \mathbf{0})$$

among the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

(2) We call S **a linear dependent set** (or **the vectors** $\mathbf{v}_1, \ldots, \mathbf{v}_m$ **are linearly dependent**) or simply **L.D.**, if there are scalars $a_1, \ldots, a_m$ in F which are not all zero (i.e., $(a_1, \ldots, a_m) \neq (0, \ldots, 0)$) such that

$$a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0},$$

or equivalently if the vector equation below

$$x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m = \mathbf{0} \quad (\#1)$$

has a non-trivial solution

$$(x_1, \ldots, x_m) = (a_1, \ldots, a_m) \neq (0, \ldots, 0)$$

of course, with $a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m = \mathbf{0}$. In this case, we say that there is a **non-trivial relation (i.e., $a_1\mathbf{v}_1+\cdots+a_m\mathbf{v}_m = \mathbf{0}$) among the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ given by the scalars $a_1, \ldots, a_m$.**

(3) The set T is a **linearly independent set** if every non-empty finite subset of T is linearly independent. The set T is a **linearly dependent set** if at least one non-empty finite subset of T is linearly dependent.

Notation 5.2. We may denote the solution to the equation (#1) in Definition 5.1 either as row vector

$$(x_1, \ldots, x_m) = (a_1, \ldots, a_n),$$

or as column vector

$$\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}.$$

Exercise 5.3. (**L.D. / L.I. inheritance**)

(1) Let $S_1 \subseteq S_2$. If the smaller set S_1 is linearly dependent then so is the larger set S_2. Equivalently, if the larger set S_2 is linearly independent then so is the smaller set S_1.

(2) $\{\mathbf{0}\}$ is a linearly dependent set.

(3) If $\mathbf{0} \in S$, then S is a linearly dependent set.

Example 5.4. (**L.D. row vectors in F^4**) Consider the row vectors

$$\mathbf{v}_1 = (2, -1, 0, 3)$$

$$\mathbf{v}_2 = (1, 2, 5, -1)$$

$$\mathbf{v}_3 = (7, -1, 5, 8)$$

in the row vector 4-space F^4.

We claim that there is a non-trivial relation

$$3\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 = 0$$

among these three vectors, and hence they are linear dependent.

To prove the claim, consider the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0} \quad (\#1)$$

as in Definition 5.1. After taking transpose, the above equation is equivalent to

$$x_1\mathbf{v}_1^t + x_2\mathbf{v}_2^t + x_3\mathbf{v}_3^t = \mathbf{0},$$

i.e.,

$$x_1 \begin{pmatrix} 2 \\ -1 \\ 0 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ 2 \\ 5 \\ -1 \end{pmatrix} + x_3 \begin{pmatrix} 7 \\ -1 \\ 5 \\ 8 \end{pmatrix} = \mathbf{0},$$

i.e.,

$$AX = 0 \quad (*1)$$

where

$$A = (\mathbf{v}_1^t, \mathbf{v}_2^t, \mathbf{v}_3^t) = \begin{pmatrix} 2 & 1 & 7 \\ -1 & 2 & -1 \\ 0 & 5 & 5 \\ 3 & -1 & 8 \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Note 1. In finding the relation among vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$, i.e., the solution to the vector equation (#1), we form a matrix A and translate the vector equation (#1) as a matrix equation (*1). The columns of A are always the initial vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$; but, if the initial vectors $\mathbf{v}_i$ are row vectors as in the current case, we have to transpose them into column vectors $\mathbf{v}_i^t$ to form

$$A = (\mathbf{v}_1^t, \cdots, \mathbf{v}_m^t).$$

Compare with **Note 1** of Example 5.5 below.

Applying Gaussian elimination to the equation $AX = 0$, we obtain its solution space or the null space of A (also being the solution space of the vector equation (#1) above):

$$\begin{aligned}\mathrm{Null}(A) &= \{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = t\begin{pmatrix} -3 \\ -1 \\ 1 \end{pmatrix}\ ;\ t \in F\} \\ &= F\begin{pmatrix} -3 \\ -1 \\ 1 \end{pmatrix}.\end{aligned}$$

We take one nonzero solution (as row vector) say $(x_1, x_2, x_3) = (3, 1, -1)$ (by letting $t = -1$) and obtain a non-trivial relation $3\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 = 0$ for the three vectors as claimed.

Example 5.5. (**L.I. column vectors in F_c^4**).

We claim that the column vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}$$

$$\mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

in the column vector 4-space F_c^4 are linearly independent.

To prove the claim, consider the vector equation

$$x_1\mathbf{v}_1 + \cdots + x_4\mathbf{v}_4 = \mathbf{0} \quad (\#1)$$

as in Definition 5.1, i.e.,

$$AX = 0$$

where

$$A = (\mathbf{v}_1, \ldots, \mathbf{v}_4) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

Note 1. In finding the relation among vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$, we form a matrix A and solve the matrix equation

$$AX = 0.$$

The columns of A are always the initial vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$, i.e.,

$$A = (\mathbf{v}_1, \cdots, \mathbf{v}_m)$$

since our vectors $\mathbf{v}_i$ are already column vectors. Compare with **Note 1** of Example 5.4 above.

Applying Gaussian elimination, or by observing that the **determinant** $\det(A) = |A| \neq 0$ and using Linear Algebra I, the equation $AX = 0$ (and hence its equivalent vector equation (#1)) has only the trivial solution $(x_1, \ldots, x_4) = (0, \ldots, 0)$. So the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are linearly independent as claimed.

Exercise 5.6. Show that the standard spanning sets below are all linearly independent:

$$\{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \subset F_c^n \text{ in } 4.5,$$

$$\{\mathbf{e}_1^t, \ldots, \mathbf{e}_n^t\} \subset F^n \text{ in } 4.5,$$

$$\{E_{ij} | 1 \le i \le m; 1 \le j \le n\} \subset M_{m\times n}(F) \text{ in } 4.6,$$

$$\{1,\ x,\ x^2,\ \ldots\} \subset F[x] \text{ in } 4.8,$$

$$\{1,\ x,\ \ldots,\ x^{n-1}\} \subset P_n[x] \text{ in } 4.8.$$

Example 5.7. (**L.D. matrices**) We claim that the three matrices

$$\mathbf{v}_1 = \begin{pmatrix} 1 & -3 & 2 \\ -4 & 0 & 5 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -3 & 7 & 4 \\ 6 & -2 & -7 \end{pmatrix},$$

$$\mathbf{v}_3 = \begin{pmatrix} -2 & 3 & 11 \\ -1 & -3 & 2 \end{pmatrix}$$

in the matrix space $M_{2\times 3}(F)$ satisfy a non-trivial relation:

$$5\mathbf{v}_1 + 3\mathbf{v}_2 - 2\mathbf{v}_3 = 0$$

and hence these three matrices are linearly dependent.

To prove the claim, we consider the vector equation:

$$(\#1)\quad \mathbf{0} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 =$$

$$\begin{pmatrix} x_1 - 3x_2 - 2x_3 & -3x_1 + 7x_2 + 3x_3 & 2x_1 + 4x_2 + 11x_3 \\ -4x_1 + 6x_2 - x_3 & 0x_1 - 2x_2 - 3x_3 & 5x_1 - 7x_2 + 2x_3 \end{pmatrix}$$

as in Definition 5.1, i.e., all entries are zero:

$$\begin{aligned} x_1 - 3x_2 - 2x_3 &= 0 \\ -3x_1 + 7x_2 + 3x_3 &= 0 \\ 2x_1 + 4x_2 + 11x_3 &= 0 \\ -4x_1 + 6x_2 - x_3 &= 0 \\ 0x_1 - 2x_2 - 3x_3 &= 0 \\ 5x_1 - 7x_2 + 2x_3 &= 0, \end{aligned}$$

i.e.,

$$AX = 0$$

where A is the coefficient matrix and X is a column vector:

$$A = \begin{pmatrix} 1 & -3 & -2 \\ -3 & 7 & 3 \\ 2 & 4 & 11 \\ -4 & 6 & -1 \\ 0 & -2 & -3 \\ 5 & -7 & 2 \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Applying Gaussian elimination to the matrix equation $AX = 0$, we obtain its solution space or the null space of A (also being the solution space of the vector equation (#1)):

$$\mathrm{Null}(A) = \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{-5t}{2} \\ \frac{-3t}{2} \\ t \end{pmatrix} ; t \in F\}$$

$$= F \begin{pmatrix} \frac{-5}{2} \\ \frac{-3}{2} \\ 1 \end{pmatrix}.$$

Note 1. We assume $2 \neq 0$ in F so that $5/2$ etc. makes sense. This is always true if $F = \mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$ (but not true if F is a field of cardinality 2^s for some

$s \geq 1$).

We take one solution $(x_1, x_2, x_3) = (5, 3, -2)$ (by letting $t = -2$) and get the non-trivial relation $5\mathbf{v}_1 + 3\mathbf{v}_2 - 2\mathbf{v}_3 = 0$ for the three vectors as claimed.

Theorem 5.8. (**Equivalent L.I./L.D. definitions**) *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a finite subset of a vector space V over a field F. Then we have:*

(1) Let $|S| \geq 2$. Then S is a linear dependent set if and only if some $\mathbf{v}_k \in S$ is a linear combination of the others, i.e. there are scalars

$$a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_m$$

in F (with all these scalars vanishing allowed) such that

$$\mathbf{v}_k = \sum_{i \neq k} a_i \mathbf{v}_i =$$

$$a_1\mathbf{v}_1 + \cdots + a_{k-1}\mathbf{v}_{k-1} + a_{k+1}\mathbf{v}_{k+1} + \cdots + a_m\mathbf{v}_m.$$

(2) Let $|S| \geq 2$*. Then* S *is linearly independent if and only if no* $\mathbf{v}_k \in S$ *is a linear combination of others.*

(3) Suppose that $S = \{\mathbf{v}_1\}$ *(a single vector). Then* S *is linearly dependent if and only if* $\mathbf{v}_1 = \mathbf{0}$*. Equivalently,* S *is linearly independent if and only if* $\mathbf{v}_1 \neq \mathbf{0}$*.*

(4) Suppose that $S = \{\mathbf{v}_1, \mathbf{v}_2\}$ *(two vectors). Then* S *is linearly dependent if and only if one of* $\mathbf{v}_1, \mathbf{v}_2$ *is a scalar multiple of the other. Equivalently,* S *is linearly independent if and only if neither one of* $\mathbf{v}_1, \mathbf{v}_2$ *is a scalar multiple of the other.*

Proof. (2) is the contraposition of (1). (3) is easy (cf. also Exercise 5.3) while (4) follows from (1) and (2).

For the direction '$\Leftarrow$' of (1), suppose that

$$\mathbf{v}_k = a_1\mathbf{v}_1 + \cdots + a_{k-1}\mathbf{v}_{k-1} + a_{k+1}\mathbf{v}_{k+1} + \cdots + a_m\mathbf{v}_m$$

for some scalars $a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_m$ in F. Then

$$\mathbf{0} = a_1\mathbf{v}_1 + \cdots + a_{k-1}\mathbf{v}_{k-1} - \mathbf{v}_k + a_{k+1}\mathbf{v}_{k+1} + \cdots + a_m\mathbf{v}_m$$

and hence the vector equation

$$x_1\mathbf{v}_1 + \cdots + x_m\mathbf{v}_m = \mathbf{0} \quad (\#1)$$

has a non-trivial (i.e., nonzero) solution

$$(x_1, \ldots, x_m) = (a_1, \ldots, a_{k-1}, -1, a_{k+1}, \ldots, a_m);$$

thus, $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent. We have proved the direction '$\Leftarrow$' of (1).

For the direction '$\Rightarrow$' of (1), suppose that S is linear dependent. Hence one has a non-trivial relation

$$b_1\mathbf{v}_1 + \cdots + b_m\mathbf{v}_m = \mathbf{0} \quad (\#1)$$

for some scalars b_i not all of which are zero, say $b_k \neq 0$.

Solving $\mathbf{v}_k$ from the equation (#1), we get

$$\begin{aligned}\mathbf{v}_k &= \frac{-1}{b_k}(b_1\mathbf{v}_1 + \cdots b_{k-1}\mathbf{v}_{k-1} + b_{k+1}\mathbf{v}_{k+1} + \cdots b_m\mathbf{v}_m) \\ &= \sum_{i\neq k} a_i\mathbf{v}_i\end{aligned}$$

where

$$a_i := \frac{-b_i}{b_k}.$$

We have proved the direction '$\Rightarrow$' of (1). □

Exercise 5.9. Give a counter example to the following false statement:

$\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if and only if $\mathbf{v}_1 = a\mathbf{v}_2$ for some $a \in F$.

Definition 5.10. (**Basis, (in)finite dimension**) Let V be a nonzero vector space over a field F. A subset B of V is called a **basis** if the following two conditions are satisfied.

(1) (Span) V is spanned by B:

$$V = \mathrm{Span}(B).$$

(2) (L.I.) B is a linearly independent set.

If V has a basis B with **cardinality**

$$|B| < \infty$$

we say that V is **finite-dimensional** and define the **dimension** $\dim_F V$ **(or simply** $\dim V$**) of** V **(over the field** F**)** as the cardinality of B:

$$\dim_F V \ := |B|.$$

(cf. Theorem 5.15 for its well definedness).

Otherwise, V is called **infinite-dimensional.**

If V equals the zero vector space $\{\mathbf{0}\}$, we define

$$\dim \{\mathbf{0}\} = 0.$$

Theorem 5.11. *(***Equivalent basis definition I***) Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ (with $\mathbf{v}_i \neq \mathbf{0}_V$) be a finite subset of a vector space V over a field F. Then the following are equivalent.*

(1) B is a basis of V.

(2) (Unique expression condition) Every vector $\mathbf{v} \in V$ can be expressed as

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$$

for some scalars $a_i \in F$ and such expression of $\mathbf{v}$ is unique: whenever $\mathbf{v} = a'_1\mathbf{v}_1 + \cdots + a'_n\mathbf{v}_n$ for some scalars $a'_i \in F$, we have $a'_i = a_i$.

(3) V has the following direct sum decomposition (cf. Notation 4.19):

$$\begin{aligned} V &= \mathrm{Span}\{\mathbf{v}_1\} \oplus \cdots \oplus \mathrm{Span}\{\mathbf{v}_n\} \\ &= F\mathbf{v}_1 \oplus \cdots \oplus F\mathbf{v}_n. \end{aligned}$$

Proof. For '(1) $\Rightarrow$ (2)', assume (1), i.e., B spans V and is L.I. (and we shall prove (2)). Then

$$V = \mathrm{Span}(B) = \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$$

and hence every $\mathbf{v} \in V$ is a linear combination of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, i.e.,

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$$

for some scalars $a_i \in F$.

To show the uniqueness of such expression, suppose

$$\mathbf{v} = a_1'\mathbf{v}_1 + \cdots + a_n'\mathbf{v}_n$$

is another expression for the same $\mathbf{v}$. Equating the same $\mathbf{v}$, we get

$$a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{v} = a_1'\mathbf{v}_1 + \cdots + a_n'\mathbf{v}_n$$

and

$$(a_1 - a_1')\mathbf{v}_1 + \cdots + (a_n - a_n')\mathbf{v}_n = \mathbf{0}.$$

Thus the vector equation

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{0} \quad (\#1)$$

has a solution

$$(x_1, \ldots, x_n) = (a_1 - a_1', \ldots, a_n - a_n').$$

Since $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent by the assumption on B, the vector equation (#1) has only trivial solution, and hence its solution

$$(a_1 - a_1', \ldots, a_n - a_n') = (0, \ldots, 0)$$

i.e.,

$$a_1 = a_1', \ldots, a_n = a_n'.$$

This shows the uniqueness required in (2). This proves (2).

The proof for '(2) $\Rightarrow$ (3)' and '(3) $\Rightarrow$ (1)' are similar. We leave the detail as an exercise for students. $\square$

Example 5.12. (**Standard spanning set being a standard basis**) Each set B below is a basis of the vector space V where B belongs to:

$$B := \{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \subset F_c^n =: V \text{ in } 4.5,$$

$$B := \{\mathbf{e}_1^t, \ldots, \mathbf{e}_n^t\} \subset F^n =: V \text{ in } 4.5,$$

$$B = \{E_{ij} | i \le m; j \le n\} \subset M_{m\times n}(F) = V \text{ in } 4.6,$$

$$B := \{1,\, x,\, x^2,\, \dots\} \subset F[x] =: V \text{ in } 4.8,$$

$$B := \{1,\, x,\, \dots,\, x^{n-1}\} \subset P_n[x] =: V \text{ in } 4.8.$$

Indeed, B is the standard spanning set of V (and hence spans V) and is linearly independent (cf. Exercise 5.6).

Each of the above B will now be called the **standard basis of V**.

In particular, by Theorem 5.11, we have (cf. Notation 4.19):

$$F_c^n = \oplus_{i=1}^n F\mathbf{e}_i,$$

$$F^n = \oplus_{i=1}^n F\mathbf{e}_i^t,$$

$$M_{m\times n}(F) = \oplus_{1\le i\le m, 1\le j\le n} FE_{ij},$$

$$F[x] = \oplus_{i=0}^{\infty} Fx^i,$$

$$P_n[x] = \oplus_{i=0}^{n-1} Fx^i.$$

The dimension $\dim_F V = |B|$ is calculated as follows (as expected) (cf. also Theorem 5.15):

$$\dim_F F_c^n = n,$$

$$\dim_F F^n = n,$$

$$\dim_F M_{m\times n}(F) = mn,$$

$$\dim_F F[x] = \infty,$$

$$\dim_F P_n[x] = n.$$

Theorem 5.13. *(**Deriving a basis from a spanning set**) Suppose that a nonzero vector space V over a field F is spanned by a finite subset $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$:*

$$V = \mathrm{Span}(B) = \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_s\}.$$

Then we have:

(1) There is a subset $B_1 \subseteq B$ such that B_1 is a basis of V. In particular,

$$\dim_F V = |B_1| \le |B|.$$

(2) Let B_2 be a maximal linearly independent subset of B: first B_2 is L.I., and secondly every subset B_3 of B larger than B_2 is L.D. Then B_2 is a basis of $V = \mathrm{Span}(B)$.

Proof. (1) Let B_1 be a subset of B with

$$V = \mathrm{Span}(B_1)$$

and $|B_1|$ the smallest. (Consider those subsets of B each of which spans V, and choose B_1 to be an one with the smallest cardinality). After relabeling, we may assume

$$B_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$$

for some $r \leq s$. It suffices to show the **claim** that B_1 is L.I.

Suppose the contrary that B_1 is L.D. instead. Then, by Theorem 5.8, some $\mathbf{v}_k \in B_1$ is a linear combination of the others:

$$\mathbf{v}_k = a_1\mathbf{v}_1 + \cdots + a_{k-1}\mathbf{v}_{k-1} + a_{k+1}\mathbf{v}_{k+1} + \cdots + a_r\mathbf{v}_r$$

for some scalars $a_i \in F$. Thus

$$\mathrm{Span}\{\mathbf{v}_k\} = F\mathbf{v}_k \subseteq \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \mathbf{v}_r\}.$$

Hence (cf. Exercises 4.16 and 4.10)

$$\begin{aligned} V &= \mathrm{Span}(B_1) = \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \\ &= \mathrm{Span}\{\mathbf{v}_k\} + \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_r\} \\ &= \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_r\}. \end{aligned}$$

Thus

$$V = \mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_r\}.$$

But then

$$B_2 := \{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_r\}$$

is a subset of B with

$$V = \mathrm{Span}(B_2), \quad |B_2| = |B_1| - 1 < |B_1|$$

contradicting the minimality assumption on $|B_1|$.

Thus B_1 is L.I. as claimed. This proves (1).

(2) is similar to (1) and left as an exercise for students. □

Remark 5.14. (**Constructing a basis from a spanning set**) The constructive proof of Theorem 5.13 gives the way to find a basis as a subset of a spanning set of a nonzero vector space V.

To be precise, suppose that $V = \text{Span}(B_0)$ for some finite set $B_0 = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$. The proof of Theorem 5.13 shows: if some vector (say $\mathbf{v}_r$) in B_0 is a linear combination of the others, then we set $B_1 := \{\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}\}$ and still have $V = \text{Span}(B_1)$.

If some vector (say $\mathbf{v}_{r-1}$) in B_1 is a linear combination of the others, then we set $B_2 := \{\mathbf{v}_1, \ldots, \mathbf{v}_{r-2}\}$ and still have $V = \text{Span}(B_2)$.

Continue this process until we cannot go further. We will get some subset B_t of B such that $V = \text{Span}(B_t)$ and no vector in B_t is a linear combination of others. Hence B_t is L.I. by Theorem 5.8 (and of course spans V). Thus B_t is a basis of V.

Theorem 5.15. (**Dimension being well defined**) *Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of a vector space V over a field F. Then we have:*

(1) Suppose that S is a subset of V with $|S| > n = |B|$. Then S is L.D.

(2) Suppose that T is a subset of V with $|T| < n$. Then T does not span V.

(3) Suppose that B' is another basis of V. Then $|B'| = |B|$. So the dimension $\dim_F V$ $(= |B|)$ of V depends only on V, but not on the choice of its basis. In other words, $\dim_F V$ is well

defined.

Proof. (1) and (2) are proved in Linear Algebra I (see e.g. [S. L. Ma, V. Tan and K. L. Ng, Linear Algebra I, Theorem 3.5.1]).

For (3), if $|B'| < |B|$ (resp. $|B| < |B'|$) then B' (resp. B) does not span V by (2) (resp. by (2) applied to the basis B') and hence B' (resp. B) is not a basis of V, a contradiction. Therefore, $|B'| = |B|$. $\square$

Theorem 5.16. *(**Expanding an L.I. set**) Let B be a L.I. subset of a vector space V over a field. Then exactly one of the following two cases is true.*

(1) B spans V and hence B is a basis of V.

(2) Let $\mathbf{w} \in V \setminus \mathrm{Span}(B)$ (and hence $\mathbf{w} \notin B$).

Then

$$B \cup \{\mathbf{w}\}$$

is an L.I. subset of V.

In particular, if V is of finite dimension n, then one can find $n - |B|$ vectors

$$\mathbf{w}_{|B|+1}, \dots, \mathbf{w}_n$$

in $V \setminus \mathrm{Span}(B)$ such that

$$B \coprod \{\mathbf{w}_{|B|+1}, \dots, \mathbf{w}_n\}$$

is a basis of V.

Proof. The first assertion is similar to that of Theorem 5.8 and left as an exercise for students.

For the last assertion, applying the first assertion to B, either Case(1) occurs, i.e., B is already a basis (and hence $|B| = n$ by Theorem 5.15), or Case(2) occurs, i.e.,

$$B \coprod \{\mathbf{w}_{|B|+1}\}$$

is linearly independent for some $\mathbf{w}_{|B|+1} \in V \backslash \mathrm{Span}(B)$. For the latter situation, we apply the first assertion again to the L.I. set $B \coprod \{\mathbf{w}_{|B|+1}\}$. Inductively, we either get a basis of V (i.e., reach Case(1) eventually):

$$B \coprod \{\mathbf{w}_{|B|+1}, \ldots, \mathbf{w}_s\}$$

(and hence $s = n$ by Theorem 5.15), or we will always encounter Case(2) and continue this process for ever, i.e. get linearly independet sets

$$B \coprod \{\mathbf{w}_{|B|+1}, \ldots, \mathbf{w}_s\} \quad (s = |B|+1, |B|+2, \ldots)$$

for *all* $s > |B|$ which is impossible by Theorem 5.15 (say taking $s = n+1$). $\square$

Remark 5.17. If V is a finite-dimensional nonzero vector space, one can use Theorem 5.16 to inductively construct a sequence

$$\mathbf{v}_1, \ldots, \mathbf{v}_n$$

of L.I. vectors with n maximal, so that they automatically form a basis of V.

Theorem 5.18. *(**Equivalent basis definition II**) Let B be a subset of a vector space V of finite dimension $\dim_F V = n \geq 1$. Then the following are equivalent.*

(1) B is a basis of V.

(2) B is L.I. and $|B| = n$.

(3) B spans V and $|B| = n$.

Proof. We shall show '(1) $\iff$ (3)' and '(1) $\iff$ (2)'.

'(1) $\Rightarrow$ (2)' and '(1) $\Rightarrow$ (3)' are from Definition 5.10 (cf. also Theorem 5.15).

For '(3) $\Rightarrow$ (1)', assume (3) (and we shall prove (1)). Since $V = \mathrm{Span}(B)$, by Theorem 5.13, there is a subset B_1 of B such that B_1 is a basis of V. Hence, by Theorem 5.15, $n = \dim_F V = |B_1| \le |B| = n$ (using (3) for the last equality). Thus they are all equal and hence B equals B_1 (a basis of V). We are done!

For '(2) $\Rightarrow$ (1)', assume (2) (and we shall prove (1)). By Theorem 5.16, either B is a basis and (1) is true, or the set

$$B \cup \{\mathbf{w}\}$$

is L.I. for some $\mathbf{w} \notin B$ and it has exactly $n + 1$ ($> n$) vectors, contradicting Theorem 5.15. So (1) is true. $\square$

Theorem 5.19. *(***Basis of a direct sum***) Let V be a (not necessarily finite-dimensional) vector space over a field F.*

(1) Suppose that B is a basis of V. Decompose it

as a disjoint union

$$B = B_1 \coprod \cdots \coprod B_s$$

of non-empty sets B_i*. Then* B_i *is a basis of* $W_i := \mathrm{Span}(B_i)$ *and*

$$V = W_1 \oplus \cdots \oplus W_s$$

is a direct sum of nonzero vector subspaces W_i *of* V*.*

(2) Conversely, suppose that

$$V = W_1 \oplus \cdots \oplus W_s$$

is a direct sum of nonzero vector subspaces W_i *of* V*. Let* B_i *be a basis of* W_i*. Then*

$$B = B_1 \coprod \cdots \coprod B_s$$

is a basis of V *and a disjoint union of non-empty sets* B_i*.*

(3) In particular, if

$$V = W_1 \oplus \cdots \oplus W_s$$

is a direct sum, then

$$\begin{aligned} \dim_F V &= \sum_{i=1}^{s} \dim_F W_i \\ &= \dim_F W_1 + \cdots + \dim_F W_s. \end{aligned}$$

Proof. (1) and (2) are tutorial questions. For (3), by (2) we have

$$\dim V = |B| = |B_1| + \cdots + |B_s| = \dim W_1 + \cdots + \dim W_s.$$

□

Exercise 5.20. (**Subspaces of the same dimension**) Suppose that W is a subspace of a **finite-dimensional** vector space V. Then

$$V = W \iff \dim V = \dim W.$$

Hint. Take a basis of W and apply Theorem 5.18.

6 Row space and Column space (T4)

In this section, we give some applications of §5 to row (resp. column) vector n-space F^n (resp. F^n_c), and find (simpler) bases of their subspaces.

We recall some results from Linear Algebra I:

Definition 6.1. (**Column / row space, Null space, Nullity, Range of A**). Let

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

be an $m \times n$ matrix with entries in a field F. Let

$$\mathrm{Col}(A) := \mathrm{Span}\{\mathbf{c}_1 := \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \mathbf{c}_n := \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}\}$$

be the **column space** of A, and let $\text{Row}(A) :=$

$$\text{Span}\{\mathbf{r}_1 := (a_{11}, .., a_{1n}), .., \mathbf{r}_m := (a_{m1}, .., a_{mn})\}$$

be the **row space** of A so that we can write

$$A = (\mathbf{c}_1, \ldots, \mathbf{c}_n) = \begin{pmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m \end{pmatrix}.$$

The **range of A** is defined as

$$R(A) = \{AX \mid X \in F_c^n\}.$$

The **nullity of A** is defined as the dimension of the **null space** (or **Kernel of A**)

$$\text{Ker}(A) := \text{Null}(A) := \{X \in F_c^n \mid AX = \mathbf{0}\}$$

i.e.,

$$\text{nullity}(A) = \dim \text{Null}(A).$$

Theorem 6.2. *(**Rank of a matrix - Matrix dimension theorem**)*

(1) The range equals the column space:

$$R(A) = \mathrm{Col}(A).$$

(2) Column and row spaces have the same dimension:

$$\dim_F \mathrm{Col}(A) = \dim_F \mathrm{Row}(A)$$

which is called the **rank of** A, *i.e.,*

$$\mathrm{rank}(A) := \dim_F \mathrm{Col}(A) = \dim_F \mathrm{Row}(A).$$

(3) There is a dimension theorem:

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = n$$

where n is the number of columns in A.

Proof. Refer to Linear Algebra I. E.g. [S. L. Ma, V. Tan and K. L. Ng, Linear Algebra I, Theorem 4.2.1]. (3) is also proved in Theorem 7.45. □

For square matrix, there is a neat result.

Theorem 6.3. (**L.I. v.s. Det.**) *In Theorem 6.2, suppose that $m = n$ so that A is a square matrix of order n. Then the following are equivalent.*

(1) A is an invertible matrix, i.e., A has a so called **inverse** *$A^{-1} \in M_n(F)$ such that*

$$AA^{-1} = I_n$$

(and hence automatically $A^{-1}A = I_n$).

(2) A has nonzero determinant

$$\det(A) = |A| \neq 0$$

in F.

(3) The column vectors

$$\mathbf{c}_1, \ldots, \mathbf{c}_n$$

of A form a basis of the column vector n-space F_c^n.

(4) The row vectors

$$\mathbf{r}_1, \, \ldots, \, \mathbf{r}_n$$

of A form a basis of the row vector n-space F^n.

(5) The column vectors

$$\mathbf{c}_1, \, \ldots, \, \mathbf{c}_n$$

of A are linearly independent in F_c^n.

(6) The row vectors

$$\mathbf{r}_1, \, \ldots, \, \mathbf{r}_n$$

of A are linearly independent in F^n.

(7) The matrix equation

$$AX = \mathbf{0}$$

where

$$X = (x_1, \ldots, x_n)^t = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

has only the trivial solution: $X = \mathbf{0}$.

Proof. This follows from Theorem 5.18 and Linear Algebra I (e.g. [S. L. Ma, V. Tan and K. L. Ng, Linear Algebra I, Theorem 2.4.5, Problem 4.2.6]). □

Extra exercise 6.4. Let $A \in M_{m\times n}(F)$. Then

$$\begin{aligned} &\operatorname{rank}(A) \\ &= \max\{s\,;\ \text{a minor of } A \text{ of order } s \text{ is invertible}\} \\ &= \max\{s\,;\ \text{a minor of } A \text{ of order } s \text{ has nonzero det.}\} \end{aligned}$$

Hint. Apply Theorems 6.2 and 6.3.

6.5. (**Row operations, elementary matrices, row equivalent matrices**)

Two matrices $A = (a_{ij})$ and $B = (b_{ij})$ (of the same size) in $M_{m\times n}(F)$ are **row equivalent** if one of them is obtained from the other by applying a sequence of elementary row operations.

There are exactly three types of **elementary row operations** (or simply **row operations**):

(I) Switch Row i with Row j :

$$R_i \longleftrightarrow R_j.$$

(II) Add a scalar multiple of Row i to Row j :

$$R_j \rightarrow R_j + \alpha R_i, \quad \text{and}$$

(III) Multiply a *nonzero* scalar to Row i :

$$R_i \rightarrow \alpha R_i.$$

If B is obtained from A by the row operation of type(I): $R_i \longleftrightarrow R_j$. then

$$B = EA$$

where

$$E \in M_m(F)$$

is obtained from I_m by performing the same row operation on I_m: $R_i \longleftrightarrow R_j$. E.g. when $m = 4$ and $(i, j) = (2, 3)$ we have

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

If B is obtained from A by the row operation of type(II): $R_j \to R_j + \alpha R_i$, then

$$B = EA$$

where E is obtained from the identity matrix $I_m \in M_m(F)$ by performing the same row operation on I_m: $R_j \to R_j + \alpha R_i$. E.g. when $m = 4$ and

$(i, j) = (1, 3)$ we have

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

If B is obtained from A by the row operation of type(III): $R_i \to \alpha R_i$, then

$$B = EA$$

where E is obtained from I_m by performing the same row operation on I_m: $R_i \to \alpha R_i$. E.g. when $m = 4$ and $i = 4$ we have

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

The matrices E above are called **elementary matrices**.

Therefore, if B is obtained from A by a sequence of elementary row operations

$$\mathrm{Op}_1, \dots, \mathrm{Op}_s$$

then

$$B = E_s \cdots E_1 A$$

where E_i are the elementary matrices corresponding to the operations Op_i as illustrated above.

Write

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = (\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

where

$$\mathbf{a}_j := \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \quad (j = 1, \dots, n)$$

are column vectors of A, and

$$B = (b_{ij}) = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ b_{21} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = (\mathbf{b}_1, \cdots, \mathbf{b}_n)$$

where

$$\mathbf{b}_j := \begin{pmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{pmatrix} \quad (j = 1, \ldots, n)$$

are column vectors of B.

Theorem 6.6. *(**Row op preserves columns relations**) Suppose that A and B are row equivalent. Then we have:*

(1) If the column vectors

$$\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_s}, \mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_t}$$

of A satisfies a relation

$$c_{i_1}\mathbf{a}_{i_1} + \cdots + c_{i_s}\mathbf{a}_{i_s} = c_{j_1}\mathbf{a}_{j_1} + \cdots + c_{j_t}\mathbf{a}_{j_t}$$

for some scalars $c_{i_k} \in F$*, then the corresponding column vectors*

$$\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_s}, \mathbf{b}_{j_1}, \ldots, \mathbf{b}_{j_t}$$

of B *satisfies exactly the same relation:*

$$c_{i_1}\mathbf{b}_{i_1} + \cdots + c_{i_s}\mathbf{b}_{i_s} = c_{j_1}\mathbf{b}_{j_1} + \cdots + c_{j_t}\mathbf{b}_{j_t}.$$

The converse is also true.

(2) The column vectors

$$\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_s}$$

of A *are linearly dependent if and only if the corresponding column vectors*

$$\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_s}$$

of B *are linearly dependent.*

(3) The column vectors

$$\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_s}$$

of A are linearly independent if and only if the corresponding column vectors

$$\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_s}$$

of B are linearly independent.

(4) The column vectors

$$\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_s}$$

of A form a basis of the column space

$$\mathrm{Col}(A) = \mathrm{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$$

if and only if the corresponding column vectors

$$\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_s}$$

form a basis of the column space

$$\mathrm{Col}(B) = \mathrm{Span}\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}.$$

(5) If

$$B_1 := \{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_s}\}$$

is a maximal L.I. subset of the set

$$C := \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$$

of all column vectors (i.e., B_1 is L.I. and every subset of C larger than B_1 is L.D.), then B_1 is a basis of the column space $\mathrm{Col}(A)$ *of A.*

(6) *Suppose that B is in row-echelon form with leading entries at* **columns**

$$i_1, \ \ldots, \ i_s.$$

Then

$$\mathbf{a}_{i_1}, \ \ldots, \ \mathbf{a}_{i_s}$$

form a basis of the column space $\mathrm{Col}(A)$ *of A.*

E.g. if

$$B = \begin{pmatrix} 2 & * & * & * \\ 0 & 0 & -3 & * \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

then the leading entries

$$2, -3, 5$$

are at the columns

$$i_1 = 1, \, i_2 = 3, \, i_3 = 4$$

respectively, and hence

$$\mathbf{a}_1, \, \mathbf{a}_3, \, \mathbf{a}_4$$

form a basis of $\mathrm{Col}(A)$.

(7) The row space of A and B are identical:

$$\mathrm{Row}(A) = \mathrm{Row}(B).$$

But the column spaces of A and B may not be the same.

(8) Suppose that

$$B = (b_{ij}) = \begin{pmatrix} \mathbf{b}'_1 \\ \vdots \\ \mathbf{b}'_m \end{pmatrix}$$

is in row-echelon form with leading entries at

rows

$$j_1, \ldots, j_t.$$

Then

$$\mathbf{b}'_{j_1}, \ldots, \mathbf{b}'_{j_t}$$

form a basis of the **row** *space* $\mathrm{Row}(A) = \mathrm{Row}(B)$ *of* A *(and also* B*).*

E.g. for the B *in* (7)*, the leading entries*

$$2, -3, 5$$

of B *are at the rows*

$$j_1 = 1, j_2 = 2, j_3 = 3$$

respectively, and hence

$$\mathbf{b}_1', \mathbf{b}_2', \mathbf{b}_3'$$

form a basis of $\text{Row}(A) = \text{Row}(B)$.

Proof. (5) is a consequence of Theorem 5.13.

(2) follows from (1).

(3) is the contraposition of (2).

(4) is a consequence of (3) and (5).

(6) follows from (4).

For (7), see e.g. [S. L. Ma, V. Tan and K. L. Ng, Linear Algebra I, Theorem 4.1.6].

(8) is a consequence of (7).

For (1), by the assumption, B is obtained from A by a sequence of elementary row operations, and hence

$$B = EA$$

where $E \in M_m(F)$ is a product of some elementary matrices (cf. 6.5). Now

$$\begin{aligned}(\mathbf{b}_1, \cdots, \mathbf{b}_n) = B &= EA \\ &= E(\mathbf{a}_1, \cdots, \mathbf{a}_n) \\ &= (E\mathbf{a}_1, \cdots, E\mathbf{a}_n)\end{aligned}$$

and hence

$$\mathbf{b}_1 = E\mathbf{a}_1, \ldots, \mathbf{b}_n = E\mathbf{a}_n.$$

Suppose that there is a relation

$$c_{i_1}\mathbf{a}_{i_1} + \cdots + c_{i_s}\mathbf{a}_{i_s} = c_{j_1}\mathbf{a}_{j_1} + \cdots + c_{j_t}\mathbf{a}_{j_t}. \ (*1)$$

Multiplying it by E, we get

$$c_{i_1}E\mathbf{a}_{i_1} + \cdots + c_{i_s}E\mathbf{a}_{i_s} = c_{j_1}E\mathbf{a}_{j_1} + \cdots + c_{j_t}E\mathbf{a}_{j_t}$$

i.e.,

$$c_{i_1}\mathbf{b}_{i_1} + \cdots + c_{i_s}\mathbf{b}_{i_s} = c_{j_1}\mathbf{b}_{j_1} + \cdots + c_{j_t}\mathbf{b}_{j_t}. \ (*2)$$

The converse is true because A is also obtained from B by a sequence of elementary row operations, or because E is invertible and we can multiplying (*2) by E^{-1} to get back (*1) (from (*2)), □

Example 6.7. (**Finding basis for a subspace of row vector space F^n, expressing vectors in a subspace as linear combinations of vectors in its basis**) Consider the row vectors

$$\mathbf{v}_1 = (1, 2, 2, -1), \ \mathbf{v}_2 = (-3, -6, -6, 3),$$
$$\mathbf{v}_3 = (4, 9, 9, -4), \ \mathbf{v}_4 = (-2, -1, -1, 2),$$
$$\mathbf{v}_5 = (5, 8, 9, -5), \ \mathbf{v}_6 = (4, 2, 7, -4).$$

in the row vector 4-space F^4.

We want to find a subset S_1 of

$$S := \{\mathbf{v}_1, \dots, \mathbf{v}_6\}$$

as a basis of the subspace $\mathrm{Span}(S)$ of the row vector 4-space F^4, and express every vector $\mathbf{v}_i$ (not in S_1) as a linear combination of the vectors in the basis S_1.

First, regard $\mathbf{v}_i$ as columns of a matrix A:

$$A = (\mathbf{v}_1^t, \ldots, \mathbf{v}_6^t) = \begin{pmatrix} 1 & -3 & 4 & -2 & 5 & 4 \\ 2 & -6 & 9 & -1 & 8 & 2 \\ 2 & -6 & 9 & -1 & 9 & 7 \\ -1 & 3 & -4 & 2 & -5 & -4 \end{pmatrix}.$$

Here, since the initial vectors $\mathbf{v}_i$ are not column vectors yet, we need to transpose them into column vectors : $\mathbf{v}_i \to \mathbf{v}_i^t$.

Applying elementary row operation on A to get its row-echelon or even reduced row-echelon form:

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

$$R = \begin{pmatrix} 1 & -3 & 0 & -14 & 0 & -37 \\ 0 & 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = (\mathbf{u}_1, \ldots, \mathbf{u}_6).$$

Since the leading entries (or rather the leading 1's now) of the echelon form R are at the columns

$$1,\ 3,\ 5$$

the subset $S_1^t = \{\mathbf{v}_1^t,\ \mathbf{v}_3^t,\ \mathbf{v}_5^t\}$ of $S^t = \{\mathbf{v}_1^t, \ldots, \mathbf{v}_6^t\}$ (i.e., columns of A) is a basis of $\mathrm{Col}(A) = \mathrm{Span}(S^t)$ by Theorem 6.6.

Taking the transpose (cf. Easy note 1 below), the subset

$$S_1 = \{\mathbf{v}_1,\ \mathbf{v}_3,\ \mathbf{v}_5\}$$

of $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_6\}$ is a basis of $\mathrm{Span}(S)$.

Easy note 1.

(1) Row vectors

$$\mathbf{w}_1,\ \ldots,\ \mathbf{w}_s$$

form a basis of a vector space

$$\mathrm{Span}\{\mathbf{v}_1,\ \ldots,\ \mathbf{v}_m\}$$

if and only if their transposes

$$\mathbf{w}_1^t,\ \ldots,\ \mathbf{w}_s^t$$

form a basis of the vector space

$$\mathrm{Span}\{\mathbf{v}_1^t,\ \ldots,\ \mathbf{v}_s^t\}.$$

(2) A relation

$$\sum_{i=1}^{m} a_i\mathbf{v}_i = \sum_{i=m+1}^{n} a_i\mathbf{v}_i$$

holds for row vectors $\mathbf{v}_i$ if and only if the same relation

$$\sum_{i=1}^{m} a_i\mathbf{v}_i^t = \sum_{i=m+1}^{n} a_i\mathbf{v}_i^t$$

holds for the transposes $\mathbf{v}_i^t$.

By inspection, one can express columns vectors $\mathbf{u}_j$ of R as linear combinations of $\mathbf{u}_1, \mathbf{u}_3, \mathbf{u}_5$:

$$\mathbf{u}_2 = -3\mathbf{u}_1$$

$$\mathbf{u}_4 = -14\mathbf{u}_1 + 3\mathbf{u}_3$$

$$\mathbf{u}_6 = -37\mathbf{u}_1 + 4\mathbf{u}_3 + 5\mathbf{u}_5,$$

hence (by Theorem 6.6) one gets exactly the same relations among column vectors $\mathbf{v}_i^t$ of A:

$$\mathbf{v}_2^t = -3\mathbf{v}_1^t$$

$$\mathbf{v}_4^t = -14\mathbf{v}_1^t + 3\mathbf{v}_3^t$$

$$\mathbf{v}_6^t = -37\mathbf{v}_1^t + 4\mathbf{v}_3^t + 5\mathbf{v}_5^t$$

or equivalently (by taking the transpose, cf. Easy note 1 above):

$$\mathbf{v}_2 = -3\mathbf{v}_1$$

$$\mathbf{v}_4 = -14\mathbf{v}_1 + 3\mathbf{v}_3$$

$$\mathbf{v}_6 = -37\mathbf{v}_1 + 4\mathbf{v}_3 + 5\mathbf{v}_5.$$

Example 6.8. (**Finding basis for a subspace**

of column vector space F^n, expressing vectors in a subspace as linear combinations of vectors in its basis) Consider the column vectors

$$\mathbf{w}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix}, \quad \mathbf{w}_2 = \begin{pmatrix} -2 \\ -5 \\ 5 \\ 6 \end{pmatrix},$$

$$\mathbf{w}_3 = \begin{pmatrix} 0 \\ -3 \\ 15 \\ 18 \end{pmatrix}, \quad \mathbf{w}_4 = \begin{pmatrix} 0 \\ -2 \\ 10 \\ 8 \end{pmatrix}, \quad \mathbf{w}_5 = \begin{pmatrix} 3 \\ 6 \\ 0 \\ 6 \end{pmatrix}.$$

in the column vector 4-space F_c^4.

We want to find a subset T_1 of

$$T := \{\mathbf{w}_1, \ldots, \mathbf{w}_5\}$$

as a basis of the subspace $\mathrm{Span}(T)$ of the column vector 4-space F_c^4, and express every vector $\mathbf{w}_i$ (not in T_1) as a linear combination of the vectors in the basis T_1.

First, regard $\mathbf{w}_i$ as columns of a matrix A_1:

$$A_1 = (\mathbf{w}_1, \ldots, \mathbf{w}_5) = \begin{pmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 6 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{pmatrix}.$$

Applying elementary row operation on A_1 to get its row-echelon or even reduced row-echelon form:

$$A_1 \longrightarrow \begin{pmatrix} 1 & -2 & 0 & 0 & 3 \\ 0 & 1 & 3 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow$$

$$R_1 = \begin{pmatrix} 1 & 0 & 0 & -2 & 3 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = (\mathbf{q}_1, \ldots, \mathbf{q}_5).$$

Since the leading entries (or rather the leading 1's now) of the echelon form R_1 are at the columns **1, 2, 3**, the subset

$$T_1 = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$$

is a basis of $\text{Col}(A_1) = \text{Span}(T)$ by Theorem 6.6.

By inspection, one can express columns vectors $\mathbf{q}_j$ of R_1 as linear combinations of $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$:

$$\mathbf{q}_4 = -2\mathbf{q}_1 - \mathbf{q}_2 + \mathbf{q}_3, \quad \mathbf{q}_5 = 3\mathbf{q}_1.$$

By Theorem 6.6, one gets exactly the same relations among the column vectors $\mathbf{w}_i$ of A_1:

$$\mathbf{w}_4 = -2\mathbf{w}_1 - \mathbf{w}_2 + \mathbf{w}_3, \quad \mathbf{w}_5 = 3\mathbf{w}_1.$$

Example 6.9. (**Simpler basis for a subspace of the row vector space F^n**) Consider the row vectors

$$\mathbf{p}_1 = (1, -3, 4, -2, 5, 4), \;\; \mathbf{p}_2 = (2, -6, 9, -1, 8, 2),$$

$$\mathbf{p}_3 = (2, -6, 9, -1, 9, 7), \;\; \mathbf{p}_4 = (-1, 3, -4, 2, -5, -4).$$

in the row vector 6-space F^6.

We shall find a simpler basis for the subspace

$$\mathrm{Span}\{\mathbf{p}_1, \ldots, \mathbf{p}_4\}$$

of F^6.

First, regard $\mathbf{p}_i$ as row vectors of a matrix A:

$$A = \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ \mathbf{p}_4 \end{pmatrix}.$$

Applying elementary row operation on A to get its reduced row-echelon form (as done in Example

6.7):

$$A \to R = \begin{pmatrix} 1 & -3 & 0 & -14 & 0 & -37 \\ 0 & 0 & 1 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \\ \mathbf{w}_4 \end{pmatrix}.$$

Since the leading 1's of R are at the rows $1, 2, 3$, the row vectors

$$\mathbf{w}_1,\ \mathbf{w}_2,\ \mathbf{w}_3$$

of R form a basis of the row space (cf. Theorem 6.6)

$$\mathrm{Row}(R) = \mathrm{Row}(A) = \mathrm{Span}\{\mathbf{p}_1,\ \ldots,\ \mathbf{p}_4\}.$$

Example 6.10. (**Simpler basis for a subspace**

of column vector space F_c^n) Consider the column vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 0 \\ 3 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ -5 \\ -3 \\ -2 \\ 6 \end{pmatrix}$$

$$\mathbf{v}_3 = \begin{pmatrix} 0 \\ 5 \\ 15 \\ 10 \\ 0 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} 2 \\ 6 \\ 18 \\ 8 \\ 6 \end{pmatrix}$$

of the column vector 5-space F_c^5. We shall find a simple basis for the subspace

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$$

of F_c^5

First, regard $\mathbf{v}_i$ as rows of a matrix

$$A_1 = \begin{pmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \mathbf{v}_3^t \\ \mathbf{v}_4^t \end{pmatrix}.$$

Here, since the initial vectors $\mathbf{v}_i$ are not row vectors yet, we need to transpose them into row vectors: $\mathbf{v}_i \rightarrow \mathbf{v}_i^t$.

Applying the elementary row operation to get its reduced row-echelon form (as done in Example 6.8):

$$A_1 \rightarrow R_1 = \begin{pmatrix} 1 & 0 & 0 & -2 & 3 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \\ \mathbf{u}_4 \end{pmatrix}.$$

Since the leading 1's of R_1 are at the rows $1, 2, 3$, the row vectors

$$\mathbf{u}_1,\ \mathbf{u}_2,\ \mathbf{u}_3$$

of R_1 form a basis of the row space (cf. Theorem 6.6):

$$\mathrm{Row}(R_1) = \mathrm{Row}(A_1) = \mathrm{Span}\{\mathbf{v}_1^t, \ldots, \mathbf{v}_4^t\}.$$

Taking the transpose (cf. Easy note 1 in Example 6.7),

$$\mathbf{u}_1^t = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -2 \\ 3 \end{pmatrix}, \mathbf{u}_2^t = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \mathbf{u}_3^t = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

form a basis of

$$\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}.$$

7 Quotient spaces and Linear transformations (T5)

In this section, we study **quotient spaces**, **linear transformations** T between vector spaces, **range = image, and kernel** of T. We will see that the concept of vector subspace is equivalent to that of the kernel of a linear transformation.

Definition 7.1. (**Sum of subsets of a space**) Let V be a vector space over a field F, and let S and T be subsets (which are not necessarily subspaces) of V. Define the **sum of S and T** as

$$S + T := \{s + t \mid s \in S,\, t \in T\}.$$

In general, given subsets S_i $(1 \leq i \leq r)$ of V, we can define the **sum of S_i** as

$$\begin{aligned}\sum_{i=1}^{r} S_i &= S_1 + \cdots + S_r \\ &:= \{\sum_{i=1}^{r} \mathbf{x}_i \,|\, \mathbf{x}_i \in S_i\} \\ &= \{\mathbf{x}_1 + \cdots + \mathbf{x}_r \,|\, \mathbf{x}_i \in S_i\}.\end{aligned}$$

If some S_i consists of a single element, say $S_2 = \{\mathbf{s}\}$, we simply write $S_1 + S_2 + S_3$ as

$$S_1 + \mathbf{s} + S_3.$$

Exercise 7.2. (**Inclusion and sum for subsets**) Let S, S_i and T_i be subsets (which are not necessarily subspaces) of a vector space V over a field F. Show the following:

(1) Associativity

$$(S_1 + S_2) + S_3 = S_1 + (S_2 + S_3).$$

(2) Commutativity

$$S_1 + S_2 = S_2 + S_1.$$

(3) If $S_1 \subseteq S_2$ and $T_1 \subseteq T_2$ then

$$S_1 + T_1 \subseteq S_2 + T_2.$$

Is the converse true?

(4) If W is a subspace of V, then

$$W + \{\mathbf{0}\} = W, \qquad W + W = W.$$

Is it still true if W is mearly a subset of V?

(5) Suppose that W is a subspace of V. Then

$$S + W = W \iff S \subseteq W.$$

Proof. (4) and (5) are more or less Exercise 4.10 = Tutorial question; the others can be proved similarly.

$\square$

Definition 7.3. (**Coset $\overline{\mathbf{v}}$**) Let V be a vector space over a field F and W a subspace of V. For any given $\mathbf{v} \in V$, the subset

$$\mathbf{v} + W := \{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in W\}$$

of V is called the **coset of W containing $\mathbf{v}$**. This subset is often denoted as

$$\overline{\mathbf{v}} := \mathbf{v} + W, \quad \text{or } [\mathbf{v}] := \mathbf{v} + W.$$

The vector $\mathbf{v}$ is called a **representative of the coset $\overline{\mathbf{v}}$**. As an example

$$\overline{\mathbf{0}} := \overline{\mathbf{0}_V} = \mathbf{0}_V + W = W.$$

Exercise 7.4. (**Coset relation; Important!**) Let W be a subspace of a vector space V. Show that the following are equivalent.

(1) $\mathbf{v} + W = W$, i.e., $\overline{\mathbf{v}} = \overline{\mathbf{0}}$.

(2) $\mathbf{v} \in W$.

(3) $\mathbf{v} + W \subseteq W$.

(4) $W \subseteq \mathbf{v} + W$.

A coset $\overline{\mathbf{v}}$ may have more than one representatives. In other words, different vectors may represent the same coset. Indeed, we have:

Theorem 7.5. *(**To be the same coset**) Let W be a subspace of V. Then for $\overline{\mathbf{v}_i} = \mathbf{v}_i + W$,*

$$\overline{\mathbf{v}_1} = \overline{\mathbf{v}_2} \iff \mathbf{v}_1 - \mathbf{v}_2 \in W \iff \mathbf{v}_2 - \mathbf{v}_1 \in W.$$

Proof. Use Exercise 7.2 or 7.4. □

The example below shows how the **quotient space of V modulo W**:

$$V/W := \{\bar{\mathbf{v}} = \mathbf{v} + W \,|\, \mathbf{v} \in V\}$$

looks like, and this phenomenon is quite general (cf. Theorem 7.40).

Example 7.6. (1) Let $W := \{(0,0,z) \,|\, z \in \mathbb{R}\}$ be the z-axis in the Euclidean 3-space $V := \mathbb{R}^3$. Show that

$$V/W = \{\,\overline{(x,y,0)} = (x,y,0) + W \,|\, x,y \in \mathbb{R}\}.$$

and

$$\overline{(x_1,y_1,0)} = \overline{(x_2,y_2,0)} \text{ in } V/W \iff (x_1,y_1,0) = (x_2,y_2,0$$

Thus, there is a bijection

$$\begin{aligned} \varphi : \mathbb{R}^2 &\to V/W \\ (x, y) &\mapsto \overline{(x, y, 0)}. \end{aligned}$$

One can verify that φ is indeed an isomorphism between vector spaces (cf. Definition 7.10).

(2) Suppose that $V = W_1 + W_2$. Show that

$$V/W_1 = \{\overline{\mathbf{w}} = \mathbf{w} + W_1 \,|\, \mathbf{w} \in W_2\}.$$

For $\mathbf{w}, \mathbf{w}' \in W_2$ can one always say that

$$\overline{\mathbf{w}} = \overline{\mathbf{w}'} \text{ in } V/W_1 \iff \mathbf{w} = \mathbf{w}' \,?$$

Remark 7.7. (**The quotient space V/W as the collapsing of W in V with the complement U of W retained**) Suppose that $V = U \oplus W$ is a direct sum of subspaces U and W. Then the map below is a bijection (and indeed a bijective linear transformation, i.e., an isomorphism to be defined below):

$$\begin{aligned} f : U &\to V/W \\ \mathbf{u} &\mapsto \bar{\mathbf{u}} = u + W. \end{aligned}$$

Definition 7.8. (**Quotient space**) Let W be a subspace of V. Let

$$V/W := \{\overline{\mathbf{v}} = \mathbf{v} + W \mid \mathbf{v} \in V\}$$

be the set of all cosets of W. It is called the **quotient space of V modulo W** (cf. Theorem 7.9 below).

We define a binary addition operation on V/W:

$$\begin{aligned} + : V/W \times V/W &\to V/W \\ (\overline{\mathbf{v}_1}, \overline{\mathbf{v}_2}) &\mapsto \overline{\mathbf{v}_1} + \overline{\mathbf{v}_2} := \overline{\mathbf{v}_1 + \mathbf{v}_2} \end{aligned}$$

and a scalar multiplication operation

$$\begin{aligned} \times : F \times V/W &\to V/W \\ (a, \overline{\mathbf{v}_1}) &\mapsto a\overline{\mathbf{v}_1} := \overline{a\mathbf{v}_1}. \end{aligned}$$

Theorem 7.9. (**Quotient space being well defined**) *Let V be a vector space over a field F and W a vector subspace of V. Then we have:*

(1) The binary addition operation and scalar multiplication operation on V/W in Definition 7.8 are well defined (i.e., if $\overline{\mathbf{v}_i'} = \overline{\mathbf{v}_i}$ then $\overline{\mathbf{v}_1' + \mathbf{v}_2'} = \overline{\mathbf{v}_1 + \mathbf{v}_2}$ and $\overline{a\mathbf{v}_1'} = \overline{a\mathbf{v}_1}$.

(2) V/W together with these binary addition and scalar multiplication operations, becomes a vector space over the same field F, with the zero vector

$$\mathbf{0}_{V/W} = \overline{\mathbf{0}_V} = \overline{\mathbf{w}}$$

for any $\mathbf{w} \in W$.

Definition 7.10. (**Linear transformation, and its Kernel and Image; Isomorphism**) Let V_i

be two vector spaces over the same field F. A map

$$\varphi : V_1 \to V_2$$

is called a **linear transformation from (or between) V_1 to V_2** if φ is compatible with the vector addition and scalar multiplication on V_1 and V_2 in the sense below:

$$\varphi(\mathbf{v}_1 + \mathbf{v}_2) = \varphi(\mathbf{v}_1) + \varphi(\mathbf{v}_2) \quad (\forall\, \mathbf{v}_i \in V)$$
$$\varphi(a\mathbf{v}) = a\varphi(\mathbf{v}), \quad (\forall\, a \in F, \forall\, \mathbf{v} \in V).$$

When $\varphi : V \to V$ is a linear transformation from V to itself, we call φ a **linear operator on V**.

A *linear transformation*

$$\varphi : V_1 \to V_2$$

as above is called an **isomorphism** if it is a bijection, i.e., injection and surjection, i.e., one to one and onto. In this case, we denote

$$V_1 \xrightarrow{\sim} V_2, \quad \text{or}$$

$$V_1 \overset{\varphi}{\simeq} V_2, \quad \text{or}$$

$$V_1 \simeq V_2, \quad \text{or}$$

$$V_1 \cong V_2.$$

and say that **V_1 is isomorphic to V_2** (via the isomorphism φ), or **V_1 and V_2 are isomorphic to each other**.

Note 1. The pre-assumption for φ to be an isomorphism is that φ is a linear transformation. In other words, a map is an isomorphism if and only if it is a bijective linear transformation.

If

$$\varphi : V_1 \to V_2$$

is an isomorphism, then its inverse

$$\varphi^{-1} : V_2 \to V_1$$

is well defined and also an isomorphism such that

$$\varphi \circ \varphi^{-1} = I_{V_2}, \quad \varphi^{-1} \circ \varphi = I_{V_1}.$$

Here

$$\begin{aligned} I_{V_i} := \mathrm{id}_{V_i} : V_i &\to V_i \\ \mathbf{v} &\mapsto \mathbf{v} \end{aligned}$$

is the identity map of V_i.

For a linear transformation $\varphi : V \to W$ as above (with $\mathbf{0}_V \in V$ and $\mathbf{0}_W \in W$ the zero vectors, respectively), we call V the **domain** of φ:

$$\mathrm{dom}(\varphi) := V,$$

W the **codomain** of φ:

$$\mathrm{codomain}(\varphi) := W,$$

and define the **Kernel of** φ as the pre-image of $\mathbf{0}_W$:

$$\begin{aligned} \mathrm{Ker}(\varphi) :&= \varphi^{-1}(\mathbf{0}_W) \\ &= \{\mathbf{v} \in V \mid \varphi(\mathbf{v}) = \mathbf{0}_W\} \subseteq V \end{aligned}$$

and the **range of φ** (or the **image of φ**) as:

$$R(\varphi) = \mathrm{Im}(\varphi) = \varphi(V)$$
$$= \{\varphi(\mathbf{v}) \,|\, \mathbf{v} \in V\} \subseteq W.$$

Example 7.11. (**Linear transformation T_A associated with a matrix A**) Let $A = (a_{ij}) \in M_{m\times n}(F)$ be an $m \times n$ matrix with entries in a field F. Define the map

$$T_A : F_c^n \to F_c^m$$
$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto AX.$$

Then T_A is a **linear transformation (associated with A)** and also called a **matrix transformation**. The kernel of T_A equals the null space (or kernel) of A:

$$\mathrm{Ker}(T_A) = \mathrm{Ker}(A) = \mathrm{Null}(A) = \{X \in F_c^n \,|, AX = 0\} \subseteq F_c^n.$$

The range (or image) of T_A equals the range of A:

$R(T_A) = T_A(F_c^n) = R(A) := \{AX \mid X \in F_c^n\} \subseteq F_c^m$.

Show the following (Tutorial question):

(1) T_A is an injection iff the null space

$$\mathrm{Null}(A) = 0$$

iff (the **nullity of A**) is zero:

$$(\dim(\mathrm{Null}(A) =:) \,\mathrm{nullity}(A) = 0.$$

(2) T_A is a surjection iff the **rank of A** is m:

$$(\dim R(A) =:) \ \mathrm{rank}(A) = m.$$

(3) T_A is an isomorphism iff A is an invertible square matrix (i.e., $m = n$ and the inverse A^{-1} exists)

Similarly, we can define another **linear transformation associated with A**:

$$T'_A : F^m \to F^n$$

$$Y = (y_1, \ldots, y_m) \mapsto YA.$$

between row vector spaces.

Remark 7.12. (**Image of $\mathbf{0}_V$**) If $T : V \to W$ is a linear transformation, then

$$T(\mathbf{0}_V) = \mathbf{0}_W.$$

Remark 7.13. (= Tutorial question) (**Direct sum vs quotient space**) Let $V = U \oplus W$, i.e., the space V is a direct sum of subspaces U, W. Show that the map below is an isomorphism

$$\begin{aligned} U &\to V/W \\ \mathbf{u} \mapsto \bar{\mathbf{u}} &= \mathbf{u} + W. \end{aligned}$$

Example 7.14. (**Zero map** 0) The map 0:

$$\begin{aligned} 0 : V &\to W \\ \mathbf{v} &\mapsto \mathbf{0}_W \end{aligned}$$

which maps every vector in V to the zero vector $\mathbf{0}_W$, is a linear transformation and called the **zero map** or **zero linear transformation from V to W**.

Example 7.15. (**Scalar map αI_V**) Let V be a vector space over a field F and let $\alpha \in F$. Set

$$I_V := \mathrm{id}_V$$

which is the identity map from V to itself.

The **scalar map**

$$\begin{aligned} \alpha I_V : V &\to V \\ \mathbf{v} &\mapsto \alpha I_V(\mathbf{v}) = \alpha \mathbf{v} \end{aligned}$$

is a linear operator on V.

When $\alpha = 0$, we get a zero map 0 as in Example 7.14.

When $\alpha = 1$,

$$1\, I_V = I_V = \mathrm{id}_V$$

is the identity map on V.

When $\alpha \neq 0$ in F, show that αI_V is an isomorphism with

$$\alpha^{-1}\, I_V$$

as its inverse.

Example 7.16. (**Inclusion map**) Let W be a subspace of a vector space V. Then the inclusion map

$$\begin{aligned} \iota : W &\to V \\ \mathbf{w} &\mapsto \mathbf{w} \end{aligned}$$

is an (injective) linear transformation, because the operations on W are borrowed from those on V.

Example 7.17. (**Restriction map**) Let $T : V \to W$ be a linear transformation. Let $V_1 \subseteq V$ be a vector subspace of V, with $\iota : V_1 \to V$ the inclusion map. Then we denote

$$\begin{aligned} T|_{V_1} = T \circ \iota : V_1 &\to W \\ \mathbf{v} &\mapsto (T \circ \iota)(\mathbf{v}) = T(\iota(\mathbf{v})) = T(\mathbf{v}). \end{aligned}$$

It is a linear transformation from V_1 to W and called the **restriction** of T to the subspace V_1.

Example 7.18. (**Linear transformation $F^3 \to$**

F^4) The map

$T_1 : F^3 \to F^4,$

$(x, y, z) \mapsto (x - y + z, y + z, x + y + 3z, x + 2z)$

is a linear transformation. Indeed, one can check that T_1 equals some T'_A (cf. Example 7.11).

Example 7.19. (**Non-linear transformations**)

The maps

$$T_2 : F^3 \to F^3$$

$$(x, y, z) \mapsto (x^2, y, z)$$

assuming $|F| \geq 3$, and

$$T_3 : F^3 \to F^3$$

$$(x, y, z) \mapsto (0, x + y + z, 1)$$

are not linear transformations (**why**?)

Example 7.20. (**Projection map**) Consider the **projection**

$$\mathrm{Pr} : F^3 \to F^2$$

$$(x, y, z) \mapsto (x, y).$$

of vectors in the xyz-space F^3 to the xy-plane F^2:

Then Pr is a linear transformation. Verify that the image

$$\Pr(F^3) = F^2$$

so Pr is onto; and

$$\mathrm{Ker}(\Pr) = \{(0, 0, z) \mid z \in \mathbf{R}\}$$

which is the z-axis.

Example 7.21. (**Trace map**) The **trace map**

$\mathrm{Tr} : M_n(F) \to F,$

$$A = (a_{ij}) \mapsto \mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + \cdots + a_{nn}.$$

is a surjective linear transformation (i.e., $\mathrm{Im}(\mathrm{Tr}) = F$). When $n \geq 2$, it is not an injection and hence is not an isomorphism.

Example 7.22. The **determinant** map

$$\det : M_n(F) \to F,$$

$$A = (a_{ij}) \mapsto \det(A) = |A|.$$

is *not* a linear transformation when $n \geq 2$ and $|F| = \infty$ (**why**?)

Theorem 7.23. *(**Equivalent linear transformation definition**) Let*

$$\varphi : V_1 \to V_2$$

be a map between two vector spaces V_i over the same field F. Then the following are equivalent.

(1) φ is a linear transformation.

(2) φ is compatible with taking linear combination in the sense below:

$$\varphi(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1\varphi(\mathbf{v}_1) + a_2\varphi(\mathbf{v}_2)$$

for all $a_i \in F$, $\mathbf{v}_i \in V$.

Proof. The proof is similar to that for 'Equivalent subspace definition' theorem 3.8. $\square$

Exercise 7.24. (**Evaluate T at a basis**) Let V be a vector space over a field F and with a basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots\}$. Let

$$T : V \to W$$

be a linear transformation. Then T is uniquely determined by its valuations

$$T(\mathbf{u}_i) \quad (i = 1, 2, \dots)$$

at the basis B. Namely, if

$$T' : V \to W$$

is another linear transformation such that

$$T'(\mathbf{u}_i) = T(\mathbf{u}_i) \quad (i = 1, 2, \dots)$$

then they are equal: $T = T'$.

Exercise 7.25. (**Composite of linear transformations**) Let

$$\varphi_1 : V_1 \to V_2$$

and

$$\varphi_2 : V_2 \to V_3$$

be two linear transformations. Then the composite map

$$\begin{aligned} \varphi_2 \circ \varphi_1 : V_1 &\to V_3 \\ \mathbf{v} &\mapsto (\varphi_2 \circ \varphi_1)(\mathbf{v}) = \varphi_2(\varphi_1(\mathbf{v})) \end{aligned}$$

is again a linear transformation.

Sometimes, we denote

$$\varphi_2 \varphi_1 := \varphi_2 \circ \varphi_1.$$

Exercise 7.26. Let

$$\begin{aligned} T_A : F_c^n &\to F_c^m \\ X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &\mapsto AX \end{aligned}$$

and

$$\begin{aligned} T_B : F_c^m &\to F_c^r \\ Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} &\mapsto BY \end{aligned}$$

be the linear transformations associated with the matrices $A \in M_{m\times n}(F)$ and $B \in M_{r\times m}(F)$. Then

$$T_B \circ T_A = T_{BA}.$$

Example 7.27. (**Quotient map**) Let V be a vector space over a field F. Let W be a subspace of V and V/W the quotient space (cf. Definition 7.8). The map

$$\begin{aligned} \gamma : V &\to V/W \\ \mathbf{v} &\mapsto \overline{\mathbf{v}} \end{aligned}$$

is a linear transformation and called the **quotient map from V to its quotient space V/W**. One verifies that γ is surjective and

$$\mathrm{Ker}(\gamma) = W.$$

Theorem 7.28. *(**Image being a vector subspace**) Let*

$$\varphi : V \to W$$

be a linear transformation between two vector spaces over the same field F. Let V_1 be a vector subspace of V. Then the image of V_1:

$$T(V_1) = \{T(\mathbf{u}) \,|\, \mathbf{u} \in V_1\}$$

is a vector subspace of W.

In particular, $T(V)$ is a vector subspace of W.

The theorem below shows that the concept of vector subspace is equivalent to that of the Kernel of a linear transformation.

Theorem 7.29. *(**Subspace v.s. Kernel**) Let V be a vector space over a field F.*

(1) Suppose that

$$\varphi : V \to U$$

is a linear transformation. Then the kernel $\mathrm{Ker}(\varphi)$ *is a vector subspace of* V*.*

(2) Conversely, suppose that W *is a vector subspace of* V*. Then there is a linear transformation*

$$\varphi : V \to U$$

such that

$$W = \mathrm{Ker}(\varphi).$$

Proof. (1) By 'Equivalent subspace definition' theorem 3.8, we only need to show: Whenever $a_i \in F$ and $\mathbf{v}_i \in \mathrm{Ker}(\varphi)$ (i.e., $\varphi(\mathbf{v}_i) = \mathbf{0}_U$), we have

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 \in \mathrm{Ker}(\varphi).$$

Equivalently, we need to show the claim:

$$\varphi(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = \mathbf{0}_U.$$

This claim is verified as follows. The map φ, being a liner transformation, is compatible with taking linear combination (cf. 'Equivalent linear transformation definition' theorem 7.23), so

$$\begin{aligned}\varphi(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) &= a_1\varphi(\mathbf{v}_1) + a_2\varphi(\mathbf{v}_2)\\ &= a_1\mathbf{0}_U + a_2\mathbf{0}_U = \mathbf{0}_U\end{aligned}$$

(cf. Exercise 2.14). This proves the claim and also (1).

(2) Suppose that W is a subspace of V. If we let

$$\begin{aligned}\gamma : V &\to V/W\\ \mathbf{v} &\mapsto \overline{\mathbf{v}}\end{aligned}$$

be the quotient map (a linear transformation), then (cf. Example 7.27)

$$\mathrm{Ker}(\gamma) = W.$$

$\square$

Extra exercise 7.30. (= Tutorial question) (**Inverse of a subspace**) Let $T : V \to W$ be a linear transformation between vector spaces over a field F. Let W_1 be a subspace of W. Show that the preimage

$$T^{-1}(W_1) = \{\mathbf{v} \in V \,|\, T(\mathbf{v}) \in W_1\}$$

of W_1 is a subspace of V.

Exercise 7.31. (**To be injective**) Let

$$\varphi : V \to W$$

be a linear transformation. Show that φ is injective if and only if $\mathrm{Ker}(\varphi) = \{\mathbf{0}\}$.

The proof is similar to Tutorial 4.5a.

Extra exercise 7.32. (**Linear combination of maps, the space $\mathrm{Hom}_F(V, W)$ of all linear maps from V to W**) Let

$$\varphi_i : V \to W$$

be linear transformations between vector spaces over the same field F, and let α, $\alpha_i \in F$. Then the **map addition**

$$\begin{aligned} \varphi_1 + \varphi_2 : V &\to W \\ \mathbf{v} &\mapsto \varphi_1(\mathbf{v}) + \varphi_2(\mathbf{v}), \end{aligned}$$

the **scalar multiplication**

$$\begin{aligned} \alpha\varphi_1 : V &\to W \\ \mathbf{v} &\mapsto \alpha\varphi_1(\mathbf{v}), \end{aligned}$$

and the **map linear combination**

$$\begin{aligned} \alpha_1\varphi_1 + \alpha_2\varphi_2 : V &\to W \\ \mathbf{v} &\mapsto \alpha_1\varphi_1(\mathbf{v}) + \alpha_2\varphi_2(\mathbf{v}). \end{aligned}$$

are all linear transformations.

Prove that the set of all linear transformations from V to W:

$$\mathrm{Hom}_F(V, W) := \{\varphi : V \to W \mid \varphi \text{ is a lin. transform.}\}$$

together with the map addition and scalar multiplication above, becomes a vector space over F, and called the **space of all linear transformations from V to W**. Its zero vector is the zero linear transformation (cf. Example 7.14).

Exercise 7.33. (**Powers of a map: T^n**) Let

$$T : V \to V$$

be a linear operator on V, i.e., a linear transformation from V to itself. Denote the composite map $T \circ T$ as T^2:

$$\begin{aligned} T^2 : V &\to V \\ \mathbf{v} &\mapsto T(T(\mathbf{v})). \end{aligned}$$

Similarly, Denote $T \circ T \circ \cdots \circ T$ (n times, $n \geq 1$) as T^n

$$T^n := T \circ T \cdots T \text{ (n times)}.$$

As convention, we set

$$T^0 = I_V = \mathrm{id}_V .$$

Thus, given a polynomial

$$f(x) = \sum_{i=0}^{n} a_i x^i = a_n x^n + \cdots + a_1 x + a_0 x^0 \ \in \ F[x]$$

we have a linear operator

$$f(T) = \sum_{i=0}^{n} a_i T^i = a_n T^n + \cdots + a_1 T + a_0 I_V$$

$$\begin{aligned} f(T) : V \ &\to \ V \\ \mathbf{v} \ &\mapsto \ f(T)(\mathbf{v}) = \sum_{i=0}^{n} a_i T^i(\mathbf{v}) \end{aligned}$$

on V (cf. Exercise 7.32).

Extra exercise 7.34. (**The ring $\mathrm{End}_F(V)$ of linear operators on V**) Let V be a vector space over a field F. Denote by

$$\mathrm{End}_F(V) := \mathrm{Hom}_F(V, V)$$

(cf. Exercise 7.32) the vector space of all linear operators on V. Define the multiplication $S \times T$ for $S, T \in \mathrm{End}_F(V)$ as the composite map:

$$S \times T := S \circ T : V \to V$$

which is again a linear operator on V (cf. Exercise 7.25) and hence is in $\mathrm{End}_F(V)$.

Show that $R := \mathrm{End}_F(V)$ together with this multiplication and the map addition in Exercise 7.32, becomes a (*not necessarily commutative*) ring (cf. Definition 2.3) with 0_R the zero map 0 and 1_R the identity map I_V.

Exercise 7.35. (**Isomorphism** $F^n \cong F^n_c$) Show that the map below is an isomorphism between the row vector n-space F^n and the column vector n-space F^n_c over the same field F:

$$\psi : F^n \to F^n_c$$
$$X = (x_1, \ldots, x_n) \mapsto X^t = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Exercise 7.36. (**Isomorphism** $M_{m\times n}(F) \cong F^{mn}$)

Show that the map below is an isomorphism between the matrix space $M_{m\times n}(F)$ and the row vector mn-space F^{mn} over the same field F:

$$\gamma : M_{m\times n}(F) \to F^{mn}$$

$$X = (x_{ij})_{1\le i\le m, 1\le j\le n} \mapsto (x_{11}, .., x_{1n}, x_{21}, .., x_{mn}).$$

Extra exercise 7.37. (**Isomorphism between subspace of polynomial ring $F[x]$ and row vector space**) Construct an isomorphism between the vector subspace (of all polynomials of degree $<$ n)

$$P_n[x] = \{g(x) \in F[x]\, ;\, \deg(g) < n\}$$

of the polynomial ring $F[x]$, and the row vector n-space F^n.

Theorem 7.38. *(**Equivalent isomorphism definition**) Let $\varphi : V \to W$ be a linear transformation. Then the following are true.*

(1) φ is an injection if and only if the kernel

$$\operatorname{Ker}(\varphi) = \{\mathbf{0}_V\}.$$

(2) φ is a surjection if and only if the image $\varphi(V) = W$.

(3) φ is an isomorphism if and only if

$$\operatorname{Ker}(\varphi) = \{\mathbf{0}_V\} \quad \text{and} \quad \varphi(V) = W.$$

Theorem 7.39. *(1st isomorphism theorem)* *Let*

$$\varphi : V \to U$$

be a linear transformation. Then there is an isomorphism

$$\begin{aligned} \overline{\varphi} : V/\operatorname{Ker}(\varphi) &\xrightarrow{\sim} \varphi(V) \subseteq U \\ \overline{\mathbf{v}} &\mapsto \varphi(\mathbf{v}). \end{aligned}$$

such that

$$\varphi = \overline{\varphi} \circ \gamma$$

where

$$\begin{aligned}\gamma : V &\to V/\operatorname{Ker}(\varphi)\\ \mathbf{v} &\mapsto \overline{\mathbf{v}}.\end{aligned}$$

is the quotient map, a linear transformation (cf. Exercise 7.27*).*

In particular, when φ is surjective, we have an isomorphism

$$\overline{\varphi} : V/\operatorname{Ker}(\varphi) \xrightarrow{\sim} U.$$

Theorem 7.40. *(**Finding basis of the quotient**) Let V be a vector space (over a field F) of finite dimension n. Let W be a subspace with a basis $B_1 = \{\mathbf{w}_1, \dots, \mathbf{w}_r\}$. Then we have:*

(1) B_1 extends to a basis

$$B := B_1 \coprod \{\mathbf{w}_{r+1}, \dots, \mathbf{w}_n\}$$

of V.

(2) The cosets

$$\{\overline{\mathbf{w}_{r+1}}, \ldots, \overline{\mathbf{w}_n}\}$$

is a basis of the quotient space V/W. In particular,

$$\dim_F V/W = \dim_F V - \dim_F W.$$

(3)

$$B_1 \coprod \{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_n\}$$

is a basis of V if and only if the cosets

$$\{\overline{\mathbf{u}_{r+1}}, \ldots, \overline{\mathbf{u}_n}\}$$

is a basis of V/W.

Proof. Since B_1 is linearly independent, (1) follows from Theorem 5.16. (2) follows from (3).

(3) is an exercise. □

Remark 7.41. Even when $\dim V = \infty$, it is still true that $\dim V = \dim W + \dim(V/W)$. It follows from Theorem 7.40 and Claim A (which follows from Theorem 7.40 (3)): if W is finite-dimensional with a basis $\{\mathbf{w}_i | i = 1, \dots r\}$ and V/W is finite-dimensional with a basis $\{\overline{\mathbf{w}_j} | j = r+1, \dots, n\}$ then $\{\mathbf{w}_k | k = 1, \dots, n\}$ is a basis of V and hence $\dim V = \dim W + \dim(V/W)$.

Exercise 7.42. (**Goodies of isomorphism**) Let $\varphi : V \to W$ be an isomorphism and let B be a subset of V. (We do not assume $\dim V$ or $\dim W$ is finite.) Then we have:

(1) If there is a relation

$$\sum_{i=1}^{r} a_i \mathbf{v}_i = \sum_{i=r+1}^{s} a_i \mathbf{v}_i \quad (*1)$$

among vectors $\mathbf{v}_i \in V$, then exactly the same relation

$$\sum_{i=1}^{r} a_i\, \varphi(\mathbf{v}_i) = \sum_{i=r+1}^{s} a_i\, \varphi(\mathbf{v}_i) \quad (*2)$$

holds among vectors $\varphi(\mathbf{v}_i) \in W$. The converse is also true.

(2) B is linearly dependent if and only if so is $\varphi(B)$.

(3) B is linearly independent if and only if so is $\varphi(B)$.

(4) We have

$$\varphi(\text{Span}(B)) = \text{Span}(\varphi(B)).$$

In particular,

$$\varphi(\text{Span}\{\mathbf{v}_1, .., \mathbf{v}_s\}) = \text{Span}\{\varphi(\mathbf{v}_1), .., \varphi(\mathbf{v}_s)\}.$$

(5) B spans V if and only if $\varphi(B)$ spans W.

(6) B is a basis of V if and only if $\varphi(B)$ is a basis of W. In particular,

$$\dim V = \dim W.$$

Hint. For (1), applying φ to (*1), we get (*2). Conversely, Applying φ^{-1} to (*2), we get back (*1).

(2) follows from (1).

(3) is the contraposition of (2).

(5) follows from (4).

(6) follows from (3) and (5).

(4) is true because linear transformation φ preserves linear combination.

Exercise 7.43. Let $A \in M_{m\times n}(F)$ and $B \in M_{n\times m}(F)$ with F a field. Then the following are equivalent.

(1) A and B are (square, i.e., $m = n$) invertible matrices, inverse to each other.

(2) $A\,B = I_m$ and $B\,A = I_n$.

Hint. (2) means $T_A \circ T_B = \mathrm{id}$ and $T_B \circ T_A = \mathrm{id}$, which imply that T_A and T_B are bijective and inverse to each other.

Warning. $A\,B = I_m$ alone does not imply that (1): A and B are invertible.

Extra exercise 7.44. (**To be isomorphic finite-dimensional spaces**) Let V and W be finite-dimensional vector spaces over the same field F. Then the following are equivalent.

(1) $\dim_F V = \dim_F W = n$.

(2) There is an isomorphism

$$\varphi : V \cong W.$$

(3) For some n, we have:

$$V \cong F^n \cong W.$$

In particular, for every given $n \geq 0$, there is, up to isomorphism, a unique vector space over F of dimension n.

Is '(1) $\Rightarrow$ (2)' still true when V and W are infinte-dimensional?

Hint. (3) $\Rightarrow$ (2) $\Rightarrow$ (1) is clear (cf. Exercise 7.42).

For (1) $\Rightarrow$ (3), take a basis

$$B_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$$

of V and

$$B_W = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$$

of W and show that

$$\begin{aligned} \varphi : V &\to W \\ \sum_{i=1}^{n} a_i \mathbf{v}_i &\mapsto \sum_{i=1}^{n} a_i \mathbf{w}_i \end{aligned}$$

is an isomorphism between vector spaces.

Theorem 7.45. *(Dimension theorem) Let $\varphi : V \to W$ be a linear transformation between vector spaces over a field F. (We do not assume $\dim V$ or $\dim W$ is finite.) Then*

$$\dim_F \operatorname{Ker}(\varphi) + \dim_F \varphi(V) = \dim_F V.$$

In particular, if $A \in M_{m\times n}(F)$ then

$$\operatorname{nullity}(A) + \operatorname{rank}(A) = n$$

where n is the number of columns of A.

Proof. The last part follows from the first part by applying to

$$\begin{aligned} \varphi = T_A : F_c^n &\to F_c^m \\ X &\mapsto AX. \end{aligned}$$

Indeed,

$$\text{nullity}(A) = \dim \text{Null}(A) = \dim \text{Ker}(T_A)$$

and (cf. Example 7.11):

$$\text{rank}(A) = \dim R(A) = \dim T_A(F_c^n).$$

For the first part, since there is an isomorphism (cf. Theorem 7.39)

$$\overline{\varphi} : V/\operatorname{Ker}(\varphi) \cong \varphi(V)$$

we have (cf. Exercise 7.42):

$$\dim V/\operatorname{Ker}(\varphi) = \dim \varphi(V) \quad (*1).$$

By Theorem 7.40 and Remark 7.41,

$$\dim V = \dim \operatorname{Ker}(\varphi) + \dim V/\operatorname{Ker}(\varphi) \quad (*2).$$

(*1) and (*2) together give the first part of the theorem. $\square$

Theorem 7.46. *(**2nd isomorphism theorem**)* *Let W_1, W_2 be vector subspaces of a vector space V. (We do not assume $\dim V$ is finite.)*

(1) Show that the map

$$\varphi : (W_1/(W_1 \cap W_2) \to (W_1 + W_2)/W_2$$
$$\mathbf{w} + (W_1 \cap W_2) = \overline{\mathbf{w}} \mapsto \overline{\overline{\mathbf{w}}} = \mathbf{w} + W_2$$

is a well defined isomorphism between vector spaces.

(2) A dimension formula:

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

Proof. (1) is a Tutorial question.

(2) follows from (1) and Exercise 7.42, and Theorem 7.40 and Remark 7.41:

$$\dim(W_1/(W_1 \cap W_2)) = \dim((W_1 + W_2)/W_2),$$

$$\dim W_1 = \dim(W_1 \cap W_2) + \dim(W_1/(W_1 \cap W_2)),$$

$$\dim(W_1 + W_2) = \dim W_2 + \dim((W_1 + W_2)/W_2).$$

□

Theorem 7.47. *(**Equivalent isomorphism definition**) Let*

$$\varphi : V \to W$$

be a linear transformation between vector spaces over a field F and of the same **finite dimension** *n. Then the following are equivalent.*

(1) φ is an isomorphism.

(2) φ is an injection (i.e., $\mathrm{Ker}(\varphi) = \{\mathbf{0}\}$; cf. Exercise 7.31)

(3) φ is a surjection (i.e., $\mathrm{Im}(\varphi) = W$).

Proof. (1) always implies (2) and (3).

We shall show '(2) $\Rightarrow$ (1)' and '(3) $\Rightarrow$ (1)'.

By Theorems 7.39 and 7.40 and Exercise 7.42,

$$\begin{aligned}\dim \varphi(V) &= \dim(V/\operatorname{Ker}(\varphi)) \\ &= \dim V - \dim \operatorname{Ker}(\varphi).\end{aligned} \qquad (*)$$

For '(2) $\Rightarrow$ (1)', assume (2) (and we will show (1)). Then

$$\dim \operatorname{Ker}(\varphi) = \dim\{\mathbf{0}\} = 0$$

and hence

$$\dim \varphi(V) = \dim V = n = \dim W$$

by (*) above. Thus W and its subspace $\varphi(V)$ have the same finite dimension n. By Exercise 5.20, $W = \varphi(V)$ and hence φ is surjective (so it is an isomorphism, using (2)). (1) is proved.

'(3) $\Rightarrow$ (1)' is similar. $\square$

Exercise 7.48. (**Importance of being finite**)

(1) Give an example to show that Theorem 7.47 and Exercise 5.20 are not true when $\dim V = \infty = \dim W$.

(2) Formulate similar theorem and exercise for a map $f : A \to B$ between sets.

8 Representation matrices of linear transformations (T6)

In this section, we describe a linear transformation between spaces, in terms of its representation matrix relative to their bases, and describe the change of bases in terms of its transition matrix.

Notation 8.1. Let

$$T : V \to W$$

be a linear transformation between vector spaces over a field F. Take a matrix

$$\mathbf{A} = (\mathbf{a}_{ij}) = \begin{pmatrix} \mathbf{a}_{11} & \cdots & \mathbf{a}_{1n} \\ \mathbf{a}_{21} & \cdots & \mathbf{a}_{2n} \\ \vdots & \vdots & \vdots \\ \mathbf{a}_{m1} & \cdots & \mathbf{a}_{mn} \end{pmatrix} \in M_{m\times n}(V)$$

of size $m \times n$ with vector entries $\mathbf{a}_{ij} \in V$. Define

$$T(\mathbf{A}) := (T(\mathbf{a}_{ij})) = \begin{pmatrix} T(\mathbf{a}_{11}) & \cdots & T(\mathbf{a}_{1n}) \\ T(\mathbf{a}_{21}) & \cdots & T(\mathbf{a}_{2n}) \\ \vdots & \vdots & \vdots \\ T(\mathbf{a}_{m1}) & \cdots & T(\mathbf{a}_{mn}) \end{pmatrix}.$$

In particular, when $m = 1$ or $n = 1$, we have

$$T(\mathbf{v}_1, \ldots, \mathbf{v}_n) := (T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)),$$

$$T(\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix}) := \begin{pmatrix} T(\mathbf{v}_1) \\ \vdots \\ T(\mathbf{v}_m) \end{pmatrix}.$$

Exercise 8.2. (**Linear transformation preserves matrix coefficient combination**) Let

$$T : V \to W$$

be a linear transformation between vector spaces over the same field F. Let

$$\mathbf{A} = (\mathbf{a}_{ij}) \in M_{m\times n}(V)$$

be a matrix with vector entries $\mathbf{a}_{ij} \in V$. Let

$$C = (c_{ij}) \in M_{r\times m}(F)$$

$$D = (d_{ij}) \in M_{n\times s}(F)$$

be matrices with scalar entries c_{ij}, $d_{ij} \in F$. Then

$$T(C\mathbf{A}D) = C(T(\mathbf{A}))D.$$

Hint. Each entry of $C\mathbf{A}D$ is a linear combination of some vectors of the form $c_{k\ell}\mathbf{a}_{ij}d_{rs}$. Use the linearity of T: the T preserves linear combination.

Notation 8.3. (**Regard a basis as a row vector in $M_{1\times n}(V)$**) In the sequel of these notes, a basis $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of a vector space V may be denoted as

$$B = (\mathbf{u}_1, \ldots, \mathbf{u}_n) \in M_{1\times n}(V)$$

and regarded as a row vector with vector entries in V. This is to simplify the presentation.

Definition 8.4. (**Coordinate vector $[\mathbf{v}]_B$**) Let V be vector space of dimension $n \geq 1$ over a field F. Let

$$B = B_V = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

be a basis of V. Every vector $\mathbf{v} \in V$ can be expressed as a linear combination

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

and this expression is unique (cf. Theorem 5.11).

We gather the coefficients c_i and form a column vector

$$[\mathbf{v}]_B := \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \in F_c^n$$

which is called the **Coordinate vector of v relative to (the basis) B.**

One can recover $\mathbf{v}$ from its coordinate vector $[\mathbf{v}]_B$:

$$\mathbf{v} = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = B\,[\mathbf{v}]_B. \quad (\#2)$$

Here (cf. Notation 8.3):

$$B = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \in M_{1\times n}(V).$$

Exercise 8.5. (**Isomorphism** $V \to F_c^n$) Let V be an n-dimensional vector space over a field F and with a basis $B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Show that the map

$$\begin{aligned} \varphi : V &\to F_c^n \\ \mathbf{v} &\mapsto [\mathbf{v}]_B \end{aligned}$$

is an isomorphism between the vector spaces V and F_c^n. Thus, for a given $n \geq 1$, there is, up to isomorphism, only one vector space of dimension n.

Theorem 8.6. *(**Representation matrix** $[T]_{B,B_W}$)*

Let

$$T : V \to W$$

be a linear transformation between vector spaces over a field F. Let

$$B = (\mathbf{v}_1, \dots, \mathbf{v}_n)$$

be a basis of V and

$$B_W = (\mathbf{w}_1, \dots, \mathbf{w}_m)$$

a basis of W. Let $A \in M_{m\times n}(F)$. Then the following three conditions on A are equivalent.

$$[T(\mathbf{v})]_{B_W} = A\,[\mathbf{v}]_B \quad (\forall\ \mathbf{v}\ \in\ V), \tag{\#3}$$

$$A = ([T(\mathbf{v}_1)]_{B_W}, .., [T(\mathbf{v}_n)]_{B_W}), \tag{\#4}$$

$$(T(\mathbf{v}_1), .., T(\mathbf{v}_n)) = (\mathbf{w}_1, .., \mathbf{w}_m)\,A. \tag{\#5}$$

We denote the above matrix A as

$$[T]_{B,B_W} := A = ([T(\mathbf{v}_1)]_{B_W}, .., [T(\mathbf{v}_n)]_{B_W})$$

and call it the **representation matrix** *or simply the* **matrix of T relative to (the bases)**

B and B_W. *Each of* $(\#3)-(\#5)$ *above is called a* **characteristic property of the representation matrix of T**.

Once basis B of V and basis B_W of W are fixed, the linear transformation T and its representation matrix $[T]_{B,B_W}$ determine each other uniquely.

In particular, T is the zero map from V to W if and only if $[T]_{B,B_W}$ is the zero matrix in $M_{m\times n}(F)$.

Proof. For '$(\#4) \Rightarrow (\#3)$', assume $(\#4)$ (and we will prove $(\#3)$). By $(\#2)$ in Definition 8.4, we have

$$\mathbf{v} = B\,[\mathbf{v}]_B,$$

$$T(\mathbf{v}) = B_W\,[T(\mathbf{v})]_{B_W}.$$

Thus we have (cf. Exercise 8.2):

$$B_W\,[T(\mathbf{v})]_{B_W} = T(\mathbf{v}) = T(B\,[\mathbf{v}]_B)$$

$$= T(B)\,[\mathbf{v}]_B = (T(\mathbf{v}_1),\,..,\,T(\mathbf{v}_n))\,[\mathbf{v}]_B$$

$$= (B_W[T(\mathbf{v}_1)]_{B_W},..,B_W[T(\mathbf{v}_n)]_{B_W})\,[\mathbf{v}]_B$$

$$= B_W\,([T(\mathbf{v}_1)]_{B_W},\,..,\,[T(\mathbf{v}_n)]_{B_W})\,[\mathbf{v}]_B.$$

Hence

$$B_W\,[T(\mathbf{v})]_{B_W} = B_W\,([T(\mathbf{v}_1)]_{B_W},..,[T(\mathbf{v}_n)]_{B_W})\,[\mathbf{v}]_B.$$

Comparing coefficients of the basis B_W, we get (cf. (#4)):

$$[T(\mathbf{v})]_{B_W} = ([T(\mathbf{v}_1)]_{B_W},\,\dots,\,[T(\mathbf{v}_n)]_{B_W})\,[\mathbf{v}]_B = A\,[\mathbf{v}]_B.$$

So (#3) is true.

For '(#3) $\Rightarrow$ (#4)', applying (#3) to $\mathbf{v} = \mathbf{v}_j$, we see that the j-th column of A is just $[T(\mathbf{v}_j)]_{B_W}$ and hence

$$A = ([T(\mathbf{v}_1)]_{B_W}, \dots, T(\mathbf{v}_n)]_{B_W}).$$

(#4) is proved.

For '(#4) $\Rightarrow$ (#5)', assume (#4) (and we will prove (#5)). We calculate (cf. (#4)):

$$\begin{aligned} & (T(\mathbf{v}_1), .., T(\mathbf{v}_n)) \\ = & \ (B_W\, [T(\mathbf{v}_1)]_{B_W}, \ \ldots, \ B_W\, [T(\mathbf{v}_n)]_{B_W}) \\ = & \ B_W([T(\mathbf{v}_1)]_{B_W}, \ \ldots, \ [T(\mathbf{v}_n)]_{B_W}) \\ = & \ (\mathbf{w}_1, \ \ldots, \ \mathbf{w}_m)\, A \end{aligned}$$

so (#5) is true.

Reversing the above process and noting that B_W is a basis (of W), we get '(#5) $\Rightarrow$ (#4)' (Exercise!).

The last assertion is true by looking at (#5) (cf. Exercise 7.24). $\square$

Note 1. Compare the location of the representation matrix $[T]_{B,B_W} \in M_{m\times n}(F)$ in (#3) and (#5) of Theorem 8.6:

$$[T(\mathbf{v})]_{B_W} = [T]_{B,B_W}[\mathbf{v}]_B \qquad (\#3)$$

$$(T(\mathbf{v}_1), \ldots, T\mathbf{v}_n) = (\mathbf{w}_1, \ldots, \mathbf{w}_m)\,[T]_{B,B_W} \quad (\#5).$$

This representation matrix appears on the left of the coordinate vector $[\mathbf{v}]_B \in M_{n\times 1}(F)$ of $\mathbf{v} \in V$, and on the right of the basis (of W)

$$B_W = (\mathbf{w}_1, \ldots, \mathbf{w}_m) \in M_{1\times m}(V).$$

Definition 8.7. (**Representation for linear operator**) When $V = W$ and $B = B_W$, and $T : V \to V$ is a linear operator, i.e., a linear transformation from V to itself, the representation matrix $[T]_{B,B}$ is simply denoted as $[T]_B$:

$$[T]_B := [T]_{B,B}.$$

By Theorem 8.6, the following are equivalent definitions of $[T]_B$:

$$[T(\mathbf{v})]_B = [T]_B\,[\mathbf{v}]_B \ \ (\forall\ \mathbf{v} \in V), \qquad (\#3')$$

$$[T]_B = ([T(\mathbf{v}_1)]_B, .., [T(\mathbf{v}_n)]_B), \qquad (\#4')$$

$$T(\mathbf{v}_1, .., \mathbf{v}_n) = (\mathbf{v}_1, \ldots, \mathbf{v}_n)\,[T]_B. \qquad (\#5')$$

Extra exercise 8.8. (**Linear transformation theory = Matrix theory**) Let V and W be vector spaces over the same field F and of dimensions n and m, respectively. Let B (resp. B_W) be a basis of V (resp. W).

We have seen in Theorem 8.6 that for every linear transformation

$$T : V \to W$$

there is a representation matrix

$$[T]_{B,B_W} \in M_{m\times n}(F)$$

satisfying the equivalent characteristic properties (#3) - (#5) there.

(1) Conversely, show that for every matrix

$$A \in M_{m\times n}(F)$$

there is a unique linear transformation

$$T : V \to W$$

such that the representation matrix

$$[T]_{B,B_W} = A.$$

(2) Consequently, show that the map

$$\begin{aligned} \varphi : \mathrm{Hom}_F(V, W) &\to M_{m\times n}(F) \\ T &\mapsto [T]_{B,B_W} \end{aligned}$$

is an isomorphism of vector spaces over F (cf. Exercise 7.32).

Example 8.9. (**Scalar maps $\longleftrightarrow$ Scalar matrices**) Suppose that

$$T = \alpha\, I_V : V \to V$$

is a scalar map on an n-dimensional space V over a field F (cf. Example 7.15). Then its representation matrix (relative to any basis B of V) is a scalar matrix:

$$[\alpha\, I_V]_B = \alpha\, I_n \in M_n(F).$$

In particular, T is the zero map on V if and only if $[T]_B = 0I_n \in M_n(F)$ (the zero matrix):

$$T = 0\, I_V \iff [T]_B = 0\, I_n \in M_n(F).$$

T is the identity map I_V on V if and only if $[T]_B = I_n \in M_n(F)$ (the identity matrix):

$$T = I_V \iff [T]_B = I_n \in M_n(F).$$

Example 8.10. (**Calculation of a representation matrix**) Consider the linear transformation

$$T : V = F_c^2 \to F_c^3 = W$$
$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} y \\ -5x + 13y \\ -7x + 16y \end{pmatrix}.$$

between the column vector 2-space and 3-space.

One can verify that

$$B := \{\mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \ \mathbf{v}_2 = \begin{pmatrix} 5 \\ 2 \end{pmatrix}\}$$

is a basis of $V = F_c^2$, and

$$B' := \{\mathbf{w}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{w}_2 = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}, \mathbf{w}_3 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}\}$$

is a basis of $W = F_c^3$.

We will find the representation matrix $[T]_{B,B'}$. We first express the images $T(\mathbf{v}_i) \in W$ as linear combinations relative to the basis B' of W (see §6):

$$T(\mathbf{v}_1) = \begin{pmatrix} 1 \\ -2 \\ -5 \end{pmatrix} = \mathbf{w}_1 - 2\mathbf{w}_3,$$

$$T(\mathbf{v}_2) = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} = 3\mathbf{w}_1 + \mathbf{w}_2 - \mathbf{w}_3.$$

Thus we get their coordinate vectors relative to the basis B' of W (cf. (#2) in Definition 8.4):

$$[T(\mathbf{v}_1)]_{B'} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \ [T(\mathbf{v}_2)]_{B'} = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix},$$

and assemble them into the representation matrix (cf. (#4) in Theorem 8.6):

$$[T]_{B,B'} = ([T(\mathbf{v}_1)]_{B'},\ [T(\mathbf{v}_2)]_{B'}) = \begin{pmatrix} 1 & 3 \\ 0 & 1 \\ -2 & -1 \end{pmatrix}$$

relative to the bases B of $V = F_c^2$ and B' of $W = F_c^3$.

To test the correctness of our calculation, let

$$\mathbf{v} := \begin{pmatrix} 2 \\ 1 \end{pmatrix} \in V = F_c^2$$

Express this $\mathbf{v} \in V$ as a linear combination relative to the basis B (cf. §6)

$$\mathbf{v} = -\mathbf{v}_1 + \mathbf{v}_2.$$

Hence we get the coordinate vector of $\mathbf{v}$ relative to the basis B (cf. (#2) in Definition 8.4):

$$[\mathbf{v}]_B = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

On the one hand, by (#3) in Theorem 8.6,

$$
\begin{aligned}
[T(\mathbf{v})]_{B'} &= [T]_{B,B'}\,[\mathbf{v}]_B \\
&= \begin{pmatrix} 1 & 3 \\ 0 & 1 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.
\end{aligned}
$$

This way, we can recover $T(\mathbf{v})$ from its coordinate vector (cf. (#2) in Definition 8.4):

$$
\begin{aligned}
T(\mathbf{v}) &= B'[T(\mathbf{v})]_{B'} \\
&= (\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}. \\
&= 2\mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3.
\end{aligned}
$$

On the other hand, we can calculate directly,

$$\begin{aligned} T(\mathbf{v}) &= T(-\mathbf{v}_1 + \mathbf{v}_2) \\ &= -T(\mathbf{v}_1) + T(\mathbf{v}_2) \\ &= -(\mathbf{w}_1 - 2\mathbf{w}_3) + (3\mathbf{w}_1 + \mathbf{w}_2 - \mathbf{w}_3) \\ &= 2\mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 \end{aligned}$$

which agrees with the calculation in the above paragraph.

Example 8.11. (**Linear operator on polynomial ring**) Consider the linear operator

$$T : V = P_3[x] \;\to\; P_3[x],$$
$$f(x) \mapsto f(3x-5).$$

on the subspace $P_3[x]$ (of degree < 3 polynomials) of $F[x]$. For a general element

$$\mathbf{v} = f(x) = c_0 + c_1x + c_2x^2 \;\in\; P_3[x]$$

we have

$$\begin{aligned} T(\mathbf{v}) &= T(f(x)) = f(3x-5) \\ &= c_0 + c_1(3x-5) + c_2(3x-5)^2. \end{aligned} \qquad (*)$$

Take the standard basis

$$B = \{1, x, x^2\}$$

of $V = P_3[x]$. We will find the representative matrix

$$[T]_B = [T]_{B,B}$$

of T relative to this basis B (cf. Definition 8.7).

First, substituting

$$f(x) = 1,\ x,\ x^2$$

one by one, into the formula (*) above, we get

$$\begin{aligned} T(1) &= 1, \\ T(x) &= 3x - 5, \\ T(x^2) &= (3x-5)^2 = 25 - 30x + 9x^2. \end{aligned}$$

By inspection, we get their coordinate vectors relative to the same standard basis B (cf. (#2) in

Definition 8.4):

$$[T(1)]_B = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ [T(x)]_B = \begin{pmatrix} -5 \\ 3 \\ 0 \end{pmatrix},$$

$$[T(x^2)]_B = \begin{pmatrix} 25 \\ -30 \\ 9 \end{pmatrix}.$$

Now assemble these column vectors into the representation matrix (cf. (#4) in Theorem 8.6):

$$\begin{aligned} [T]_B &= ([T(1)]_B, \ [T(x)]_B, \ [T(x^2)]_B) \\ &= \begin{pmatrix} 1 & -5 & 25 \\ 0 & 3 & -30 \\ 0 & 0 & 9 \end{pmatrix} \end{aligned}$$

relative to the basis B.

Again, to test the correctness of our calculation, let

$$\mathbf{v} = f(x) := 1 + 2x + 3x^2 \in P_3[x].$$

By inspection, we get its coordinate vector relative to the standard basis B (cf. (#2) in Definition 8.4):

$$[\mathbf{v}]_B = [f(x)]_B = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

On the one hand, by (#3) in Theorem 8.6,

$$\begin{aligned} [T(\mathbf{v})]_B &= [T]_B \, [\mathbf{v}]_B \\ &= \begin{pmatrix} 1 & -5 & 25 \\ 0 & 3 & -30 \\ 0 & 0 & 9 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 66 \\ -84 \\ 27 \end{pmatrix}. \end{aligned}$$

This way, we can recover $T(\mathbf{v})$ from its coordinate vector (cf. (#2) in Definition 8.4):

$$\begin{aligned} T(\mathbf{v}) &= B\,[T(\mathbf{v})]_B \\ &= (1, x, x^2) \begin{pmatrix} 66 \\ -84 \\ 27 \end{pmatrix} \\ &= 66 - 84x + 27x^2. \end{aligned}$$

On the other hand, we can calculate directly (using the formula (*) above)

$$\begin{aligned} T(1 + 2x + 3x^2) &= 1 + 2(3x - 5) + 3(3x - 5)^2 \\ &= 66 - 84x + 27x^2 \end{aligned}$$

which agrees with the calculation in the above paragraph.

Example 8.12. (**A and the representation matrix of T_A**) Take a matrix $A = (a_{ij}) \in M_{m\times n}(F)$

and consider the linear transformation

$$T_A : F_c^n \to F_c^m$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = X \mapsto AX = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

associated with A.

For $r > 0$, let

$$B_r := \{\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \ \mathbf{e}_r = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}\}$$

be the standard basis of the column r-space F_c^r (cf. Example 5.12).

Verify that

$$[\mathbf{v}]_{B_r} = \mathbf{v} \ \ (\forall\, \mathbf{v} \in F_c^r),$$

and for $A \in M_{m\times n}(F)$ the representation matrix $[T_A]_{B_n,B_m}$ of T_A relative to bases B_n and B_m, equals A:

$$[T_A]_{B_n,B_m} = A.$$

Extra exercise 8.13. (**Close relation between the space of vectors and the space of their coordinates**). Let

$$T : V \to W$$

be a linear transformation between the vector spaces V and W over the same field F, of dimensions n and m, respectively. Let

$$B_V := (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

be a basis of V, and

$$B_W = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$$

a basis of W. Let

$$M_{m\times n}(F) \ni A := [T]_{B_V,B_W} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$$

(with $\mathbf{a}_i$ the i-th column of A) be the representation matrix of T relative to B_V and B_W. Then the following are isomorphisms (if you show the compositions $\varphi \circ \psi = \mathrm{id}$, $\psi \circ \varphi = \mathrm{id}$, $\xi \circ \eta = \mathrm{id}$ and $\eta \circ \xi = \mathrm{id}$, then φ and ψ are bijections and inverse to each other; and ξ and η are bijections and inverse to each other; of course you also have to show that the 4 maps are well defined and (at least φ and ξ) are linear transformations):

$$
\begin{aligned}
\varphi : \operatorname{Ker}(T) &\to \operatorname{Null}(A)\\
\mathbf{v} &\mapsto [\mathbf{v}]_{B_V},\\
\psi : \operatorname{Null}(A) &\to \operatorname{Ker}(T)\\
X &\mapsto B_V X = (\mathbf{v}_1, \ldots, \mathbf{v}_n)X,\\
\xi : R(T) &\to R(T_A) = \text{col.sp.of}\, A = \operatorname{Span}\{\mathbf{a}_1, .., \mathbf{a}_n\}\\
\mathbf{w} &\mapsto [\mathbf{w}]_{B_W},\\
\eta : R(T_A) &\to R(T)\\
Y &\mapsto B_W Y = (\mathbf{w}_1, \ldots, \mathbf{w}_m)Y.
\end{aligned}
$$

Below are some consequences of the isomorphisms above (cf. Exercise 7.42).

(1) The subset

$$\{X_1, \ldots, X_s\}$$

of F_c^n is a basis of $\operatorname{Null}(A)$ if and only if the vectors

$$B_V X_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_n)X_1, \quad B_V X_2, \quad \ldots, \quad B_V X_s$$

of V form a basis of $\mathrm{Ker}(T)$.

(2) The subset

$$\{Y_1, \ldots, Y_t\}$$

of F_c^m is a basis of $R(T_A)$ if and only if the vectors

$$B_W Y_1 = (\mathbf{w}_1, \ldots, \mathbf{w}_m) Y_1, \quad B_W Y_2, \quad \ldots, \quad B_W Y_t$$

of W form a basis of $R(T)$.

(3) The range of T is given by

$$R(T) = \mathrm{Span}\{B_W \mathbf{a}_1, \ldots, B_W \mathbf{a}_n\}.$$

(4) $T : V \to W$ is an isomorphism if and only if its representation matrix $A = [T]_{B_V, B_W}$ is an invertible (square) matrix in $M_n(F)$.

Theorem 8.14. *(**Representation matrix of a composite map**) Let V_1, V_2, V_3 be vector spaces of finite dimension over the same field F and let*

$$B_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\},$$

$$B_2 = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\},$$

$$B_3 = \{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$$

be their respective bases. Let

$$T_1 : V_1 \to V_2$$

and

$$T_2 : V_2 \to V_3$$

be linear transformations. Then we have:

$$[T_2 \circ T_1]_{B_1,B_3} = [T_2]_{B_2,B_3} [T_1]_{B_1,B_2}.$$

Namely, the representation matrix of the composite map

$$T_2 \circ T_1 : V_1 \to V_3$$

relative to the bases B_1 and B_3 equals the representation matrix $[T_2]_{B_2,B_3}$ times the representation matrix $[T_1]_{B_1,B_2}$.

Proof. We use the characteristic property (#5) in Theorem 8.6 of the representation matrices of T_1,

T_2, and $T_2 \circ T_1$:

$$T_1(\mathbf{u}_1, \ldots, \mathbf{u}_m) = (\mathbf{v}_1, \ldots, \mathbf{v}_n)\, [T_1]_{B_1,B_2},$$

$$T_2(\mathbf{v}_1, \ldots, \mathbf{v}_n) = (\mathbf{w}_1, \ldots, \mathbf{w}_r)\, [T_2]_{B_2,B_3}, \qquad (*)$$

$$(T_2 \circ T_1)(\mathbf{u}_1, .., \mathbf{u}_m) = (\mathbf{w}_1, .., \mathbf{w}_r)\, [T_2 \circ T_1]_{B_1,B_3}.$$

On the other hand, substituting the first two equalities above, we get (with Exercise 8.2 used for the third equality below):

$$\begin{aligned}
&(T_2 \circ T_1)(\mathbf{u}_1, \ldots, \mathbf{u}_m)\\
=&T_2((T_1(\mathbf{u}_1), \ldots, T_1(\mathbf{u}_1))\\
=&T_2((\mathbf{v}_1, \ldots, \mathbf{v}_n)\, [T_1]_{B_1,B_2})\\
=&(T_2(\mathbf{v}_1), \ldots, T_2(\mathbf{v}_n))[T_1]_{B_1,B_2}\\
=&(\mathbf{w}_1, \ldots, \mathbf{w}_r)\, [T_2]_{B_2,B_3}[T_1]_{B_1,B_2}.
\end{aligned}$$

Thus we have

$$(T_2 \circ T_1)(\mathbf{u}_1, .., \mathbf{u}_m) = (\mathbf{w}_1, .., \mathbf{w}_r)\, [T_2]_{B_2,B_3}[T_1]_{B_1,B_2}.$$

Comparing this with the last equation in (*) above, i.e., comparing the coefficient matrix on the right of $(\mathbf{w}_1, \ldots, \mathbf{w}_r)$, we get the equality in the theorem (Exercise!). $\square$

Extra exercise 8.15. (**Representation matrix of inverse of an isomorphism**) Let

$$T : V \to W$$

be an isomorphism between vector spaces over the same field F and of the (necessarily same, cf. Exercise 7.42) finite dimension. Let

$$T^{-1} : W \to V$$

be the inverse isomorphism of T. Let B_V (resp. B_W) be a basis of V (resp. W). Then

$$[T^{-1}]_{B_W,B_V} = [T_{B_V,B_W}]^{-1}.$$

Namely, the representation matrix of the inverse of T equals the inverse of the representation matrix of T.

Hint. Apply T^{-1} to the characteristic property (#5) in Theorem 8.6. You may also use Exercises 8.2 and 8.13.

Extra exercise 8.16. (**Representation matrix of map combination**) Let

$$T_i : V \to W$$

be two linear transformations between finite-dimensional vector spaces over the same field F. Let B (resp. B_W) be a basis of V (resp. W). Then for any $a_i \in F$, the map linear combination $a_1T_1 + a_2T_2$ (cf. Exercise 7.32) has the representation matrix

$$[a_1T_1 + a_2T_2]_{B,B_W} = a_1[T_1]_{B,B_W} + a_2[T_2]_{B,B_W}.$$

Hint. Use characteristic property (#5) in Theorem 8.6.

Theorem 8.17. *(**Equivalent transition matrix definition**) Let V a vector space over a field*

F and of finite dimension $n \geq 1$. Let

$$B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

and

$$B' := (\mathbf{v}'_1, \ldots, \mathbf{v}'_n)$$

be two bases of V. Let $P \in M_n(F)$. Then the following are equivalent.

(1)

$$P = ([\mathbf{v}'_1]_B, \ \ldots, \ [\mathbf{v}'_n]_B).$$

(2)

$$B' = BP.$$

(3) For any $\mathbf{v} \in V$, we have

$$P\,[\mathbf{v}]_{B'} = [\mathbf{v}]_B.$$

Proof. For '(1) $\Longleftrightarrow$ (2)', assume (1) (and we shall show (2)). Then (by (#2) in Definition 8.4)

$$
\begin{aligned}
BP &= B\,([\mathbf{v}'_1]_B, \ldots, [\mathbf{v}'_n]_B) \\
&= (B\,[\mathbf{v}'_1]_B, \ldots, B\,[\mathbf{v}'_n]_B) \\
&= (\mathbf{v}'_1, \ldots, \mathbf{v}'_n) = B'.
\end{aligned}
$$

So (2) is true.

Conversely, assume (2) (and we shall show (1)). Then (with (#2) in Definition 8.4 used again):

$$
\begin{aligned}
BP = B' &= (\mathbf{v}'_1, \ldots, \mathbf{v}'_n) \\
&= (B\,[\mathbf{v}'_1]_B, \ldots, B\,[\mathbf{v}'_n]_B) \\
&= B\,([\mathbf{v}'_1]_B, \ldots, [\mathbf{v}'_n]_B).
\end{aligned}
$$

Hence

$$BP = B\,([\mathbf{v}'_1]_B, \ldots, [\mathbf{v}'_n]_B).$$

Since B is a basis, this implies (1) (by sort of canceling B) (Exercise!).

For '(3) $\Longleftrightarrow$ (2)', assume (3) (and we shall prove (2)). Then (by (#2) in Definition 8.4), we have

$$B'\,[\mathbf{v}]_{B'} = \mathbf{v} = B\,[\mathbf{v}]_B = BP\,[\mathbf{v}]_{B'}$$

and hence

$$B' [\mathbf{v}]_{B'} = BP [\mathbf{v}]_{B'} \ \ (\forall\, \mathbf{v} \in V).$$

Since the above is true for arbitrary $\mathbf{v} \in V$, we get (2) (cf. the proof of Theorem 8.20).

Conversely, assume (2) (and we shall prove (3)). Then we have:

$$B' = BP$$

and hence (by (#2) in Definition 8.4)

$$B\,[\mathbf{v}]_B = \mathbf{v} = B'\,[\mathbf{v}]_{B'} = B\,P\,[\mathbf{v}]_{B'}.$$

Thus

$$B\,[\mathbf{v}]_B = B\,P\,[\mathbf{v}]_{B'}.$$

Hence (3) is true as remarked at the end of the proof of '(2) $\Rightarrow$ (1)'. $\square$

Definition 8.18. (**Transition matrix**) The matrix in Theorem 8.17:

$$P_{B' \to B} := P = ([\mathbf{v}_1']_B, .., [\mathbf{v}_n']_B) \ \in \ M_n(F)$$

is (denoted as above and) called the **transition matrix from (new) basis B' to (old) basis B**. It is invertible (Exercise!).

Remark 8.19. In the notation of Theorem 8.17, if we let $T : V \to V$ be the linear transformation such that

$$T(B) = T(\mathbf{v}_1, \dots, \mathbf{v}_n) = (\mathbf{v}'_1, \dots, \mathbf{v}'_n) = B'$$

i.e., $T(\mathbf{v}_i) = \mathbf{v}'_i$ (use Theorem 8.6 to check that such T exists and unique), then the representation matrix $[T]_B$ equals the transition matrix $P_{B' \to B}$ and hence the latter is invertible since T is obviously an isomorphism; see Exercise 8.13.

Theorem 8.20. *(**Basis change theorem for representation matrix**) Let V be a vector space over a field F and of finite dimension $n \geq 1$. Let*

$$B = (\mathbf{v}_1, \dots, \mathbf{v}_n)$$

and

$$B' := (\mathbf{v}'_1, \ldots, \mathbf{v}'_n)$$

be two bases of V. *Then*

$$[T]_{B'} = P^{-1}\,[T]_B\,P$$

where

$$P := P_{B' \to B}$$

is the transition matrix from (new basis) B' *to (old basis)* B *(cf. Definition* 8.18*).*

Proof. Let $\mathbf{v} \in V$ be an arbitrary vector. Then by Theorem 8.17,

$$P\,[\mathbf{v}]_{B'} = [\mathbf{v}]_B$$

and (applying Theorem 8.17 to $T(\mathbf{v})$)

$$P\,[T(\mathbf{v})]_{B'} = [T(\mathbf{v})]_B$$

or equivalently,

$$[T(\mathbf{v})]_{B'} = P^{-1}\,[T(\mathbf{v})]_B.$$

These, together with (#2') and (#3') in Definitions 8.7, imply:

$$\begin{aligned}[T]_{B'}\,[\mathbf{v}]_{B'} &= [T(\mathbf{v})]_{B'} \\ &= P^{-1}\,[T(\mathbf{v})]_B \\ &= P^{-1}\,[T]_B\,[\mathbf{v}]_B \\ &= P^{-1}\,[T]_B\,P\,[\mathbf{v}]_{B'}.\end{aligned}$$

So

$$[T]_{B'}[\mathbf{v}]_{B'} = (P^{-1}\,[T]_B P)\,[\mathbf{v}]_{B'} \quad (*)$$

This implies the equality in the theorem, since $[\mathbf{v}]_{B'}$ is arbitrary and can attain any column vector value by varying $\mathbf{v}$.

An alternative and more rigorous proof is as follows. (*) above implies:

$$0 = ([T]_{B'} - P^{-1}\,[T]_B\,P)\,[\mathbf{v}]_{B'} \quad (\forall\,\mathbf{v}\,\in\,V).$$

Now substituting

$$\mathbf{v} = \mathbf{v}_1',\,\ldots,\,\mathbf{v}_n'$$

one by one, into the above to get n equalities and assembling these into:

$$\begin{aligned} 0 &= \left([T]_{B'} - P^{-1}\,[T]_B\,P\right)\left([\mathbf{v}'_1]_{B'}, .., [\mathbf{v}'_1]_{B'}\right) \\ &= [T]_{B'} - P^{-1}\,[T]_B\,P. \end{aligned}$$

Here we used Exercise 8.21 in the last equality. $\square$

Exercise 8.21. (**Coordinate vectors of the basis vectors**) Let

$$B' = \{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$$

be a basis of a vector space V. Show that

$$([\mathbf{v}'_1]_{B'}, \ldots, [\mathbf{v}'_1]_{B'}) = I_n.$$

Example 8.22. (**Basis change**) Consider the linear transformation

$$\begin{aligned} T : V = F_c^2 &\to F_c^2 \\ \begin{pmatrix} x \\ y \end{pmatrix} &\mapsto \begin{pmatrix} x+y \\ -2x+4y \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} \end{aligned}$$

where

$$A = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}.$$

Namely, $T = T_A$ (the matrix linear transformation associated with A).

Let

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

be the standard basis of $V = F_c^2$, and let

$$B' = \{\mathbf{v}_1' = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{v}_2' = \begin{pmatrix} 1 \\ 2 \end{pmatrix}\}$$

be another basis of $V = F_c^2$. By Example 8.12, the representation matrix

$$[T]_B = A.$$

By inspection,

$$\mathbf{v}_1' = \mathbf{e}_1 + \mathbf{e}_2, \quad \mathbf{v}_2' = \mathbf{e}_1 + 2\mathbf{e}_2.$$

Hence

$$[\mathbf{v}_1']_B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad [\mathbf{v}_2']_B = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Thus the transition matrix from B' to B is

$$\begin{aligned} P = P_{B' \to B} &= ([\mathbf{v}_1']_B, [\mathbf{v}_2']_B) \\ &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \end{aligned}$$

By Theorem 8.20, the representation matrix $[T]_{B'}$ of T relative to B' is as follows:

$$[T]_{B'} = P^{-1} [T]_B P =$$

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

Alternatively, one calculates directly:

$$T(\mathbf{v}_1') = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\mathbf{v}_1', \quad T(\mathbf{v}_2') = \begin{pmatrix} 3 \\ 6 \end{pmatrix} = 3\mathbf{v}_2',$$

$$[T(\mathbf{v}_1')]_{B'} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad [T(\mathbf{v}_2')]_{B'} = \begin{pmatrix} 0 \\ 3 \end{pmatrix},$$

$$[T]_{B'} = ([T(\mathbf{v}_1')]_{B'}, [T(\mathbf{v}_2')]_{B'}) = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

which agrees with the calculation in the previous paragraph.

Definition 8.23. (**Similar matrices**) Two square matrices (of the same order) $A_1, A_2 \in M_n(F)$ are **similar** if there is an invertible matrix $P \in M_n(F)$ such that

$$A_2 = P^{-1}A_1P.$$

In this case, we denote

$$A_1 \sim A_2.$$

Exercise 8.24. (**Similarity relation**) Show that the similarity property in Definition 8.23 is an equivalence relation, i.e. to show:

(1) **Reflexivity:**

$$A \sim A,$$

(2) **Symmetry:**

$$A_1 \sim A_2 \;\Rightarrow\; A_2 \sim A_1, \quad \text{and}$$

(3) **Transitivity:**

$$A_1 \sim A_2,\; A_2 \sim A_3 \;\Rightarrow\; A_1 \sim A_3.$$

Exercise 8.25. (**Determinant of similar matricies**) Similar matrices have the same determinant:

$$A_1 \sim A_2 \;\Rightarrow\; |A_1| = |A_2|.$$

Definition 8.26. (**Determinant/Trace of a linear operator**) Let

$$T : V \to V$$

be a linear operator on a finite-dimensional vector space V.

We define the **determinant** $\det(T)$ (or simply $|T|$) of T as

$$\det(T) := \det([T]_B)$$

and the **trace** of T as

$$\mathrm{Tr}(T) = \mathrm{Tr}([T]_B)$$

where B is any basis of V.

Theorem 8.20 and Exercises 8.28 and 8.29 show that $\det(T)$ and $\mathrm{Tr}(T)$ do not depend on the choice of the basis B of V, and hence are well defined (i.e., uniquely determined by T alone).

Definition 8.27. (**Characteristic polynomial $p_A(x)$, $p_T(x)$**)

(1) Let $A \in M_n(F)$.

$$\begin{aligned} p_A(x) :&= |xI_n - A| \\ &= x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0 \end{aligned}$$

is called the **characteristic polynomial** of A, which is of degree n (= the order of A).

(2) Let

$$T : V \to V$$

be a linear operator on an n-dimensional vector space V. Set

$$A := [T]_B$$

where B is any basis of V. Then

$$\begin{aligned} p_T(x) : &= p_A(x) = |xI_n - A| \\ &= x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0 \end{aligned}$$

is called the **characteristic polynomial** of T, which is of degree n $(= \dim V)$.

Exercise 8.28. (**Similar matrices have equal characteristic polynomial**)

$$A_1 \sim A_2 \Rightarrow p_{A_1}(x) = p_{A_2}(x).$$

This and Theorem 8.20 imply that $p_T(x)$ in Definition 8.27 does not depend on the choice of the basis of B and hence is well defined (i.e., is uniquely determined by T alone).

Exercise 8.29. (**Characteristic polynomial v.s. determinant and trace**) Show that for $A = (a_{ij}) \in M_n(F)$, we have

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + \cdots + a_{nn} = -b_{n-1}$$

$$\det(A) = (-1)^n b_0 = (-1)^n p_A(0).$$

where b_i are coefficients of $p_A(x)$ as in Definition 8.27.

9 Eigenvalue and Cayley-Hamilton theorem

In this section, we study linear transformations T, and their eigenvalues, eigenvectors, eigenspaces, characteristic polynomials, T-invariant subspaces and T-cyclic subspaces. We will conclude with the proof of Cayley-Hamilton theorem.

Definition 9.1. (**Eigenvalue, Eigenvector**) Assume that

$$\lambda \in F.$$

(1) Let V be a vector space over a field F. Let

$$T : V \to V$$

be a linear operator. A *nonzero* vector $\mathbf{v}$ in V is called an **eigenvector of T corresponding to the eigenvalue $\lambda \in F$ of T** if:

$$T(\mathbf{v}) = \lambda \mathbf{v}.$$

(2) For an $n \times n$ matrix A in $M_n(F)$, a nonzero column vector $\mathbf{u}$ in F_c^n is called an **eigenvector of A corresponding to the eigenvalue $\lambda \in F$ of A** if:

$$A\mathbf{u} = \lambda\mathbf{u}.$$

Exercise 9.2. (**Complex zero of $p_A(x)$**)

(1) Find a real matrix $A \in M_3(\mathbf{R})$ such that its characteristic polynomial $p_A(x)$ (cf. Definition 8.27) has one real zero λ_1, and two complex zeros

$$\lambda_2 = a + b\sqrt{-1}, \quad \lambda_3 = a - \sqrt{-1}b = \overline{\lambda}_2$$

where $b \neq 0$. The above λ_2 and λ_3 are called **complex eigenvalues of A** and they are conjugate to each other.

(2) In general, for an $A \in M_n(F)$ or a linear operator T, the characteristic polynomial $p_A(x)$ or $p_T(x)$ (cf. Definition 8.27) may not have all (or

any) zeros in F.

(3) Of course, if the field F is so called **algebraically closed**, eg. $F = \mathbb{C}$, then by the **fundamental theorem of algebra**, every zero of $p_A(x)$ is in $F = \mathbb{C}$.

Theorem 9.3. *(**Equivalent definition of eigenvalue and eigenvector**) Let V be a vector space of dimension n over a field F and with a basis B. Let*

$$T : V \to V$$

be a linear operator. Assume that

$$\lambda \in F.$$

Then the following are equivalent (for a matrix version of the theorem, let $V = F_c^n$, $T = T_A$ with A a matrix in $M_n(F)$, and B the standard basis so that $[T_A]_B = A$ and $[\mathbf{v}]_B = \mathbf{v}$; cf. Example

8.12*).*

(1) λ *is an eigenvalue of* T *(corresponding to an eigenvector* $\mathbf{0} \neq \mathbf{v} \in V$ *of* T*, i.e.,* $T(\mathbf{v}) = \lambda\mathbf{v}$*).*

(2) λ *is an eigenvalue of* $[T]_B$ *(corresponding to an eigenvector* $\mathbf{0} \neq [\mathbf{v}]_B \in F_c^n$ *of* $[T]_B$*, i.e.,* $[T]_B\,[\mathbf{v}]_B = \lambda[\mathbf{v}]_B$*).*

(3) The linear operator (cf. Example 7.15*, Exercise* 7.32*)*

$$\lambda I_V - T \,:\, V \to V,$$
$$\mathbf{x} \mapsto \lambda\mathbf{x} - T(\mathbf{x})$$

is not an isomorphism, i.e., there is some (cf. Theorem 7.47*):*

$$0 \neq \mathbf{v} \in \mathrm{Ker}(\lambda I_V - T).$$

(4) The matrix $\lambda I_n - [T]_B$ *is not invertible, i.e., the matrix equation*

$$(\lambda I_n - [T]_B)X = 0$$

has a non-trivial solution (equal to $[\mathbf{v}]_B$ *in* (2) *above).*

(5) λ *is a zero of the characteristic polynominal* $p_T(x)$ *of* T *(and also of* $[T]_B$*; cf. Definition* 8.27*), i.e.,*

$$p_T(\lambda) = |\lambda I_n - [T]_B| = 0.$$

Proof. We observe (cf. (#3') in Definition 8.7):

$$\mathbf{0} = (T - \lambda I_V)\,\mathbf{v}\,(= T(\mathbf{v}) - \lambda\mathbf{v})$$

$$\Longleftrightarrow$$

$$T(\mathbf{v}) = \lambda\mathbf{v}$$

$$\Longleftrightarrow$$

$$([T]_B\,[\mathbf{v}]_B =)\,[T(\mathbf{v})]_B = [\lambda\mathbf{v}]_B\,(= \lambda[\mathbf{v}]_B)$$

This implies the equivalence of (1) - (4). Theorem 6.3 implies the equivalence of (4) and (5). $\square$

Convention. (**Eigenvalue in an over field of F**) Most of the time, when we say that λ is an eigenvalue of a linear operator T on a vector space V defined over a field F (or a matrix $A \in M_n(F)$), we assume that

$$\lambda \in F.$$

However, we may occasionally loosely say that λ is an eigenvalue of T (or A) when λ is a zero in an over field $\widetilde{F}$ of F of the characteristic polynomial $p(x) = p_T(x)$ or $p(x) = p_A(x)$, i.e., when

$$p(\alpha) = 0, \quad \text{with} \quad \alpha \in \widetilde{F}.$$

For instance, λ_2 and λ_3 in Exercise 9.2 (1) are eigenvalues of A in the over field $\mathbb{C}$ of $F = \mathbb{R}$.

Exercise 9.4. (**Determinant $|A|$ as product of eigenvalues**) Let $A \in M_n(F)$ (or let T be a linear operator on an n-dimensional vector space V

over a field F). Let $p(x) = p_T(x)$ or $p(x) = p_A(x)$ be the characteristic polynomial. Factorize

$$p(x) = (x - \lambda_1) \dots (x - \lambda_n)$$

in some over field of F (i.e., some large enough field containing all zeros of $p(x)$, e.g. a so called **algebraic closure of F**). Then the determinant of T (or of A) equals

$$\prod_{i=1}^{n} \lambda_i = \lambda_1 \cdots \lambda_n.$$

In particular, if $F = \mathbb{C}$, then the determinant $|A|$ (or $|T|$) is the product of its n eigenvalues (not necessarily all distinct).

Indeed, by Exercise 8.29,

$$\begin{aligned} |A| &= (-1)^n p_A(0) \\ &= (-1)^n \prod_{i=1}^{n}(-\lambda_i) = \prod_{i=1}^{n} \lambda_i. \end{aligned}$$

Definition 9.5. (**Eigenspace of an eigenvalue**)
Let $\lambda \in F$ be an eigenvalue of a linear operator

$$T : V \to V$$

on an n-dimensional vector space V over the field F.

The subspace (of all eigenvectors corresponding to the eigenvalue λ, plus $\mathbf{0}_V$):

$$\begin{aligned} V_\lambda : &= V_\lambda(T) \\ : &= \mathrm{Ker}(\lambda I_V - T) \\ &= \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda \mathbf{v}\} \end{aligned}$$

of V is called the **eigenspace of T corresponding to the eigenvalue λ**. This definition is valid even when $\dim_F V = \infty$.

For a matrix $A \in M_n(F)$ with an eigenvalue λ, the subspace (of all eigenvectors corresponding to the eigenvalue λ, plus $\mathbf{0}_V$):

$$\begin{aligned} V_\lambda :&= V_\lambda(A) \\ :&= \mathrm{Ker}(\lambda I_n - A) \\ &= \mathrm{Null}(\lambda I_n - A) \\ &= \mathrm{Null}(A - \lambda I_n) \\ &= \{X \in F_c^n \,|\, (A - \lambda I_n)X = \mathbf{0}\}. \end{aligned}$$

of F_c^n is called the **eigenspace of A corresponding to the eigenvalue λ**.

One can check that

$$V_\lambda(A) = V_\lambda(T_A).$$

Definition 9.6. (**Geometric / Algebraic multiplicity**) Let $\lambda \in F$ and $T : V \to V$ (or $A \in M_n(F)$) be as in Definition 9.5.

(1) The dimension

$$\dim V_\lambda$$

of the eigenspace V_λ of T (or A) is called the

geometric multiplicity of the eigenvalue λ of T.. We have

$$n \geq \dim V_\lambda \geq 1.$$

(2) The **algebraic multiplicity** of the eigenvalue λ of T (or A) is defined to be the largest positive integer k such that $(x-\lambda)^k$ is a **factor** of the characteristic polynomial $p_T(x)$ (or $p_A(x)$), i.e., $p_T(x)$ (or $p_A(x)$) is divisible by $(x-\lambda)^k$ but not by $(x-\lambda)^{k+1}$, i.e.,

$$(x-\lambda)^k \mid p_T(x), \ (x-\lambda)^{k+1} \nmid p_T(x).$$

We shall see that (cf. Definition 10.11 and Remark 10.19):

$$(\text{geom.multiplicity of } \lambda) \leq (\text{alg.multiplicity of } \lambda).$$

Exercise 9.7. (= Tutorial question) (**Eigenspaces of T and $[T]_B$**) Let

$$T : V \to V$$

be a linear operator on an n-dimensional vector space V with a basis B. Set

$$A := [T]_B.$$

Use Theorem 9.3 or Exericse 8.13 to show that the map below

$$\begin{aligned} f : \operatorname{Ker}(T - \lambda I_V) &\to \operatorname{Null}(A - \lambda I_n) \\ \mathbf{w} &\mapsto [\mathbf{w}]_B \end{aligned}$$

gives an isomorphism (between eigenspaces, when λ is an eigenvalue). In particular (cf. Exercise 7.42):

$$\dim V_\lambda(T) = \dim V_\lambda(A).$$

Remark 9.8. (**Eigenspace of T v.s. that of $[T]_B$**) By Exercise 8.13, the following are equivalent.

(1) The subset

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$$

of V is a basis of the eigenspace $V_\lambda(T)$ of T.

(2) The subset

$$\{[\mathbf{u}_1]_B, \ldots, [\mathbf{u}_s]_B\}$$

of F_c^n is a basis of the eigenspace $V_\lambda([T]_B)$ of the representation matrix $[T]_B$ of T relative to a basis B of V.

By the same reasoning, the following are equivalent.

(1) The subset

$$\{X_1, \ldots, X_s\}$$

of F_c^n is a basis of the eigenspace $V_\lambda([T]_B)$ of the representation matrix $[T]_B$ of T relative to a basis $B = (\mathbf{u}_1, \ldots, \mathbf{u}_s)$ of V.

(2) The subset

$$\{B\,X_1, \ldots, B\,X_s\}$$

of V is a basis of the eigenspace $V_\lambda(T)$ of T.

Extra exercise 9.9. (**Eigenspaces of similar matrices**) Let $A \in M_n(F)$. Suppose that

$P^{-1}AP = C$. Show that

$$F_c^n \supseteq V_\lambda(A) = PV_\lambda(C) := \{P\,X \mid X \in V_\lambda(C)\}.$$

In particular, $\dim V_\lambda(A) = \dim V_\lambda(C)$.

Example 9.10. (**Eigenvalue of differential operator on $C^\infty[x]$**) Let

$$C^\infty[x]$$

be the set of all real-valued functions $f(x)$ having derivatives $f^{(n)}(x)$ of order n for all $n \geq 1$. Then the derivative operator

$$\begin{aligned} D: C^\infty[x] &\to C^\infty[x], \\ f(x) &\mapsto \frac{df(x)}{dx} \end{aligned}$$

is a linear operator.

Verify that every real number λ is an eigenvalue of D and the eigenspace is (cf. e.g. Howard Anton [Calculus, §7.7, First-order differential equations and applications]):

$$V_\lambda(D) = \{c\, e^{\lambda x} \,|\, c \in \mathbf{R}\}.$$

In particular,

$$V_0(D) = \{c \,|\, c \in \mathbf{R}\}$$

is the set of all constant functions.

Example 9.11. (**Eigenspace of a matrix in $M_3(F)$**) Consider the matrix

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix} \in M_3(F).$$

We calculate the characteristic polynomial of A:

$$\begin{aligned}
p_A(x) = |xI_3 - A| &= \begin{vmatrix} x & 0 & 2 \\ -1 & x-2 & -1 \\ -1 & 0 & x-3 \end{vmatrix} \\
&= \begin{vmatrix} 0 & 0 & x^2-3x+2 \\ 0 & x-2 & 2-x \\ -1 & 0 & x-3 \end{vmatrix} \\
&= (-1)^{3+1}(-1) \begin{vmatrix} 0 & x^2-3x+2 \\ x-2 & 2-x \end{vmatrix} \\
&= (x-1)(x-2)^2.
\end{aligned}$$

So there are two distinct eigenvalues

$$\lambda_1 = 1, \quad \lambda_2 = 2$$

of A. The algebraic multiplicity of λ_1 (resp. λ_2) is 1 (resp. 2).

To find the eigenspace $V_{\lambda_1}(A)$, we use Gaussian elimination and solve the matrix equation:

$$0 = (A - \lambda_1 I_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Its solution space is the eigenspace $V_{\lambda_1}(A)$. So

$$V_{\lambda_1}(A) = \{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2t \\ t \\ t \end{pmatrix} = t\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} \,|\, t \in F\}$$

$$= \mathrm{Span}\{\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}\} = F\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

To find the eigenspace $V_{\lambda_2}(A)$, we use Gaussian elimination and solve the matrix equation:

$$0 = (A - \lambda_2 I_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Its solution space is the eigenspace $V_{\lambda_2}(A)$ $(\subset F_c^3)$. So

$$V_{\lambda_2}(A) = \{\begin{pmatrix} -t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} | s, t \in F\}$$

$$= \mathrm{Span}\{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\}$$

$$= F\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + F\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Thus

$$\dim V_{\lambda_1}(A) = 1, \quad \dim V_{\lambda_2}(A) = 2,$$

i.e., the geometric multiplicities of λ_1 and λ_2 are 1 and 2, respectively.

In the present case, the algebraic multiplicity and geometric multiplicity of λ_i are identical. Indeed, A is diagonalizable (cf. Theorem 10.24)

Example 9.12. (**Eigenspace of a linear operator on $F[x]$**) Consider the linear operator

$$\begin{aligned} T : P_3[x] &\to P_3[x] \\ a_0 + a_1 x + a_2 x^2 &\mapsto T(a_0 + a_1 x + a_2 x^2) \end{aligned}$$

where

$$T(a_0+a_1x+a_2x^2) = -2a_2+(a_0+2a_1+a_2)x+(a_0+3a_2)x^2,$$

on the subspace $P_3[x]$ (of all polynomials of degree < 3) of $F[x]$.

Let

$$B = (1,\, x,\, x^2)$$

be the standard basis of $P_3[x]$. One sees that the representation matrix $[T]_B$ equals the matrix A in Example 9.11. So the characteristic polynomial of

T is (cf. Definition 8.27):

$$p_T(x) = p_A(x) = (x-1)(x-2)^2$$

and

$$\lambda_1 = 1, \quad \lambda_2 = 2$$

are the only eigenvalues of T.

Since

$$X_1 := \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

alone forms a basis of $V_{\lambda_1}(A)$,

$$B\,X_1 = (1,\ x,\ x^2) \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = -2 + x + x^2$$

alone forms a basis of $V_{\lambda_1}(T)$ (cf. Remark 9.8).

Since

$$Y_1 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad Y_2 := \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

form a basis of $V_{\lambda_2}(A)$,

$$B\,Y_1 = x, \quad B\,Y_2 = -1 + x^2$$

form a basis of $V_{\lambda_2}(T)$ (cf. Remark 9.8).

Example 9.13. (**Eigenvalue of upper triangular matrices**) Consider the upper triangular matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1,n} \\ 0 & a_{22} & \cdots & a_{2,n-1} & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \cdots & 0 & a_{n,n} \end{pmatrix}.$$

We calculate the characteristic polynomial of A:

$$\begin{aligned}
p_A(x) &= |xI_n - A| \\
&= \begin{vmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1,n-1} & -a_{1,n} \\ 0 & x - a_{22} & \cdots & -a_{2,n-1} & -a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & x - a_{n-1,n-1} & -a_{n-1,n} \\ 0 & 0 & \cdots & 0 & x - a_{n,n} \end{vmatrix} \\
&= (x - a_{11})(x - a_{22}) \ldots (x - a_{nn}).
\end{aligned}$$

So the eigenvalues of A are

$$\lambda_1 = a_{11},\ \lambda_2 = a_{22},\ \ldots, \lambda_n = a_{nn}$$

which are the entries on the diagonal and are not necessarily distinct.

The same is true for a lower triangular matrix.

Theorem 9.14. *(***Sum of eigenspaces***) Let*

$$\lambda_1, \ldots, \lambda_k$$

be some distinct eigenvalues of a linear operator T on a vector space V over a field F. Then the sum of eigenspaces

$$W := \sum_{i=1}^{k} V_{\lambda_i}(T)$$
$$= V_{\lambda_1}(T) + \cdots + V_{\lambda_k}(T)$$

is a direct sum:

$$W = \oplus_{i=1}^{k} V_{\lambda_i}(T)$$
$$= V_{\lambda_1}(T) \oplus \cdots \oplus V_{\lambda_k}(T).$$

Proof. Set

$$W_i := V_{\lambda_i}(T).$$

We use induction on k. When $k = 1$, it is clear.

Assume that

$$W_1 + \cdots + W_r$$

is already a direct sum. We need to show that

$$W_1 + \cdots + W_{r+1}$$

is also a direct sum.

Suppose the contrary that

$$W_1 + \cdots + W_{r+1}$$

is not a direct sum. Then by Definition 4.26 and the induction hypothesis, we have

$$W_{r+1} \cap \sum_{i=1}^{r} W_i \neq \{\mathbf{0}\}.$$

Let

$$\mathbf{0} \neq \mathbf{w}_{r+1} = \sum_{i=1}^{r} \mathbf{w}_i \quad (*)$$

for some $\mathbf{w}_i \in W_i$. Applying T on (*), we get

$$\lambda_{r+1}\mathbf{w}_{r+1} = \sum_{i=1}^{r} \lambda_i \mathbf{w}_i. \quad (**)$$

Multiplying λ_{r+1} to (*), we get

$$\lambda_{r+1}\mathbf{w}_{r+1} = \sum_{i=1}^{r} \lambda_{r+1} \mathbf{w}_i. \quad (***)$$

Using (***) to subtract (**), we get

$$\mathbf{0} = \sum_{i=1}^{r} (\lambda_i - \lambda_{r+1}) \mathbf{w}_i.$$

Since

$$W_1 + \cdots + W_r$$

is a direct sum by the inductive hypothesis (and hence the expression of $\mathbf{0}$ is unique: $\mathbf{0} = \mathbf{0} + \cdots + \mathbf{0}$, cf. Theorem 4.31), we have

$$(\lambda_i - \lambda_{r+1})\mathbf{w}_i = \mathbf{0} \quad (\forall\ i \le r).$$

Hence

$$\mathbf{w}_i = \mathbf{0} \quad (\forall\ i \le r)$$

since $\lambda_i - \lambda_{r+1} \neq 0$ by the assumption. This and (*) implies

$$\mathbf{w}_{r+1} = \mathbf{0},$$

which contradicts the choice of $\mathbf{w}_{r+1}$ in (*). Thus the sum is a diret sum, by induction. $\square$

Definition 9.15. (**Multiplication of linear operators $S_1 \ldots S_r$**) Let

$$T : V \to V$$

be a linear operator on a vector space V over a field F. Define

$$T^2 := T \circ T : V \to V$$

and (for $s \geq 1$)

$$T^s := T \circ \cdots \circ T \ \ \text{(s times)}$$

which is a linear operator on V (cf. Exercise 7.25):

$$\begin{aligned} T^s : V &\to V \\ \mathbf{v} &\mapsto T^s(\mathbf{v}). \end{aligned}$$

By convention, set

$$T^0 := I_V = \mathrm{id}_V \, .$$

More generally, for a polynomial

$$\begin{aligned} f(x) &= \sum_{i=0}^{r} a_i x^i \\ &= a_0 x^0 + a_1 x + \cdots + a_r x^r \ \in \ F[x] \end{aligned}$$

define

$$\begin{aligned} f(T) &:= \sum_{i=0}^{r} a_i T^i \\ &= a_0 I_V + a_1 T + \cdots + a_r T^r \ \in \ F[T]. \end{aligned}$$

Then $f(T)$ is a linear operator on V (cf. Exercise 7.33):

$$\begin{aligned} f(T) : V &\to V \\ \mathbf{v} &\mapsto f(T)(\mathbf{v}) = \sum_{i=0}^{r} a_i T^i(\mathbf{v}). \end{aligned}$$

Similarly, for linear operators

$$S_i : V \to V$$

we define

$$S_1 S_2 \cdots S_r := S_1 \circ S_2 \circ \cdots S_r$$

which is a linear operator (cf. Exercise 7.25):

$$\begin{aligned} S_1 S_2 \cdots S_r : V &\to V \\ \mathbf{v} &\mapsto S_1(S_2(\ldots(S_r(\mathbf{v})))). \end{aligned}$$

For a matrix $A \in M_n(F)$, we define

$$\begin{aligned} f(A) &:= \sum_{i=0}^{r} a_i A^i \\ &= a_0 I_n + a_1 A + \cdots + a_r A^r \in M_n(F). \end{aligned}$$

Exercise 9.16. (**Polynomials in T**) Let

$$T : V \to V$$

be a linear operator on a vector space V over a field F and with a basis

$$B = (\mathbf{u}_1, \ldots, \mathbf{u}_n).$$

Let

$$f(x),\ g(x) \in F[x]$$

be polynomials. Prove:

(1) We have

$$[f(T)]_B = f([T]_B).$$

In particular (cf. Theorem 8.6 or Example 8.9),

$$f(T) = 0\, I_V \text{ (the zero map on } V) \iff f([T]_B) = 0 I_n.$$

(2) The multiplication $f(T)\,g(T)$ as polynomials in T equals the composite $f(T) \circ g(T)$ as linear operators:

$$f(T)\,g(T) = f(T) \circ g(T).$$

(3) Commutativity:

$$f(T)\,g(T) = g(T)\,f(T).$$

(4) (proved in Tutorial T8.4) If $P \in M_n(F)$ is invertible, then

$$f(P^{-1}\,A\,P) = P^{-1}\,f(A)\,P.$$

(5) If

$$S : V \to V$$

is an isomorphism with inverse isomorphism

$$S^{-1} : V \to V$$

then

$$f(S^{-1}\,T\,S) = S^{-1}\,f(T)\,S.$$

Hint. For (1), apply Theorem 8.14 and Exercise 8.16, show

$$[aT^r + bT^s]_B = a([T]_B)^r + b([T]_B)^s$$

and use induction on the number of terms in $f(T)$.

For (2), note: $T^r\,T^s = T^r \circ T^s$ by the definition of the powers of T.

For (3), use $f(x)\,g(x) = g(x)\,f(x)$.

For (4) and (5), observe (and use (1)):

$$(P^{-1}\,A\,P)^r = P^{-1}\,A^r\,P.$$

Definition 9.17. (T**-invariant subspace**) Let

$$T : V \to V$$

be a linear operator on a vector space V. A subspace W of V is called T**-invariant** if the image of W under the map T is included in W:

$$T(W) := \{T(\mathbf{w}) \,|\, \mathbf{w} \in W\} \;\subseteq\; W$$

i.e.,

$$T(\mathbf{w}) \in W \quad (\forall\, \mathbf{w} \in W).$$

In this case, we have a well defined restriction map:

$$\begin{aligned} T|_W = T \circ \iota : W &\to W \\ \mathbf{w} &\mapsto T(\mathbf{w}) \end{aligned}$$

which is a linear operator on W. Here $\iota : W \to V$ is the inclusion map.

Exercise 9.18. (= Tutorial question) (**T-invariant subpaces** is printed as **Obvious T-invariant subpaces**) Let

$$T : V \to V$$

be a linear operator on a vector space V over a field F. Then the following are T-invariant subspaces of V.

(1) The zero vector space $\{0_V\}$.

(2) The whole space V.

(3) The image $T(V) = \{T(\mathbf{v}) \,|\, \mathbf{v} \in V\}$.

(4) The kernel

$$\mathrm{Ker}(T) = \{\mathbf{v} \in V \,|\, T(\mathbf{v}) = \mathbf{0}_V\}.$$

(5) The eigenspace

$$V_\lambda(T) = \{\mathbf{v} \in V \,|\, T(\mathbf{v}) = \lambda\mathbf{v}\}$$

of T corresponding to a fixed eigenvalue λ.

(6)

$$\mathrm{Span}\{\mathbf{v}_1\} = F\mathbf{v}_1$$

where $\mathbf{v}_1 \neq \mathbf{0}_V$ is an eigenvector of T corresponding to an eigenvalue λ.

Exercise 9.19. (**Sum of T-invariant subspaces**) If

$$W_1, \;\; W_2$$

are two T-invariant subspaces of a vector space V, then the sum

$$W_1 + W_2$$

is also a T-invariant subspace of V.

Example 9.20. (**T-(non)invariant subspace in F^3, adopted from S. L. Ma's notes**) Consider the linear operator

$$\begin{aligned} T : F^3 &\to F^3 \\ (x, y, z) &\mapsto (y, x, z) \end{aligned}$$

on the row vector 3-space F^3. Geometrically, this T gives the reflection about the following plane in F^3:

$$x - y = 0.$$

(1) The z-axis

$$W_1 = \{(0, 0, z) \,|\, z \in F\}$$

of F^3 is a T-invariant subspace of F^3.

(2) The xy-plane W_2 given by the equation $z = 0$, i.e.,

$$W_2 := \{(x, y, 0) \,|\, x, y \in F\}$$

is a T-invariant subspace of F^3.

Reason: for any $\mathbf{w} = (x_1, y_1, 0) \in W_2$, we have $T(\mathbf{w}) = (y_1, x_1, 0) \in W_2$.

(3) The yz-plane W_3 given by the equation $x = 0$, i.e.,

$$W_3 := \{(0, y, z) \mid y, z \in F\}$$

is not a T-invariant subspace of F^3, because $\mathbf{w} := (0, 1, 0) \in W_3$ but $T(\mathbf{w}) = (1, 0, 0) \notin W_3$.

Example 9.21. (**T-(non)invariant subspace in F^3**) Consider the linear operator

$$\begin{aligned} T : F^3 &\to F^3 \\ (x,\, y,\, z) &\mapsto (x + y,\, y + z,\, z + x) \end{aligned}$$

Then the subspace

$$\begin{aligned} W_1 : &= \{(u, u, u) \mid u \in F\} \\ &= \operatorname{Span}\{(1, 1, 1)\} \end{aligned}$$

is T-invariant, but the subspace

$$\begin{aligned} W_2 :&= \{(u, v, 0) \mid u, v \in F\} \\ &= \mathrm{Span}\{(1, 0, 0), (0, 1, 0)\} \end{aligned}$$

is not T-invariant.

Example 9.22. (**D-(non)invariant subspaces of $C^\infty[x]$**) Let

$$\begin{aligned} D : C^\infty[x] &\to C^\infty[x] \\ f &\mapsto D(f) = \frac{df}{dx} \end{aligned}$$

be the derivative linear operator in Example 9.10.

Then the subspace

$$U_1 := \mathbf{R}[x]$$

of real polynomials is a D-invariant subspace of $C^\infty[x]$, because "if $f(x)$ is a polynomial in $U_1 = \mathbf{R}[x]$ then

$$D(f) = \frac{df(x)}{dx}$$

is again a polynomial in $U_1 = \mathbf{R}[x]$, whence $D(U_1) \subseteq U_1$."

The subspace

$$\begin{aligned} U_2 : &= \mathrm{Span}\{\sin x, \cos x\} \\ &= \{c_1 \sin x + c_2 \cos x \,|\, c_1, c_2 \in \mathbf{R}\} \end{aligned}$$

of $C^\infty[x]$ is also D-invariant.

However, the subspace

$$\begin{aligned} U_3 : &= \mathrm{Span}\{1, \sin x\} \\ &= \{c_1 + c_2 \sin x \,|\, c_1, c_2 \in \mathbf{R}\} \end{aligned}$$

of C^∞ is not D-invariant.

Extra exercise 9.23. (= Tutorial question) (**Kernels and Images of Commutative operators**) Let $T_i : V \to V$ be two commutative to each other linear operators, i.e.,

$$T_1 \circ T_2 = T_2 \circ T_1$$

as maps. Namely,

$$T_1(T_2(\mathbf{v})) = T_2(T_1(\mathbf{v})) \quad (\forall\, \mathbf{v} \in V).$$

Show that both

$$\mathrm{Ker}(T_2), \quad \mathrm{Im}(T_2)$$

are T_1-invariant subspaces of V.

Give an example to show the necessity of the commutativity assumption on T_i.

Extra exercise 9.24. (**Evaluate T on a basis of a subspace**) Let

$$T : V \to V$$

be a linear operator on a vector space V over a field F. Let W be a subspace of V with a basis $B_W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots\}$. Then W is T-invariant, if and only if

$$T(B_W) \subseteq W,$$

i.e., $T(\mathbf{w}_i) \in W$ for all i.

Exercise 9.25. (**T-cyclic subspace**) Let

$$T : V \to V$$

be a linear operator on a vector space V over a field F. Fix a vector

$$\mathbf{0} \neq \mathbf{w}_1 \in V.$$

(1) Show that the subspace

$$\begin{aligned} W &:= \mathrm{Span}\{T^s(\mathbf{w}_1) \mid s \geq 0\} \\ &= \mathrm{Span}\{\mathbf{w}_1,\, T(\mathbf{w}_1),\, T^2(\mathbf{w}_1),\, \dots\} \end{aligned}$$

of V is T-invariant.

The above W is called the **T-cyclic subspace of V generated by $\mathbf{w}_1$**.

(2) Suppose that V is finite-dimensional. Let s be the smallest positive integer such that

$$T^s(\mathbf{w}_1) \in \mathrm{Span}\{\mathbf{w}_1,\, T(\mathbf{w}_1),\, \dots,\, T^{s-1}(\mathbf{w}_1)\}.$$

Show that

$$\dim_F W = s$$

and

$$B := \{\mathbf{w}_1,\, T(\mathbf{w}_1),\, \dots,\, T^{s-1}(\mathbf{w}_1)\}$$

is a basis of W.

(3) In (2), if

$$T^s(\mathbf{w}_1) = c_0\mathbf{w}_1 + c_1T(\mathbf{w}_1) + \cdots + c_{s-1}T^{s-1}(\mathbf{w}_1)$$

for some scalars $c_i \in F$, then the characteristic polynomial of the restriction operator $T|W$ on W is:

$$p_{T|W}(x) = -c_0 - c_1x - \cdots - c_{s-1}x^{s-1} + x^s.$$

Proof. (1) is easy. For (2) (= Tutorial question), you may borrow some argument from the proof of Theorem 5.8.

For (3), we use (2). Then the representation matrix

$$[T]_B = ([T(\mathbf{w}_1)]_B, [T^2(\mathbf{w}_1)]_B, .., [T^s(\mathbf{w}_1)]_B)$$
$$= \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & c_0 \\ 1 & 0 & 0 & \cdots & 0 & c_1 \\ 0 & 1 & 0 & \cdots & 0 & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & c_{s-1} \end{pmatrix}.$$

Thus the characteristic polynomial can be calculated by induction or co-factor expansion (as an exercise):

$$p_{T|W}(x) = |xI_s - [T]_B|$$
$$= \begin{vmatrix} x & 0 & 0 & \cdots & 0 & -c_0 \\ -1 & x & 0 & \cdots & 0 & -c_1 \\ 0 & -1 & x & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & x - c_{s-1} \end{vmatrix}$$
$$= -c_0 - c_1 x - \cdots - c_{s-1}x^{s-1} + x^s.$$

□

Theorem 9.26. *(**Characteristic polynomial of the restriction operator**) Let*

$$T : V \to V$$

be a linear operator on a vector space V over a field F and of dimension $n \geq 1$. Let W be a T-invariant subspace of V. Then the characteristic polynomial $p_{T|W}(x)$ of the restriction operator $T|W$ on W is a factor of the characteristic polynomial $p_T(x)$ of T, i.e.,

$$p_{T|W}(x) \,|\, p_T(x).$$

Namely,

$$p_T(x) = q(x)\, p_{T|W}(x)$$

for some polynomial $q(x) \in F[x]$.

Proof. Let

$$B_W = (\mathbf{w}_1, \ldots, \mathbf{w}_s)$$

be a basis of W. Extend it to a basis of V:

$$B := (\mathbf{w}_1, \ldots, \mathbf{w}_s, \ldots, \mathbf{w}_n)$$

(cf. Theorem 5.16). Then (cf. Definition 8.7)

$$\begin{aligned}[T]_B &= ([T(\mathbf{w}_1)]_B, .., [T(\mathbf{w}_s)]_B, .., [T(\mathbf{w}_n)]_B) \\ &= \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}\end{aligned}$$

where

$$\begin{aligned}\begin{pmatrix} A_1 \\ 0 \end{pmatrix} &= ([T(\mathbf{w}_1)]_B, \ldots, [T(\mathbf{w}_s)]_B) \\ &= \begin{pmatrix} [T|W]_{B_W} \\ 0 \end{pmatrix}\end{aligned}$$

and

$$\begin{pmatrix} A_2 \\ A_3 \end{pmatrix} = ([T(\mathbf{w}_{s+1})]_B, .., [T(\mathbf{w}_n)]_B).$$

with $A_3 \in M_{n-s}(F)$ and $A_2 \in M_{s\times(n-s)}(F)$.

Now the characteristic polynomial of T is (cf. Exercise 9.27 for the 4th equality below)

$$\begin{aligned} p_T(x) &= |xI_n - [T]_A| \\ &= |xI_n - \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}| \\ &= \begin{vmatrix} xI_s - A_1 & -A_2 \\ 0 & xI_{n-s} - A_3 \end{vmatrix} \\ &= |xI_s - A_1|\ |xI_{n-s} - A_3| \\ &= |xI_s - [T|W]_{B_W}|\ |xI_{n-s} - A_3| \\ &= p_{T|W}(x)\, q(x) \end{aligned}$$

with $q(x) := |xI_{n-s} - A_3|$. □

Exercise 9.27. (**Determinat of upper triangular matrix blocks**) If

$$C_1 \in M_s(F), \quad C_3 \in M_{n-s}(F)$$

are square matrices, then we can calulate the following determinant (of a square matrix in $M_n(F)$)

$$\begin{vmatrix} C_1 & C_2 \\ 0 & C_3 \end{vmatrix} = |C_1|\ |C_3|.$$

Extra exercise 9.28. (**To be T-invariant in terms of $[T]_B$**) Use Exercise 9.24 and the argument of the proof of Theorem 9.26 to show that a subspace W of an n-dimensional space V is T-invariant for a linear operator T on V, if and only if every basis B_W of W can be extended to a basis

$$B = (B_W,\ B_2)$$

of V such that the representation matrix of T relative to B, is of the form:

$$[T]_B = \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}$$

for some square matrices A_1, A_3 (automatically with $A_1 = [T|W]_{B_W}$).

In this case, show also that the matrix

$$A_2 = 0$$

if and only if

$$W_2 := \mathrm{Span}(B_2)$$

is a T-invariant subspace of V (automatically with $[T|W_2] = A_3$).

Extra exercise 9.29. (**Upper triangular form of a matrix**) Let $A \in M_n(F)$ (or let $T : V \to V$ be a linear operator on an n-dimensional vector space V over a field F). Suppose that the characteristic polynomial $p(x) = p_A(x)$ or $p(x) = p_T(x)$ is factorized as

$$p(x) = (x - \lambda_1)^{n_1} \dots (x - \lambda_k)^{n_k}$$

for some $\lambda_i \in F$. Show that there is an invertible $P \in M_n(F)$ (resp. a basis B of V) such that $P^{-1}AP$ (resp. the representation matrix $[T]_B$) is upper triangular.

Hint. We only consider T. For A, just consider $T = T_A$ and use Basis change theorem 8.20.

We use the proof and notation of Theorem 9.26 and Exercise 9.28. Let

$$W := V_{\lambda_1}(T)$$

be the eigenspace of T with a basis B_1. Set $W_2 :=$ Span(B_2) as in Exercise 9.28, and consider the linear transformation

$$T' : W_2 \to W_2$$

with the representation matrix $[T']_{B_2} = A_3$. This means $T' = T$ on $W_2 = \operatorname{Span} B_2$ (modulo error terms in W).

Warning: W_2 may not be T-invariant; hence the restriction map $T_{|W_2} : W_2 \to T(W_2)$ is not a self-map or a linear operator on W_2. Namely, in general,

$T(W_2) \neq W_2$, So we cannot simply consider the map T from W_2 (to $T(W_2)$) only and apply inductive hypothesis to reduce to the problem on W_2, which is, though, lower-dimensional.

Since $\dim W_2 < \dim V$, by induction on the dimension, we may assume that the representation matrix $[T]_{B_2'}$ equals an upper triangular A_3' for some new basis B_2' of W_2. Show that (B_1, B_2') is a basis of V so that $[T]_B$ is of upper triangular form

$$\begin{pmatrix} \lambda_1 I_{|B_1|} & A_2' \\ 0 & A_3' \end{pmatrix}.$$

Alternatively, you may show that the map

$$\begin{aligned} \overline{T} : V/W &\to V/W \\ \overline{\mathbf{v}} &\mapsto \overline{T(\mathbf{v})} \end{aligned}$$

is a well defined linear operator on the quotient space V/W so that $[\overline{T}]_{\overline{B}}$ is upper triangular for some basis

$$\overline{B} = \{\overline{\mathbf{w}_{|B_1|+1}}, \ldots, \overline{\mathbf{w}_n}\}$$

of V/W (by the dimension induction again). Then show that

$$B = B_1 \cup \{\mathbf{w}_{|B_1|+1}, \ldots, \mathbf{w}_n\}$$

is a basis of V and does the job (cf. Theorem 7.40).

Exercise 9.30. (**Characteristic polynomials of direct sums**) Let

$$T : V \to V$$

be a linear operator on an n-dimensional vector space V over a field F. Suppose that there are T-invariant subspaces

$$W_i \ \ (1 \le i \le r)$$

of V such that V is the direct sum

$$V = \oplus_{i=1}^r W_i = W_1 \oplus \cdots \oplus W_r$$

of W_i.

Show that the characteristic polynomial $p_T(x)$ of

T is the product:

$$\begin{aligned} p_T(x) &= \prod_{i=1}^{r} p_{T|W_i}(x) \\ &= p_{T|W_1}(x) \cdots p_{T|W_r}(x) \end{aligned}$$

of the characteristic polynomials of the restriction operators $T|W_i$ on W_i.

Hint. Let B_i be a basis of W_i. Then

$$B = (B_1, \dots, B_r)$$

is a basis of V (cf. Theorem 5.19) so that (as in Theorem 9.26):

$$[T]_B = \begin{pmatrix} [T|W_1]_{B_1} & 0 & \cdots & 0 & 0 \\ 0 & [T|W_2]_{B_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & [T|W_1]_{B_r} \end{pmatrix}$$

Now as in Theorem 9.26, we use Exercise 9.27 to calculate $p_T(x)$.

By the same reasoning, i.e., using Exercise 9.27, if

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 & 0 \\ 0 & A_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & A_r \end{pmatrix}$$

where $A_i \in M_{n_i}(F)$ are square matrices, then the characteristic polynomial of A is:

$$\begin{aligned} p_A(x) &= \prod_{i=1}^{r} p_{A_i}(x) \\ &= p_{A_1}(x) \cdots p_{A_r}(x). \end{aligned}$$

Extra exercise 9.31. (**T-invariant subspaces** is preceded by **To be direct sum of**) Use Exercise 9.24 and the argument of Exercise 9.30 to show that an n-dimensional vector space V with a linear operator T, is a direct sum

$$V = \oplus_{i=1}^{r} W_i = W_1 \oplus \cdots \oplus W_r$$

of some T-invariant subspaces W_i, if and only if every set of bases B_i of T_i gives rise to a basis

$$B = (B_1, \ldots, B_r)$$

of V (with $|B| = \sum_{i=1}^{r} |B_i|$) such that the representation matrix of T relative to B is of the form:

$$[T]_B = \begin{pmatrix} A_1 & 0 & \cdots & 0 & 0 \\ 0 & A_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & A_r \end{pmatrix}$$

(automatically with $[T|W_i]_{B_i} = A_i$) with A_i of order $|B_i| = \dim W_i$.

Theorem 9.32. *(Cayley-Hamilton theorem)* *Let (cf. Definition 8.27)*

$$p_T(x) = |xI_n - [T]_B| = \sum_{i=0}^{n} b_i x^i$$
$$= b_0 x^0 + b_1 x + \cdots + b_n x^n \in F[x]$$

be the characteristic polynomial of a linear operator

$$T : V \to V$$

on an n-dimensional vector space V over a field F and with a basis B. Then T satisfies the equation $p_T(x) = 0$, i.e.,

$$\begin{aligned} p_T(T) &= \sum_{i=0}^{n} b_i T^i \\ &= b_0 I_V + b_1 T + \cdots + b_n T^n = 0\, I_V \end{aligned}$$

which is the zero map on V.

Taking $T = T_A$ and B the standard basis so that $[T_A]_B = A$ (cf. Exercise 8.12), we get (cf. Exercise 9.16):

Matrix version of the theorem: *For a matrix $A \in M_n(F)$, its characteristic polynomial*

$$|xI_n - A| = \sum_{i=0}^{n} b_i x^i$$

satisfies

$$p_A(A) = \sum_{i=0}^{n} b_i A^i = 0\, I_n$$

which is the zero matrix in $M_n(F)$.

Proof. To show that the operator

$$p_T(T) = \sum_{i=0}^{n} b_i T^i = 0\, I_V$$

equivalently, we need to show the **claim** that

$$\begin{aligned} p_T(T)(\mathbf{v}_1) &= \sum_{i=0}^{n} b_i T^i(\mathbf{v}_1) \\ &= 0\, I_V(\mathbf{v}_1) \; = \; \mathbf{0}_V \; \in \; V \end{aligned}$$

for all $\mathbf{v}_1 \in V$.

Since $p_T(T)$ is a linear operator (cf. Definition 9.15) and hence (cf. Remark 7.12)

$$p_T(T)(\mathbf{0}_V) = \mathbf{0}_V,$$

to prove the claim, we may assume $\mathbf{v}_1 \neq \mathbf{0}_V$.

Now consider the T-cyclic subspace

$$W := \mathrm{Span}\{\mathbf{v}_1, T(\mathbf{v}_1), \dots\}$$

of V generated by $\mathbf{v}_1$. By Exercise 9.25, we have

$$T^s(\mathbf{v}_1) = c_0\mathbf{v}_1 + c_1T(\mathbf{v}_1) + \cdots + c_{s-1}T^{s-1}(\mathbf{v}_1) \quad (*)$$

for some scalars $c_i \in F$ so that the characteristic polynomial of the restriction operator $T|W$ is

$$p_{T|W}(x) = -c_0 - c_1x - \cdots - c_{s-1}x^{s-1} + x^s.$$

Setting $x = T|W$ in the above, we get

$$p_{T|W}(T|W) = -\sum_{i=0}^{s-1} c_i(T|W)^i + (T|W)^s.$$

Applying the above to $\mathbf{v}_1 \in W \subseteq V$, we get (using (*) above for the last equality):

$$p_{T|W}(T|W)(\mathbf{v}_1) = -\sum_{i=0}^{s-1} c_iT^i(\mathbf{v}_1) + T^s(\mathbf{v}_1) = \mathbf{0}. \quad (**)$$

On the other hand, by Theorem 9.26, the characteristic polynomial of T is

$$p_T(x) = q(x)\, p_{T|W}(x)$$

for some polynomial $q(x) \in F[x]$. Setting $x = T|W$

in the above, we get (cf. Exercise 9.16):

$$\begin{aligned} p_T(T|W) &= q(T|W)\, p_{T|W}(T|W) \\ &= q(T|W) \circ p_{T|W}(T|W). \end{aligned}$$

Applying the above to $\mathbf{v}_1 \in W \subseteq V$, we get (using (**) above and Remark 7.12 for the last two equalities):

$$\begin{aligned} p_T(T)(\mathbf{v}_1) &= q(T|W)(p_{T|W}(T|W)(\mathbf{v}_1)) \\ &= q(T|W)(\mathbf{0}) = \mathbf{0}. \end{aligned}$$

This proves the claim and also the theorem. $\square$

Example 9.33. (**Cayley-Hamilton for an operator on a $3D$ space**) Verify:

$$p_T(T) = 0, \quad p_A(A) = 0,$$

respectively, for the T in Example 9.12 and the A in Example 9.11.

In these early examples we have seen that

$$[T]_B = A$$

for the standard basis B and hence (cf. Definition 8.27):

$$\begin{aligned} p_T(x) &= p_A(x) \\ &= (x-1)(x-2)^2 =: f(x). \end{aligned}$$

Thus, by Exercise 9.16, checking

$$0 = p_T(T) = f(T)$$

is equivalent to checking

$$0 = f(A) = p_A(A)$$

which is to be verified by students as an exercise.

10 Minimal polynomial and Jordan canonical form

In this section, we study an arbitrary linear transformation T and its minimal polynomial, and Jordan canonical form relative to its Jordan canonical basis. Diagonalizability criteria will also be introduced. We will conclude the section with the additive Jordan decomposition.

Definition 10.1. (**Minimal polynomial**)

(1) A poylynomial $f(x) \in F[x]$ of degree n, is **monic** if its leading coefficient (i.e., the coefficient of x^n) is 1, i.e.,

$$f(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0.$$

(2) Let

$$T : V \to V$$

be a linear operator on an n-dimensional vector space over a field F. A nonzero polynomial

$$m(x) \in F[x]$$

is a **minimal polynomial of T** if it satisfies:

(2a) $m(x)$ is monic, i.e., its leading coefficient is 1.

(2b) Vanishing condition:

$$m(T) = 0I_V, \quad \text{and}$$

(2c) Minimality degree condition: Whenever $f(x) \in F[x]$ is another nonzero polynomial such that $f(T) = 0\, I_V$, we have

$$\deg(f(x)) \geq \deg(m(x)).$$

Similarly, for a matrix $A \in M_n(F)$, we can define a **minimal polynomial of A** as a monic polynomial $m(x)$ with the smallest positive degree such that

$$m(A) = 0I_n \in M_n(F).$$

Remark 10.2. (**Existence of minimal polynomial**) By Cayley-Hamilton theorem, the characteristic polynomial $p_T(x)$ satisfies

$$p_T(T) = 0\, I_V.$$

Thus there does exist a minimal polynomial $m(x)$ of T and

$$\deg(m(x)) \le \deg(p_T(x)) = n.$$

Indeed, we have (cf. Theorem 10.3):

$$m(x) \,|\, p_T(x).$$

Theorem 10.3. *(**Uniqueness of a minimal polynomial** $m_T(x)$) Let $T : V \to V$ be a linear operator on an n-dimensional vector space V over a field F (resp. $A \in M_n(F)$). Let $m(x)$ be a minimal polynomial of T (resp. A). Let $f(x) \in F[x]$. Then the following are equivalent.*

(1) $f(T) = 0\, I_V$ (resp. $f(A) = 0\, I_n$).

(2) $m(x)$ *is a factor of* $f(x)$*, i.e.,* $m(x) \mid f(x)$*.*

Namely,

$$f(x) = q(x)\, m(x)$$

for some polynomial $q(x) \in F[x]$*.*

In particular, there is exactly one minimal polynomial of T *(resp.* A*) and will be denoted as*

$$m_T(x) := m(x) \quad (\textit{resp. } m_A(x) := m(x))$$

Further, if $A = [T]_B$ *(the representation matrix), then* $m_T(x) = m_A(x)$*.*

Proof. For '(2) $\Rightarrow$ (1)', (2) implies (cf. ExerciseEnd(V))

$$\begin{aligned} f(T) &= q(T)\, m(T) \\ &= q(T)\, 0I_V = 0I_V. \end{aligned}$$

For '(1) $\Rightarrow$ (2)', assume (1) (and we will show (2)). By Euclidean division algorithm, we can write

$$f(x) = q(x)\, m(x) + r(x) \quad (*)$$

for polynomials

$$q(x),\, r(x) \in F[x]$$

such that either

$$r(x) = 0, \quad \text{or} \;\; 0 \le \deg(r(x)) < \deg(m(x)).$$

If $r(x) = 0$ then (2) is true. We are done.

Suppose the contrary that $r(x) \ne 0$. Setting $x = T$ in the relation

$$r(x) = f(x) - q(x)\, m(x)$$

(deduced from (*)), we get (using (1)):

$$\begin{aligned} r(T) &= f(T) - q(T)\, m(T) \\ &= 0 I_V - q(T)\, 0 I_V = 0\, I_V. \end{aligned}$$

So $r(x)$ is a polynomial with $r(T) = 0$ and with positive degree smaller than that of $m(x)$.

This contradicts the minimality condition of $m(x)$ (cf. Definition 10.1).

Thus, $r(x) = 0$ and (2) is true.

For the uniqueness, suppose that $n(x)$ is another (monic) minimal polynomial of T. Then $n(T) = 0\,I_V$ and hence the above proved equivalence (1) and (2) imply that

$$n(x) = m_1(x)\,m(x) \quad (**)$$

for some polynomial $m_1(x) \in F[x]$.

Similarly (interchanging the role of $m(x)$ and $n(x)$), we have

$$m(x) = m_2(x)\,n(x) \quad (***)$$

for some $m_2(x) \in F[x]$.

Combining (**) and (***), we get

$$m(x) = m_2(x)\,m_1(x)\,m(x)$$

and hence (cancelling the nonzero $m(x)$),

$$1 = m_2(x)m_1(x).$$

Thus both $m_i(x)$ equal to some constants c_i (Exercise! Hint. $0 = \deg(m_2 m_1) = \deg(m_2)+\deg(m_1) \geq 0 + 0$). Now

$$m(x) = c_2 n(x)$$

and the monic assumption on both $m(x)$ and $n(x)$ imply that $c_2 = 1$ (Exercise! Hint. Compare the leading coefficients of the two sides). Hence

$$m(x) = n(x).$$

This proves the uniqueness of a minimal polynomial $m(x)$.

The final assertion

$$m_T(x) = m_A(x)$$

follows from the observation below (cf. Exercise 9.16) and is left as an exercise for students:

$$[f(T)]_B = f([T]_B).$$

$\square$

Example 10.4. ($m_A(x)$ **for a** 3×3 **matrix**) For the matrix A in Exercise 9.11, one can calculate that

$$m(A) = 0\, I_3 \in M_3(F)$$

where

$$m(x) := (x-1)(x-2) = x^2 - 3x + 2.$$

Indeed,

$$\begin{aligned} m(A) &= (A - I_3)(A - 2I_3) \\ &= \begin{pmatrix} -1 & 0 & -2 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} -2 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} = 0. \end{aligned}$$

You may also calculate like (check it!):

$$m(A) = A^2 - 3A + 2I_3 = 0.$$

One can show that the minimal polynomial $m_A(x)$ of A equals $m(x)$.

Indeed, since

$$m_A(x) \mid p_A(x)$$

i.e.,

$$m_A(x) \mid (x-1)(x-2)^2$$

(cf. Remark 10.2), the possible candidates for monic $m_A(x)$ are :

$$\begin{aligned} &(x-1), \quad (x-2) \\ m(x) =&(x-1)(x-2) \\ p_A(x) =&(x-1)(x-2)^2. \end{aligned}$$

The first two candidates can not be $m_A(x)$, because they do not kill A (Exercise!), or because Remark 10.19 below shows that $m_A(x)$ and $p_A(x)$ should have the same zero set: $\{1,2\}$).

Thus $m(x)$ is the monic polynomial killing A and with the least positive degree, so $m(x)$ equals the minimal polynomial $m_A(x)$ of A.

Note that $\deg(m(x)) = 2 < 3 = \deg(p_A(x)$ (compare with Remark 10.2).

For the T in Example 9.12, we have

$$m_T(x) = m_A(x) = m(x)$$

since $[T]_B = A$ and by Theorem 10.3.

Exercise 10.5. (**Minimal polynomials of similar matrices**) If two matrices A_i are similar: $A_1 \sim A_2$, then they have the same minimal polynomial:

$$m_{A_1}(x) = m_{A_2}(x).$$

Hint. Write $A_2 = P^{-1}A_1P$ and see the Hint to Exercise 9.16.

Exercise 10.6. (**Minimal polynomials of direct sums**) Consider the matrix

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 & 0 \\ 0 & A_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & A_r \end{pmatrix}$$

where $A_i \in M_{n_i}(F)$ are square matrices. Show (as a Tutorial question) that the minimal polynomial $m_A(x)$ of A is equal to the (monic) **least common mulitple** of the minimal polynomials $m_{A_i}(x)$ (of A_i), i.e.,

$$m_A(x) = \operatorname{lcm}\{m_{A_1}(x), \ldots, m_{A_r}(x)\}.$$

Let

$$T : V \to V$$

be a linear operator on an n-dimensional vector space V over a field F. Suppose that there are T-invariant subspaces

$$W_i \ (1 \le i \le r)$$

of V such that V is the direct sum

$$V = \oplus_{i=1}^{r} W_i = W_1 \oplus \cdots \oplus W_r$$

of W_i. Then the minimal polynomial of T is:

$$m_T(x) = \mathrm{lcm}\{m_{T|W_1}(x), \ldots, m_{T|W_r}(x)\}.$$

Hint. Find $A := [T]_B$ as in Exercise 9.30 and note that $m_T(x) = m_A(x)$ (cf. Theorem 10.3).

Remark 10.7. (**How to find the minimal polynomial $m_T(x)$?**) On the one hand (cf. Remark 10.2)

$$m_T(x) \,|\, p_T(x).$$

On the other hand, we will see late (cf. Tutorial 8, Q6, or Remark 10.19) that the set of zeros (in an over field $\overline{F}$ of F) of $p_T(x)$ and that of $m_T(x)$ are identical;

$$\{\alpha \in \overline{F} \mid m_T(\alpha) = 0\} = \{\alpha \in \overline{F} \mid p_T(\alpha) = 0\}.$$

So $p_T(x)$ and $m_T(x)$ are different only in their multiplicities of zeros. Thus if

$$p_T(x) = (x - \lambda_1)^{n_1} \dots (x - \lambda_k)^{n_k}$$

where $\lambda_i \in \overline{F}$, and n_i are the so called algebraic multiplicities of λ_i, then

$$m_T(x) = (x - \lambda_1)^{m_1} \dots (x - \lambda_k)^{m_k}$$

for some $1 \le m_i \le n_i$ $(1 \le i \le k)$.

Definition 10.8. (**Jordan Block**) Let λ be a scalar in a field F. The matrix below

$$J := J_s(\lambda) := \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}_{s \times s} \in M_s(F)$$

is called the **Jordan Block of order s with eigenvalue λ**.

The characteristic polynomial and minimal polynomial of J are identical (cf. Theorem 9.26 for the calculation of $p_J(x)$):

$$m_J(x) = (x - \lambda)^s = p_J(x).$$

The eigenspace

$$V_\lambda(J) = \text{Span}\{\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\}$$

has dimension 1, i.e., the geometric multiplicity of λ is 1. But the algebraic multiplicity of λ is s.

Example 10.9. (**A Jordan canonical form of size 6×6**). Let λ be a scalar in a field F. The matrix below

$$A = \begin{pmatrix} J_2(\lambda) & 0 & 0 \\ 0 & J_1(\lambda) & 0 \\ 0 & 0 & J_3(\lambda) \end{pmatrix}$$

$$= \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & \lambda \end{pmatrix}_{6\times 6} \in M_6(F)$$

is a matrix in so called Jordan canonical form (see Definition 10.11 below for the general definition).

The characteristic polynomial of A is

$$p_A(x) = (x - \lambda)^6$$

while the minimal polynomial of A is (cf. Exercise

10.6):

$$\begin{aligned}m_A(x) &= \mathrm{lcm}(m_{J_2(\lambda)}(x),\, m_{J_1(\lambda)}(x),\, m_{J_3(\lambda)}(x))\\ &= \mathrm{lcm}((x-\lambda)^2,\, (x-\lambda),\, (x-\lambda)^3)\\ &= (x-\lambda)^3.\end{aligned}$$

Verify that the eigenspace $V_\lambda(A) =$

$$\mathrm{Span}\{\mathbf{e}_1 = \begin{pmatrix}1\\0\\0\\0\\0\\0\end{pmatrix}, \mathbf{e}_3 = \begin{pmatrix}0\\0\\1\\0\\0\\0\end{pmatrix}, \mathbf{e}_4 = \begin{pmatrix}0\\0\\0\\1\\0\\0\end{pmatrix}\}$$

has dimension 3.

We now use the notation in Definition 10.11. The geometric multiplicity of λ is

$$e(\lambda) := \dim V_\lambda(A) = 3$$

while the algebraic multiplicity of λ (i.e., the multiplicity of λ in $p_A(x)$) is

$$s(\lambda) = 6.$$

The multiplicity of λ in the minimal polynomial $m_A(x)$ is

$$s_e(\lambda) = 3.$$

There are exactly

$$r = 3$$

Jordan blocks in the Jordan canonical form A.

The number of distinct eigenvalues in A is:

$$k = 1.$$

You may either find a basis of $V_\lambda(A)$ directly, or use the **correspondence** between a **basis** of the eigenspace $V_\lambda(A)$ and the **first columns** of the Jordan blocks with eigenvalue λ in the Jordan canonical form

$$A = \mathrm{diag}[J_2(\lambda),\ J_1(\lambda),\ J_3(\lambda)]$$

as detailed in Defintion 10.11:

The **first** column of the Jordan block $J_2(\lambda)$ (resp. $J_1(\lambda)$, or $J_3(\lambda)$) is the **1**st (resp. **3**rd, or **4**th) column of the Jordan canonical form A, so we obtain a basis of $V_\lambda(A)$:

$$(\mathbf{e_1},\ \mathbf{e_3},\ \mathbf{e_4})$$

as already mentioned above.

Exercise 10.10. For the A in Exercise 10.9 and each j, express

$$A\mathbf{e}_j$$

as a linear combination of vectors in the standard basis $(\mathbf{e}_1, \ldots, \mathbf{e}_6)$.

Definition 10.11. (**Jordan canonical form**). Let λ be a scalar in a field F. We have seen in

Definition 10.8 that the Jordan Block

$$J_s(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}_{s \times s} \in M_s(F)$$

satisfies:

$$p_{J_s(\lambda)}(x) = m_{J_s(\lambda)}(x) = (x - \lambda)^s;$$

the eigenspace

$$V_\lambda(J_s(\lambda)) = \mathrm{Span}\{\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}\}$$

has dimension 1, i.e., the geometric multiplicity of λ is 1; the algebraic multiplicity of λ is s.

Let λ be a nonzero scalar in a field F. Let

$$s_1 \leq s_2 \leq \cdots \leq s_e.$$

The following **Block Diagonal**

$$A(\lambda) = \begin{pmatrix} J_{s_1}(\lambda) & 0 & 0 & \cdots & 0 & 0 \\ 0 & J_{s_2}(\lambda) & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & J_{s_{e-1}}(\lambda) & 0 \\ 0 & 0 & 0 & \cdots & 0 & J_{s_e}(\lambda) \end{pmatrix}$$

is called a **Jordan canonical form with eigenvalue λ**.

The order of $A(\lambda)$ is

$$s = \sum_{i=1}^{e} s_i$$

i.e., $A(\lambda) \in M_s(F)$.

The characteristic polynomial and minimal polynomial of A are (cf. Exercise 10.6):

$$p_{A(\lambda)}(x) = (x-\lambda)^s, \quad m_{A(\lambda)}(x) = (x-\lambda)^{s_e}$$

where s is the so called algebraic multiplicity of λ of $A(\lambda)$.

Let

$$B := (\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_s = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix})$$

be the standard basis of F_c^s. Verify that the eigenspace of A:

$$V_\lambda(A(\lambda)) =$$

$$\mathrm{Span}\{\mathbf{e}_1, \mathbf{e}_{1+s_1}, \mathbf{e}_{1+s_1+s_2}, .., \mathbf{e}_{1+s_1+\cdots+s_{e-1}}\},$$

has dimension equal to $e = e(\lambda)$ (the number of Jordan blocks in $A(\lambda)$).

Indeed, there is a **correspondence** between a **basis** of the eigenspace $V_\lambda(A(\lambda))$ and the **first columns** of the Jordan blocks with eigenvalue λ in the Jordan canonical form

$$A(\lambda) = \mathrm{diag}[J_{s_1}(\lambda),\ J_{s_2}(\lambda),\ \ldots,\ J_{s_e}(\lambda)]$$

as detailed below (compare with Example 10.9):

The **first** column of the Jordan block $J_{s_1}(\lambda)$ (resp. $J_{s_2}(\lambda)$, or $J_{s_3}(\lambda)$, ..., or $J_{s_e}(\lambda)$) is the **1**st (resp. $(1+s_1)$-th, or $(1+s_1+s_2)$-th, ..., or $(1+s_1+\cdots+s_{e-1})$-th) column of the Jordan canonical form $A(\lambda)$, so we obtain the basis of $V_\lambda(A(\lambda))$:

$$(\mathbf{e_1}, \mathbf{e}_{1+s_1}, \mathbf{e}_{1+s_1+s_2}, .., \mathbf{e}_{1+s_1+\cdots+s_{e-1}}).$$

Finally, since the geometric (resp. algebraic) multiplicity of the eigenvalue λ of $A(\lambda)$ is $\dim V_\lambda(A_\lambda) = e$ (reap. $s = \sum_{i=1}^{e} s_i$), we have

$$(\text{Geom.multiplicity of eigenvalue}\,\lambda) \leq (\text{Alg.mult.of}\,\lambda).$$

For late use, we denote by

$$s = s(\lambda)$$

the order of $A(\lambda)$, i.e., $A(\lambda) \in M_{s(\lambda)}(F)$,

$$e = e(\lambda)$$

the number of Jordan blocks $J_{s_i}(\lambda)$'s in $A(\lambda)$, and

$$s_e = s_e(\lambda)$$

the multiplicity of λ in the minimal polynomial $m_{A(\lambda)}(x)$.

More generally, let

$$\lambda_1, \ldots, \lambda_k$$

be distinct scalars in F. Usually we also assume

$$\lambda_1 < \lambda_2 < \cdots < \lambda_k$$

when the field $F = \mathbf{R}$.

Then the block diagonal

$$J = \begin{pmatrix} A(\lambda_1) & 0 & 0 & \cdots & 0 & 0 \\ 0 & A(\lambda_2) & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A(\lambda_{k-1}) & 0 \\ 0 & 0 & 0 & \cdots & 0 & A(\lambda_k) \end{pmatrix}$$

is called a **Jordan canonical form** or simply a **canonical form**, where $A(\lambda_i)$ is a Jordan canonical form with eigenvalue λ_i as shown above.

Each $A(\lambda_i)$ is of order

$$s(\lambda_i),$$

and $s(\lambda_i)$ is also the number of times the same scalar λ_i appears on the diagonal of J and also the algebraic multiplicity of the eigenvalue λ_i of J (i.e., the multiplicity of λ_i as a zero of $p_J(x)$).

So J is of order equal to

$$\sum_{i=1}^{k} s(\lambda_i).$$

There are exactly

$$e(\lambda_i)$$

Jordan blocks (with eigenvalue λ_i) in $A(\lambda_i)$, the largest of which is of order

$$s_e(\lambda_i).$$

This $s_e(\lambda_i)$ is also the multiplicity of λ_i in the minimal polynomial $m_J(x)$.

There are exactly

$$\sum_{i=1}^{k} e(\lambda_i)$$

Jordan blocks in J.

Now one can check (cf. Exercises 9.30, 10.6 and 10.8):

$$
\begin{aligned}
p_J(x) &= \prod_{i=1}^{k}(x-\lambda_i)^{s(\lambda_i)} \\
&= (x-\lambda_1)^{s(\lambda_1)}..(x-\lambda_k)^{s(\lambda_k)}, \\
m_J(x) &= \prod_{i=1}^{k}(x-\lambda_i)^{s_e(\lambda_i)} \\
&= (x-\lambda_1)^{s_e(\lambda_1)}..(x-\lambda_k)^{s_e(\lambda_k)}.
\end{aligned}
$$

The eigenspace $V_{\lambda_i}(J)$ has dimension $e(\lambda_i)$ and is spanned by the $e(\lambda_i)$ vectors 'corresponding' to the first columns of the $e(\lambda_i)$ Jordan blocks in $A(\lambda_i)$; see Examples 10.9 and 10.12 and the case of the matrix $A(\lambda)$ above for the 'corresponding'.

As in the case of $A(\lambda)$, for the matrix J, we have

$$\dim V_{\lambda_i}(J) = e(\lambda_i) \ \le\ s(\lambda_i) \ \text{ i.e.,}$$

(Geom.multiplicity of eigenvalue λ) $\le$ (Alg.mult.of λ).

Sometimes, a block diagonal J below

$$J = \begin{pmatrix} J_{s_1}(\lambda_1) & 0 & 0 \cdots & 0 & 0 \\ 0 & J_{s_2}(\lambda_2) & 0 \cdots & 0 & 0 \\ \vdots & \vdots & \vdots \ \vdots & 0 & 0 \\ 0 & 0 & 0 \cdots & J_{s_{r-1}}(\lambda_{r-1}) & 0 \\ 0 & 0 & 0 \cdots & 0 & J_{s_r}(\lambda_r) \end{pmatrix}$$

is also called a **Jordan canonical form**, where each $J_{s_i}(\lambda_i)$ is a Jordan block with eigenvalue $\lambda_i \in F$, but these λ_i's may not be distinct.

Assume that there are exactly k distinct elements in the set

$$\{\lambda_1, \ldots, \lambda_r\}$$

and we assume that

$$\lambda_{m_i} \ \ (1 \le i \le k)$$

are these k distinct ones. These k of λ_{m_i} are just the distinct eigenvalues of J.

Let

$$s(\lambda_{m_i})$$

be the number of times the same scalar λ_{m_i} appears on the diagonal of J.

Let

$$e(\lambda_{m_i})$$

be the number of Jordan blocks (among the r such in J) with eigenvalue the same λ_{m_i}; among these $e(\lambda_{m_i})$ Jordan blocks, the largest is of order say

$$s_e(\lambda_{m_i}).$$

The eigenspace $V_{\lambda_{m_i}}(J)$ has dimension $e(\lambda_{m_i})$ and is spanned by $e(\lambda_{m_i})$ vectors 'corresponding' to the first columns of these $e(\lambda_{m_i})$ Jordan blocks (cf. Examples 10.9 and 10.12 and the case of the matrix $A(\lambda)$ for the 'corresponding').

Then

$$\begin{aligned} p_J(x) &= \prod_{i=1}^{k} (x - \lambda_{m_i})^{s(\lambda_{m_i})} \\ &= (x - \lambda_{m_1})^{s(\lambda_{m_1})} .. (x - \lambda_{m_k})^{s(\lambda_{m_k})}, \\ m_J(x) &= \prod_{i=1}^{k} (x - \lambda_{m_i})^{s_e(\lambda_{m_i})} \\ &= (x - \lambda_{m_1})^{s_e(\lambda_{m_1})} .. (x - \lambda_{m_k})^{s_e(\lambda_{m_k})}. \end{aligned}$$

As in the case of $A(\lambda)$, for the matrix J, we have

$$\dim V_{\lambda_{m_i}}(J) = e(\lambda_{m_i}) \ \leq \ s(\lambda_{m_i}) \ \text{ i.e.,}$$

$$(\text{Geom.multiplicity of eigenvalue}\, \lambda) \leq (\text{Alg.mult.of}\, \lambda).$$

To double check the calculation, one may use the relations:

J has order equal to

$$\sum_{i=1}^{r} s_i = \sum_{i=1}^{k} s(\lambda_{m_i}), \quad \text{and}$$

the number of Jordan blocks in J is

$$r = \sum_{i=1}^{k} e(\lambda_{m_i}).$$

Example 10.12. (**A Jordan canonical form of size** 8×8) Consider the Jordan canonical form below

$$J = \begin{pmatrix} J_2(\lambda_1) & 0 & 0 & 0 \\ 0 & J_1(\lambda_2) & 0 & 0 \\ 0 & 0 & J_3(\lambda_1) & 0 \\ 0 & 0 & 0 & J_2(\lambda_2) \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & & \lambda_2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix}_{8\times 8} \in \ M_8(F)$$

where $\lambda_1 \neq \lambda_2$ in a field F.

Let

$$B := (\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_8 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix})$$

be the standard basis of F_c^8. Then we can calculate (cf. Exercise 10.6):

$$p_J(x) = (x - \lambda_1)^5 (x - \lambda_2)^3,$$
$$m_J(x) = (x - \lambda_1)^3 (x - \lambda_2)^2,$$
$$V_{\lambda_1}(J) = \mathrm{Span}\{\mathbf{e}_1, \mathbf{e}_4\},$$
$$V_{\lambda_2}(J) = \mathrm{Span}\{\mathbf{e}_3, \mathbf{e}_7\}.$$

Exercise 10.13. (**Finding Jordan canonical forms with given $p_A(x)$, $m_A(x)$**) Find all Jordan canonical forms J (up to re-ordering of Jordan blocks in J) such that the characteristic polynomial $p_J(x)$ and minimal polynomial $m_J(x)$ are the same as those in Example 10.12.

For each of such J, find also the eigenspaces $V_{\lambda_i}(J)$.

Theorem 10.14. *(**Jordan canonical form of a linear operator**). Let V be a vector space of dimension n over a field F and*

$$T : V \to V$$

a linear operator with characteristic polynomial $p_T(x)$ and minimal polynomial $m_T(x)$ as follows

$$p_T(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k},$$
$$m_T(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

where

$$\lambda_1, \dots, \lambda_k$$

are distinct scalars in F.

Then there is a basis B of V such that the representative matrix $[T]_B$ equals a Jordan canonical form $J \in M_n(F)$ given in Definition 10.11, with

$$n_i = s(\lambda_i), \quad m_i = s_e(\lambda_i).$$

Such a block diagonal J is called a **Jordan canonical form of** *T. It is unique up to re-ordering of λ_i, i.e., the re-ordering of the Jordan blocks in J. The basis B of V is called a* **Jordan canonical basis for** *T.*

Proof. We give a sketch of the proof, the details of which can be found in Friedberg, Insel and Spence [§7.1 The Jodan canonical form I].

Step 1. Show that V is a direct sum

$$V = K_1 \oplus \cdots \oplus K_k$$

where

$$K_i := \mathrm{Ker}(T - \lambda_i I_V)^{m_i}.$$

This K_i (resp. its nonzero vector) is called the **generalized eigenspace (resp. generalized eigenvector) corresponding to the eigenvalue** λ_i.

For comparison, recall that the eigenspace is

$$V_{\lambda_i}(T) = \mathrm{Ker}(T - \lambda I_V).$$

It can be proved that

$$\begin{aligned} K_i : &= \mathrm{Ker}(T - \lambda_i I_V)^{m_i} \\ &= \mathrm{Ker}(T - \lambda_i I_V)^{r} \end{aligned}$$

for all $r \geq m_i$, especially for $r = n_i$. Hence, once we get $p_T(x)$ we can already calculate K_i, even without knowing $m_T(x)$.

Since T commutes with $T - \lambda_i I_V$ (cf. Exercise 9.16), the subspace

$$K_i = \mathrm{Ker}(T - \lambda_i I_V)^{m_i}$$

is T-invariant (cf. Exercise 9.23).

Step 2. Show that K_i has a basis B_i such that $[T|K_i]_{B_i}$ equals some Jordan canonical form $A(\lambda_i)$ as in Definition 10.11 with eigenvalue λ_i:

$$[T|K_i]_{B_i} = A(\lambda_i).$$

Now take the basis

$$B = (B_1, \ldots, B_k)$$

and the theorem follows (cf. Exercise 9.31).

One can also show that

$$p_{T|K_i}(x) = (x - \lambda_i)^{n_i},$$

$$m_{T|K_i}(x) = (x - \lambda_i)^{m_i},$$

$$\dim K_i = n_i.$$

$\square$

Theorem 10.15. *(***Matrix version of Jordan canonical form***) Let A be a matrix in $M_n(F)$ with characteristic polynomial $p_A(x)$ and minimal polynomial $m_A(x)$ as follows*

$$p_A(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k},$$

$$m_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

where

$$\lambda_1, \ldots, \lambda_k$$

are distinct scalars in F.

Then there is an invertible matrix P in $M_n(F)$ such that $P^{-1}AP$ equals some Jordan canonical form $J \in M_n(F)$ as in Definition 10.11, *with*

$$n_i = s(\lambda_i), \quad m_i = s_e(\lambda_i).$$

Such a block diagonal J is called a **Jordan canonical form of** *A. It is unique up to re-ordering of λ_i, i.e., the re-ordering of the Jordan blocks in J.*

Proof. Applying Theorem 10.14 to

$$\begin{aligned} T = T_A : F_c^n &\to F_c^n \\ X &\mapsto AX \end{aligned}$$

we get a so called Jordan canonical basis B such that

$$[T]_B = J$$

which is a Jordan canonical form.

If $\widetilde{B}$ is the standard basis of F_c^n, then

$$[T]_{\widetilde{B}} = [T_A]_{\widetilde{B}} = A$$

(cf. Example 8.12). Let P be the transition matrix from the new basis B to the old basis $\widetilde{B}$. Then (cf. Basis change theorem 8.20):

$$\begin{aligned} P^{-1} \, A \, P &= P^{-1} \, [T]_{\widetilde{B}} \, P \\ &= [T]_B = J. \end{aligned}$$

This proves the theorem.

We remark that the above P is characterized by the equality below (cf. Theorem 8.17):

$$B = \widetilde{B} \, P.$$

$\square$

The argument in the proof of Theorem 10.15 also proves:

Theorem 10.16. *(***A canonical form of T is a canonical form of $[T]_{\widetilde{B}}$ and vice versa***) Let*

V be an n-dimensional vector space over a field F and

$$T : V \to V$$

a linear operator. Let

$$A := [T]_{\widetilde{B}}$$

be the representation matrix of T relative to a basis

$$\widetilde{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

of V. Let J be a Jordan canonical form. Then the following are equivalent.

(1) There is an invertible matrices $P \in M_n(F)$ such that

$$P^{-1}AP = J.$$

(2) There is an invertible matrix $P \in M_n(F)$ such that the representation matrix $[T]_B$ relative to

the new basis

$$B := \widetilde{B}\, P = (\mathbf{v}_1, \dots, \mathbf{v}_n)\, P$$

is J, i.e.,

$$[T]_B = J.$$

Remark 10.17. (**Zeros of $p_T(x)$**)

(1) For a matrix $A \in M_n(F)$ or a linear operator T on an n-dimensional vector space V over a field F, the zeros of $p_A(x)$ (or $p_T(x)$) may not be in F. For instance when $F = \mathbb{R}$, we can find examples of $p_A(x)$ or $p_T(x)$ with non-real complex zeros (cf. Exercise 9.2).

(2) Thus when we write the characteristic polynomial of $p(x) := p_A(x)$ (or $p(x) := p_T(x)$) as

$$p(x) = (x - \lambda_1) \dots (x - \lambda_n)$$

there is no guarantee that

$$\lambda_i \in F \ (\forall\ i) \ (*)$$

while this condition (*) is assumed in Theorems 10.14 and 10.15.

Theorem 10.18. *(**Existence of Jordan canonical form**) Let F be a field. Let A be a matrix in $M_n(F)$ (or let T be a linear operator on an n-dimensional vector space V over F). Let $p(x) := p_A(x)$ (or $p(x) := p_T(x)$) be the characteristic polynomial. Then the following are equivalent.*

(1) A (or T) has a Jordan canonical form $J \in M_n(F)$.

(2) Every zero (or root) of the characteristic polynomial $p(x)$ belongs to F.

(3) We can factor $p(x)$ as

$$p(x) = (x - \lambda_1) \dots (x - \lambda_n)$$

where all $\lambda_i \in F$.

In particular, if F is so called **algebraically closed** *(e.g. $F = \mathbb{C}$, cf. the fundamental theorem of algebra), then every matrix $A \in M_n(F)$ and every T on an n-dimensional vector space V over F, have a Jordan canonical form $J \in M_n(F)$.*

Proof. This follows from Theorems 10.14 and 10.15 and the fact that $p_J(x) = p(x)$ (cf. Exercise 8.28).

$\square$

Below are some consequences of Theorems 10.14 and 10.15.

Remark 10.19. (**Consequences of Jordan canonical forms**) Let A be a matrix in $M_n(F)$ (or let T be a linear operator on an n-dimensional vector space V over F). Set $p(x) := p_A(x)$ or $p_T(x)$, and $m(x) := m_A(x)$ or $m_T(x)$.

(1) (Also a Tutorial question) The characteristic polynomial $p(x)$ and the minimal polynomial $m(x)$ have the same zero sets:

$$\{\alpha \in F \mid p(\alpha) = 0\} = \{\alpha \in F \mid m(\alpha) = 0\}.$$

Indeed, Theorems 10.14 and 10.15 say that the multiplicity n_i and m_i of a zero λ_i of $p(x)$ and $m(x)$ satisfy

$$n_i \geq m_i \geq 1.$$

(2) If $J \in M_n(F)$ is a Jordan canonical form of T or of $T = T_A$ (i.e., of A), then, in notation of Definition 10.11, we have (cf. Exercise 9.7):

$$\dim V_{\lambda_i}(T) = \dim V_{\lambda_i}(J) = e(\lambda_i) \ \leq \ s(\lambda_i)$$

i.e.,

(Geom.multiplicity of eigenvalue λ) $\leq$ (Alg.mult.of λ).

Theorem 10.20. *(**Canonical forms of similar matrices**) Let $A_i \in M_n(F)$ and $J_i' \in M_n(F)$*

be its Jordan canonical form. Then the following are equivalent.

(1) A_1 and A_2 are similar (i.e., $A_1 \sim A_2$).

(2) We have $J_1' = J_2'$, after re-ordering of their Jordan blocks.

Proof. If the representation matrix of T relative to a basis

$$B = (B_1, \ldots, B_a, \ldots, B_b, \ldots B_k)$$

of V, is given by

$$[T]_B = \mathrm{Diag}[J_1, \ldots, J_a, \ldots, J_b, \ldots, J_k]$$

where J_i's are Jordan blocks, then under the new basis

$$B' := (B_1, \ldots, B_b, \ldots, B_a, \ldots, B_k)$$

(with B_a and B_b switched), we have

$$[T]_{B'} = \mathrm{Diag}[J_1, \ldots, J_b, \ldots, J_a, \ldots, J_k]$$

(with the Jodan blocks J_a and J_b switched).

These two Jordan canonical forms $[T]_B$ and $[T]_{B'}$ are similar matrices (cf. Basis change theorem 8.20).

This argument shows that: if (2) is true then $J'_1 \sim J'_2$ (which is proven in the special case with $J'_1 = \mathrm{Diag}[J_1, \ldots, J_a, \ldots, J_b, \ldots, J_k]$ and $J'_2 = \mathrm{Diag}[J_1, \ldots, J_b, \ldots,$ while the general case can be reduced to the composite of such special cases) and hence

$$A_1 \sim J'_1 \sim J'_2 \sim A_2$$

which implies (1) (Exercise!).

Conversely, if (1) is true, then the proof of Basis change theorem 8.20 shows that A_1 and A_2 are representation matrices of the same linear operator T on F_c^n, relative two different bases, i.e., $A_i = [T]_{B_i}$. By the same reasoning, $A_i \sim J'_i$ implies that $J'_i = [T]_{C_i}$. Thus both J'_i are Jordan canonical forms of the same T relative to some new bases C_i (cf. Theorem 10.16). Hence (2) is true by the uniqueness of the canonical forms (up to ordering) in Theorem 10.14. $\square$

Definition 10.21. (**Diagonalizable operator**). Let V be an n-dimensional vector space over a field F. A linear operator

$$T : V \to V$$

is **diagonalizable over F** or simply **diagonalizable**, if the representation matrix $[T]_B$ relative to some basis B of V is a diagonal matrix in $M_n(F)$:

$$[T]_B = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix} =: J \in M_n(F)$$

where λ_i are scalars in F. This J is then automatically a Jordan canonical form of T.

Clearly,

$$\lambda_1, \ldots, \lambda_n$$

exhaust all zeros of $p_T(x)$ and the characteristic polynomial of T is

$$p_T(x) = (x - \lambda_1) \cdots (x - \lambda_n).$$

A square matrix $A \in M_n(F)$ is **diagonalizable over F** or simply **diagonalizable**, if A is similar to a diagonal matrix in $M_n(F)$, i.e.,

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix} =: J \ \in \ M_n(F) \quad (*).$$

for some invertible $P \in M_n(F)$, where λ_i are scalars in F. This J is then automatically a Jordan canonical form of A.

Write

$$P = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$$

with $\mathbf{p}_j$ the j-th column of P.

The diagonalizability condition (*) on A is equivalent to

$$AP = P \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix},$$

i.e.,

$$\begin{aligned} (A\mathbf{p}_1, \ \ldots, \ A\mathbf{p}_n) &= A(\mathbf{p}_1, \ldots, \mathbf{p}_n) \\ &= (\lambda_1\mathbf{p}_1, \ \ldots, \ \lambda_n\mathbf{p}_n) \end{aligned}$$

i.e.,

$$A\mathbf{p}_1 = \lambda_1\mathbf{p}_1, \ \ldots, \ A\mathbf{p}_n = \lambda\mathbf{p}_n, \quad (**)$$

i.e., each $\mathbf{p}_i$ is an eigenvector of A corresponding to the eigenvalue λ_i.

Suppose that $A = [T]_{B'}$. Then the condition (**) above is equivalent to (cf. Definition 8.7):

$$[T(\mathbf{v}_i)]_{B'} = [T]_{B'}[\mathbf{v}_i]_{B'} = \lambda_i[\mathbf{v}_i]_{B'} \quad (\forall\, i)$$

where $\mathbf{v}_i := B'\,\mathbf{p}_i \in V$ with (cf. coordinate Definition 8.4):

$$[\mathbf{v}_i]_{B'} = \mathbf{p}_i,$$

i.e. (cf. Theorem 9.3)

$$T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i \quad (\forall\, i),$$

i.e., each $\mathbf{v}_i$ is an eigenvector of T corresponding to the eigenvalue λ_i, i.e.,

$$\begin{aligned} T(\mathbf{v}_1, \ldots, \mathbf{v}_n) &= (T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)) \\ &= (\mathbf{v}_1, \ldots, \mathbf{v}_n) \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \end{aligned}$$

i.e. (cf. Definition 8.7):

$$[T]_B = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

where

$$B := (\mathbf{v}_1, \ldots, \mathbf{v}_n) = (B'\,\mathbf{p}_1, \ldots, B'\,\mathbf{p}_n) = B'\,P$$

is a basis of V since

$$([\mathbf{v}_1]_{B'}, \ldots, [\mathbf{v}_n]_{B'}) = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$$

is a basis of the column vector space F_c^n (cf. Exercises 8.5 and 7.42).

Exercise 10.22. (= Tutorial question) (**Diagonal linear operators v.s. matrices**) Use Basis change theorem 8.20 to show:

(1) T is diagonalizable if and only if the representation matrix $[T]_{B'}$ relative to one (and hence

every) basis B' is diagonalizable.

(2) A matrix $A \in M_n(F)$ is diagonalizable if and only if the matrix transformation T_A on the column n-space F_c^n is diagonalizable (cf. also Exercise 8.12).

Summarize the discussion above we get (cf. also Theorem 9.14):

Theorem 10.23. *(**Equivalent diagonalizable conditions**) Let V be an n-dimensional vector space over a field F, and*

$$T : V \to V$$

a linear operator. Then the following are equivalent:

(1) T is diagonalizable over F, i.e., the representation matrix of T relative to some basis B of

V is a diagonal matrix in $M_n(F)$:

$$[T]_B = \mathrm{Diag}[\lambda_1, \ldots, \lambda_n]$$
$$= \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix} \in M_n(F).$$

(2) $[T]_{B'}$ is diagonalizable over F for one (and hence every) basis B' of V, i.e., there is an invertible $P \in M_n(F)$ such that

$$P^{-1}\,[T]_{B'}\,P = \mathrm{Diag}[\lambda_1, \ldots, \lambda_n]$$

for some scalars $\lambda_i \in F$ (automatically being eigenvalues of T).

(3) A basis

$$B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

of V is formed by eigenvectors $\mathbf{v}_i$ of T (corresponding to some eigenvalue λ_i, with $[\mathbf{v}_i]_{B'}$

equal to the $\mathbf{p}_i$ *in* (5) *and* (6)*).*

(4) There are n *linearly independent eigenvectors* $\mathbf{v}_i$ *of* T*.*

(5) For the representation matrix $[T]_{B'}$ *relative to one (and hence every) basis* B' *of* V*, a basis*

$$P := (\mathbf{p}_1, \ldots, \mathbf{p}_n)$$

of the column n*-space* F_c^n *is formed by eigenvectors* $\mathbf{p}_i$ *of* $[T]_{B'}$ *(corresponding to some eigenvalue* λ_i *so that* $P^{-1}\,[T]_{B'}\,P = \mathrm{Diag}[\lambda_1, \ldots, \lambda_n]$*).*

(6) For the representation matrix $[T]_{B'}$ *relative to one (and hence every) basis* B' *of* V*, there are* n *linearly independent eigenvectors* $\mathbf{p}_i$ *of* $[T]_{B'}$*.*

(7) Let

$$\lambda_{m_1}, \ldots, \lambda_{m_k}$$

be the only distinct eigenvalues of T *and let* B_i *be a basis of the eigenspace* $V_{\lambda_i}(T)$. *Then*

$$B = (B_1, \ldots, B_k)$$

is a basis of V, *automatically with* $|B| = \sum_{i=1}^{k} |B_k|$ *and*

$$[T]_B = \begin{pmatrix} \lambda_{m_1} I_{|B_1|} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_{m_2} I_{|B_2|} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_{m_k} I_{|B_k|} \end{pmatrix}.$$

(8) Let

$$\lambda_{m_1}, \ldots, \lambda_{m_k} \in F$$

be the only distinct eigenvalues of T. *Then* V *is a direct sum of the eigenspaces:*

$$V = V_{\lambda_{m_1}}(T) \oplus \cdots \oplus V_{\lambda_{m_k}}(T).$$

(9) Let

$$\lambda_{m_1}, \ldots, \lambda_{m_k} \in F$$

be the only distinct eigenvalues of T. Then

$$\sum_{i=1}^{k} \dim V_{\lambda_i}(T) = \dim V.$$

(10) T has a Jordan canonical form $J \in M_n(F)$, and J is a diagonal matrix.

Theorem 10.24. *(**Minimal polynomial and diagonalizability**). Let F be a field. Let A be a matrix in $M_n(F)$ (or let V be an n-dimensional vector space over F). Let $m(x) := m_A(x)$ (or $m(x) := m_T(x)$) be minimal polynomial of A (or of T). Then the following are equivalent.*

(1) A (or T) is diagonalizable over F.

(2) The minimal polynomial $m(x)$ is a product of distinct linear polynomials in $F[x]$:

$$m(x) = (x - \lambda_1) \cdots (x - \lambda_k)$$

where λ_i are distinct scalars in F.

(3) We can factor $m(x)$ over F as:

$$m(x) = (x - \lambda_1) \cdots (x - \lambda_k)$$

for some scalars $\lambda_i \in F$ and $m(x)$ has only simple zeros (i.e., no multiple zeros).

(4) Let $p(x) = p_A(x)$ (or $p(x) = p_T(x)$) be the characteristic polynomial. Then we can factorize $p(x)$ over F as

$$p(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k}$$

where λ_i are distinct scalars in F. The dimension of the eigenspace satisfies:

$$\dim V_{\lambda_i} = n_i \ \ (\forall\ i)$$

i.e., the geometric multiplicity coincides with the algebraic multiplicity of λ_i for all i.

Proof. We consider T only. For A, you may consider $T := T_A$. The equivalence of (2) and (3) are clear.

Let B be a basis of V such that $J = [T]_B$ is a Jordan canonical form of T (cf. Theorem 10.18 and Remark 10.19). In notation of Definition 10.11, we write (cf. Theorem 10.3))

$$m_T(x) = m_J(x) = \prod_{i=1}^{k} (x - \lambda_i)^{s_e(\lambda_i)}$$

where λ_i exhaust all distinct eigenvalues of J (i.e., of T, cf. Theorem 9.3), and $s_e(\lambda_i)$ is the order of the largest Jordan blocks with eigenvalue λ_i.

First, we show the equivalence of (1) and (2). By Theorem 10.23, T is diagonal, iff

the Jordan canonical form J of T is in $M_n(F)$ and diagonal, iff

$s_e(\lambda_i) = 1$ for all i, by the maximality of $s_e(\lambda_i)$, iff

$m_T(x) = (x-\lambda_1) \cdots (x-\lambda_k)$ where λ_i are distinct scalars in F.

This proves the equivalence of (1) and (2).

By Theorem 9.14, we have a direct sum

$$W := \oplus_{i=1}^{k} V_{\lambda_i}.$$

Next, we show the equivalence of (1) and (4). By Theorem 10.23, (1) holds, iff

$W = V$, iff (cf. Exercise 5.20):

$$(\sum_{i=1}^{k} \dim V_{\lambda_i} =) \ \dim W$$

$$= \dim V \ (= \deg(p(x)) = \sum_{i=1}^{k} n_i)$$

(cf. Definition 8.27), iff

$$\dim V_{\lambda} = n_i \ \ (\forall\ i),$$

since

$$\dim V_{\lambda_i} \ \leq \ n_i,$$

i.e., (Geometric multiplicity of λ_i) $\leq$ (Algebraic multiplicity of λ_i) (cf. Remark 10.19).

This proves the equivalene of (1) and (4). $\square$

Remark 10.25. (**Finding P to diagonalize A**) Suppose that $A \in M_n(F)$ is diagonalizable over F. Write

$$p_A(x) = (x - \lambda_1)^{n_1} \dots (x - \lambda_k)^{n_k}$$

with $\lambda_i \in F$ all distinct. Take any basis

$$B_i = (\mathbf{w}_{i1}, \mathbf{w}_{i2}, \dots)$$

of the eigenspace $V_{\lambda_i}(A)$ (a subspace of F_c^n). Set

$$P = (B_1, \dots, B_k) \in M_n(F)$$

(regarded as a square matrix). Then the discussion preceding Theorem 10.23 implies that

$$AB_i = \lambda_i B_i, \quad AB = BJ$$

or equivalently

$$AP = PJ$$

(if you regard B as P, the collection of column vectors in B, i.e., a matrix), or equivalently

$$P^{-1}AP = J$$

where J is the Jordan canonical form:

$$J := \begin{pmatrix} \lambda_1 I_{n_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 I_{n_2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_k I_{n_k} \end{pmatrix}.$$

Remark 10.26. (**Finding a basis to diagonalize T**) Let V be an n-dimensional vector space over a field F with a basis B'. Let

$$T : V \to V$$

be a linear operator on V. Set

$$A := [T]_{B'}.$$

Assume that T is diagonalizable (over F), i.e., A is diagonalizable (cf. Theorem 10.23).

We use the notation in Remark 10.25 and also write (cf. Definition 8.27):

$$\begin{aligned} p_T(x) &= p_A(x) \\ &= (x - \lambda_1)^{n_1} \dots (x - \lambda_k)^{n_k}; \end{aligned}$$

by Remark 10.25, $P^{-1}\, A\, P$ is equal to some diagonal matrix J there.

By the discussion in Definition 10.21, $[T]_B$ is equal to this diagonal

$$J = \begin{pmatrix} \lambda_1 I_{n_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 I_{n_2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_k I_{n_k} \end{pmatrix},$$

where

$$B := (B'\, \mathbf{p}_1, \dots, B'\, \mathbf{p}_n) = B'\, P$$

is the basis of V.

Example 10.27. (**Finding P for some $A \in M_3(F)$**) For the matrix A in Example 9.11 or 10.4, its characteristic polynomial is

$$(x-\lambda_1)^{n_1}\dots(x-\lambda_k)^{n_k} = p_A(x) = (x-1)(x-2)^2$$

where

$$k=2,\ \lambda_1=1,\ \lambda_2=2$$

$$n_1=1,\ n_2=2,$$

and the minimal polynomial factors as

$$m_A(x) = (x-\lambda_1)(x-\lambda_2)$$

and has only simple zeros. So A is diagonalizable over F by Theorem 10.24.

Indeed, to diagonalize A, we just need to find the eigenspaces and their bases (cf. Theorem 10.23) which has been done in Exercise 9.11:

$$V_{\lambda_1}(A) = \mathrm{Span}(B_1)$$

where

$$B_1 = \{\mathbf{p}_1 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}\},$$

$$V_{\lambda_2}(A) = \mathrm{Span}(B_2)$$

where

$$B_2 = \{\mathbf{p}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{p}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}\}.$$

Following Remark 10.25, set

$$\begin{aligned} P := (B_1, B_2) &= (\mathbf{p}_1,\ \mathbf{p}_2,\ \mathbf{p}_3) \\ &= \begin{pmatrix} -2 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Then:

$$(*) \quad P^{-1}\,A\,P = \begin{pmatrix} \lambda_1 I_{n_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 I_{n_2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_k I_{n_k} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 I_{n_1} & 0 \\ 0 & \lambda_2 I_{n_2} \end{pmatrix}$$

$$= \begin{pmatrix} 1\,I_1 & 0 \\ 0 & 2I_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Remark. You do not have to compute $P^{-1}\,A\,P$ directly. The (*) above should be regarded as a formula, so long the columns in the matrix P are put in orderly: basis of V_{λ_1}, basis of V_{λ_2}, $\cdots$, basis of V_{λ_k}.

Example 10.28. For the linear operator T on the subspace $V = P_3[x]$ of $F[x]$ in Example 9.12, we note that the representation matrix $[T]_{B'}$ relative to the standard basis

$$B' = (1, x, x^2)$$

of V, is equal to the A in Example 10.27 or 9.11. So A and hence T are diagonalizable over F.

First, $P^{-1}AP$ is equal to the J in Example 10.27, where

$$P = (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$$

is given there.

Following Remark 10.26, we set

$$\mathbf{v}_1 = B'\,\mathbf{p}_1 = (1,\ x,\ x^2)\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = -2 + x + x^2,$$

$$\mathbf{v}_2 = B'\,\mathbf{p}_2 = (1,\ x,\ x^2)\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x,$$

$$\mathbf{v}_3 = B'\,\mathbf{p}_3 = (1,\ x,\ x^2)\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = -1 + x^2.$$

By Remark 10.26, if we choose a new basis

$$B := B'P = (B'\mathbf{p}_1, B'\mathbf{p}_2, B'\mathbf{p}_3) = (\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3)$$

of $V = P_3[x]$, then $[T]_B$ is equal to the J in Example 10.27, i.e.,

$$[T]_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Extra exercise 10.29. Let V be an n-dimensional vector space over a field F.

A linear operator

$$T_1 : V \to V$$

on V is **nilpotent** if

$$T_1^m = 0\, I_V$$

for some positive integer m.

Show that the derivative linear operator

$$\begin{aligned} D : P_n[x] &\to P_n[x] \\ f(x) &\mapsto \frac{df(x)}{dx} \end{aligned}$$

on the subspace $V = P_n[x]$ (of all polynomials of degree $< n$) of $F[x]$, is nilpotent.

Suppose that T_1 has a Jordan canonical form $J \in M_n(F)$. Show that the following are equivalent (cf. Tutorial question):

(1) T_1 is nilpotent.

(2) J equals some $A(\lambda)$ with $\lambda = 0$ in Definition 10.11.

(3) Every eigenvalue of T_1 is zero.

(4) The characteristic polynomial of T_1 is: $p_{T_1}(x) = x^n$.

(5) The minimal polynomial of T_1 is: $m_{T_1}(x) = x^s$ for some $s \geq 1$.

A linear operator T_2 on V is **semi-simple** if it is diagonalizable over F.

Suppose that a linear operator

$$T : V \to V$$

has a Jordan canonical form in $M_n(F)$. Show that there are linear operators

$$T_s : V \to V$$

and

$$T_n : V \to V$$

satisfying the following:

(1) A decomposition (cf. Exercise 7.32):

$$T = T_s + T_n$$

(2) T_s is semi-simple, i.e., it is diagonalizable over F.

(3) T_n is nilpotent.

(4) The commutativity:

$$T_s \circ T_n = T_n \circ T_s.$$

(5) (hard!) There are polynomials $f(x)$, $g(x)$ in $F[x]$ such that

$$T_s = f(T), \quad T_n = g(T).$$

In particular, if another linear operator

$$S : V \to V$$

commutes with T then S commutes also with both T_s and T_n.

Show that the above decomposition

$$T = T_s + T_n$$

is unique: if $T = T_s' + T_n'$ satisfying the conditions (2) - (4) above, then $T_s' = T_s$ and $T_n' = T_n$.

We call the above decomposition $T = T_s + T_n$ the **(additive) Jordan decomposition.**

Hint for the proof. Reduce to the case where $T = T_A$ for some matrix. Let $J \in M_n(F)$ be a Jordan canonical form of A. So there is an invertible $P \in M_n(F)$ such that $P^{-1} A P = J$. We can write $J =$

$J_s + J_n$ where

$$J_s = \begin{pmatrix} \lambda_1 I_{n_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 I_{n_2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_k I_{n_k} \end{pmatrix}$$

and J_n equals some Jordan canonical form $A(\lambda)$ with $\lambda = 0$ in Definition 10.11. Indeed, to get J_s, just keep all the entries on the diagonal of J and throw away the $1's$ lying above the diagonal. Keeping those $1's$ lying above the diagonal of J and throwing away all the entries on the diagonal of J, we get J_n.

You can check that

$$J = J_s + J_n$$

satisfies the conditions (2) - (4) above. Now simply put

$$A_s := PJ_sP^{-1}, \ \ A_n := PJ_nP^{-1}.$$

Then

$$\begin{aligned} A_s + A_n &= P(J_s + J_n)P^{-1} \\ &= PJP^{-1} = A \end{aligned}$$

and the decomposition

$$A = A_s + A_n$$

satisfies the conditions (2) - (4) above.

11 Quadratic forms, Inner product spaces and Conics

In this section, we discuss solutions of differential equations via Jordan canonical forms, bilinear forms, quadratic forms, Principal axis theorem for symmetric matrix or adjoint matrix (or linear operator), and the classification of conics in $\mathbf{R}^2$.

Example 11.1. (**Solve differential equations**). Let

$$y_i = y_i(x) \ \ (i = 1, 2, 3)$$

be differentiable functions in x. Set

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \ Y' = \begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix}.$$

We will solve the following differential equations, where $A = (a_{ij})$ is in $M_3(\mathbf{R})$:

$$\begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = Y' = AY = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

or equivalently

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3 \\ y_2' &= a_{21}y_1 + a_{22}y_2 + a_{23}y_3 \\ y_3' &= a_{31}y_1 + a_{32}y_2 + a_{33}y_3. \end{aligned}$$

Let P be an invertible matrix in $M_3(\mathbf{R})$ such that

$$J := P^{-1} A\, P$$

is a Jordan canonical form of A. Write

$$Y = PZ = P \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

i.e.,

$$Z = P^{-1} Y.$$

Set

$$Z' = \begin{pmatrix} z_1' \\ z_2' \\ z_3' \end{pmatrix}.$$

Since the entries of $P = (p_{ij})$ are all constants in $\mathbf{R}$, we have

$$PZ' = (PZ)' = Y' = AY = APZ.$$

Multiplying the equation by P^{-1} on the left, we get

$$Z' = P^{-1}APZ = JZ.$$

If we can solve the simpler equation

$$Z' = J\,Z$$

then

$$Y = P\,Z$$

will be the solution to the original equation

$$Y' = A\,Y.$$

For instance, suppose the Jordan canonical form J of A is given by

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

Then

$$\begin{pmatrix} z_1' \\ z_2' \\ z_3' \end{pmatrix} = Z' = JZ$$

$$= \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} \lambda z_1 + z_2 \\ \lambda z_2 + z_3 \\ \lambda z_3. \end{pmatrix},$$

or equivalently

$$z_1' = \lambda z_1 + z_2 \qquad (1)$$

$$z_2' = \lambda z_2 + z_3 \qquad (2)$$

$$z_3' = \lambda z_3. \qquad (3)$$

We apply the following known procedure of solving differential equation of first order (cf. e.g. Howard Anton [Calculus, §7.7, First-order differential equations and applications]):

$$z'(x) + p(x)\,z = q(x),$$
$$\mu := e^{\int p(x)\,dx},$$
$$z = \frac{1}{\mu}\Big(\int \mu\, q(x)\,dx + C\Big).$$

Applying the above formula and solving the equation(3) above (with $p = -\lambda$, $q = 0$), we obtain:

$$z_3 = c_1\, e^{\lambda x}.$$

Substituing this z_3 into the equation(2) above (with $p = -\lambda$, $q = z_3 = c_1 e^{\lambda x}$), we obtain:

$$z_2 = (c_1 x + c_2)\, e^{\lambda x}.$$

Substituing this z_2 into the equation(1) above (with $p = -\lambda$, $q = z_2 = (c_1 x + c_2)e^{\lambda x}$), we obtain:

$$z_1 = (\frac{1}{2}c_1 x^2 + c_2 x + c_3)\, e^{\lambda x}.$$

Now the solution Y to the original equation $Y' = AY$ is as follows, where c_1, c_3, c_3 are constants:

$$Y = PZ = P \begin{pmatrix} (\frac{1}{2}\, c_1\, x^2 \,+\, c_2\, x \,+\, c_3)\, e^{\lambda x} \\ (c_1\, x \,+\, c_2)\, e^{\lambda x} \\ c_1\, e^{\lambda x} \end{pmatrix}.$$

Definition 11.2. (**Bilnear forms**) Let V be a vector space over a field F. Consider the map (or function) H below:

$$H : V \times V \;\to\; F,$$
$$(\mathbf{x}, \mathbf{y}) \mapsto H(\mathbf{x}, \mathbf{y}).$$

(1) H is called a **bilinear form on V** if H is linear in both variables, i.e., for all

$$\mathbf{x}_i,\ \mathbf{y}_j,\ \mathbf{x},\ \mathbf{y} \in V,\ a_i,\ b_i \in F$$

we have:

$$H(a_1\mathbf{x}_1 + a_2\mathbf{x}_2,\ \mathbf{y}) = a_1 H(\mathbf{x}_1,\ \mathbf{y}) + a_2 H(\mathbf{x}_2,\ \mathbf{y}),$$
$$H(\mathbf{x},\ b_1\mathbf{y}_1 + b_2\mathbf{y}_2) = b_1 H(\mathbf{x},\ \mathbf{y}_1) + b_2 H(\mathbf{x},\ \mathbf{y}_2).$$

(2) A bilinear form H on V is **symmetric** if

$$H(\mathbf{x}, \mathbf{y}) = H(\mathbf{y}, \mathbf{x}) \ \ (\forall\ \mathbf{x}, \mathbf{y} \in V).$$

Exercise 11.3. (**Representation matrix**) Suppose that

$$B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

is a basis of a vector space V over a field F. Let

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

be a matrix in $M_n(F)$.

We define the function

$$H_A : V \times V \ \to \ F$$
$$(\sum_{i=1}^n x_i\mathbf{v}_i, \sum_{j=1}^n y_j\mathbf{v}_j) \mapsto H_A(\sum_{i=1}^n x_i\mathbf{v}_i, \sum_{j=1}^n y_j\mathbf{v}_j)$$

where

$$
\begin{aligned}
& H_A\Big(\sum_{i=1}^{n} x_i\mathbf{v}_i, \sum_{j=1}^{n} y_j\mathbf{v}_j\Big) \\
:= & \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_j \\
= & (x_1, \ldots, x_n)\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
= & X^t\, A\, Y
\end{aligned}
$$

with

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

(1) Then H_A is a bilinear form on V and called the **bilinear form associated with A** (and relative to the basis B of V).

(2) Conversely, every bilinear form H on V is of the

form H_A for some A in $M_n(F)$. Indeed, just set

$$(*) \quad a_{ij} := H(\mathbf{v}_i, \mathbf{v}_j), \quad A := (a_{ij}).$$

Then one can use the bilinearity of H, to show that $H = H_A$.

The matrix A is called the **representation matrix of H relative to the basis $B = (\mathbf{v}_1, \dots, \mathbf{v}_n)$**.

(3) Show that H_A is a symmetric bilinear form if and only if A is a **symmetric matrix**, i.e., the **transpose** A^t equals A:

$$A^t = A.$$

Hint. Use the relation (*) in (2).

Example 11.4. (**Bilinear forms on F_c^n**) Let

$A = (a_{ij}) \in M_n(F)$. Then the map on $V := F_c^n$:

$$H : V \times V \rightarrow F$$
$$(X, Y) \mapsto X^t\, A\, Y$$

is a bilinear form. Here, as usual,

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

are column vectors in $V = F_c^n$.

The representation matrix of H relative to the standard basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_n)$ of $V = F_c^n$ is just A. Thus $H = H_A$.

Definition 11.5. (**Non-degenerate bilinear forms**)

A bilinear form H on V is **non-degenerate** if for every $\mathbf{y}_0 \in V$ we have:

$$H(\mathbf{x}, \mathbf{y}_0) = 0 \; (\forall\, \mathbf{x} \in V) \;\Rightarrow\; \mathbf{y}_0 = \mathbf{0}.$$

One can verify that a bilinear form $H = H_A$ is non-degenerate if and only if its representation matrix A (relative to one and hence all bases of V) is invertible.

Example 11.6. (**A non-symmetric bilinear form**) Consider the map on $V = F_c^2$:

$$\begin{aligned} H : V \times V &\to F \\ \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) &\mapsto 2x_1y_1 + 3x_1y_2 + 4x_2y_1 - x_2y_2. \end{aligned}$$

Then H is a bilinar form on V, and its representation matrix relative to the standard basis of V, is

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

so that $H = H_A$.

This H is not symmetric since its representation matrix A is not symmetric.

Since the determinant $|A| = -14$, our H is non-degenerate when $F = \mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$ (or when the so called **characteristic** of the field F is not 2 or 7, so that $-14 \neq 0_F$).

Example 11.7. (**A symmetric bilinear form**)
Consider the map on $V = F_c^2$:

$$H : V \times V \ \to \ F,$$
$$(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}) \mapsto x_1 y_2 + x_2 y_1.$$

Then H is a bilinar form on $V = F_c^2$, and its representation matrix relative to the standard basis of $V = F_c^2$ is

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

so that $H = H_A$.

This H is symmetric since its representation matrix A is symmetric. It is also non-degenerate since the determinant

$$|A| = -1 \neq 0_F.$$

Definition 11.8. (**Congruent matrices**) Two matrices A and B in $M_n(F)$ are **congruent** if there is an invertible matrix $P \in M_n(F)$ such that

$$B = P^t AP.$$

Verify that being congruent is an equivalent relation (Exercise!).

Consider the bilinear form

$$\begin{aligned} H : F_c^n \times F_c^n &\rightarrow F \\ (X, Y) &\mapsto X^t\, A\, Y \end{aligned}$$

in Example 11.4, where $A = (a_{ij}) \in M_n(F)$.

If we write

$$X = P\,Y$$

with an invertible matrix $P \in M_n(F)$ and introduce

Y as a new coordinate system for F_c^n, then

$$\begin{aligned} H(X_1, X_2) &= X_1^t\, A\, X_2 \\ &= (PY_1)^t\, A\,(PY_2) \\ &= Y_1^t\,(P^t A P)\, Y_2 \end{aligned}$$

where we have done the substitution

$$X_i := P\,Y_i.$$

Thus, the bilinear form above would have a simpler form in new coordinates Y, if $P^t AP$ (which is congruent to A) is simpler. This simplification is very useful in classifying all conics (cf. Theorem 11.41).

To be precise, we have the following theorem.

Theorem 11.9. *(**Weak version of Principal axis theorem**) Let $A \in M_n(F)$ be a symmetric matrix. Assume that $1 + 1 \neq 0$ in F (e.g., $F = \mathbf{Q}, \mathbf{R}, \mathbf{C}$). Then there is an invertible matrix P in $M_n(F)$ such that the matrix $P^t\, A\, P$ is* *diagonal:*

$$P^t\,AP = \mathrm{Diag}[d_1, \ldots, d_n] =: D,$$

i.e., A is congruent to a diagonal matrix D.

In this case, as calculated in Definition 11.8*, the bilinear form*

$$H(X_1, X_2) = X_1^t\,A\,X_2$$

(in coordinates X) on $V = F_c^n$ has a simpler

form (in coordinates Y):

$$\begin{aligned}
H(X_1, X_2) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i y_j \\
&= X_1^t\, A\, X_2 \\
&= Y_1^t\, D\, Y_2 \\
&= (y_{11}, .., y_{1n}) \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \vdots & d_n \end{pmatrix} \begin{pmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2n} \end{pmatrix} \\
&= d_1 y_{11} y_{21} + d_2 y_{12} y_{22} + \cdots + d_n y_{1n} y_{2n} \\
&= \sum_{j=1}^{n} d_j y_{1j} y_{2j}
\end{aligned}$$

where we have used the substitution :

$$X_i = P\,Y_i, \quad Y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix}.$$

Proof. We omit the proof, since we will introduce a stronger version in Theorem 11.33. □

Definition 11.10. (**Inner product, Orthogonal, Norm**)

We start with the real version.

Consider a function H:

$$H : V \times V \to \mathbf{R},$$
$$(\mathbf{x}, \mathbf{y}) \mapsto H(\mathbf{x}, \mathbf{y})$$

on a vector space V over the field $\mathbb{R}$ of real numbers.

The function H is called a **real inner product** (or simply an **inner product**) and V a **real inner product space** (or simply an **inner product space**), if the following three conditions are satisfied, where we denote

$$\langle \mathbf{x}, \mathbf{y} \rangle := H(\mathbf{x}, \mathbf{y}).$$

(1) H is a bilinear form, i.e., for all

$$\mathbf{x}_i,\ \mathbf{y}_j,\ \mathbf{x},\ \mathbf{y} \in V,\ a_i,\ b_i \in C$$

we have

$$\langle a_1\mathbf{x}_1 + a_2\mathbf{x}_2, \mathbf{y}\rangle = a_1\langle \mathbf{x}_1, \mathbf{y}\rangle + a_2\langle \mathbf{x}_2, \mathbf{y}\rangle,$$

$$\langle \mathbf{x}, b_1\mathbf{y}_1 + b_2\mathbf{y}_2\rangle = b_1\langle \mathbf{x}, \mathbf{y}_1\rangle + b_2\langle \mathbf{x}, \mathbf{y}_2\rangle$$

(2) H is symmetric, i.e., for all $\mathbf{x}$, $\mathbf{y} \in V$, we have

$$\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle.$$

(3) The positivity:

for all $\mathbf{0} \neq \mathbf{x} \in V$, we have

$$\langle \mathbf{x}, \mathbf{x}\rangle > 0.$$

Next is the complex version.

For a complex number

$$\alpha = r_1 + r_2\sqrt{-1} \in \mathbb{C}$$

with $r_i \in \mathbb{R}$, we denote by

$$\overline{\alpha} := r_1 - r_2\sqrt{-1}$$

the **conjugate** of α. The **modulus** of α is defined and denoted as

$$|\alpha| = \sqrt{r_1^2 + r_2^2}.$$

Remark.

$$|\alpha|^2 = \alpha\,\overline{\alpha}.$$

Exercise. Can you prove that

$$\mathbb{C} \to \mathbb{C}$$

$$\alpha \mapsto \overline{\alpha}$$

is an isomorphism from $\mathbb{C}$ to itself, regarded as a 2-dimensional vector space over $\mathbb{R}$ (with $(1, \sqrt{-1})$ a basis)?

Consider a function H:

$$H : V \times V \to \mathbf{C},$$

$$(\mathbf{x}, \mathbf{y}) \mapsto H(\mathbf{x}, \mathbf{y})$$

on a vector space V over the field $\mathbb{C}$ of complex numbers.

The function H is called a **complex inner product** (or simply an **inner product**) and V a **complex inner product space** (or simply an **inner product space**), if the following three conditions are satisfied, where we denote

$$\langle \mathbf{x},\, \mathbf{y}\rangle := H(\mathbf{x}, \mathbf{y}).$$

(1) For all

$$\mathbf{x}_i,\, \mathbf{y}_j,\, \mathbf{x},\, \mathbf{y} \in V,\, a_i,\, b_i \in \mathbb{C}$$

we have

$$\langle a_1\mathbf{x}_1 + a_2\mathbf{x}_2, \mathbf{y}\rangle = a_1\langle \mathbf{x}_1,\, \mathbf{y}\rangle + a_2\langle \mathbf{x}_2, \mathbf{y}\rangle,$$

$$\langle \mathbf{x},\, b_1\mathbf{y}_1 + b_2\mathbf{y}_2\rangle = \overline{b_1}\langle \mathbf{x}, \mathbf{y}_1\rangle + \overline{b_2}\langle \mathbf{x},\, \mathbf{y}_2\rangle$$

(2) For all $\mathbf{x},\, \mathbf{y} \in V$, we have

$$\langle \mathbf{y},\, \mathbf{x}\rangle = \overline{\langle \mathbf{x},\, \mathbf{y}\rangle}.$$

(3) The positivity:

for all $\mathbf{0} \neq \mathbf{x} \in V$, we have

$$\langle \mathbf{x},\, \mathbf{x}\rangle > 0.$$

On a real or complex inner product space V, there are three more definitions:

(1) The **norm** of a vector $\mathbf{x} \in V$ is denoted and defined as:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

One can verify that the

$$\|\mathbf{x}\| \geq 0$$

and

$$\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}_V.$$

(2) Two vectors $\mathbf{x}$, $\mathbf{y}$ in V are **orthogonal** to each other and denoted as

$$\mathbf{x} \perp \mathbf{y}$$

if their inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0.$$

(3) Sometimes, we use

$$(V, \langle , \rangle)$$

to denote a vector space V with an inner product $\langle , \rangle$.

Exercise 11.11. (**Inner product is non-degenerate**) Let $(V.\langle , \rangle)$ be an inner product space over a field F with $F = \mathbb{R}$ or $F = \mathbb{C}$. Then the product $\langle , \rangle$ is **non-degenerate** in the sense:

for every $\mathbf{u}_0 \in V$,

$$\langle \mathbf{u}_0, \mathbf{y} \rangle = 0 \ (\forall \ \mathbf{y} \ \in \ V) \ \Rightarrow \ \mathbf{u}_0 = \mathbf{0}_V$$

and for every $\mathbf{v}_0 \in V$,

$$\langle \mathbf{x}, \mathbf{v}_0 \rangle = 0 \ (\forall \ \mathbf{x} \ \in \ V) \ \Rightarrow \ \mathbf{v}_0 = \mathbf{0}_V.$$

Definition 11.12. (**Orthonormal basis**) Let $(V, \langle , \rangle)$ be a real or complex inner product space. A basis $B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is called an **orthonormal basis**

of the inner product space V (relative to the inner product $\langle , \rangle$) if it satisfies the following two conditions:

(1) **Orthogonality**:

for all $i \neq j$, we have:

$$\mathbf{v}_i \perp \mathbf{v}_j \quad \text{i.e.,} \quad \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0.$$

(2) **Normalized**:

for all i, we have

$$\|\mathbf{v}_i\| = 1 \quad \text{i.e.,} \quad \langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1.$$

Namely, $\mathbf{v}_i$ is a **unit** vector.

Example 11.13. (**Standard inner product on the column n-space $\mathbf{R}_c^n$ and $\mathbf{C}_c^n$**)

First is the real version.

Setting $V := \mathbb{R}_c^n$ and $A = I_n$ in Example 11.4, the bilinear form

$$\begin{aligned} H_{I_n} : V \times V &\to \mathbb{R} \\ (X, Y) &\mapsto \langle X, Y \rangle := H_{I_n}(X, Y) \\ &= X^t\, Y \\ &= (x_1, \ldots, x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= x_1 y_1 + \cdots + x_n y_n \end{aligned}$$

on V is an inner product and called the **standard inner product on V**.

We call this pair $(V, \langle\,\rangle)$ or simply $V = \mathbb{R}^n_c$, the **inner product space with a standard inner product $\langle\,\rangle$**.

Next is the complex version.

For complex column vector

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{C}_c^n$$

or row vector

$$X^t = (x_1, \ldots, x_n) \in \mathbb{C}^n$$

we define their **conjugates** as

$$\overline{X} = \begin{pmatrix} \overline{x_1} \\ \vdots \\ \overline{x_n} \end{pmatrix} \in \mathbb{C}_c^n$$

$$\overline{X^t} = (\overline{x_1}, \ldots, \overline{x_n}) \in \mathbb{C}^n.$$

Setting $V := \mathbb{C}_c^n$, the map

$$
\begin{aligned}
H : V \times V &\to \mathbb{C} \\
(X, Y) &\mapsto \langle X, Y \rangle := X^t\, \overline{Y} \\
&= (x_1, \ldots, x_n) \begin{pmatrix} \overline{y_1} \\ \vdots \\ \overline{y_n} \end{pmatrix} \\
&= x_1\overline{y_1} + \cdots + x_n\overline{y_n}
\end{aligned}
$$

on V is an inner product and called the **standard inner product on V**.

We call this pair $(V, \langle , \rangle)$ or simply $V = \mathbb{C}_c^n$, the **inner product space with a standard inner product $\langle , \rangle$**.

(1) For the standard inner product space $(V, \langle , \rangle)$ with $V = \mathbb{R}_c^n$ or $V = \mathbb{C}_c^n$, one can verify that

the standard basis of V:

$$B = (\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix})$$

is an orthonormal basis (relative to the standard inner product of V).

More generally, let $s_1, \ldots, s_n$ be integers such that

$$\{s_1, \ldots, s_n\} = \{1, \ldots, n\}$$

as set. Then the new basis

$$B' := \{\mathbf{e}_{s_1}, \ldots, \mathbf{e}_{s_n}\}$$

is also an orthonormal basis of the standard inner product space V.

For instance,

$$B' := \{\mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_2\}$$

is an orthonormal basis of $V = \mathbb{R}^3_c$ or $V = \mathbb{C}^3_c$.

(2) If W is a subspace of V, then the **restriction** of the inner product $\langle , \rangle$ to W gives W an inner product structure, simply denoted as

$$(W, \langle , \rangle).$$

Example 11.14. (**A non-standard orthonormal basis of F^2_c**)

$$(\mathbf{v}_1 = \begin{pmatrix} \frac{3}{5} \\ \\ \\ \frac{4}{5} \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} \frac{4}{5} \\ \\ \\ \frac{-3}{5} \end{pmatrix})$$

is a (non-standard) orthonormal basis of the standard inner product space $V = \mathbb{R}^2_c$ or $V = \mathbb{C}^2_c$.

Indeed, for both $V = \mathbb{R}^2_c$ and $V = \mathbb{C}^2_c$, one can calculate the standard inner products (noting that

$\overline{\mathbf{v}_i} = \mathbf{v}_i$ now, even when $V = \mathbb{C}_c^2$):

$$\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \frac{3}{5} \times \frac{3}{5} + \frac{4}{5} \times \frac{4}{5} = 1,$$
$$\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \frac{3}{5} \times \frac{4}{5} + \frac{4}{5} \times \frac{-3}{5} = 0,$$
$$\langle \mathbf{v}_2, \mathbf{v}_2 \rangle = \frac{4}{5} \times \frac{4}{5} + \frac{-3}{5} \times \frac{-3}{5} = 1.$$

A standard inner product space (of dimension $\geq$ 2) have many (indeed infinitely many) different orthonormal bases.

11.15. (**Gram-Schmidt process**)

Let $V = F_c^n$ with $F = \mathbb{R}_c^n$ or $F = \mathbb{C}_c^n$. Let us employ the standard inner product $\langle , \rangle$ for V.

Let

$$(\mathbf{u}_1, \ldots, \mathbf{u}_r)$$

be a basis of a subspace W of V. Then one can apply the following **Gram-Schmidt process** to get an orthonormal basis

$$(\mathbf{v}_1, \ldots, \mathbf{v}_r)$$

of W (relative to the standard inner product of V):

$$\begin{aligned}
\widetilde{\mathbf{v}}_1 &:= \mathbf{u}_1, \\
\widetilde{\mathbf{v}}_2 &:= \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \widetilde{\mathbf{v}}_1 \rangle}{\|\widetilde{\mathbf{v}}_1\|^2} \widetilde{\mathbf{v}}_1, \\
\widetilde{\mathbf{v}}_3 &:= \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \widetilde{\mathbf{v}}_1 \rangle}{\|\widetilde{\mathbf{v}}_1\|^2} \widetilde{\mathbf{v}}_1 - \frac{\langle \mathbf{u}_3, \widetilde{\mathbf{v}}_2 \rangle}{\|\widetilde{\mathbf{v}}_2\|^2} \widetilde{\mathbf{v}}_2, \\
\widetilde{\mathbf{v}}_k &:= \mathbf{u}_k - \sum_{i=1}^{k-1} \frac{\langle \mathbf{u}_k, \widetilde{\mathbf{v}}_i \rangle}{\|\widetilde{\mathbf{v}}_i\|^2} \widetilde{\mathbf{v}}_i \qquad (k \geq 3), \\
\mathbf{v}_j &:= \frac{\widetilde{\mathbf{v}}_j}{\|\widetilde{\mathbf{v}}_j\|} \qquad (j \geq 1).
\end{aligned}$$

Definition 11.16. (**Adjoint matrices** A^*)

For a matrix $A = (a_{ij}) \in M_n(\mathbb{C})$, the **adjoint** or **conjugate transpose** of A is defined and denoted as:

$$A^* = (\overline{A})^t = (\overline{a_{ij}})^t$$

i.e., the (i, j)-entry of A^* equals $\overline{a_{ji}}$ (the conjugate of the (j, i)-entry of A). Note that

$$A^* = \overline{(A^t)}.$$

Exercise 11.17. (**Adjoint matrix A^* and inner product**) Let $V = F_c^n$ with $F = \mathbb{R}_c^n$ or $F = \mathbb{C}_c^n$. Let us employ the standard inner product $\langle , \rangle$ for V as in Example 11.13. Show that for a matrix $A \in M_n(F)$ we have

$$\langle AX, Y\rangle = \langle X, A^*Y\rangle \ (\forall\ X, Y \in V).$$

Theorem 11.18. *(**Adjoint linear operator**) Let $T : V \to V$ be a linear operator on an n-dimensional inner product space V over a field F ($F = \mathbf{R}$, or $F = \mathbf{C}$). Then we have:*

(1) There is a **unique** *linear operator*

$$T^* : V \to V$$

on V such that

$$\langle T(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, T^*(\mathbf{v})\rangle \ (\forall\ \mathbf{u}, \mathbf{v} \in V).$$

Such T^* is called the **adjoint linear operator of T.**

(2) Let $B = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ be an orthonormal basis of the inner product space V. Then

$$[T^*]_B = ([T]_B)^*$$

i.e., the representation matrix of the adjoint T^* equals the adjoint of the representation matrix $[T]_B$ (all relative to the same orthonormal basis B).

Proof. For the standard inner product and the matrix linear operator $T = T_A$, Exercise 11.17 says that $T^* = T_{A^*}$.

For the general case, set

$$(a_{ij}) = A := [T]_B.$$

By Definition 8.7

$$A = [T]_B = ([T(\mathbf{w}_1)]_B, .., [T(\mathbf{w}_n)]_B)$$

and

$$(T(\mathbf{w}_1), .., T(\mathbf{w}_n)) = (\mathbf{w}_1, .., \mathbf{w}_n)A.$$

Then for arbitrary

$$\mathbf{u} = \sum x_i \mathbf{w}_i, \ \mathbf{v} = \sum y_j \mathbf{w}_j \ \in \ V$$

we have

$$\begin{aligned}
\langle T(\mathbf{u}), \ \mathbf{v} \rangle &= \langle T(\sum x_i \mathbf{w}_i), \ \sum y_j \mathbf{w}_j \rangle \\
&= \langle \sum x_i T(\mathbf{w}_i), \ \sum y_j \mathbf{w}_j \rangle \\
&= \langle \sum_i x_i \sum_k a_{ki} \mathbf{w}_k, \ \sum_j y_j \mathbf{w}_j \rangle \\
&= \sum_i x_i \sum_k a_{ki} \overline{y_k} \, \langle \mathbf{w}_k, \mathbf{w}_k \rangle \\
&= \sum_i \sum_k \ a_{ki} x_i \overline{y_k} \\
&= X^t \ A^t \, \overline{Y} \\
&= \overline{Y}^t \ A \, X.
\end{aligned}$$

Hence

$$\langle T(\sum x_i \mathbf{w}_i), \ \sum y_j \mathbf{w}_j \rangle = X^t \ A^t \, \overline{Y}. \ \ (*)$$

If we define

$$T^* : V \to V$$

such that

$$[T^*]_B = A^*$$

i.e. (cf. Definition 8.7)

$$(T^*(\mathbf{w}_1), .., T^*(\mathbf{w}_n)) = (\mathbf{w}_1, .., \mathbf{w}_n)\ A^*$$

then a similar calculation shows

$$\begin{aligned}\langle \mathbf{u}, T^*(\mathbf{v})\rangle &= \overline{Y}^t\ A\ X \\ &= \langle T(\mathbf{u}), \mathbf{v}\rangle.\end{aligned}$$

This shows (1) and (2) except for the uniqueness of T^* which is easy to check using Exercise 11.11. □

Exercise 11.19. (**Adjoint of adjoint**) We have

$$(T^*)^* = T.$$

Indeed,

$$\langle T^*(\mathbf{u}), \mathbf{v}\rangle = \overline{\langle \mathbf{v}, T^*(\mathbf{u})\rangle} = \overline{\langle T(\mathbf{v}), \mathbf{u}\rangle} = \langle \mathbf{u}, T(\mathbf{v})\rangle.$$

Hence

$$\langle T^*(\mathbf{u}), \mathbf{v}\rangle = \langle \mathbf{u}, T(\mathbf{v})\rangle \quad (\forall\ \mathbf{u}, \mathbf{v}\ \in\ V).$$

Thus $T = (T^*)^*$ (cf. Theorem 11.18).

Exercise 11.20. (**Adjoint of map linear combination**) Let T, T_i be linear operators on an inner product space V. Let $a_i \in \mathbb{C}$. Use solely the definition of adjoint in Theorem 11.18 to show:

(1) Suppose that $T = \alpha\, I_V$ is a scalar map. Then

$$T^* = \overline{\alpha}\, I_V.$$

(2) (cf. Exercise 7.32)

$$(a_1 T_1 + a_2 T_2)^* = \overline{a_1}\, T_1^* + \overline{a_2}\, T_2^*.$$

(3) Use the definition of adjoint to show:

$$(T_1 \circ T_2)^* = T_2^* \circ T_1^*.$$

Hint. Note that for $A = \alpha\, I_n$ we have $A^* = \overline{\alpha}\, I_n$. Use Theorem 11.18 to reduce to (representation) matrix case.

Definition 11.21. (**Orthogonal, Unitary, Hermitian = Self-adjoint, or Normal linear operators**) Let $A \in M_n(\mathbb{C})$ (or let $T : V \to V$ be a linear operator on an n-dimensional inner product space which is over a field F with $F = \mathbb{R}$ or $F = \mathbb{C}$ and with an **orthonormal basis** B). Let A^* (resp. T^*) be the adjoint of A (resp. T) as in Definition 11.16 (resp. Theorem 11.18).

As usual, for linear operators S, T, we denote

$$S\,T := S \circ T.$$

(1) A linear operator T over a real inner product space is **orthogonal** if

$$T\,T^* = I_V \text{ (or equivalently, } \; T^*\,T = I_V).$$

(2) A real matrix A in $M_n(\mathbf{R})$ is **orthogonal** if

$$A\,A^t = I_n \text{ (or equivalently, } \; A^t\,A = I_n).$$

(3) A linear operator T over a complex inner product space is **unitary** if

$$T\,T^* = I_n \text{ (or equivalently, } \; T^*\,T = I_V).$$

(4) A complex matrix A in $M_n(\mathbf{C})$ is **unitary** if

$$A\,A^* = I_n \text{ (or equivalently, } A^*\,A = I_n).$$

(5) T is **Hermitian** (or **self-adjoint**) if its adjoint T^* equals itself:

$$T^* = T.$$

When the field $F = \mathbf{R}$, a self-adjoint operator is also called a **symmetric** operator.

(6) A complex matrix $A \in M_n(\mathbf{C})$ is **Hermitian** (or **self-adjoint**) if the adjoint matrix A^* of A equals itself:

$$A^* = A.$$

When A is a real matrix in $M_n(\mathbf{R})$, the matrix A is self-adjoint if and only if it is symmetric.

(7) A linear operator T over a complex inner product

space is **normal** if

$$T\,T^* = T^*\,T.$$

(8) A complex matrix $A \in M_n(\mathbb{C})$ is **normal** if

$$A\,A^* = A^*\,A.$$

One sees that orthogonal, unitary and self-adjoint linear operators (or matrices) are normal.

Exercise 11.22. (**Definitions : Linear operators v.s. representation matrices**) Show that T is orthogonal, unitary, self-adjoint or normal if and only if its representation matrix $A := [T]_B$ (relative to one and hence every orthonormal basis B) is respectively orthogonal, unitary, self-adjoint or normal.

Hint. T is orthogonal iff T is unitary and T (and also V) is defined over $F = \mathbb{R}$.

T is unitary, self-adjoint or normal, iff:

$$T\,T^* - I_V = 0\,I_V,$$

$$T - T^* = 0 I_V,$$

$$T\,T^* - T^*\,T = 0\,I_V,$$

respectively, iff (cf. Theorem 8.6 or Example 8.9)

$$[T]_B\,[T^*]_B - I_n = 0\,I_n,$$

$$[T]_B - [T^*]_B = 0\,I_n,$$

$$[T]_B\,[T^*]_B - [T^*]_B\,[T]_B = 0\,I_n,$$

respectively, iff (cf. Theorem 11.18)

$$A\,A^* - I_n = 0\,I_n,$$

$$A - A^* = 0\,I_n,$$

$$A\,A^* - A^*\,A = 0\,I_n,$$

respectively.

Theorem 11.23. *(***Equivalent orthogonal matrix definition***) For a real matrix P in $M_n(\mathbf{R})$, the following are equivalent, where we employ the standard inner product on $\mathbb{R}^n_c$.*

(1) P is orthogonal, i.e., $P\,P^t = I_n$.

(2) Write

$$P = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$$

where the $\mathbf{p}_j$ are the column vectors of P. Then the column vectors $(\mathbf{p}_1, \ldots, \mathbf{p}_n)$ form an orthonormal basis of $\mathbf{R}^n_c$.

(3) The matrix transformation

$$\begin{aligned} T_P : \mathbf{R}^n_c &\to \mathbf{R}^n_c \\ X &\mapsto PX \end{aligned}$$

preserves the standard inner product, i.e., for all X, Y in $\mathbf{R}^n_c$, we have

$$\langle PX,\, PY\rangle = \langle X,\, Y\rangle.$$

(4) The matrix transformation

$$\begin{aligned} T_P : \mathbf{R}^n_c &\to \mathbf{R}^n_c \\ X &\mapsto PX \end{aligned}$$

preserves the distance, i.e., for all X, Y *in* $\mathbf{R}_c^n$*, we have:*

$$\|PX - PY\| = \|X - Y\|.$$

(5) The matrix transformation

$$\begin{aligned} T_P : \mathbf{R}_c^n &\to \mathbf{R}_c^n \\ X &\mapsto PX \end{aligned}$$

preserves the norm, i.e., for all X *in* $\mathbf{R}_c^n$*, we have:*

$$\|PX\| = \|X\|.$$

(6) For one (and hence every) orthonormal basis

$$B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

of $\mathbf{R}_c^n$*, the new basis*

$$B' := B\,P$$

is again an orthonormal basis of $\mathbb{R}_c^n$*.*

Proof. Similar to Theorem 11.24 and left as an exercise for students. $\square$

Theorem 11.24. *(= Tutorial question) (***Equivalent unitary matrix definition***) For a complex matrix P in $M_n(\mathbb{C})$, the following are equivalent, where we employ the standard inner product on $\mathbb{C}_c^n$.*

(1) P is unitary, i.e., $P\,P^* = I_n$.

(2) Write

$$P = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$$

where the $\mathbf{p}_j$ are the column vectors of P. Then the column vectors $(\mathbf{p}_1, \ldots, \mathbf{p}_n)$ form an orthonormal basis of $\mathbf{C}_c^n$.

(3) The matrix transformation

$$\begin{aligned} T_P : \mathbf{C}_c^n &\to \mathbf{C}_c^n \\ X &\mapsto PX \end{aligned}$$

preserves the standard inner product, i.e., for *all X, Y in $\mathbf{C}_c^n$, we have*

$$\langle PX, PY\rangle = \langle X, Y\rangle.$$

(4) The matrix transformation

$$\begin{aligned} T_P : \mathbf{C}_c^n \;&\to\; \mathbf{C}_c^n \\ X \;&\mapsto\; PX \end{aligned}$$

preserves the distance, i.e., for all X, Y in $\mathbf{C}_c^n$, we have:

$$\|PX - PY\| = \|X - Y\|.$$

(5) The matrix transformation

$$\begin{aligned} T_P : \mathbf{C}_c^n \;&\to\; \mathbf{C}_c^n \\ X \;&\mapsto\; PX \end{aligned}$$

preserves the norm, i.e., for all X in $\mathbf{C}_c^n$, we have:

$$\|PX\| = \|X\|.$$

(6) For one (and hence every) orthonormal basis

$$B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$$

of $\mathbf{C}_c^n$*, the new basis*

$$B' := B\,P$$

is again an orthonormal basis of $\mathbb{C}_c^n$*.*

Exercise 11.25. (**Eigenvalues of orthogonal or unitary matrices**)

(1) If a real matrix $P \in M_n(\mathbf{R})$ is orthogonal, then every zero of $p_P(x)$ has modulus equal to 1. In particular, the determinant (cf. Exercise 9.4):

$$|P| = \pm 1.$$

(2) If a complex matrix $P \in M_n(\mathbf{C})$ is unitary, then every eigenvalue

$$\lambda = r_1 + \sqrt{-1}\,r_2$$

of P has **modulus**

$$|\lambda| = \sqrt{r_1^2 + r_2^2} = 1.$$

In particular, the determinant $|P| \in \mathbb{C}$ has modulus 1 (cf. Exercise 9.4).

Hint. Apply Exercise 11.17 to an eigenvector X of A and to $Y = A\,X$.

Exercise 11.26. (**Eigenvalue of self-adjoint linear operators**)

(1) Suppose that a real matrix $A \in M_n(\mathbf{R})$ is symmetric. Show that every zero of $p_A(x)$ is a real number.

(2) Suppose that a complex matrix $A \in M_n(\mathbf{C})$ is self-adjoint. Show that every zero of $p_A(x)$ is a real number.

(3) More generally, suppose that T is a self-adjoint linear operator. Show that every zero of $p_T(x)$ is a real number.

(4) Suppose that a T is a self-adjoint linear operator

(or $A \in M_n(\mathbb{C})$ is self-adjoint, or $A \in M_n(\mathbb{R})$ is symmetric). Let $\mathbf{v}_i$ $(i = 1, 2)$ be two eigenvectors corresponding to two distinct eigenvalues λ_i of T (or A). Show that

$$\mathbf{v}_1 \perp \mathbf{v}_2, \text{ i.e., } \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0.$$

Hint. To say the eigenvalue to be real, apply Theorem 11.18 or Exercise 11.17 to $X =$ (an eigenvector of A) and $Y = AX$. To say the orthogonality of $\mathbf{v}_i$, apply Theorem 11.18 or Exercise 11.17 to

$$X := \mathbf{v}_1, \quad Y := \mathbf{v}_2.$$

Exercise 11.27. (**Congruent/Similar matrices**) Let $U \in M_n(\mathbb{C})$ be a unitary matrix and $A \in M_n(\mathbb{C})$. Then A is unitary (resp. self-adjoint, or normal) if and only if $U^* A U$ is unitary (resp. self-adjoint, or normal).

Exercise 11.28. (**Diagonal matrices**)

(1) Every diagonal complex matrix $C \in M_n(\mathbb{C})$ is self-adjoint and also normal.

(2) Let $D = \mathrm{Diag}[\lambda_1, \ldots, \lambda_n] \in M_n(\mathbb{C})$ be a diagonal complex matrix. Then D is unitary if and only if the modulus $|\alpha_i| = 1$ for all $1 \leq i \leq n$.

Definition 11.29. (**Positive / negative definite linear operators**) Let $A \in M_n(\mathbb{C})$ (or let V be an n-dimensional inner product space which is over a field F with $F = \mathbb{R}$ or $F = \mathbb{C}$).

(1) T is **positive definite** if T is self-adjoint and

$$\langle T(\mathbf{v}), \mathbf{v} \rangle > 0 \ (\forall\ \mathbf{0} \neq \mathbf{v}\ \in\ V).$$

(2) T is **negative definite** if T is self-adjoint and

$$\langle T(\mathbf{v}), \mathbf{v} \rangle < 0 \ (\forall\ \mathbf{0} \neq \mathbf{v}\ \in\ V).$$

Thus T is negative definite if and only if $-T$ is positive definite.

(3) A is **positive definite** if A is self-adjoint and

$$(AX)^t\,\overline{X} = X^t\; A^t\,\overline{X} > 0\ (\forall\ X\ \neq \mathbf{0}).$$

(4) A is **negative definite** if A is self-adjoint and

$$(AX)^t\,\overline{X} = X^t\; A^t\,\overline{X} < 0\ (\forall\ X\ \neq \mathbf{0}).$$

Thus, A is negative definite if and only if $-A$ is positive definite.

Remark 11.30. (**Equivalent positive-definite definition**) Let

$$A = (a_{ij}) \in M_n(\mathbb{R})$$

be a symmetric real matrix. Then A is positive definite if and only if all its **principal minors**

$$(a_{ij})_{1\leq i,j\leq r}\ (1 \leq r \leq n)$$

of order r have positive determinants (cf. W. Keith Nicholson [Elementary Linear Algebra, Chapter 4.8.2]).

Let T be a self-adjoint linear operator on an inner product space V which is over F (with $F = \mathbb{R}$ or $\mathbb{C}$) and with an orthonormal basis B. Set

$$A := [T]_B \in M_n(F).$$

Then the following are equivalent.

(1) T is positive definite.

(2) A is positive definite.

(3) Every eigenvalue of T is positive.

(4) Every eigenvalue of A is positive.

(5) One can write A as

$$A = C^*C$$

for some invertible complex matrix $C \in M_n(\mathbb{C})$.

Hint. T is adjoint and hence normal and one can apply Theorem 11.33. Use (*) in the proof of Theorem 11.18.

If $\lambda_i > 0$ then $\mathrm{Diag}[\lambda_1, \ldots, \lambda_n] = C\,C^*$ where $C = \mathrm{Diag}[\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}]$.

Remark 11.31. In advanced calculus, we learnt that given a differentiable real-valued function $f(x, y)$, the function $f(x, y)$ attains a relative maximum (resp. relative minimum) at a stationary point P_0 (i.e., the point with vanishing derivatives: $f_x(P_0) = f_y(P_0) = 0$), if the matrix

$$M = \begin{pmatrix} f_{xx}(P_0) & f_{xy}(P_0) \\ f_{xy}(P_0) & f_{yy}(P_0) \end{pmatrix}$$

is negative (resp. positive) definite.

In one variable case, $f(x)$ attains a relative maximum (resp. relative minimum) at a stationary point P_0 of $f(x)$, if

$$f''(P_0) < 0 \ \ (\text{resp. } f''(P_0) > 0).$$

Extra exercise 11.32. Let $A \in M_n(\mathbb{C})$. Show that the function H on $V := \mathbb{C}_c^n$ below

$$H : V \times V \to \mathbf{C},$$
$$(X, Y) \mapsto \langle X, Y \rangle := (AX)^t\,\overline{Y} = X^t\,A^t\,\overline{Y}$$

defines an inner product on V if and only if A is positive definite.

.

Theorem 11.33. *(Principal Axis Theorem)*

(1) Let $T : V \to V$ be a linear operator on a real inner product space V of dimension n. Then T is self-adjoint (i.e., $T^* = T$) if and only if there is an orthonormal basis B such that

$$[T]_B$$

is a diagonal matrix in $M_n(\mathbf{R})$.

(2) A real matrix $A \in M_n(\mathbf{R})$ is self-adjoint (i.e., $A^* = A$) if and only if there is an orthogonal

matrix P *such that*

$$P^{-1}\,A\,P = P^t\,A\,P$$

is a diagonal matrix in $M_n(\mathbf{R})$.

(3) *Let* $T : V \to V$ *be a linear operator on a complex inner product space* V *of dimension* n. *Then* T *is normal (i.e.,* $T\,T^* = T^*\,T$*) if and only if there is an orthonormal basis* B *such that*

$$[T]_B$$

is a diagonal matrix in $M_n(\mathbf{C})$.

(4) *A complex matrix* $A \in M_n(\mathbf{C})$ *is normal (i.e.,* $A^*A = AA^*$ *), if and only if there is a unitary matrix* U *such that*

$$U^{-1}\,A\,U = U^*\,A\,U$$

is a diagonal matrix in $M_n(\mathbb{C})$.

Proof. A diagonal matrix is clearly normal (and even self-adjoint when it is a real matrix), so the 'if parts' of (1) - (4) are clear (cf. Exercise 11.22).

We now prove the 'only if' part of (1). Let $T : V \to V$ be a self-adjoint operator on the space V equipped with an inner product $\langle , \rangle$. We prove by the induction on n.

If $n = 1$, it is clear. Suppose it is true for self-adjoint operators on $(n-1)$-dimensional inner product spaces. Take an eigenvalue $\lambda_1 \in \mathbf{R}$ of T (cf. Exercise 11.26). Let $\widetilde{X}_1$ be an eigenvector corresponding to λ_1. Set

$$X_1 := \widetilde{X}_1 / \|\widetilde{X}_1\|.$$

Then $\|X_1\| = 1$.

Denote by

$$W := \mathrm{Span}\{X_1\}.$$

Set

$$W^{\perp} := \{Y \in V \mid \langle Y, X_1 \rangle = 0\}.$$

Then both W and $W^{\perp}$ are T-invariant (because T is symmetric; cf. the proof of **T10.6e**) and also (cf. Exercise 11.34):

$$V = W \oplus W^{\perp}.$$

The restriction $T_{|W^{\perp}}$ is again a self-adjoint linear operator (Exercise!). So by the induction, there is an orthonormal basis

$$B_2 = \{X_2, \ldots, X_n\}$$

of the $(n-1)$-dimensional space $W^{\perp}$ (whose inner product is the restriction of that on V) such that the matrix $[T|W^{\perp}]_{B_2}$ relative to B_2 equals an diagonal matrix

$$\operatorname{diag}[\lambda_2, \ldots, \lambda_n].$$

Now let $B_1 = \{X_1\}$ and

$$B = (B_1, B_2) = (X_1, \ldots, X_n)$$

which is an orthonormal basis of $V = \mathbf{R}^n_c$ (Exercise!). Then (cf. Exercise 9.31):

$$[T]_B = \begin{pmatrix} [T|W]_{B_1} & 0 \\ 0 & [T|W^{\perp}]_{B_2} \end{pmatrix}$$
$$= \mathrm{diag}[\lambda_1, \ldots, \lambda_n].$$

This proves the 'only if' part of (1). Thus (1) is proved.

For the 'only if' part of (2), Applying (1) to $T = T_A$ and $V = \mathbf{R}^n_c$ with the standard inner product, we can find an orthonormal basis

$$B = (X_1, \ldots, X_n)$$

of V such that the representation matrix $[T_A]_B$ equals a diagonal $J \in M_n(\mathbb{R})$, i.e.,

$$[T_A]_B = J.$$

Let

$$\widetilde{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n)$$

be the standard basis. Then (cf. Exercise 8.12)

$$[T_A]_{\widetilde{B}} = A.$$

By Basis change Theorem 8.20,

$$J = [T_A]_B = P^{-1}\,[T_A]_{\widetilde{B}}\,P = P^{-1}\,A\,P$$

where P is the transition matrix (from new basis B to old basis $\widetilde{B}$; cf. Theorem 8.17):

$$P = ([X_1]_{\widetilde{B}}, .., [X_n]_{\widetilde{B}}) = (X_1, .., X_n) = B.$$

Since the column vectors in P (i.e., X_i's) form an orthonormal basis of $V = \mathbb{R}^n_c$, the matrix P is orthogonal (cf. Theorem 11.23). This proves the 'only if' part of (2). We have proved (2).

For the 'only if' part of (3), we refer to Friedberg, Insel and Spencer [§6.4. Normal and self-adjoint operators]. An almost full proof is also given in Tutorial 10.

The 'only if' part of (4) can be proved as in the assertion (2), but we use Theorem 11.24 (instead of Theorem 11.23). $\square$

Exercise 11.34. (**Orthogonal complement**) Let W be a subspace of an inner product space V. Take an orthonormal basis B_W of W. Show:

(1) One can extend B_W to an orthonormal basis $B = (B_W, B_2)$ of V.

Hint. Extend B_W to a basis B' of V and apply Gram-Schmidt process.

(2) B_2 is an orthonormal basis of the so called **orthogonal complement of W**:

$$W^{\perp} := \{\mathbf{x} \in V \mid \langle \mathbf{x}, \mathbf{w} \rangle = 0, \ \forall\, \mathbf{w} \in W\}.$$

(3)

$$V = W \oplus W^{\perp}.$$

In the rest of the section, as an application of the theory so far, we will find standard forms of quadratic forms and classify conics in the Euclidean plane $\mathbb{R}^2$.

Definition 11.35. (**Quadratic forms**) Let V be a vector space over a field F. A function

$$K : V \rightarrow F$$

or simply $K(\mathbf{x})$ $(\mathbf{x} \in V)$ is a **quadratic form** if there is a symmetric bilinear form

$$H : V \times V \rightarrow F$$

such that

$$K(\mathbf{x}) = H(\mathbf{x}, \mathbf{x}).$$

When $V = F_c^n$, a quadratic form is always of the form (cf. Exercise 11.3):

$$\begin{aligned} K : V &\to F \\ X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &\mapsto K(X) := H_A(X, X) \\ &= X^t \, A \, X \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \end{aligned}$$

for some symmetric matrix $A = (a_{ij}) \in M_n(F)$.

A polynomial $f(x_1, \ldots, x_n)$ in n variables x_i, is **homogeneous of degree** d if every term in f is of degree d. For example,

$$3x_1 \, x_3^2 \, x_n^3 - 10x^6 + 4x_2 \, x_{n-1}^5$$

is a homogeneous polynomial of degree 6.

Every quadratic form

$$H_A(X, X) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

above is a homogeneous polynomial of degree 2.

Conversely, every homogeneous polynomial $f(x_1, \ldots, x_n)$ of degree 2 is a quadratic form on F_c^n.

For example, the homogeneous polynomial of degree 2 below is a quadratic form on F_c^3:

$$\begin{aligned} f(x_1, x_2, x_3) &= 2x_1^2 - x_2^2 + 6x_1x_2 - 4x_2x_3 \\ &= 2x_1^2 - x_2^2 + (3x_1x_2 + 3x_2x_1) \\ &\quad + (-2x_2x_3 - 2x_3x_2) \\ &= X^t \, A \, X = H_A(X, X) \end{aligned}$$

where

$$X := \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \ A := \begin{pmatrix} 2 & 3 & 0 \\ 3 & -1 & -2 \\ 0 & -2 & 0 \end{pmatrix}.$$

In general, if

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} x_i x_j$$

is a degree 2 homogeneous polynomial in n variables

x_i, we set

$$a_{ij} = a_{ji} := \frac{1}{2}(b_{ij} + b_{ji}).$$

Then

$$\begin{aligned} f(x_1, \ldots, x_n) &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \\ &= X^t\, A\, X = H_A(X, X) \end{aligned}$$

is a quadratic form, where

$$A = (a_{ij}) \in M_n(F)$$

is a symmetric matrix.

Theorem 11.36. *(***Quadratic form version of Principal Axis Theorem***). Let*

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

be a quadratic form in coordinates

$$X := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

with

$$A = (a_{ij}) \in M_n(\mathbb{R})$$

a symmetric matrix. Then there is an orthogonal matrix P such that f has the following **standard form**

$$f(x_1, \ldots, x_n) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

in the new coordinates

$$Y := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = P^{-1} X$$

where $\lambda_i \in \mathbf{R}$ are the eigenvalues of A.

This standard form is unique up to relabelling of $\lambda_i y_i$.

Proof. By Principal axis theorem 11.33, there is an orthogonal matrix P such that

$$P^{-1} A P = P^t A P$$

equals a diagonal matrix

$$J := \mathrm{diag}[\lambda_1, \ldots, \lambda_n].$$

By Definition 8.27, the characteristic polynomial $p_A(x)$ equals that of J:

$$p_A(x) = p_J(x) = (x - \lambda_1) \cdots (x - \lambda_n).$$

In particular, these λ_i are the eigenvalues of A.

Set

$$X := P\,Y \text{ (or } Y = P^{-1}\,X).$$

Then

$$\begin{aligned} f(x_1, \ldots, x_n) &= X^t\ A\,X \\ &= Y^t\,(P^t\ A\,P)Y \\ &= Y^t\,J\,Y \\ &= \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2. \end{aligned}$$

$\square$

Remark 11.37. (**Find the orthogonal matrix P diagonalizing a symmetric real matrix**) Let $A \in M_n(\mathbf{R})$ be a real symmetric matrix. We will show the way to find an orthogonal matrix P such that

$$P^{-1} \, A \, P = P^t \; A \, P = J \quad (*)$$

for some diagonal matrix $J \in M_n(\mathbb{R})$.

By Principal axis theorem 11.33, A is diagonalizable. Let

$$\lambda_1, \ldots, \lambda_k \, \in \, F$$

exhaust all distinct eigenvalues of A. Then

$$V := \mathbf{R}_c^n = V_{\lambda_1}(A) \oplus \cdots \oplus V_{\lambda_k}(A)$$

(cf. Theorem 10.23).

Use Gram-Schmidt process to find an orthonormal basis B_i for each eigenspace $V_{\lambda_i}(A)$. Then

$$B := (B_1, \, \ldots \; B_k)$$

is an orthonormal basis of $V = \mathbf{R}_c^n$ (cf. Exercise 11.26).

If we write $B_i = (\mathbf{w}_{i1}, \mathbf{w}_{i2}, \dots)$ then

$$\begin{aligned} AB_i &= (A\mathbf{w}_{i1}, A\mathbf{w}_{i2}, \dots) \\ &= (\lambda_i \mathbf{w}_1, \lambda_i \mathbf{w}_2, \dots) = \lambda_i B_i. \end{aligned}$$

Regard

$$P := (B_1, \dots, B_k)$$

as an $n \times n$ matrix whose columns are the vectors in B_i. Then

$$\begin{aligned} AP &= (AB_1, \dots, AB_k) \\ &= (\lambda_1 B_1, \dots, \lambda_k B_k) \\ &= (B_1, .., B_k) \begin{pmatrix} \lambda_1 I_{|B_1|} & 0 & \cdots & 0 \\ 0 & \lambda_2 I_{|B_2|} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_k I_{|B_k|} \end{pmatrix} \\ &= P\,J \end{aligned}$$

where

$$J := \mathrm{diag}[\lambda_1 I_{n_1}, \ldots, \lambda_k I_{n_k}]$$

is a diagonal matrix and

$$n_i := |B_i| = \dim V_{\lambda_i}$$

is the geometric multiplicity of λ_i (and also the algebraic multiplicity of λ_i, since A is diagonalizable as seen below or by Theorem 11.33).

So

$$A\,P = P\,J \ \ (\text{i.e., } \ P^{-1}\,AP = J).$$

Since the column vectors of P form the orthonormal basis B of $V = \mathbb{R}^n_c$, our P is orthogonal (cf. Theorem 11.23).

Thus we have found an orthogonal matrix P satisfying (*) above.

Example 11.38. (**Finding diagonalizing P for some 3×3 matrix A**) Consider the symmetric matrix

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

in $M_3(\mathbb{R})$. We will find an orthogonal matrix P such that

$$P^{-1}\,A\,P = P^t\;A\,P$$

is diagonal.

We calculate the characteristic polynomial

$$\begin{aligned} &(x-\lambda_1)^{n_1}\dots(x-\lambda_k)^{n_k} \\ =\,&p_A(x) = |xI_3 - A| \\ =\,&(x-2)^2(x-8). \end{aligned}$$

So

$$\lambda_1 := 2, \quad \lambda_2 := 8$$

are the only distinct eigenvalues of A and

$$k = 2, \quad n_1 = 2, \quad n_2 = 1.$$

One can calculate the eigenspaces:

$$V_{\lambda_1}(A) = \mathrm{Span}\{\mathbf{u}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}\}$$

and

$$V_{\lambda_2}(A) = \mathrm{Span}\{\mathbf{u}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\}.$$

Applying Gram-Schmidt process as in 11.15, we obtain orthonormal bases

$$B_1 := (\mathbf{v}_1 = \begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \ \mathbf{v}_2 = \begin{pmatrix} \frac{-1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix})$$

of $V_{\lambda_1}(A)$ and

$$B_2 := (\mathbf{v}_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix})$$

of $V_{\lambda_2}(A)$.

Now, by the formula in Remark 11.37,

$$P := (B_1, B_2) = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$$
$$= \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

is an orthogonal matrix in $M_3(\mathbb{R})$ such that

$$
\begin{aligned}
(*)\quad P^{-1}\,A\,P = P^t\ A\,P &= \mathrm{diag}[\lambda_1 I_{n_1}, \ldots, \lambda_k I_k] \\
&= \mathrm{diag}[\lambda_1 I_2, \lambda_2 I_1] \\
&= \begin{pmatrix} 2I_2 & 0 \\ 0 & 8I_1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix}
\end{aligned}
$$

Remark. You do not have to compute $P^{-1}\,A\,P$ directly. The (*) above should be regarded as a formula, so long the columns in the matrix P are put in orderly: orthonormal basis of V_{λ_1}, orthonormal basis of V_{λ_2}, $\cdots$, orthonormal basis of V_{λ_k}.

11.39. We now apply the orthogonal diagonalization in Theorem 11.36 to the classification of conics in the Euclidean plane $\mathbf{R}_c^2$.

Let $f(x_1, x_2)$ be a real polynomial of degree 2 (i.e., a quadratic polynomial) in the polynomial ring $\mathbf{R}[x_1, x_2]$. After a suitable composition of rotations:

$$X := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = R_\theta \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}$$

(with a rotation matrix R_θ as displayed below) and translations:

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix},$$

our f will become a degree 2 polynomial in x_1', x_2' with no degree 1 terms, **unless** f is as follows

$$f(x_1, x_2) = a_1 x_1^2 + a_2 x_2$$

in new coordinates, where $a_i \neq 0$.

Therefore, we consider first the polynomials of the form

$$\begin{aligned} f(x_1,\, x_2) &= a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 - d \\ &= (x_1,\, x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - d \\ &= X^t\, A\, X - d \end{aligned}$$

where $d \in \mathbf{R}$ is a constant,

$$X := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

and

$$A := (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in M_2(\mathbb{R})$$

is symmetric, i.e.,

$$a_{12} = a_{21}.$$

By Principal axis theorem 11.33, there is an orthogonal matrix $P \in M_2(\mathbb{R})$ such that

$$\begin{aligned} P^{-1}\, A\, P &= P^t\ A\, P \\ &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} =: J\ \in\ M_2(\mathbb{R}) \end{aligned}$$

is a diagonal matrix.

Set

$$X := P\, Y$$

with

$$Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Then

$$\begin{aligned} X^t\ A\, X &= (P\, Y)^t\ A\, P\, Y \\ &= Y^t\, (P^t\ A\, P)\, Y \\ &= Y^t\ J\, Y = \lambda_1 y_1^2 + \lambda_2 y_2^2 \end{aligned}$$

Thus, our initial polynomial

$$\begin{aligned} f(x_1,\, x_2) &= X^t\, A\, X - d \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 - d. \end{aligned}$$

Since P is orthogonal, we have $|P| = \pm 1$ (cf. Exercise 11.25). Moreover, for some $0 \leq \theta < 2\pi$ we have (cf. Friedberg, Insel and Spencer [§6.5. Unitary and orthogonal operations and their matrices]):

$$P = R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad \text{if } |P| = 1$$

$$P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{if } |P| = -1.$$

Therefore, the matrix transformation

$$\begin{aligned} T_P : \mathbf{R}_c^2 &\to \mathbf{R}_c^2 \\ X &\mapsto P\, X \end{aligned}$$

with P orthogonal, is either a rotation R_θ through the angle θ (counter-clock wise), or the composition

$$R_\theta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

of a rotation R_θ and a reflection about the x_1-axis:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}.$$

Conversely, rotation matrix R_θ and reflection matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and their multiplications are always orthogonal matrices.

Definition 11.40. (**Rigid motions in Euclidean spaces**) Let V be an inner product space over $F = \mathbf{R}$ or $F = \mathbf{C}$.

(1) A map

$$R : V \to V$$

is a **rigid motion** if it preserves the distance:

$$\|f(X) - f(Y)\| = \|X - Y\| \ (\forall\ X, Y \in V).$$

A rigid motion of the plane $\mathbf{R}_c^2$ (resp. the 3-space $\mathbf{R}_c^3$) does not change the shape of an object in the plane (resp. in the 3-space).

(2) A translation

$$\begin{aligned} V &\to V \\ X &\mapsto X + c \end{aligned}$$

is a rigid motion.

If $A \in M_n(\mathbf{R})$ is an orthogonal matrix, then the **orthogonal matrix transformation**

$$\begin{aligned} T_A : \mathbf{R}_c^n &\to \mathbf{R}_c^n \\ X &\mapsto A\,X \end{aligned}$$

is a rigid motion (cf. Theorem 11.23).

The composition of any rigid motions (including translations and orthogonal matrix transformations) is also a rigid motion.

(3) Conversely, a rigid motion of the Euclidean plane

$\mathbb{R}^2$ is a composition of some of the translations and orthogonal matrix transformations, i.e., the composition of some of translations, rotations and reflection (cf. Friedberg, Insel and Spencer [§6.5]).

Now we are ready to classify the **conics** in $\mathbb{R}^2$ (i.e., a curve defined by a degree 2 polynomial).

First, as mentioned above, we have:

Theorem 11.41. *(**Standard form of quadratic polynomials**) Let*

$$f(x_1, x_2)$$

be a real polynomial of degree 2 (i.e., a quadratic polynomial). Then in new coordinates (y_1, y_2) (obtained from (x_1, x_2) by a rigid motion, i.e., a composition of some of translations, rotations and reflection), our f can be expressed as one of *the following* **standard forms** *in* y_1, y_2, *where* $a_i \neq 0$*:*

$$f(x_1,\, x_2) = \lambda_1 y_1^2 + \lambda_2 y_2^2 - d,$$

$$f(x_1,\, x_2) = a_1 y_1^2 + a_2 y_2.$$

11.42. (**Classification of the shapes of plane conics**) Next we want to classify the shape of a **conic** C defined as follows

$$C: \;\; f(x_1, x_2) = 0$$

where f is a real polynomial of degree 2 (i.e., a quadratic polynomial).

Since a rigid motion of the plane $\mathbf{R}_c^2$ does not change the shape of an object, in view of Theorem 11.41, we may assume that C or its defining equation f is already of one of the following forms where $a_i \neq$

0:

$$C \ : \ f(x,\, y) = \lambda_1 x^2 + \lambda_2 y^2 - d = 0,$$
$$C \ : \ f(x,\, y) = a_1 x^2 + a_2 y = 0.$$

According to the values of

$$\lambda_i, \quad d, \quad a_i$$

the shape of the plane conic C is classified into the following seven types.

(1) **Double line.**

$d = 0$, one of λ_i $(i = 1, 2)$ is zero, say $\lambda_2 = 0$ but $\lambda_1 \neq 0$.

Then C is as follows

$$C : \quad x^2 = 0$$

and is the 'double' y-axis, noting that $x = 0$ defines the y-axis.

(2) **Intersecting lines.**

$d = 0$, λ_i $(i = 1, 2)$ have different signs (i.e., one positive and the other negative), say

$$\lambda_1 = a_1^2 > 0, \quad \lambda_2 = -a_2^2 < 0.$$

Then C is as follows

$$C: \quad (a_1 x + a_2 y)(a_1 x - a_2 y) = 0$$

and is a union of two intersecting lines

$$a_1 x + a_2 y = 0$$

and

$$a_1 x - a_2 y = 0.$$

(3) **A point.**

$d = 0$, λ_i $(i = 1, 2)$ have the same sign (i.e., both positive or both negative). Then C is a point:

$$C = \{(0, 0)\}.$$

(4) **The empty set $\emptyset$.**

$d \neq 0$, $\lambda_i \neq 0$, λ_i $(i = 1, 2)$ have the same sign but different from d, say

$$\lambda_i = a_i^2 > 0, \quad d = -c^2 < 0.$$

Then C is the empty set as explained below:

$$C: \quad a_1^2 x^2 + a_2^2 y^2 = -c^2.$$

(5) **Ellipse.**

$d \neq 0$, $\lambda_i \neq 0$, and three constants d, λ_1, λ_2 have the same sign, say

$$d/\lambda_1 = a^2 > 0, \quad d/\lambda_2 = b^2 > 0.$$

Then C is an ellipse as follows (when $a = b$, it is a circle):

$$C: \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

(6) **Hyperbola.**

$d \neq 0$, $\lambda_i \neq 0$, and the two λ_i $(i = 1, 2)$ have different signs, say

$$d/\lambda_1 = a^2 > 0, \quad d/\lambda_2 = -b^2 < 0.$$

Then C is a hyperbola as follows:

$$C: \quad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

(7) **Parabola.**

$a_i \neq 0$ $(i = 1, 2)$ and write

$$\frac{-a_1}{a_2} = \pm\, e^2.$$

Then C is a parabola as follows:

$$C: \quad y = \pm\, e^2 x^2.$$

End of the course

Thanks for your patient reading!