

Linear Algebra

*Prof. Allen Yuan**Scribe: Ekene Ezeunala***Contents**

1	Lecture 01—18th January, 2022	3
1.1	Introduction	3
2	Lecture 02—20th January, 2022	6
2.1	Vector spaces	6
3	Lecture 03—25th January, 2022	9
3.1	Subspaces	9
4	Lecture 04—27th January, 2022	10
4.1	Linear combinations and spans	10
5	Lecture 05—01st February, 2022	12
5.1	Linear independence	12
6	Lecture 05—03rd February, 2022	15
6.1	Bases and dimension	15
7	Lecture 07—08th February, 2022	19
7.1	Linear transforms (I)	19
8	Lecture 08—10th February, 2022	23
8.1	Linear transforms (II): dimension theorem	23
9	Lecture 09—15th February, 2022	26
9.1	The ordered basis and coordinates	26
10	Lecture 10—17th February, 2022	28
10.1	The matrix representation of a linear transformation	28
11	Lecture 11—22nd February, 2022	31
11.1	Composition of linear transformations	31
12	Lecture 12—24th February, 2022	33
12.1	Invertibility and isomorphisms (I)	33
13	Lecture 13—01st March, 2022	36
13.1	Invertibility and isomorphisms (II)	36

14 Lecture 14—03rd March, 2022	38
14.1 The change of coordinate matrix	38
15 Lecture 15—08th March, 2022	41
15.1 The similar matrices and eigendecompositions	41
16 Lecture 16—10th March, 2022	43
16.1 Diagonalisation	43
17 Lecture 17—22nd March, 2022	45
17.1 Eigenvalues and eigenvectors (II)	45
18 Lecture 18—24th March, 2022	47
18.1 Diagonalisation of linear operators (I): distinct and complex eigenvalues	47
19 Lecture 19—29th March, 2022	49
19.1 Diagonalisation of linear operators (II): repeated eigenvalues	49
20 Lecture 20—31st March, 2022	51
20.1 Applications: homogeneous linear system with constant coefficients	51
21 Lecture 21—05th April, 2022	52
21.1 Inner product spaces	52
22 Lecture 22—07th April, 2022	54
22.1 The Gram-Schmidt orthogonalisation process	54
23 Lecture 23—12th April, 2022	58
23.1 Adjoint operators	58
24 Lecture 24—14th April, 2022	64
24.1 The singular value decomposition (SVD)	64
25 Lecture 25—19th April, 2022	66
25.1 The singular value decomposition: forms and subspaces	66
26 Lecture 26—21st April, 2022	68
26.1 Low-rank approximation and geometric interpretation	68
27 Lecture 27—26th April, 2022	70
27.1 The singular value theorem for linear transforms	70
27.2 Applications of the Singular Value Decomposition	72

§1 Lecture 01—18th January, 2022

Course logistics This is an introductory course on abstract linear algebra, although we will also focus on the diagonalisation of linear operators between abstract vector spaces and their applications. We will discuss general vector spaces, linear transformations between them, diagonalisation of these transformations via matrix representation, inner product spaces and the concept of orthogonality, linear operators on inner product spaces and their singular value decompositions, as well as some applications of the abstract theory. The main texts for this course will be [Axl15, FIS14, Hal13, Lan87, Str06]. 20% of the final grade will depend on homework assignments, with the rest of the grade equally dispersed between the midterm and final exams.

§1.1 Introduction

We begin by going through the basics of matrix theory.

Definition 1.1. A field $\mathbb{F}$ is a set on which two operations $+$ and $\cdot$ (addition and multiplication, respectively) are defined so that, for each pair of elements $a, b \in \mathbb{F}$, there are unique elements $a + b$ and $a \cdot b$ (the sum and product of a and b , respectively) in $\mathbb{F}$ for which the following conditions hold for all elements $a, b, c \in \mathbb{F}$:

F1 (Commutativity) $a + b = b + a$ and $a \cdot b = b \cdot a$.

F2 (Associativity) $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

F3 (Additive identity) There exists $0 \in \mathbb{F}$ such that $a + 0 = a$.

F4 (Multiplicative identity) There exists $1 \in \mathbb{F}$ such that $a \cdot 1 = a$.

F5 (Additive inverse) For each $a \in \mathbb{F}$, there is $-a \in \mathbb{F}$ with $a + (-a) = 0$.

F6 (Multiplicative inverse) For each $a \neq 0 \in \mathbb{F}$, there is $a^{-1} \in \mathbb{F}$ with $a \cdot a^{-1} = 1$.

F7 (Distributivity) $a \cdot (b + c) = a \cdot b + a \cdot c$.

We call elements of a field *scalars*.

- Example 1.2.**
1. The set of real numbers $\mathbb{R}$ with the usual definitions of addition and multiplication is a field.
 2. The set of positive numbers $\mathbb{N}$ with the usual definitions of addition and multiplication is not a field, as it does not satisfy the additive inverse property.

We will typically consider $+$ and $\cdot$ defined differently from usual, as in the following field instances.

- Example 1.3.**
1. The set of complex numbers $\mathbb{C}$ with addition defined as

$$c_1 + c_2 = (a_1 + b_1i) + (a_2 + b_2i) = (a_1 + a_2) + (b_1 + b_2)i$$

where $c_1 = a_1 + b_1i$ and $c_2 = a_2 + b_2i$ are in $\mathbb{C}$, and multiplication defined as

$$c_1 \cdot c_2 = (a_1 + b_1i) \cdot (a_2 + b_2i) = (a_1a_2 - b_1b_2) + (a_1b_2 + a_2b_1)i.$$

2. The set $\mathbb{Z}_2 = \{0, 1\}$ with addition and multiplication respectively defined as

$$\begin{array}{ll} 0 + 0 = 0, & 0 \cdot 0 = 0, \\ 0 + 1 = 1, & 0 \cdot 1 = 0, \\ 1 + 0 = 1, & 1 \cdot 0 = 0, \\ 1 + 1 = 0, & 1 \cdot 1 = 1, \end{array}$$

is a field.

We will usually write ab for the product of a and b in a field.

Definition 1.4 (Matrix). *Let $m, n \geq 1$ be two integers. A $m \times n$ matrix over $\mathbb{F}$ is an m -row and n -column array of scalars in $\mathbb{F}$. The set of $m \times n$ matrices over $\mathbb{F}$ is denoted by $\mathbb{F}^{m \times n}$.*

Definition 1.5 (Column and row vectors). *Fix $m, n \geq 2$. A matrix of size $m \times 1$ is called a column vector, and a matrix of size $1 \times n$ is called a row vector. Any vector by default is a column vector.*

An $m \times n$ matrix A over a field $\mathbb{F}$ is written in the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

where $a_{ij} \in \mathbb{F}$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$ are the scalar entries of A . Depending on the context, we will use $\mathbb{F}^n$ to denote either the set of column vectors of size n or the set of row vectors of size n , or the set of n -tuple of scalars in $\mathbb{F}$.

Definition 1.6. *For $A, B \in \mathbb{F}^{m \times n}$, we say that $A = B$ if $a_{ij} = b_{ij}$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$.*

Definition 1.7. *On the set of $m \times n$ matrices, we define the operations of addition (+) and scalar multiplication ($\cdot$) as follows: For any matrices $A, B \in \mathbb{F}^{m \times n}$ and any scalar $c \in \mathbb{F}$,*

$$(A + B)_{ij} = a_{ij} + b_{ij}, \quad (cA)_{ij} = ca_{ij}.$$

Theorem 1.1. *The following statements hold:*

1. *Commutativity of addition: $A + B = B + A$ for all $A, B \in \mathbb{F}^{m \times n}$.*
2. *Associativity of addition: $(A + B) + C = A + (B + C)$ for all $A, B, C \in \mathbb{F}^{m \times n}$.*
3. *Additive identity: There exists a matrix $0_{mn} \in \mathbb{F}^{m \times n}$ such that $A + 0 = A$ for all $A \in \mathbb{F}^{m \times n}$. The matrix 0_{mn} is the matrix with all entries equal to 0.*
4. *Additive inverse: For each $A \in \mathbb{F}^{m \times n}$, there is a matrix $-A \in \mathbb{F}^{m \times n}$ such that $A + (-A) = 0$.*
5. *Distributivity: $c(A + B) = cA + cB$ and $(c + d)A = cA + dA$ for all $A, B \in \mathbb{F}^{m \times n}$ and $c, d \in \mathbb{F}$.*

6. *Multiplicative identity:* There exists a matrix $I_{nn} \in \mathbb{F}^{n \times n}$ such that $AI = A$ for all $A \in \mathbb{F}^{m \times n}$. The matrix I_{nn} is the identity matrix with 1's on the diagonal and 0's elsewhere.
7. *Associativity of scalar multiplication:* $c(dA) = (cd)A$ for all $A \in \mathbb{F}^{m \times n}$ and $c, d \in \mathbb{F}$.

Proof. These are all easily checked. □

The following facts about matrix multiplications are usually counterintuitive for beginners:

1. In general, $AB \neq BA$ for matrices $A, B \in \mathbb{F}^{m \times n}$, even if both products are defined. We do have $A^{-1}A = AA^{-1} = I$ for square matrices A with an inverse A^{-1} , and $AI = IA = A$ for all $A \in \mathbb{F}^{m \times n}$.
2. In general, $AB = AC \not\Rightarrow B = C$. If A is invertible, then $A^{-1}AB = A^{-1}AC \Rightarrow B = C$.
3. In general, $AB = 0 \not\Rightarrow A = 0$ or $B = 0$. If A is invertible, then $A^{-1}AB = 0 \Rightarrow B = 0$.

Example 1.8. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 0 \\ 4 & 5 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$. Then $AB = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ and $AC = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, but $B \neq C$.

Now let $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Then $AB = 0$, but $A \neq 0$ and $B \neq 0$.

Definition 1.9 (Transpose matrix). For any matrix $A \in \mathbb{F}^{m \times n}$, the transpose of A , denoted by $A^\top$, is the $n \times m$ matrix obtained by interchanging the rows and columns of A . That is, if $A = [a_{ij}]$, then $A^\top = [a_{ji}]$.

Definition 1.10 (Symmetric matrix). A matrix $A \in \mathbb{F}^{n \times n}$ is symmetric if $A = A^\top$.

- Theorem 1.2.**
1. For all $A, B \in \mathbb{F}^{m \times n}$ and $c \in \mathbb{F}$, $(A^\top)^\top = A$, $(A + B)^\top = A^\top + B^\top$ and $(cA)^\top = cA^\top$.
 2. For all $A \in \mathbb{F}^{m \times n}$ and $B \in \mathbb{F}^{n \times p}$, $(AB)^\top = B^\top A^\top$.
 3. For all $A \in \mathbb{F}^{n \times n}$, if $AA^\top = 0$, then $A = 0$.

Proof.

1. By definition of the transpose, $(A^\top)^\top = A$, and similarly $(A + B)^\top = A^\top + B^\top$ and $(cA)^\top = cA^\top$ by entrywise addition and scalar multiplication.
2. The (i, j) -entry of $(AB)^\top$ is $\sum_k a_{ik}b_{kj}$, which matches the (j, i) -entry of $B^\top A^\top$, so $(AB)^\top = B^\top A^\top$.
3. If $AA^\top = 0$, then $\text{trace}(AA^\top) = \sum_{i,j} (A)_{ij}^2 = 0$, so all $A_{ij} = 0$ and $A = 0$. □

Example 1.11. $\mathbb{F}^{m \times n}$ is the set of all $m \times n$ matrices over $\mathbb{F}$. Is $\mathbb{F}^{m \times n}$ a field with matrix addition and multiplication?

1. Matrix multiplication is not well-defined for all pairs of matrices in $\mathbb{F}^{m \times n}$, as the number

of columns of the first matrix must equal the number of rows of the second matrix. So $\mathbb{F}^{m \times n}$ is not a field.

2. If $m = n$, then $AB = C \in \mathbb{F}^{n \times n}$ is well-defined for all $A, B \in \mathbb{F}^{n \times n}$, so $\mathbb{F}^{n \times n}$ is a ring. However, $AB \neq BA$ in general, so $\mathbb{F}^{n \times n}$ is not a commutative ring, and hence not a field.
3. If we define scalar multiplication as $cA = cI_{mn}A$ for all $c \in \mathbb{F}$ and $A \in \mathbb{F}^{m \times n}$, then $\mathbb{F}^{m \times n}$ is a vector space over $\mathbb{F}$.

§2 Lecture 02—20th January, 2022

§2.1 Vector spaces

Definition 2.1 (Vector space). *Let V be a set on field $\mathbb{F}$. We define two binary operations on V , namely addition and scalar multiplication, denoted by $+$ and $\cdot$, respectively as follows:*

Addition *For any $u, v \in V$, there exists a unique element $u + v \in V$.*

Scalar multiplication *For any $c \in \mathbb{F}$ and $v \in V$, there exists a unique $cv \in V$.*

If the addition and scalar multiplication operations satisfy the following conditions for all $u, v, w \in V$ and $c, d \in \mathbb{F}$, then V is called a vector space over $\mathbb{F}$:

A1 (Commutativity) $u + v = v + u$ for all $u, v \in V$.

A2 (Associativity) $(u + v) + w = u + (v + w)$ and $(c + d)v = cv + dv$ for all $u, v, w \in V$ and $c, d \in \mathbb{F}$.

A3 (Additive identity) *There exists $0 \in V$ such that $v + 0 = v$ for all $v \in V$.*

A4 (Additive inverse) *For each $v \in V$, there exists $-v \in V$ such that $v + (-v) = 0$.*

M1 (Distributivity) $c(u + v) = cu + cv$ and $(c + d)v = cv + dv$ for all $u, v \in V$ and $c, d \in \mathbb{F}$.

M2 (Associativity) $c(dv) = (cd)v$ for all $v \in V$ and $c, d \in \mathbb{F}$.

M3 (Multiplicative identity) *There exists $1 \in \mathbb{F}$ such that $1v = v$ for all $v \in V$.*

We call elements of V vectors, and the elements of $\mathbb{F}$ scalars.

Example 2.2. 1. Let $\mathbb{F}$ be a field. Then $\mathbb{F}^{3 \times 2}$, the set of matrices with 3 rows and 2 columns over $\mathbb{F}$, is a vector space over $\mathbb{F}$ with the usual definitions of addition and scalar multiplication.

2. Let $A \in \mathbb{R}^{n \times n}$ be a matrix. Then $W = \{v \in \mathbb{R}^n \mid Av = 0\}$ is a vector space over $\mathbb{R}$.

1. The set $W = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$ is a vector space over $\mathbb{R}$.

2. The set $W = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1\}$ is not a vector space over $\mathbb{R}$.

Proof. Let $x = (1, 0, 0)$ and $y = (0, 1, 0)$ be in W . Then $x + y = (1, 1, 0)$ is not in W . $\square$

3. The set $W = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i < 0\}$ is not a vector space over $\mathbb{R}$.

Proof. Let $x = (-1, -1)$ be in W . Then $-2x = (2, 2)$ is not in W . Hence not closed under scalar multiplication. $\square$

Example 2.3. The set $\mathbb{P}_n(\mathbb{R})$ for $n \geq 0$, the set of polynomials of degree at most n with coefficients in $\mathbb{R}$, consists of all polynomials of the form

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where the scalars a_i are in $\mathbb{R}$, and the degree of $p(x)$ is at most n . Then $\mathbb{P}_n(\mathbb{R})$ is a vector space over $\mathbb{R}$.

Proof. Let $p(t), q(t), g(t) \in \mathbb{P}_n(\mathbb{R})$ and $a_i, b_i, c_i \in \mathbb{R}$ for $0 \leq i \leq n$, so

$$\begin{aligned} p(t) &= a_0 + a_1t + a_2t^2 + \cdots + a_nt^n, \\ q(t) &= b_0 + b_1t + b_2t^2 + \cdots + b_nt^n, \\ g(t) &= c_0 + c_1t + c_2t^2 + \cdots + c_nt^n. \end{aligned}$$

1. (Closed under addition.) $p(t) + q(t) = (a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2 + \cdots + (a_n + b_n)t^n \in \mathbb{P}_n(\mathbb{R})$.
 - a) (Commutativity.) $p(t) + q(t) = q(t) + p(t) = (a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2 + \cdots + (a_n + b_n)t^n$.
 - b) (Associativity.) $(p(t) + q(t)) + g(t) = p(t) + (q(t) + g(t)) = (a_0 + b_0 + c_0) + (a_1 + b_1 + c_1)t + (a_2 + b_2 + c_2)t^2 + \cdots + (a_n + b_n + c_n)t^n$.
 - c) (Additive identity.) $0(t) = 0 + 0t + 0t^2 + \cdots + 0t^n$, so $p(t) + 0(t) = p(t)$.
 - d) (Additive inverse.) $-p(t) = -a_0 - a_1t - a_2t^2 - \cdots - a_nt^n$, so $p(t) + (-p(t)) = 0(t)$.
2. (Closed under scalar multiplication.) $\beta p(t) = \beta a_0 + \beta a_1t + \beta a_2t^2 + \cdots + \beta a_nt^n \in \mathbb{P}_n(\mathbb{R})$.
 - a) (Distributivity.) $\alpha(p(t) + q(t)) = \alpha(a_0 + b_0) + \alpha(a_1 + b_1)t + \alpha(a_2 + b_2)t^2 + \cdots + \alpha(a_n + b_n)t^n = \alpha p(t) + \alpha q(t)$.
 - b) (Distributivity.) $(\alpha + \beta)p(t) = (\alpha + \beta)a_0 + (\alpha + \beta)a_1t + (\alpha + \beta)a_2t^2 + \cdots + (\alpha + \beta)a_nt^n = \alpha p(t) + \beta p(t)$.
 - c) (Multiplicative identity.) $1p(t) = 1a_0 + 1a_1t + 1a_2t^2 + \cdots + 1a_nt^n = p(t)$, so $\exists 1 \in \mathbb{R}$ such that $1p(t) = p(t)$.
 - d) (Associativity.) $(\alpha\beta)p(t) = \alpha(\beta(a_0 + a_1t + a_2t^2 + \cdots + a_nt^n)) = (\alpha\beta)a_0 + (\alpha\beta)a_1t + (\alpha\beta)a_2t^2 + \cdots + (\alpha\beta)a_nt^n = \alpha(\beta p(t))$.

Therefore $\mathbb{P}_n(\mathbb{R})$ is a vector space over $\mathbb{R}$. $\square$

Example 2.4. Let $C([0, 1])$ be the set of continuous functions on $[0, 1]$. Then $C([0, 1])$ is a vector space over $\mathbb{R}$.

Proof. Let $f, g, h \in C([0, 1])$ and $a, b, c \in \mathbb{R}$.

1. (Closed under addition.) $(f + g)(x) = f(x) + g(x) \in C([0, 1])$.
 - a) (Commutativity.) $(f + g)x = f(x) + g(x) = g(x) + f(x) = (g + f)(x)$.
 - b) (Associativity.) $((f + g) + h)(x) = (f + g)(x) + h(x) = f(x) + g(x) + h(x) = f(x) + (g + h)(x) = (f + (g + h))(x)$.
 - c) (Additive identity.) $0(x) = 0$ for all $x \in [0, 1]$, so $f(x) + 0(x) = f(x)$.
 - d) (Additive inverse.) $-f(x) = -f(x)$, so $f(x) + (-f(x)) = 0(x)$.
2. (Closed under scalar multiplication.) $(af)(x) = af(x) \in C([0, 1])$.
 - a) (Distributivity.) $(a(f + g))(x) = a(f(x) + g(x)) = af(x) + ag(x) = (af)(x) + (ag)(x)$.
 - b) (Distributivity.) $((a + b)f)(x) = (a + b)f(x) = af(x) + bf(x) = (af)(x) + (bf)(x)$.
 - c) (Multiplicative identity.) $1f(x) = f(x)$, so $\exists 1 \in \mathbb{R}$ such that $1f(x) = f(x)$.
 - d) (Associativity.) $(ab)f(x) = a(bf(x)) = a(bf)(x) = (abf)(x)$.

Therefore $C([0, 1])$ is a vector space over $\mathbb{R}$. □

Theorem 2.1. *The following properties hold for any vector space V over $\mathbb{F}$:*

1. (Cancellation law) For any $u, v, w \in V$ and $c \in \mathbb{F}$, if $u + v = u + w$, then $v = w$.
2. The additive identity, i.e. the zero vector $0 \in V$ is uniquely defined.
3. For any $x \in V$, the additive inverse element $-x$ is unique.
4. For any $\alpha \in F$, $\alpha 0 = 0$.
5. For any $x \in V$, $0x = 0$.
6. For all $\alpha \in F$ and $x \in V$, $(-\alpha)x = -(\alpha x) = \alpha(-x)$.

Proof.

1. Let $u + v = u + w$. Then $v = 0 + v = (-u + u) + v = -u + (u + v) = -u + (u + w) = (-u + u) + w = 0 + w = w$.
2. Suppose $0, 0' \in V$ are both distinct additive identities. Then $0 = 0 + 0' = 0'$, a contradiction.
3. Suppose $-x, -x' \in V$ are both additive inverses of x . Then $x + (-x) = 0 = x + (-x')$, so $-x = -x'$ by the cancellation law.
4. For all $\alpha \in F$, $\alpha x = \alpha(x + 0) = \alpha x + \alpha 0$, so $\alpha 0 = 0$ by the cancellation law.
5. For all $x \in V$, $0 = 0x = (0 + 0)x = 0x + 0x$, so $0x = 0$ by the cancellation law.
6. For all $\alpha \in F$ and $x \in V$, $0 = 0x = \alpha x + (-\alpha x) = (\alpha + (-\alpha))x = 0x = 0$, so $-\alpha x = -(\alpha x)$. Similarly, $0 = 0x = \alpha x + (-\alpha x) = (\alpha + (-\alpha))x = 0x = 0$, so $-\alpha x = \alpha(-x)$. □

§3 Lecture 03—25th January, 2022

§3.1 Subspaces

Definition 3.1 (Subspace, 1st definition). *A subset W of a vector space V is called a subspace of V if W is a vector space over $\mathbb{F}$ under the operations of addition and scalar multiplication defined in V .*

Fortunately it is not necessary to verify all of the vector space properties to show that a subset is a subspace, because commutativity, associativity, and distributivity hold for all vectors in the vector space. Thus a subset W of a vector space V is a subspace if and only if the following four properties hold:

1. It is closed under addition.
2. It is closed under scalar multiplication.
3. The additive identity of V , namely 0 , is in W .
4. Every element of W has an additive inverse in W .

Definition 3.2 (Subspace, 2nd definition). *Let V be a vector space and W a subset of V . Then W is a subspace of V if and only if (i) $0 \in W$, (ii) $u, v \in W \implies u + v \in W$, and (iii) $u \in W, c \in \mathbb{F} \implies cu \in W$.*

Example 3.3. 1. Let V be a vector space. Then both V and $\{0\}$ are subspaces of V .
2. The set $\{(x_1, 0) : x_1 \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.
3. Let $\mathbb{F}$ be a field. The set $W = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + 2x_2 = 0\}$ is a subspace of $\mathbb{R}^3$.

Example 3.4. $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$. Although $\mathbb{R}^2$ includes the zero vector $(0 \ 0)^\top$, is closed under addition, and is closed under scalar multiplication, it is not a subspace because it is not a subset of $\mathbb{R}^3$. To see this, take the vector $(1 \ 1)^\top$ in $\mathbb{R}^2$. This vector is not in $\mathbb{R}^3$, so $\mathbb{R}^2$ is not a subset of $\mathbb{R}^3$.

However,

$$H = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

is a subspace of $\mathbb{R}^3$. H looks and acts like $\mathbb{R}^2$, but is a logically distinct object.

Example 3.5 (Trace). The trace of an $n \times n$ matrix A is defined as the sum of the diagonal elements of A . The trace is a linear map from the vector space of $n \times n$ matrices to $\mathbb{R}$:

$$\text{tr}(M) = A_{11} + A_{22} + \cdots + A_{nn} = \sum_{i=1}^n A_{ii}.$$

The set of $n \times n$ matrices with trace 0 is a subspace of the vector space of $n \times n$ matrices.

Proof. Let A, B be $n \times n$ matrices with trace 0. Then $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B) = 0 + 0 = 0$, so $A + B$ has trace 0. Similarly, for any scalar $c \in \mathbb{R}$, $\text{tr}(cA) = c\text{tr}(A) = c \cdot 0 = 0$, so cA has trace 0. Thus the set of $n \times n$ matrices with trace 0 is a subspace of the vector space of $n \times n$ matrices. $\square$

Theorem 3.1. *Let V be a vector space, and let W_1 and W_2 be two subspaces of V . Then the intersection $W_1 \cap W_2$ is a subspace of V .*

Proof. Let $W = W_1 \cap W_2$. Since $0 \in W_1$ and $0 \in W_2$, we have $0 \in W_1 \cap W_2 = W$. Now note that $W \subseteq W_1 \subseteq V$ and $W \subseteq W_2 \subseteq V$. For any $x, y \in W$, it follows that $x, y \in W_1$ and $x, y \in W_2$, so then $x + y \in W_1$ and $x + y \in W_2$, and hence $x + y \in W_1 \cap W_2 = W$. Similarly, for any $x \in W$ and $c \in \mathbb{F}$, we have $cx \in W_1$ and $cx \in W_2$, so $cx \in W_1 \cap W_2 = W$. Thus W is a subspace of V . $\square$

Example 3.6. Let V be a vector space, and let W_1 and W_2 be two subspaces of V . Is $W_1 \cup W_2$ a subspace of V ? No! Let $W = W_1 \cup W_2$. Then:

1. The subset property is satisfied: For all $x \in W$, $x \in W_1 \subseteq V$ or $x \in W_2 \subseteq V$; either way, $x \in W \implies x \in V$, so $W \subseteq V$.
2. The zero property is satisfied: Since W_1 and W_2 are subspaces, $0 \in W_1$ and $0 \in W_2$, so $0 \in W_1 \cup W_2 = W$.
3. The closure under scalar multiplication is satisfied: For any $\alpha \in \mathbb{F}$ and $x \in W$, if $x \in W_1$, then $\alpha x \in W_1$; if $x \in W_2$, then $\alpha x \in W_2$. So $\alpha x \in W_1 \cup W_2 = W$.
4. However, the closure under addition is not necessarily satisfied; consider the following counterexample:

$$W_1 = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} : x \in \mathbb{R} \right\}, \quad W_2 = \left\{ \begin{pmatrix} 0 \\ y \end{pmatrix} : y \in \mathbb{R} \right\}.$$

Then $W_1 \cup W_2 = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ y \end{pmatrix} \right\}$, which is not closed under addition, because

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin W_1 \cup W_2.$$

§4 Lecture 04—27th January, 2022

§4.1 Linear combinations and spans

Definition 4.1 (Linear combination). *Let V be a vector space over some field $\mathbb{F}$. Let $B = \{b_1, \dots, b_n\}$ be a set in V . The linear combination of the vectors in B is defined as the following form:*

$$x = a_1 b_1 + a_2 b_2 + \dots + a_n b_n,$$

where $a_i \in \mathbb{F}$ for $i = 1, \dots, n$ are arbitrary scalars.

Remark 1. 0 is a linear combination of any set of vectors in V . Just take all the scalars to be 0 .

Definition 4.2 (Span). Let B be a set of vectors. The span of B , denoted by $\text{span}(B)$, is the set of all linear combinations of the vectors in B .

If $B = \emptyset$, then $\text{span}(B) = \{0\}$.

If $\text{span}(B) = V$, then we say that B spans V .

Theorem 4.1 (Constructing subspaces). Let $B = \{b_1, b_2, \dots, b_n\}$ be a subset of a vector space V . Then $\text{span}(B)$ is a subspace of V .

Proof. 1. First we prove that $\text{span}(B) \subseteq V$. For all $x \in \text{span}(B)$, $x = a_1b_1 + a_2b_2 + \dots + a_nb_n$ for some $a_i \in \mathbb{F}$ since $b_i \in V$ for all i . Since V is a vector space, each of the a_ib_i is in V , and hence $x \in V$, so $\text{span}(B) \subseteq V$.

2. Then we prove that the subset $\text{span}(B)$ is a subspace of V . Let $x, y \in \text{span}(B)$ be two vectors. Then $x = a_1b_1 + a_2b_2 + \dots + a_nb_n$ and $y = c_1b_1 + c_2b_2 + \dots + c_nb_n$ for some $a_i, c_i \in \mathbb{F}$. Then

$$\begin{aligned} 0 &= 0b_1 + 0b_2 + \dots + 0b_n \in \text{span}(B), \\ x + y &= (a_1 + c_1)b_1 + (a_2 + c_2)b_2 + \dots + (a_n + c_n)b_n \in \text{span}(B), \\ \alpha x &= (\alpha a_1)b_1 + (\alpha a_2)b_2 + \dots + (\alpha a_n)b_n \in \text{span}(B) \quad \text{for all } \alpha \in \mathbb{F}. \end{aligned}$$

Hence $\text{span}(B)$ is a subspace of V .

3. Finally, we show that $\text{span}(B)$ is the smallest subspace of V containing B . Suppose not, and let W be a subspace of V containing B such that $W \neq \text{span}(B)$. Then there exists a vector $x \in W$ such that $x \notin \text{span}(B)$. But this is a contradiction, since $B \subseteq \text{span}(B)$ and W is a subspace containing B . Hence $\text{span}(B)$ is the smallest subspace of V containing B .

Note that if $B = \emptyset$, then $\text{span}(B) = \{0\}$, which is a subspace of V . □

It is clear that the statement in the theorem above fails to hold if V is not a vector space; $\text{span}(B)$ is not necessarily a subspace of V if V is not a vector space.

Example 4.3. Let

$$V = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \quad B = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}.$$

Then $\text{span}(B) = \left\{ \begin{pmatrix} a \\ 0 \end{pmatrix} : a \in \mathbb{R} \right\}$. Now:

1. $B \subset V$.

2. $0 \in \text{span}(B)$.

3. For all $x, y \in \text{span}(B)$, $x = \begin{pmatrix} a \\ 0 \end{pmatrix}$ and $y = \begin{pmatrix} b \\ 0 \end{pmatrix}$ for some $a, b \in \mathbb{R}$. Then $x + y = \begin{pmatrix} a + b \\ 0 \end{pmatrix} \in \text{span}(B)$, hence $\text{span}(B)$ is closed under addition.

4. For all $x \in \text{span}(B)$ and $\alpha \in \mathbb{R}$, $x = \begin{pmatrix} a \\ 0 \end{pmatrix}$ for some $a \in \mathbb{R}$. Then $\alpha x = \begin{pmatrix} \alpha a \\ 0 \end{pmatrix} \in \text{span}(B)$, hence $\text{span}(B)$ is closed under scalar multiplication.

So $\text{span}(B)$ is a vector space but not a subspace of V , because V is not a vector space.

Definition 4.4 (Generating set). Let V be a vector space and B be a subset of vectors. If $\text{span}(B) = V$, then B is said to be a generating set for V , or B generates V , or B spans V .

Example 4.5. Let $V = \mathbb{P}_2(\mathbb{R})$ be the space of polynomials of degree less than or equal to 2 with coefficients in $\mathbb{R}$. Let

$$B = \{1, x, x^2\}.$$

Then $\text{span}(B) = \mathbb{P}_2(\mathbb{R})$, so B is a generating set for V .

§5 Lecture 05—01st February, 2022

§5.1 Linear independence

Definition 5.1 (Linear independence). Let $S = \{x_1, x_2, \dots, x_n\}$ be a set of vectors in V . We say that the vectors in S are linearly independent, or S is a linearly independent set of vectors, if

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = 0 \implies a_1 = a_2 = \dots = a_n = 0.$$

If the vectors $x_1, x_2, \dots, x_n$ are not linearly independent, then they are linearly dependent.

Remark 2. The empty set is linearly independent. This is because the statement “ $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0 \implies a_1 = a_2 = \dots = a_n = 0$ ” is vacuously true when $n = 0$.

On the flip side, linearly dependent sets are nonempty.

Example 5.2. The set $\left\{ \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 4 & 6 \end{pmatrix} \right\}$ is linearly independent.

Proof. Let

$$a \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} + b \begin{pmatrix} 3 & 1 \\ 4 & 6 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then we obtain the system

$$\begin{cases} a + 3b = 0, \\ 2a + b = 0, \\ 4b = 0, \\ a + 6b = 0. \end{cases}$$

Solving this system, we find that $a = b = 0$, so the set is linearly independent. $\square$

Example 5.3. Let $\mathbb{P}_3(\mathbb{R})$ be the real-valued polynomials of degree at most 3. Show that the set $\{x^3 + 2x^2, x^2 + 1, -x^3 + 5\}$ in $\mathbb{P}_3(\mathbb{R})$ is linearly independent.

Proof. We have

$$\begin{aligned} a(x^3 + 2x^2) + b(x^2 + 1) + c(-x^3 + 5) = 0 &\implies (a - c)x^3 + (2a + b)x^2 + b + 5c = 0 \\ &\implies \begin{cases} a - c = 0, \\ 2a + b = 0, \\ b + 5c = 0. \end{cases} \\ &\implies a = b = c = 0. \end{aligned}$$

□

Example 5.4. Is the singletons set $\{x\}$ linearly independent in V ? It depends.

1. If $x = 0$, then for any scalar a , $ax = 0$, so the set is linearly dependent.
2. If $x \neq 0$, then suppose that $\{x\}$ were linearly dependent, so that $ax = 0$ for some nonzero scalar $a \neq 0$. Then

$$x = \frac{1}{a}ax = \frac{1}{a}0 = 0,$$

which is a contradiction. Hence the set is linearly independent.

Theorem 5.1 (Characterisation of linear independence). *A set of $n \geq 2$ vectors $\{v_1, \dots, v_n\}$ is linearly independent if and only if one of the vectors, say v_i , can be written as a linear combination of the other vectors, say $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n$.*

Proof. First we prove the forward direction. If $\{v_1, \dots, v_n\}$ is a linearly dependent set, then by definition, there exist scalars $a_1, \dots, a_n$ not all zero such that

$$a_1v_1 + a_2v_2 + \dots + a_nv_n = 0.$$

Suppose without loss of generality that $a_i \neq 0$. Then we can write

$$v_i = -\frac{a_1}{a_i}v_1 - \dots - \frac{a_{i-1}}{a_i}v_{i-1} - \frac{a_{i+1}}{a_i}v_{i+1} - \dots - \frac{a_n}{a_i}v_n = -\frac{1}{a_i} \sum_{\substack{j=1 \\ j \neq i}}^n a_j v_j.$$

Now we prove the reverse direction. Suppose that v_i can be written as a linear combination of the other vectors, say $v_i = \sum_{\substack{j=1 \\ j \neq i}}^n a_j v_j$. Then

$$\sum_{\substack{j=1 \\ j \neq i}}^n a_j v_j - v_i = \sum_{\substack{j=1 \\ j \neq i}}^n a_j v_j + (-1)v_i = 0 \implies a_i = -1 \neq 0,$$

so the set $\{v_1, \dots, v_n\}$ is linearly dependent. This completes the proof. □

Theorem 5.2. Let V be a vector space, and $S_1 \subseteq S_2 \subseteq V$ be two sets of vectors. If S_2 is linearly independent, then S_1 is linearly independent.

Proof. We shall prove the contrapositive of the theorem statement. Let N_1 and N_2 be the number of vectors in S_1 and S_2 respectively, with $N_1 \leq N_2$. If S_1 is not linearly independent, then there exist scalars $a_1, \dots, a_{N_1}$ not all zero such that

$$a_1 v_1 + \dots + a_{N_1} v_{N_1} = 0,$$

where $v_1, \dots, v_{N_1}$ are the vectors in S_1 . Therefore,

$$a_1 v_1 + \dots + a_{N_1} v_{N_1} + 0v_{N_1+1} + \dots + 0v_{N_2} = 0,$$

where $v_{N_1+1}, \dots, v_{N_2}$ are the vectors in S_2 but not in S_1 . Hence S_2 is linearly dependent, so the contrapositive holds. $\square$

Theorem 5.3. Let S be a linearly independent subset of the vector space V . Let $v \neq 0$ be an element of V that is not in S . Then $S \cup \{v\}$ is linearly independent if and only if $v \notin \text{span}(S)$.

Proof. The proof is trivial.

$\implies$ If $S \cup \{v\}$ is linearly independent, then v cannot be written as a linear combination of the vectors in S , so $v \notin \text{span}(S)$.

$\impliedby$ If $v \notin \text{span}(S)$, then v cannot be written as a linear combination of the vectors in S , so $S \cup \{v\}$ is linearly independent. $\square$

Example 5.5. Let $V = \mathbb{R}^3$ and $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$. Then $\text{span}(S)$ is the xy -plane in $\mathbb{R}^3$.

Suppose we wished to add a third vector v to S to S and keep S linearly independent.

1. $v \neq 0$

2. $v \in V$, but $v \notin S$

3. $v \notin \text{span}(S) \implies v \neq a \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ for any $a, b \in \mathbb{R}$. Taking $v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ satisfies all the conditions, so

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

is linearly independent.

Example 5.6. Let $V = \mathbb{R}^3$; find a linearly independent set in V .

1. Take $v_1 \in \mathbb{R}^3$ and $v_1 \neq 0$, then $S = \{v_1\}$ is linearly independent.

2. Take $v_2 \in \mathbb{R}^3$ and $v_2 \neq 0$ and $v_2 \notin \text{span}(\{v_1\})$, then $S = \{v_1, v_2\}$ is linearly independent.
3. Take $v_3 \in \mathbb{R}^3$ and $v_3 \neq 0$ and $v_3 \notin \text{span}(\{v_1, v_2\})$, then $S = \{v_1, v_2, v_3\}$ is linearly independent.
4. If V is finite, then we can continue this process until we have a linearly independent set of vectors in V .

§6 Lecture 05—03rd February, 2022

§6.1 Bases and dimension

One important concept in the study of vector spaces is that of a basis. This is a concept that allows us to make connections between an abstract vector space with $\mathbb{R}^n$; indeed, we will see that any vector space is “equivalent” to $\mathbb{R}^n$ in some sense.

Definition 6.1 (Basis). *Let V be a vector space and $B = \{v_1, v_2, \dots, v_n\}$ be a set of vectors in V . If (i) B is linearly independent and (ii) $\text{span}(B) = V$, then B is called a basis for V .*

Example 6.2. In $\mathbb{R}^{2 \times 2}$, the set

$$B = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

is a basis for $\mathbb{R}^{2 \times 2}$.

Proof. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2 \times 2}$. Then

$$A = a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

so B spans $\mathbb{R}^{2 \times 2}$. To show that B is linearly independent, suppose

$$a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = 0.$$

Then $a = b = c = d = 0$, so B is linearly independent. □

In $\mathbb{R}^{2 \times 2}$, the set

$$B = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 3 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} \right\}$$

is a basis for $\mathbb{R}^{2 \times 2}$.

Proof. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2 \times 2}$. Then

$$A = a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{b}{2} \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} + \frac{c}{3} \begin{pmatrix} 0 & 0 \\ 3 & 0 \end{pmatrix} + \frac{d}{4} \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix},$$

so B spans $\mathbb{R}^{2 \times 2}$. To show that B is linearly independent, suppose

$$a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 3 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix} = 0.$$

Then $a = b = c = d = 0$, so B is linearly independent. $\square$

Example 6.3. 1. In $\mathbb{P}_3(\mathbb{R})$, the set of polynomials of degree no greater than 3, the set $B = \{1, x, x^2, x^3\}$ is a basis.

Proof. Let $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \in \mathbb{P}_3(\mathbb{R})$. Then

$$p(x) = a_0(1) + a_1(x) + a_2(x^2) + a_3(x^3),$$

so B spans $\mathbb{P}_3(\mathbb{R})$. To show that B is linearly independent, suppose

$$a_0(1) + a_1(x) + a_2(x^2) + a_3(x^3) = 0.$$

Then $a_0 = a_1 = a_2 = a_3 = 0$, so B is linearly independent. Hence B is a basis for $\mathbb{P}_3(\mathbb{R})$. $\square$

2. In $\mathbb{P}_3(\mathbb{R})$, the set of polynomials of degree no greater than 3, the set $B = \{1, 2x, x^2, 4x^3\}$ is a basis.

Proof. Let $p(x) = a_0 + 2a_1x + a_2x^2 + 4a_3x^3 \in \mathbb{P}_3(\mathbb{R})$. Then

$$p(x) = a_0(1) + 2a_1(x) + a_2(x^2) + 4a_3(x^3),$$

so B spans $\mathbb{P}_3(\mathbb{R})$. To show that B is linearly independent, suppose

$$a_0(1) + 2a_1(x) + a_2(x^2) + 4a_3(x^3) = 0.$$

Then $a_0 = a_1 = a_2 = a_3 = 0$, so B is linearly independent. Hence B is a basis for $\mathbb{P}_3(\mathbb{R})$. $\square$

It is clear from the above examples that it is possible for a vector space to have more than one basis. We also emphasise that the linear independence of a basis is crucial; if a basis is not linearly independent, then it is not a basis. We also note that the number of vectors in a basis is not unique; a basis can have any number of vectors.

Example 6.4. Take

$$B = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}, \quad B' = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Here $\text{span}(B) = \text{span}(B') = \mathbb{R}^2$. B' is a basis for $\mathbb{R}^2$, but B is not a basis for $\mathbb{R}^2$ because B is linearly dependent.

Theorem 6.1. Let V be a vector space and $B = \{v_1, v_2, \dots, v_n\}$ be a subset of V . Then B is a

basis of V if and only if every $v \in V$ can be uniquely expressed as a linear combination of vectors in B .

Proof. First we prove the forward direction. If B is a basis, then $\text{span}(B) = V$, and therefore every $v \in V$ has $v \in \text{span}(B)$, i.e. v can be written as linear combinations of vectors in B . For uniqueness, let

$$v = a_1v_1 + a_2v_2 + \cdots + a_nv_n = b_1v_1 + b_2v_2 + \cdots + b_nv_n.$$

Then $0 = (a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \cdots + (a_n - b_n)v_n$, and since $v_1, \dots, v_n$ are linearly independent, it follows that $a_1 - b_1 = a_2 - b_2 = \cdots = a_n - b_n = 0$, i.e. $a_i = b_i$ for all i .

For the reverse direction, observe that if every $v \in V$ can be expressed as a linear combination of vectors in B , then $v \in \text{span}(B)$, and hence $V \subseteq \text{span}(B)$. Also because $B \subseteq V$ and V is a vector space, $\text{span}(B)$ is a subspace of V . Therefore, $V \supseteq \text{span}(B)$ hence $V = \text{span}(B)$. Since V is a vector space, $0 \in V$, and 0 can be written as a linear combination of vectors in B such that $0 = a_1v_1 + a_2v_2 + \cdots + a_nv_n$, so that $a_1 = a_2 = \cdots = a_n = 0$ by uniqueness. Therefore, B is linearly independent. Hence B is a basis for V . $\square$

The following result says that if a vector space V is the span of a finite number of vectors, then V has a finite basis.

Theorem 6.2. *Let B be a finite subset of a vector space V such that $\text{span}(B) = V$. Then there exists $B' \subseteq B$ such that B' is a basis for V , and hence V has a finite basis.*

Proof. Consider the case $V = \{0\}$. Then the subset $B = \emptyset$ or $B = \{0\}$ and $\text{span}(B) = \{0\} = V$, hence there exists $B' = \emptyset \subseteq B$ such that B' is a basis for V .

If B is linearly independent and $\text{span}(B) = V$, then there exists $B' = B \subseteq B$ such that B' is a basis for V . So assume B is linearly dependent with n vectors. Choose $v_1, \dots, v_n$ and let $B' = \{v_1, \dots, v_{n'}\}$ with $n' < n$ be the largest set of linearly independent vectors in B . It is clear that B' contains at least one nonzero vector v . Since $V = \text{span}(B)$, then every $v \in V$ can be represented as

$$v = a_1v_1 + a_2v_2 + \cdots + a_{n'}v_{n'} + a_{n'+1}v_{n'+1} + \cdots + a_nv_n = b_1v_1 + b_2v_2 + \cdots + b_{n'}v_{n'} \implies \text{span}(B') = V.$$

Thus B' is a basis for V . $\square$

Example 6.5. Take $V = \mathbb{R}^2$,

$$B = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}.$$

Here B is not a basis of $\mathbb{R}^2$ since it is linearly dependent. Thus there exists $B' \subseteq B$ such that B' is a basis for $\mathbb{R}^2$. For instance, we can take

$$B' = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \quad \text{or} \quad B' = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}, \quad \text{or} \quad B' = \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}.$$

Definition 6.6 (Dimension of a vector space). *A vector space V is said to be finite-dimensional if it has a basis consisting of a finite number of vectors. The unique number of vectors in each basis for V is called the dimension of V , denoted $\dim(V)$.*

Remark 3. 1. If V is not spanned by a finite set, then V is said to be infinite-dimensional.

2. The dimension of $Z = \{0\}$ is 0, with basis $B = \emptyset$. The zero-point is zero-dimensional.

3. The dimension of $\{v\}$ where $v \neq 0$ is undefined, as $\{v\}$ is not a vector space as $0 \notin \{v\}$.

4. The dimension of $V = \text{span}(\{v\})$ where $v \neq 0$ is 1, with basis $B = \{v\}$.

Example 6.7. 1. $\dim(Z) = 0$.

2. $\dim(\mathbb{R}) = 1$, $\dim(\mathbb{R}^n) = n$, $\dim(\mathbb{R}^{m \times n}) = mn$.

3. $\dim(\mathbb{P}_n(\mathbb{R})) = n + 1$, and $\dim(\mathbb{P}(\mathbb{R})) = \infty$.

Theorem 6.3 (Independence of dimension on bases). *Let V be a vector space that has a finite basis. Then every basis for V has the same number of vectors.*

The proof is rather complicated in the general case but it basically says this: the dimension of a finite-dimensional vector space is independent of the choice of basis. That is, all finite bases of a vector space have the same number of linearly independent vectors.

Example 6.8. 1. Over the field of real numbers, the vector space of complex numbers $\mathbb{C}(\mathbb{R})$ has dimension 2, with basis $B = \{1, i\}$; any complex number $c \in \mathbb{C}(\mathbb{R})$ can be expressed as $c = 1a + bi$ for unique $a, b \in \mathbb{R}$.

2. Over the field of complex numbers, the vector space of complex numbers $\mathbb{C}(\mathbb{C})$ has dimension 1, with basis $B = \{1\}$; any complex number $c \in \mathbb{C}(\mathbb{C})$ can be expressed as $c = 1a$ for unique $a \in \mathbb{C}$.

3. Over $\mathbb{C}^2(\mathbb{R})$ the vector space of complex vectors $\left\{ \begin{bmatrix} a_1 + b_1 i \\ a_2 + b_2 i \end{bmatrix} : a_1, a_2, b_1, b_2 \in \mathbb{R} \right\}$ has dimension 4, with basis

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} i \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ i \end{bmatrix} \right\},$$

and for all $v \in \mathbb{C}^2(\mathbb{R})$, v can be expressed as

$$v = a_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + a_3 \begin{bmatrix} i \\ 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 \\ i \end{bmatrix} = \begin{bmatrix} a_1 + a_3 i \\ a_2 + a_4 i \end{bmatrix}.$$

4. $\mathbb{C}(\mathbb{R}^2)$ is not defined because $\mathbb{R}^2$ is not a field; multiplication is not well-defined.

§7 Lecture 07—08th February, 2022

§7.1 Linear transforms (I)

We now study a special type of operators between vector spaces.

Definition 7.1. Let V and W be two vector spaces over $\mathbb{F}$. A function $T: V \rightarrow W$ is called a linear transform if it satisfies the following two properties:

- (i) $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$.
- (ii) $T(\alpha v) = \alpha T(v)$ for all $v \in V$ and $\alpha \in \mathbb{F}$.

Example 7.2. 1. Let $V = \mathbb{R}^2$ and $W = V$. Define $T: V \rightarrow W$ by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 3x \\ 6y - 4x \end{bmatrix}.$$

Then T is a linear transform.

Proof. (i) Let $v_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$. Then

$$\begin{aligned} T(v_1 + v_2) &= T\left(\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}\right) = \begin{bmatrix} 3(x_1 + x_2) \\ 6(y_1 + y_2) - 4(x_1 + x_2) \end{bmatrix} \\ &= \begin{bmatrix} 3x_1 + 3x_2 \\ 6y_1 + 6y_2 - 4x_1 - 4x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 \\ 6y_1 - 4x_1 \end{bmatrix} + \begin{bmatrix} 3x_2 \\ 6y_2 - 4x_2 \end{bmatrix} \\ &= T(v_1) + T(v_2). \end{aligned}$$

(ii) Let $v = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\alpha \in \mathbb{R}$. Then

$$\begin{aligned} T(\alpha v) &= T\left(\begin{bmatrix} \alpha x \\ \alpha y \end{bmatrix}\right) = \begin{bmatrix} 3(\alpha x) \\ 6(\alpha y) - 4(\alpha x) \end{bmatrix} = \begin{bmatrix} \alpha(3x) \\ \alpha(6y - 4x) \end{bmatrix} \\ &= \alpha \begin{bmatrix} 3x \\ 6y - 4x \end{bmatrix} = \alpha T(v). \end{aligned}$$

□

2. Let $V = \mathbb{R}^n$ and $W = \mathbb{R}$. Define $L(v) = v_1$ for any $v = (v_1 \ v_2 \ \cdots \ v_n)^\top \in V$. Then L is a linear transform.

Proof. (i) Let $v_1 = (v_{11} \ v_{12} \ \cdots \ v_{1n})^\top$ and $v_2 = (v_{21} \ v_{22} \ \cdots \ v_{2n})^\top$. Then

$$\begin{aligned} L(v_1 + v_2) &= L\left((v_{11} + v_{21} \ v_{12} + v_{22} \ \cdots \ v_{1n} + v_{2n})^\top\right) \\ &= v_{11} + v_{21} \\ &= L(v_1) + L(v_2). \end{aligned}$$

(ii) Let $v = (v_1 \ v_2 \ \cdots \ v_n)^\top$ and $\alpha \in \mathbb{R}$. Then

$$L(\alpha v) = L\left((\alpha v_1 \ \alpha v_2 \ \cdots \ \alpha v_n)^\top\right) = \alpha v_1 = \alpha L(v).$$

□

Example 7.3. Let $V = \mathbb{R}^m$ and $W = \mathbb{R}^n$. A matrix $A \in \mathbb{R}^{n \times m}$ defines a linear transform $T: V \rightarrow W$ by $T(v) = Av$ for any $v \in V$.

Proof. For $A: V \rightarrow W$, $A(v) = Av = w$ where $v \in V$ and $w \in W$. Then:

1. $A(v_1 + v_2) = Av_1 + Av_2 = w_1 + w_2 = w_1 + w_2 = A(v_1) + A(v_2)$.
2. $A(\alpha v) = A(\alpha v) = \alpha Av = \alpha w = \alpha A(v)$. □

The map $L_A: \mathbb{F}^n \rightarrow \mathbb{F}^m$ defined by $L_A(v) = Av$ is a linear transform for $v \in \mathbb{F}^n$ and $A \in \mathbb{F}^{m \times n}$. Then L_A is a linear transform, called the *left multiplication transformation*.

Example 7.4. For any angle θ , define $T_\theta: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by the following rule: $T_\theta(v)$ is the vector obtained by rotating v counterclockwise by an angle θ if $v \neq 0$, and $T_\theta(0) = 0$. Then T_θ is a linear transform, called the “rotation by θ ” transformation.

Proof. Let α be the angle that $v = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$ makes with the positive x -axis, and let $r = \|v\| = \sqrt{a_1^2 + a_2^2}$. Then

$$v = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = r \begin{bmatrix} \cos(\alpha) \\ \sin(\alpha) \end{bmatrix}.$$

Then v and $T_\theta(v)$ have the same length r , and indeed

$$\begin{aligned} T_\theta \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right) &= \begin{bmatrix} r \cos(\alpha + \theta) \\ r \sin(\alpha + \theta) \end{bmatrix} = \begin{bmatrix} r \cos(\alpha) \cos(\theta) - r \sin(\alpha) \sin(\theta) \\ r \sin(\alpha) \cos(\theta) + r \cos(\alpha) \sin(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} r \cos(\alpha) \\ r \sin(\alpha) \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}}_{\text{rotation matrix}} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \end{aligned}$$

Furthermore, $T_\theta(0) = 0$ since the zero vector has no direction. Thus, $T_\theta: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a left multiplication, hence a linear transformation. □

Example 7.5. Let $V = \mathbb{P}_n(\mathbb{R})$ and $W = \mathbb{P}_{n-1}(\mathbb{R})$. Define $T(f(x)) = f'(x)$ for any $f(x) \in V$. Then T is a linear transformation.

Proof. For all $f(x), g(x) \in V$, we have

$$T(f(x) + g(x)) = (f(x) + g(x))' = f'(x) + g'(x) = T(f(x)) + T(g(x)).$$

For all $\alpha \in \mathbb{R}$ and $f(x) \in V$, we have

$$T(\alpha f(x)) = (\alpha f(x))' = \alpha f'(x) = \alpha T(f(x)).$$

Thus, T is a linear transformation. □

Theorem 7.1. Let V and W be vector spaces and $T: V \rightarrow W$ be a function. Then the following statements hold:

1. If T is linear, then $T(0) = 0$.
2. T is linear if and only if $T(\alpha v_1 + v_2) = \alpha T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$ and $\alpha \in \mathbb{F}$.
3. If T is linear, then $T(x - y) = T(x) - T(y)$ for all $x, y \in V$.
4. T is linear if and only if, for $x_1, x_2, \dots, x_n \in V$ and $a_1, a_2, \dots, a_n \in \mathbb{F}$, we have

$$T\left(\sum_{k=1}^n a_k x_k\right) = \sum_{k=1}^n a_k T(x_k).$$

The proof is easy and left as an exercise for the reader.

Definition 7.6 (Null space and range). Let $T: V \rightarrow W$ be a linear transform. The null space or kernel of T is the set

$$\mathcal{N}(T) = \{v \in V : T(v) = 0\},$$

and the range or image of T is the set

$$\mathcal{R}(T) = \{y \in W : y = T(v) \text{ for some } v \in V\}.$$

Example 7.7. 1. Let $V = \mathbb{P}_2(\mathbb{R})$ and $W = \mathbb{P}_1(\mathbb{R})$. Define $T(f(x)) = f'(x)$ for any $f(x) \in V$. Then $\mathcal{N}(T)$ contains the polynomials whose first derivative is zero, i.e., the constant polynomials. Furthermore, $\mathcal{R}(T) = W$ since the derivative of any quadratic polynomial is a linear polynomial:

$$T(f(x)) = (ax^2 + bx + c)' = 2ax + b.$$

2. Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^2$. Let $T = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$ be a linear transform. Then:

$$\mathcal{N}(T) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \iff \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies x_1 = 0, 2x_1 + 3x_2 = 0 \implies x_2 = 0,$$

$$\mathcal{R}(T) = \mathbb{R}^2 \iff \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \implies \begin{cases} x_1 = y_1, \\ 2x_1 + 3x_2 = y_2 \implies x_2 = \frac{y_2 - 2y_1}{3}. \end{cases}$$

3. Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^2$. Let $T = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ be a linear transform.

$$\mathcal{N}(T) = \left\{ \begin{bmatrix} 0 \\ a \end{bmatrix} : a \in \mathbb{R} \right\} \iff \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies x_1 = 0, x_2 = \text{free},$$

$$\mathcal{R}(T) = \left\{ \begin{bmatrix} a \\ a \end{bmatrix} : a \in \mathbb{R} \right\} \iff \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}.$$

What these examples show is that if T is a matrix, then $\mathcal{R}(T) = \text{col}(T)$.

Theorem 7.2. Let V and W be vector spaces, and let $T: V \rightarrow W$ be a linear transform. Then $\mathcal{N}(T)$ is a subspace of V and $\mathcal{R}(T)$ is a subspace of W .

Proof. Obvious from the definitions. $\square$

Theorem 7.3. Let V and W be vector spaces, and let $T: V \rightarrow W$ be a linear transform. Let $B = \{v_1, \dots, v_n\}$ be a basis for V . Then

$$\mathcal{R}(T) = \text{span}\{T(B)\} \equiv \text{span}\{T(v_1), T(v_2), \dots, T(v_n)\}.$$

Proof. Only a matter of checking that

$$\begin{aligned} \mathcal{R}(T) &= \{T(x) : x \in V\} = \left\{ T \left(\sum_{k=1}^n a_k v_k \right) : a_k \in \mathbb{F} \right\} \\ &= \left\{ \sum_{k=1}^n a_k T(v_k) : a_k \in \mathbb{F} \right\} = \text{span}\{T(v_1), T(v_2), \dots, T(v_n)\} \equiv \text{span}\{T(B)\}. \end{aligned} \quad \square$$

In [Theorem 7.3](#), $B = \{v_1, \dots, v_n\}$ is a basis, therefore $\{v_1, \dots, v_n\}$ is linearly independent and spans V . However, $\{T(v_1), \dots, T(v_n)\}$ may not be linearly independent and may not span $\mathcal{R}(T)$. For example, consider

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad T = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{with basis} \quad B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

$$\text{Then } \mathcal{R}(T) = \text{span}\{T(v_1), T(v_2)\} = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

Example 7.8. Let $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_3(\mathbb{R})$ be the linear transform defined by $T(f(x)) = 2f'(x) + 3 \int_0^x f(t) dt$. Find $\mathcal{R}(T)$.

1. By definition, $f(x) = ax^2 + bx + c \in \mathbb{P}_2(\mathbb{R})$, so

$$\begin{aligned} T(f(x)) &= 2f'(x) + 3 \int_0^x f(t) dt = 2(2ax + b) + 3 \int_0^x (at^2 + bt + c) dt \\ &= 4ax + 2b + 3 \left[\frac{at^3}{3} + \frac{bt^2}{2} + ct \right]_0^x = 4ax + 2b + ax^3 + \frac{3bx^2}{2} + 3cx \\ &= ax^3 + \frac{3b}{2}x^2 + (4a + 3c)x + 2b. \end{aligned}$$

2. By [Theorem 7.3](#): a basis for $\mathbb{P}_2(\mathbb{R})$ is $\{1, x, x^2\}$, so

$$\begin{aligned} T(1) &= 2(0) + 3 \int_0^x 1 dt = 3x, \\ T(x) &= 2(1) + 3 \int_0^x t dt = 2 + 3 \left[\frac{t^2}{2} \right]_0^x = 2 + \frac{3}{2}x^2, \\ T(x^2) &= 2(2x) + 3 \int_0^x t^2 dt = 4x + 3 \left[\frac{t^3}{3} \right]_0^x = 4x + x^3. \end{aligned}$$

Thus, $\mathcal{R}(T) = \text{span}\{3x, 2 + \frac{3}{2}x^2, 4x + x^3\}$.

§8 Lecture 08—10th February, 2022

§8.1 Linear transforms (II): dimension theorem

Definition 8.1 (Nullity and rank). Let V and W be vector spaces, and $T: V \rightarrow W$ be a linear transform. If $\mathcal{N}(T)$ and $\mathcal{R}(T)$ are finite-dimensional, we define the nullity of T , denoted by $\text{nullity}(T)$, as the dimension of $\mathcal{N}(T)$, and the rank of T , denoted by $\text{rank}(T)$, as the dimension of $\mathcal{R}(T)$.

Theorem 8.1 (Dimension theorem). Let V and W be vector spaces, and $T: V \rightarrow W$ be a linear transform. If V is finite-dimensional, then

$$\text{nullity}(T) + \text{rank}(T) = \dim(V).$$

Proof. Let $\dim V = n$ and $\dim(\mathcal{N}(T)) = k$; we will show that $\text{rank}(T) = n - k$. Assume $\{v_1, v_2, \dots, v_k\}$ is a basis for $\mathcal{N}(T)$, so that $T(v_j) = 0$ for $1 \leq j \leq k$. Since $\mathcal{N}(T)$ is a subspace of V , we can extend the basis of $\mathcal{N}(T)$ to a basis of V , say $\{v_1, v_2, \dots, v_k, v_{k+1}, \dots, v_n\}$. We now prove that $\{T(v_{k+1}), T(v_{k+2}), \dots, T(v_n)\}$ forms a basis of $\mathcal{R}(T)$ which means that $\text{rank}(T) = n - k$ and completes the proof.

1. The basis of $V = \{v_1, v_2, \dots, v_n\}$, so that

$$\mathcal{R}(T) = \text{span}(T(v_1), \dots, T(v_k), \dots, T(v_n)) = \text{span}(T(v_{k+1}), \dots, T(v_n)),$$

since $T(v_j) = 0$ for $1 \leq j \leq k$.

2. Now let $\sum_{j=k+1}^n a_j T(v_j) = 0$, since T is a linear transform, so that $T\left(\sum_{j=k+1}^n a_j v_j\right) = 0$, and therefore $\sum_{j=k+1}^n a_j v_j \in \mathcal{N}(T)$. That is, $\sum_{j=k+1}^n a_j v_j$ can be written as a linear combination of $\{v_1, v_2, \dots, v_k\}$ since it is a basis of $\mathcal{N}(T)$, so we write

$$\sum_{j=k+1}^n a_j v_j = \sum_{j=1}^k c_j v_j \implies \sum_{j=k+1}^n (-a_j) v_j + \sum_{j=1}^k c_j v_j = 0.$$

Observe that $\{v_1, v_2, \dots, v_n\}$ is linearly independent since it is a basis of V . This implies $c_1 = \dots = c_k = a_{k+1} = \dots = a_n = 0$. Therefore $\{T(v_{k+1}), T(v_{k+2}), \dots, T(v_n)\}$ is linearly independent.

Hence $\{T(v_{k+1}), T(v_{k+2}), \dots, T(v_n)\}$ is a basis of $\mathcal{R}(T)$, so that $\text{rank}(T) = n - k$, and the dimension theorem is proved. $\square$

Recall that for vector spaces V, W , we say that

1. $T: V \rightarrow W$ is *injective* (or *one-to-one*) if $T(v) = T(w) \implies v = w$ for all $v, w \in V$.
2. $T: V \rightarrow W$ is *surjective* (or *onto*) if for all $w \in W$, there exists $v \in V$ such that $T(v) = w$.

Example 8.2. 1. Consider $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2 \times 2}$. Then T is injective and

surjective:

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

2. $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $T = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2 \times 2}$. Then T is not injective:

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ a \end{pmatrix}$$

and also is not surjective:

$$\mathcal{R}(T) = \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \neq \mathbb{R}^2.$$

Example 8.3 (Nonlinear transform). 1. The map $T: \mathbb{R} \rightarrow \mathbb{R}$ defined by $T(x) = x^2$ is not linear, since $T(0) = 0$ but $T(0) \neq 0$. It is also not injective, since $T(-1) = T(1) = 1$, and not surjective, since $T(x) \geq 0$ for all $x \in \mathbb{R}$.

2. For $\mathbb{R}_0 = [0, \infty)$, the map $T: \mathbb{R} \rightarrow \mathbb{R}_0$ defined by $T(x) = x^2$ is not linear, since $T(0) = 0$ but $T(0) \neq 0$. It is also not injective, since $T(-1) = T(1) = 1$, but it is surjective, since for all $y \in \mathbb{R}_0$, there exists $x = \sqrt{y} \in \mathbb{R}$ such that $T(x) = y$.

Theorem 8.2. Let V and W be vector spaces, and $T: V \rightarrow W$ be a linear transform. Then T is injective if and only if $\mathcal{N}(T) = \{0\}$.

Proof. $\implies$ T is linear, so $T(0) = 0 \implies 0 \in \mathcal{N}(T)$. If T is one-to-one, then $x \neq 0$ means $T(x) \neq T(0) = 0$, so $\mathcal{N}(T) = \{0\}$.

$\impliedby$ If $\mathcal{N}(T) = \{0\}$, then $T(x) = 0$ and $T(x) \neq 0$ if $x \neq 0$. Let $x = x_1 - x_2 \neq 0$, so that $T(x_1 - x_2) \neq 0 \implies T(x_1) \neq T(x_2)$. Hence T is injective. $\square$

Example 8.4. Take $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2 \times 2}$. We claim this is both injective and surjective.

Proof. For any $T \in \mathbb{R}^{n \times n}$, if T is invertible, then

$$Tx = 0 \implies x = T^{-1}0 = 0, \implies \mathcal{N}(T) = \{0\} \iff T \text{ is injective by Theorem 8.2.}$$

On the other hand,

$$\begin{aligned} Tx = b &\implies x = T^{-1}b : \text{ for all } b \in W = \mathbb{R}^n, \text{ there exists } x \in V = \mathbb{R}^n, \\ &\text{s.t. } Tx = b \implies T \text{ is onto.} \end{aligned}$$

$\square$

Theorem 8.3. Let V and W be finite-dimensional vector spaces of the same dimension, and let $T: V \rightarrow W$ be a linear transform. Then the following statements are equivalent:

1. T is injective.
2. T is surjective.
3. $\text{rank}(T) = \dim(V)$.

Proof. We proceed directly.

$$\begin{array}{llll}
T \text{ is one to one} & \iff \mathcal{N}(T) = \{0\} & & \text{by Theorem 8.2} \\
& \iff \text{nullity}(T) = 0 & & \text{by definition of nullity} \\
& \iff \text{rank}(T) = \dim(V) & & \text{by dimension theorem} \\
& \iff \dim(\mathcal{R}(T)) = \dim(V) & & \text{by definition of rank} \\
& \iff \mathcal{R}(T) = W & & \text{by } \begin{cases} \mathcal{R}(T) \text{ is a subspace of } W, \\ \dim(\mathcal{R}(T)) = \dim(W) \end{cases} \\
& \iff T \text{ is onto} & & \text{by definition of surjective.} \quad \square
\end{array}$$

Example 8.5. Consider the two maps:

1. $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $T = A \in \mathbb{R}^{2 \times 2}$, where T is surjective but not injective.
2. $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $T = B \in \mathbb{R}^{2 \times 2}$, where T is injective but not surjective.

Since $V = W = \mathbb{R}^n$, and T is a linear transform, there is no such A or B that satisfies the conditions, by Theorem 8.3.

In Theorem 8.3, the result says that being injective and being surjective are equivalent in the finite-dimensional case. It is very important to note that this finite-dimension requirement cannot be removed; in the infinite-dimensional case, injective linear transforms can fail to be surjective, and surjective linear transforms can fail to be injective.

Example 8.6. Consider $T: \mathbb{P}(\mathbb{R}) \rightarrow \mathbb{P}(\mathbb{R})$ defined by $T(f(x)) = \int_0^x f(t) dt$. Then T is linear and injective, but not surjective. (This also gives an example where $\mathcal{R}(T)$ is infinite-dimensional.)

The following result essentially says that a linear transform is completely determined by its action on bases.

Theorem 8.4. Let V and W be vector spaces over $\mathbb{F}$, and suppose that $\{v_1, \dots, v_n\}$ is a basis for V . Then for $w_1, \dots, w_n \in W$, there exists exactly one linear transform $T: V \rightarrow W$ such that $T(v_i) = w_i$ for all $i = 1, \dots, n$.

Proof. We first construct a linear transform that satisfies the required property, and then show that this linear transform is unique. For any $v \in V$, there exist unique scalars $\{a_1, \dots, a_n\}$ such that $v = \sum_{i=1}^n a_i v_i$. Construct $T(v) = \sum_{i=1}^n a_i w_i$. Take $v = v_i$ (with $a_i = 1$ and $a_j = 0$ for $j \neq i$), then $T(v_i) = w_i$ for all $i = 1, \dots, n$. Thus, T is uniquely determined by its action on the basis vectors.

For linearity, check that for $v = \sum_{i=1}^n a_i v_i \in V$ and $u = \sum_{i=1}^n b_i v_i \in V$, $\alpha \in \mathbb{F}$, and thus $T(v) = \sum_{i=1}^n a_i w_i$ and $T(u) = \sum_{i=1}^n b_i w_i$, after which we get

$$T(v+u) = T\left(\sum_{i=1}^n (a_i + b_i) v_i\right) = \sum_{i=1}^n (a_i + b_i) w_i = \sum_{i=1}^n a_i w_i + \sum_{i=1}^n b_i w_i = T(v) + T(u),$$

and $T(\alpha v) = T(\sum_{i=1}^n (\alpha a_i) v_i) = \sum_{i=1}^n (\alpha a_i) w_i = \alpha \sum_{i=1}^n a_i w_i = \alpha T(v)$. Thus, T is linear. Along the same light, one may also prove that $T(\alpha v + w) = \alpha T(v) + T(w)$ for all $w \in W$.

So we only need prove uniqueness. Let T and T' be two linear transforms satisfying the required property. Then $T(v_i) = T'(v_i) = w_i$ for all $i = 1, \dots, n$. For any $v \in V$, there exist unique scalars $\{a_1, \dots, a_n\}$ such that $v = \sum_{i=1}^n a_i v_i$. Thus,

$$T(v) = T\left(\sum_{i=1}^n a_i v_i\right) = \sum_{i=1}^n a_i T(v_i) = \sum_{i=1}^n a_i w_i \quad T'(v) = T'\left(\sum_{i=1}^n a_i v_i\right) = \sum_{i=1}^n a_i T'(v_i) = \sum_{i=1}^n a_i w_i,$$

so that $T(v) = T'(v)$ for all $v \in V$, and hence $T = T'$. $\square$

§9 Lecture 09—15th February, 2022

§9.1 The ordered basis and coordinates

Definition 9.1. Let V be a finite-dimensional vector space. An ordered basis for V is a basis for V endowed with a total ordering on its elements.

The definition basically says that an ordered basis for V is a finite sequence of linearly independent vectors in V that generates V .

Example 9.2. Let $V = \mathbb{R}^3$. Then $\alpha = (e_1, e_2, e_3)$ and $\beta = (e_2, e_1, e_3)$ are two ordered bases for V , and $\alpha \neq \beta$ as ordered bases.

Definition 9.3. Let $B = (v_1, v_2, \dots, v_n)$ be an ordered basis for an n -dimensional vector space V over a field $\mathbb{F}$. Then every $v \in V$ has the unique representation $v = \sum_{k=1}^n a_k v_k$ for some $a_k \in \mathbb{F}$. The coordinates of $v \in V$ related to the basis B , denoted $[v]_B$, is the vector in $\mathbb{F}^n$ defined by

$$[v]_B = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Take B an ordered basis for V over $\mathbb{F}$. Then for $v \in V$, we have $v = B[v]_B = \sum_{k=1}^n a_k v_k$.

Example 9.4. Let $V = \mathbb{P}_2(\mathbb{R})$ and $\beta = \{x, 1, x^2\}$ be an ordered basis for V . Let $f(x) =$

$4 + 6x + 2x^2$. Then

$$[f(x)]_\beta = \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix}, \quad [f(x)]_\beta = \beta[f]_\beta = \begin{bmatrix} x & 1 & x^2 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \\ 2 \end{bmatrix} = 6x + 4 + 2x^2.$$

Example 9.5. 1. $B = \{e_1, e_2, e_3\}$ is a basis for $\mathbb{R}^3$. Then $v = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}$ has coordinates

$$[v]_B = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}.$$

2. $B' = \left\{ \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is another basis for $\mathbb{R}^3$. Then

$$v = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \implies [v]_{B'} = \begin{bmatrix} \frac{1}{2} \\ 1 \\ 3 \end{bmatrix}.$$

3. Let B and B' be the matrices with basis vectors as columns. Then we have

$$B[v]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = v = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1 \\ 3 \end{bmatrix} = B'[v]_{B'}.$$

Theorem 9.1. Let B be an ordered basis for a vector space V over a field $\mathbb{F}$.

(i) For all $u, v \in V$ and $a \in \mathbb{F}$, we have

$$[u + v]_B = [u]_B + [v]_B, \quad [au]_B = a[u]_B,$$

and hence $[\cdot]_B$ is a linear transform from $V(\mathbb{F})$ to $\mathbb{F}^{\dim(V)}$.

- (ii) Let $v_1, \dots, v_n \in V$ and $u \in V$. Then u is a linear combination of v_i 's if and only if $[u]_B$ is a linear combination of $[v_i]_B$'s.
- (iii) Let $v_1, \dots, v_n \in V$. Then $\{v_1, \dots, v_n\}$ is linearly independent in V iff $\{[v_1]_B, \dots, [v_n]_B\}$ is linearly independent in $\mathbb{F}^{\dim(V)}$.

Proof. (i) The proof is trivial and follows from the definitions.

- (ii) If u is a linear combination of the v_i , then $u = a_1v_1 + \dots + a_nv_n$ for some $a_i \in \mathbb{F}$. Thus $[u]_B = [a_1v_1 + \dots + a_nv_n]_B = a_1[v_1]_B + \dots + a_n[v_n]_B$, so $[u]_B$ is a linear combination of the $[v_i]_B$'s. Conversely, if $[u]_B$ is a linear combination of the $[v_i]_B$'s, then $[u]_B = a_1[v_1]_B + \dots + a_n[v_n]_B$ for some $a_i \in \mathbb{F}$. Thus $u = B[u]_B = B(a_1[v_1]_B + \dots + a_n[v_n]_B) = a_1v_1 + \dots + a_nv_n$, so u is a linear combination of the v_i 's.

- (iii) For the forward direction, suppose $\{v_1, \dots, v_n\}$ is linearly independent in V . If $a_1[v_1]_B + \dots + a_n[v_n]_B = 0$ for some $a_i \in \mathbb{F}$, then $B(a_1[v_1]_B + \dots + a_n[v_n]_B) = a_1v_1 + \dots + a_nv_n = 0$. Since $\{v_1, \dots, v_n\}$ is linearly independent, we have $a_i = 0$ for all i . Thus $\{[v_1]_B, \dots, [v_n]_B\}$ is linearly independent in $\mathbb{F}^{\dim(V)}$. Conversely, suppose $\{[v_1]_B, \dots, [v_n]_B\}$ is linearly independent in $\mathbb{F}^{\dim(V)}$. If $a_1v_1 + \dots + a_nv_n = 0$ for some $a_i \in \mathbb{F}$, then $[a_1v_1 + \dots + a_nv_n]_B = a_1[v_1]_B + \dots + a_n[v_n]_B = 0$. Since $\{[v_1]_B, \dots, [v_n]_B\}$ is linearly independent, we have $a_i = 0$ for all i . Thus $\{v_1, \dots, v_n\}$ is linearly independent in V . $\square$

Example 9.6. Consider $f, g, q \in \mathbb{P}_2(\mathbb{R})$ where $f(t) = 1 + 2t$, $g(t) = 1 + t + t^2$, and $q(t) = t^2$. Determine if these vectors are linearly independent.

Solution. 1. By definition,

$$\begin{aligned} a(1 + 2t) + b(1 + t + t^2) + c(t^2) = 0 &\iff (a + b) + (2a + b)t + (b + c)t^2 = 0 \\ &\implies a + b = 0, 2a + b = 0, b + c = 0. \end{aligned}$$

Solving this system of equations, we find that $a = 0$, $b = 0$, and $c = 0$. Thus, the vectors are linearly independent.

2. By [Theorem 9.1](#). Let $B = \{1, t, t^2\}$. Then

$$[f]_B = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad [g]_B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad [q]_B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Since $[f]_B, [g]_B, [q]_B$ are linearly independent in $\mathbb{R}^3$, it follows that f, g, q are also linearly independent in $\mathbb{P}_2(\mathbb{R})$. $\square$

Remark 4. The basis is essentially a concept that can be used to make a connection between a vector space $V(\mathbb{F})$ and its coordinate representation in $\mathbb{F}^n$.

§10 Lecture 10—17th February, 2022

§10.1 The matrix representation of a linear transformation

Definition 10.1. Let $\alpha = (v_1, \dots, v_n)$ be an ordered basis for an n -dimensional vector space V , $\beta = (w_1, \dots, w_m)$ be an ordered basis for an m -dimensional vector space W , and $T: V \rightarrow W$ be a linear transformation. Then the matrix representation of T in the ordered basis α and β , denoted by $[T]_{\alpha}^{\beta}$, is the matrix whose elements are determined by $T(v_j) = \sum_{i=1}^m a_{ij}w_i$ for $j = 1, \dots, n$.

These coefficients form the j -th column of the matrix $[T]_{\alpha}^{\beta}$.

Example 10.2. Let $V = W = \mathbb{R}^2$ and define T by $Tx = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Let $\beta = (e_1, e_2)$ and

$\gamma = (e_2, e_1)$ be ordered bases for $\mathbb{R}^2$. Find $[T]_\beta^\beta, [T]_\beta^\gamma, [T]_\gamma^\beta, [T]_\gamma^\gamma$.

Solution. We have

$$T(e_1) = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = e_1 + 3e_2, \quad T(e_2) = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2e_1 + e_2.$$

Thus,

$$[T]_\beta^\beta = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}, \quad [T]_\beta^\gamma = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}, \quad [T]_\gamma^\gamma = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}, \quad [T]_\gamma^\beta = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}. \quad \square$$

Write $[T]_\beta^\beta = [T]_\beta$.

Theorem 10.1. $[T]_\alpha^\beta[v]_\alpha = w_\beta$.

Proof. For $T: V \rightarrow W$, we have $T(v) = w$, where $v = \sum_{j=1}^n \alpha_j v_j$ and $w = \sum_{i=1}^m \beta_i w_i$. Thus

$$\begin{aligned} T\left(\sum_{j=1}^n \alpha_j v_j\right) = w &\implies \sum_{j=1}^n \alpha_j T(v_j) = w \implies \sum_{j=1}^n \alpha_j \left(\sum_{i=1}^m a_{ij} w_i\right) = w \\ &\implies \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \alpha_j\right) w_i = \sum_{i=1}^m \beta_i w_i. \end{aligned}$$

The coefficient for each w_i must match, so we have $\sum_{j=1}^n a_{ij} \alpha_j = \beta_i$ for all $i = 1, \dots, m$. This gives us the desired result. $\square$

Example 10.3. Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ be the linear transform defined by

$$T\left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\right) = \begin{bmatrix} a_1 + 3a_2 \\ 0 \\ 2a_1 - 4a_2 \end{bmatrix}.$$

Let β and γ be the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Find $[T]_\beta^\gamma$ and verify that $[T]_\beta^\gamma[v]_\beta = w_\gamma$.

Solution. We first compute

$$\begin{cases} T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = 1e_1 + 0e_2 + 2e_3, \\ T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 0 \\ -4 \end{bmatrix} = 3e_1 + 0e_2 - 4e_3. \end{cases} \implies [T]_\beta^\gamma = \begin{bmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{bmatrix}.$$

For the rest, just use the other ordered bases. It's pretty standard and obvious. $\square$

Theorem 10.2. Let V and W be vector spaces over a field $\mathbb{F}$. For linear transforms $T, \tilde{T}: V \rightarrow W$ and a scalar $\alpha \in \mathbb{F}$, we define the linear transforms $T + \tilde{T}$ and αT by

$$(T + \tilde{T})(v) = T(v) + \tilde{T}(v), \quad (\alpha T)(v) = \alpha T(v)$$

for all $v \in V$. Then the following statements hold:

1. for all $\alpha \in \mathbb{F}$, $\alpha T + \tilde{T}$ is linear.
2. the set of all linear transforms from V to W , denoted by $\mathcal{L}(V, W)$, is a vector space over $\mathbb{F}$.

Proof. Obvious, as it is a straightforward application of the definitions. □

The following result can be used to simplify some calculations.

Theorem 10.3. Let V and W be finite-dimensional vector spaces with ordered bases β and γ respectively. Let $T, \tilde{T}: V \rightarrow W$ be linear transforms. Then

1. $[T + \tilde{T}]_{\beta}^{\gamma} = [T]_{\beta}^{\gamma} + [\tilde{T}]_{\beta}^{\gamma}$,
2. $[\alpha T]_{\beta}^{\gamma} = \alpha [T]_{\beta}^{\gamma}$ for all $\alpha \in \mathbb{F}$.

Proof. Obvious, as it is a straightforward application of the definitions. □

Example 10.4. Let $T, U: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ be linear transforms respectively defined by

$$T \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right) = \begin{bmatrix} a_1 + 3a_2 \\ 0 \\ 2a_1 - 4a_2 \end{bmatrix}, \quad \text{and} \quad U \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right) = \begin{bmatrix} a_1 - a_2 \\ 2a_1 \\ 3a_1 + 2a_2 \end{bmatrix}.$$

Let β and γ be the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Find $[T + U]_{\beta}^{\gamma}$.

Solution. By definition,

$$(T + U) \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right) = \begin{bmatrix} a_1 + 3a_2 \\ 0 \\ 2a_1 - 4a_2 \end{bmatrix} + \begin{bmatrix} a_1 - a_2 \\ 2a_1 \\ 3a_1 + 2a_2 \end{bmatrix} = \begin{bmatrix} 2a_1 + 2a_2 \\ 2a_1 \\ 5a_1 - 2a_2 \end{bmatrix},$$

and evaluating at the bases gives

$$(T + U) \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 2 \\ 5 \end{bmatrix}, \quad (T + U) \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \implies [T + U]_{\beta}^{\gamma} = \begin{bmatrix} 2 & 2 \\ 2 & 0 \\ 5 & -2 \end{bmatrix}.$$

But by the above theorem, we get

$$[T + U]_{\beta}^{\gamma} = \begin{bmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 0 \\ 5 & -2 \end{bmatrix}. \quad \square$$

§11 Lecture 11—22nd February, 2022

§11.1 Composition of linear transformations

We will now examine the composition of linear transforms. The following result is natural:

Theorem 11.1. *Let V, W , and Z be vector spaces over the same field $\mathbb{F}$. Let $T: V \rightarrow W$ and $U: W \rightarrow Z$ be linear transforms. Then the composition $U \circ T: V \rightarrow Z$ is also a linear transform, where $(U \circ T)(v) = U(T(v))$ for all $v \in V$.*

Proof. This follows from the fact that

$$\begin{aligned}(U \circ T)(cx + y) &= U(T(cx + y)) = U(cT(x) + T(y)) = cU(T(x)) + U(T(y)) \\ &= c(U \circ T)(x) + (U \circ T)(y).\end{aligned}$$

□

Theorem 11.2. *Let V be a vector space and $T, U_1, U_2 \in \mathcal{L}(V)$ linear transforms on V . Then we have*

1. $T(U_1 + U_2) = TU_1 + TU_2$ and $(U_1 + U_2)T = U_1T + U_2T$.
2. $T(U_1U_2) = (TU_1)U_2 = U_1(TU_2)$.
3. $TI_V = I_VT = T$, where $I_V: V \rightarrow V$ is the identity transform that takes each vector to itself.
4. For every $\lambda \in \mathbb{F}$, we have $\lambda(U_1U_2) = (\lambda U_1)U_2 = U_1(\lambda U_2)$.

Proof. Each result follows from the definitions:

1. For all $v \in V$, we have

$$\begin{aligned}T(U_1 + U_2)(v) &= T(U_1(v) + U_2(v)) = TU_1(v) + TU_2(v) = (TU_1 + TU_2)(v), \\ (U_1 + U_2)T(v) &= U_1(T(v)) + U_2(T(v)) = U_1T(v) + U_2T(v) = (U_1T + U_2T)(v).\end{aligned}$$

2. For all $v \in V$, we have

$$\begin{aligned}T(U_1U_2)(v) &= T(U_1(U_2(v))) = TU_1(U_2(v)) = (TU_1)(U_2(v)) = (TU_1)U_2(v), \\ (U_1U_2)T(v) &= U_1(U_2(T(v))) = U_1(U_2T(v)) = U_1(TU_2(v)) = U_1(TU_2)(v).\end{aligned}$$

3. For all $v \in V$, we have $TI_V(v) = T(v) = (TI_V)(v)$ since $I_V(v) = v$ for all $v \in V$.

4. For every $\lambda \in \mathbb{F}$, we have $\lambda(U_1U_2) = (\lambda U_1)U_2 = U_1(\lambda U_2)$.

□

Theorem 11.3 (Matrix representation of compositions). *Let V, W , and Z be finite-dimensional (with dimensions n, m , and ℓ respectively) vector spaces over the same field $\mathbb{F}$. Let $T: V \rightarrow W$ and $U: W \rightarrow Z$ be linear transforms, and let α, β, γ be ordered bases of V, W , and Z respectively. Then $[U \circ T]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta$.*

Sketch of proof. Let $[T]_\alpha^\beta = (t_{ij}) \in \mathbb{F}^{m \times n}$ and $[U]_\beta^\gamma = (u_{jk}) \in \mathbb{F}^{\ell \times m}$. Then for all $v \in V$, we have

$$[U \circ T]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta = (u_{jk})(t_{ij}) = \sum_{r=1}^m u_{jr} t_{ir} \in \mathbb{F}^{\ell \times n}.$$

The matrix representation of the composition $U \circ T$ with respect to the bases α and γ is given by the matrix product of the individual representations. $\square$

Corollary 11.1. *Let V be a finite-dimensional vector space with ordered basis β . Let $T, U \in \mathcal{L}(V, V)$. Then $[UT]_\beta = [U]_\beta [T]_\beta$.*

Proof. We have $[UT]_\beta = [U]_\beta [T]_\beta$ by the previous theorem, since U and T are both linear transforms on V . $\square$

Corollary 11.2. *Let V, W be finite-dimensional vector spaces over $\mathbb{F}$ with ordered bases β and γ respectively. Let $T \in \mathcal{L}(V, W)$. Then $[T(v)]_\gamma = [T]_\beta^\gamma [v]_\beta$ for all $v \in V$.*

Proof. Fix $v \in V$, and define linear transforms

$$\begin{aligned} f: \mathbb{F} &\rightarrow V, & f(a) &= av, & a \in \mathbb{F}, v \in V \\ g: \mathbb{F} &\rightarrow W, & g(a) &= T(f(a)) = T(av) = aT(v), & a \in \mathbb{F}, v \in V \end{aligned}$$

Clearly $g = T \circ f$. Let $\alpha = \{1\}$ be the standard basis for $\mathbb{F}$. Then we have

$$[T(v)]_\gamma = [g(1)]_\gamma = [g]_\alpha^\gamma [1]_\alpha = [T \circ f]_\alpha^\gamma [1]_\alpha = [T]_\beta^\gamma [f]_\alpha^\beta [1]_\alpha = [T]_\beta^\gamma [v]_\beta. \quad \square$$

Example 11.3. Let $U: \mathbb{P}_3(\mathbb{R}) \rightarrow \mathbb{P}_2(\mathbb{R})$ and $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_3(\mathbb{R})$ be the linear transforms defined by $U(f) = f'$ and $T(g) = \int_0^x g(t)dt$ respectively. Let α and β be the standard ordered basis for $\mathbb{P}_3(\mathbb{R})$ and $\mathbb{P}_2(\mathbb{R})$ respectively, i.e. $\alpha = \{1, x, x^2, x^3\}$ and $\beta = \{1, x, x^2\}$. Find $[UT]_\beta^\beta$, and $[TU]_\alpha^\alpha$.

Solution. It is easy to show that

$$[T]_\beta^\alpha = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}, \quad [U]_\alpha^\beta = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

This gives

$$[UT]_\beta^\beta = [U]_\alpha^\beta [T]_\beta^\alpha = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3,$$

which is consistent with our intuition that the differentiating an integrated polynomial should

yield the original polynomial. However,

$$[TU]_{\alpha}^{\alpha} = [T]_{\beta}^{\alpha} [U]_{\alpha}^{\beta} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad \square$$

Theorem 11.4. Take $A \in \mathbb{F}^{m \times n}$, $L_A: \mathbb{F}^n \rightarrow \mathbb{F}^m$, $B \in \mathbb{F}^{m \times n}$, and let β and γ be standard ordered bases of $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively. Then:

1. L_A is linear.
2. $[L_A]_{\gamma}^{\beta} = A$.
3. $L_A = L_B$ if and only if $A = B$.
4. $L_{A+B} = L_A + L_B$, and $L_{\alpha \cdot A} = \alpha \cdot L_A$ for all $\alpha \in \mathbb{F}$.
5. If E is an $n \times p$ matrix, then $L_{AE} = L_A \circ L_E$.
6. If $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$ is linear, then there exists a unique $C \in \mathbb{F}^{m \times n}$ such that $T = L_C$. In fact, we have $[T]_{\gamma}^{\beta} = C$.
7. If $m = n$, then $L_{I_n} = I_{\mathbb{F}^n}$.

The proof is straightforward and left as a homework exercise.

§12 Lecture 12—24th February, 2022

§12.1 Invertibility and isomorphisms (I)

Definition 12.1 (Invertible matrix and inverse). Let $A \in \mathbb{R}^{n \times n}$. If there exists an $n \times n$ matrix B such that $AB = BA = I_n$, then B is the inverse of A , denoted A^{-1} , and A is invertible (or non-singular). If no such B exists, then A is non-invertible (or singular).

Proposition 12.2 (Properties of inverse matrices). 1. A singular matrix has no inverse.

2. Both A and A^{-1} are square matrices.
3. If A^{-1} exists, then it is unique.
4. Both A and A^{-1} are invertible, and $(A^{-1})^{-1} = A$.
5. If A and B are invertible matrices of the same size, then so is AB , and $(AB)^{-1} = B^{-1}A^{-1}$.

How might we calculate the inverse? One way is called *Gauss-Jordan elimination*, described below:

1. Construct an augmented matrix $[A \ I_n]$.
2. Use elementary row operations to get the reduced row echelon form $[I_n \ A^{-1}]$, so that

$$A^{-1} [A \ I_n] = [A^{-1}A \ A^{-1}I_n] = [I_n \ A^{-1}].$$

From this, the proof of correctness of the Gauss-Jordan elimination method is immediate.

Example 12.3. Find the inverse of the following matrix, if it exists:

$$A = \begin{bmatrix} 1 & 2 & -3 \\ 0 & 1 & 4 \\ 5 & -2 & -3 \end{bmatrix}.$$

Solution. The elimination technique here gives:

$$\begin{aligned} \left[\begin{array}{ccc|ccc} 1 & 2 & -3 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 & 1 & 0 \\ 5 & -2 & -3 & 0 & 0 & 1 \end{array} \right] & \xrightarrow{\substack{R_2 \leftarrow R_2 - R_1, \\ R_3 \leftarrow R_3 - 5R_1}} \left[\begin{array}{ccc|ccc} 1 & 2 & -3 & 1 & 0 & 0 \\ 0 & -4 & 4 & -1 & 1 & 0 \\ 0 & -12 & 12 & -5 & 0 & 1 \end{array} \right] \\ & \xrightarrow{R_3 \leftarrow R_3 - 3R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & -3 & 1 & 0 & 0 \\ 0 & -4 & 4 & -1 & 1 & 0 \\ 0 & 0 & 0 & -2 & -3 & 1 \end{array} \right]. \end{aligned}$$

Since the left 3×3 block has a zero row, we conclude that $\text{rank}(A) < 3$. Therefore, A is singular and A^{-1} does not exist. $\square$

Example 12.4. Find the inverse of the following matrix, if it exists:

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

Solution. The elimination technique here gives:

$$\begin{aligned} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 3 & 0 & 1 & 0 \\ 5 & 5 & 1 & 0 & 0 & 1 \end{array} \right] & \xrightarrow{R_3 \leftarrow R_3 - 5R_1} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 3 & 0 & 1 & 0 \\ 0 & 0 & -4 & -5 & 0 & 1 \end{array} \right] \\ & \xrightarrow{\substack{R_2 \leftarrow \frac{1}{2}R_2, \\ R_3 \leftarrow -\frac{1}{4}R_3}} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & \frac{3}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{5}{4} & 0 & -\frac{1}{4} \end{array} \right] \\ & \xrightarrow{\substack{R_2 \leftarrow R_2 - \frac{3}{2}R_3, \\ R_1 \leftarrow R_1 - R_3}} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 & -\frac{15}{8} & \frac{1}{2} & \frac{3}{8} \\ 0 & 0 & 1 & \frac{5}{4} & 0 & -\frac{1}{4} \end{array} \right] \\ & \xrightarrow{R_1 \leftarrow R_1 - R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{13}{8} & -\frac{1}{2} & -\frac{1}{8} \\ 0 & 1 & 0 & -\frac{15}{8} & \frac{1}{2} & \frac{3}{8} \\ 0 & 0 & 1 & \frac{5}{4} & 0 & -\frac{1}{4} \end{array} \right]. \end{aligned}$$

Thus, we have

$$A^{-1} = \begin{bmatrix} \frac{13}{8} & -\frac{1}{2} & -\frac{1}{8} \\ -\frac{15}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{5}{4} & 0 & -\frac{1}{4} \end{bmatrix}.$$

$\square$

Now we will discuss several questions related to the invertibility of linear transforms.

Definition 12.5. Let $T: V \rightarrow W$ be a linear transform. If there exists a transform $T^{-1}: W \rightarrow V$ such that $T^{-1} \circ T = I_V$ and $T \circ T^{-1} = I_W$, then T^{-1} is the inverse of T , and T is invertible.

Remark 5. Here's one way to see this: If T is invertible, then $T^{-1} \circ T(v) = T^{-1}(T(v)) = v$ for all $v \in V$, and so we have the maps $V \rightarrow W \rightarrow V$. Likewise, $T \circ T^{-1}(w) = T(T^{-1}(w)) = w$ for all $w \in W$, giving the maps $W \rightarrow V \rightarrow W$. So the inverse effectively “undoes” the action of T , and vice versa.

Theorem 12.1. Let $T: V \rightarrow W$ and $U: W \rightarrow Z$ be two linear transforms. Then the following statements hold:

1. Not every linear transform has an inverse; in particular, a linear transform has an inverse if and only if it is both one-to-one and onto.
2. If a linear transform is invertible, then $\dim(V) = \dim(W) = \text{rank}(T)$.
3. If T^{-1} exists, then it is unique.
4. $(T^{-1})^{-1} = T$.
5. $(U \circ T)^{-1} = T^{-1} \circ U^{-1}$, provided that T^{-1} and U^{-1} exist.

Proof. 1. ($\implies$) If T is invertible with inverse U , then $T \circ U = I_W$ implies T is onto, because for every $w \in W$, $w = (T \circ U)(w) = T(U(w))$; also, $U \circ T = I_V$ implies $\mathcal{N} = \{0\}$, because $T(v) = 0 \implies v = (U \circ T)(v) = U(T(v)) = 0$. Thus, T is one-to-one and onto.

($\impliedby$) If T is one-to-one and onto, then the set-theoretic inverse $T^{-1}: W \rightarrow V$ exists. Furthermore it is linear; in particular, for $w_1, w_2 \in W$ and scalars a, b ,

$$T(T^{-1}(aw_1 + bw_2)) = aw_1 + bw_2 = aT(T^{-1}(w_1)) + bT(T^{-1}(w_2)) = T(aT^{-1}(w_1) + bT^{-1}(w_2)),$$

and injectivity of T implies that $T^{-1}(aw_1 + bw_2) = aT^{-1}(w_1) + bT^{-1}(w_2)$. Thus, T^{-1} is a linear transform, and T is invertible.

(For the “not every” condition, we can take, for example the zero map $0: V \rightarrow W$, which is neither one-to-one nor onto.)

2. Invertibility $\implies$ surjectivity by the above, so $\text{im}(T) = W$ and hence $\text{rank}(T) = \dim(\text{im}(T)) = \dim(W)$. Furthermore, invertibility implies injectivity, so $\mathcal{N}(T) = \{0\}$ and hence the rank-nullity theorem gives $\dim(V) = \text{rank}(T) + \dim(\mathcal{N}(T)) = \text{rank}(T) + 0 = \text{rank}(T)$.
3. Suppose $U_1, U_2: W \rightarrow V$ are both inverses of T . Then $U_1 = (U_1 \circ T) \circ U_2 = I_V \circ U_2 = U_2$.

The last two are left as easy exercises. □

The following result says that the inverse of a linear transform is also linear.

Theorem 12.2. Let V and W be vector spaces and let $T: V \rightarrow W$ be an invertible linear transform. Then the inverse $T^{-1}: W \rightarrow V$ is also linear.

Proof. Let $w_1, w_2 \in W$ and $a \in \mathbb{F}$. Consider $T^{-1}(w_1 + aw_2)$. Since T is invertible, it is one-to-one and onto, so there exist unique $v_1, v_2 \in V$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Then

$$T^{-1}(w_1 + aw_2) = T^{-1}(T(v_1) + aT(v_2)) = T^{-1}(T(v_1 + av_2)) = v_1 + av_2 = T^{-1}(w_1) + aT^{-1}(w_2),$$

and so T^{-1} is linear. □

§13 Lecture 13—01st March, 2022

§13.1 Invertibility and isomorphisms (II)

Theorem 13.1. *Let V and W be finite-dimensional vector spaces with ordered bases β and γ , respectively. Let $T: V \rightarrow W$ be linear. Then T is invertible if and only if $[T]_{\beta}^{\gamma}$ is invertible, and in this case, $[T^{-1}]_{\gamma}^{\beta} = ([T]_{\beta}^{\gamma})^{-1}$.*

Proof. ($\implies$) Suppose that T is invertible, then $\dim(V) = \dim(W) = n$ and consequently $[T]_{\beta}^{\gamma}$ is an $n \times n$ matrix. Now $T^{-1}: W \rightarrow V$ is linear, $TT^{-1} = I_W$. and $T^{-1}T = I_V$. Thus,

$$\begin{cases} I_n = [I_W]_{\gamma}^{\gamma} = [TT^{-1}]_{\gamma}^{\gamma} = [T]_{\beta}^{\gamma}[T^{-1}]_{\gamma}^{\beta}, \\ I_n = [I_V]_{\beta}^{\beta} = [T^{-1}T]_{\beta}^{\beta} = [T^{-1}]_{\gamma}^{\beta}[T]_{\beta}^{\gamma}. \end{cases}$$

Thus $[T]_{\beta}^{\gamma}$ is invertible and $[T^{-1}]_{\gamma}^{\beta} = ([T]_{\beta}^{\gamma})^{-1}$.

($\impliedby$) Suppose that $[T]_{\beta}^{\gamma}$ is invertible. Then there exists an $n \times n$ matrix B such that $AB = BA = I_n$. Now we use B to construct n vectors $\{x_j\}_{j=1}^n$ in V as follows:

$$x_j = \sum_{i=1}^n b_{ij}v_i, \quad \text{where } \beta = \{v_1, \dots, v_n\} \text{ is the basis for } V.$$

Thus by [Theorem 8.4](#), there exists a unique $U \in \mathcal{L}(W, V)$ such that $U(w_j) = x_j$ for all $j \in [n]$, where $\gamma = \{w_1, \dots, w_n\}$ is the basis for W . Thus $U(w_j) = x_j = \sum_{i=1}^n B_{ij}v_i$, which implies $[U]_{\gamma}^{\beta} = B$. In this way,

$$\begin{aligned} [UT]_{\beta}^{\beta} &= [U]_{\gamma}^{\beta}[T]_{\beta}^{\gamma} = BA = I_n = [I_V]_{\beta}^{\beta} \implies UT = I_V, \\ [TU]_{\beta}^{\beta} &= [T]_{\beta}^{\gamma}[U]_{\gamma}^{\beta} = [I_W]_{\gamma}^{\gamma} = I_n \implies TU = I_W. \end{aligned}$$

Therefore T is invertible, and $T^{-1} = U$. Hence $[T^{-1}]_{\gamma}^{\beta} = [U]_{\gamma}^{\beta} = ([T]_{\beta}^{\gamma})^{-1}$. □

This theorem is pretty nice because it offers us a “sure banker” for computing the inverse of an invertible linear transform: given $T: V \rightarrow W$ invertible with bases α and β for V and W , respectively, we can compute the inverse via:

$$T \longrightarrow [T]_{\alpha}^{\beta} \xrightarrow{\text{Gauss-Jordan elim.}} ([T]_{\alpha}^{\beta})^{-1} \xrightarrow{\text{Theorem 13.1}} [T^{-1}]_{\beta}^{\alpha} \longrightarrow T^{-1}.$$

Example 13.1. Let $T: \mathbb{P}_1(\mathbb{R}) \rightarrow \mathbb{R}^2$ and $T(a + bx) = \begin{bmatrix} a \\ a + b \end{bmatrix}$. Let β and γ be the standard ordered bases of $\mathbb{P}_1(\mathbb{R})$ and $\mathbb{R}^2$, respectively. Find T^{-1} .

Solution. By observation,

$$T^{-1}: \mathbb{R}^2 \rightarrow \mathbb{P}_1(\mathbb{R}), \quad T^{-1} \begin{bmatrix} c \\ d \end{bmatrix} = c + (d - c)x.$$

By [Theorem 13.1](#),

$$\beta = \{1, x\}, \quad \gamma = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \quad [T]_{\beta}^{\gamma} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Now we have that

$$\begin{aligned} \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array} \right] &\sim \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 1 \end{array} \right] \implies ([T]_{\beta}^{\gamma})^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = [T^{-1}]_{\gamma}^{\beta} \\ &\implies \begin{cases} T^{-1} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) = 1 - x \\ T^{-1} \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = x \end{cases} \end{aligned}$$

Thus we get that

$$T^{-1} \begin{bmatrix} c \\ d \end{bmatrix} = T^{-1} \left(c \begin{bmatrix} 1 \\ 0 \end{bmatrix} + d \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = cT^{-1} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) + dT^{-1} \left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = c(1 - x) + dx = c + (d - c)x.$$

□

[Theorem 13.1](#) can also serve as a test for invertibility.

Example 13.2. Let $U: \mathbb{P}_3(\mathbb{R}) \rightarrow \mathbb{P}_2(\mathbb{R})$ be defined by $U(f(x)) = f'(x)$, the derivative of f , and let $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_3(\mathbb{R})$ be defined by $T(g(x)) = \int g(x) dx$. Let $\alpha = \{1, x, x^2, x^3\}$ and $\beta = \{1, x, x^2\}$ be the standard ordered bases for $\mathbb{P}_3(\mathbb{R})$ and $\mathbb{P}_2(\mathbb{R})$, respectively. Are T and U invertible?

Solution. The linear transforms T and U are not invertible because $\dim(\mathbb{P}_3(\mathbb{R})) \neq \dim(\mathbb{P}_2(\mathbb{R}))$. It is also easy to show that

$$[T]_{\beta}^{\alpha} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}, \quad [U]_{\alpha}^{\beta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Both of these are non-square matrices, and hence not invertible. Thus by [Theorem 13.1](#), T and U are not invertible. □

Theorem 13.2 (Isomorphism). Let V and W be vector spaces. We say that V and W are isomorphic if there exists a linear transform $T: V \rightarrow W$ that is invertible. Such a linear transform is called an isomorphism from V to W .

Remark 6. A basis is used to build a map between $V(\mathbb{F})$ (with $\dim V = n$) and $\mathbb{F}^n$ to make the two isomorphic to each other. More precisely, $v \rightarrow [v]_{\alpha}$ is precisely the isomorphism we are looking

for between V and $\mathbb{F}^n$. In general, however, there are many bases for V , so the isomorphism is not unique.

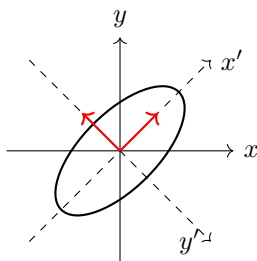
§14 Lecture 14—03rd March, 2022

§14.1 The change of coordinate matrix

Today we start with the equation of the ellipse $2x^2 - 4xy + 5y^2 = 1$. Here the basis can be taken as $\beta = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. So let

$$\begin{cases} x = \frac{1}{\sqrt{5}}(2x' - y') \\ y = \frac{1}{\sqrt{5}}(x' + 2y') \end{cases} \implies (x')^2 + 6(y')^2 = 1,$$

and so we have the following picture to keep in mind:



Then the $x'y'$ -coordinates of points have the basis $\beta' = \left\{ \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \right\}$, with these as the unit vectors along the x' and y' axes.

So the question is this: how can a coordinate vector relative to one basis be changed into a coordinate vector relative to the other? So for instance, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top$ in β' is $\begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$ in β , and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}^\top$ in β' is $\begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$ in β . Thus we can write:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}.$$

Notice now that $\begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = [I_{\mathbb{R}^2}]_{\beta'}^\beta = Q$, so then we have for all $v \in \mathbb{R}^2$ that $[v]_\beta = Q[v]_{\beta'}$.

Theorem 14.1. Let β and β' be two ordered bases for a finite-dimensional vector space V , let $Q = [I_V]_{\beta'}^\beta$, then

- (i) Q is invertible (since I_V is invertible),
- (ii) for any $v \in V$, we have $[v]_\beta = Q[v]_{\beta'}$.

Definition 14.1 (Change of coordinates matrix). The matrix $Q = [I_V]_{\beta'}^\beta$ above is called the change of coordinates matrix from β' to β .

If $\beta = \{x_1, \dots, x_n\}$ and $\beta' = \{y_1, \dots, y_n\}$, then

$$I_V(y_j) = y_j = \sum_{i=1}^n Q_{ij}x_i \implies Q = \begin{bmatrix} \left| \begin{array}{c} [y_1]_{\beta} \\ \vdots \end{array} \right| & \left| \begin{array}{c} [y_2]_{\beta} \\ \vdots \end{array} \right| & \cdots & \left| \begin{array}{c} [y_n]_{\beta} \\ \vdots \end{array} \right| \end{bmatrix},$$

that is, the j -th column of Q is given by the coordinates of y_j with respect to the basis β .

Remark 7. In this discussion, we are only going to consider linear transformations from vector spaces V to itself, i.e., $T: V \rightarrow V$. Such a linear transform is called a linear operator on V .

A linear operator T on V can be represented by $[T]_{\beta}$ and $[T]_{\beta'}$. What is the relationship between these two matrices?

Theorem 14.2. Let T be a linear operator on a finite-dimensional vector space V , and let β and β' be two ordered bases for V . Suppose that Q is the change of coordinate matrix that changes β' -coordinates to β -coordinates. Then

$$[T]_{\beta'} = Q^{-1}[T]_{\beta}Q.$$

Proof. Set $Q = [I_V]_{\beta'}^{\beta}$. Then for any $v \in V$, we have

$$Q[T]_{\beta'} = [I_V]_{\beta'}^{\beta}[T]_{\beta'} = [I_V T]_{\beta'}^{\beta} = [T I_V]_{\beta'}^{\beta} = [T]_{\beta}[I_V]_{\beta'}^{\beta} = [T]_{\beta}Q,$$

and left-multiplying by Q^{-1} gives the desired result. $\square$

Example 14.2. Take $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $T \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3a - b \\ a + 3b \end{bmatrix}$. Let $\beta = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ be the standard basis for $\mathbb{R}^2$, and let $\beta' = \left\{ \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$. Find $[T]_{\beta'}$.

Solution. By definition,

$$T \left(\begin{bmatrix} 2 \\ 4 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 14 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 4 \end{bmatrix} - 2 \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad T \left(\begin{bmatrix} 3 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 8 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Thus $[T]_{\beta'} = \begin{bmatrix} 4 & -2 \\ 1 & 2 \end{bmatrix}$. But we can also find $[T]_{\beta'}$ using [Theorem 14.2](#). Check that

$$\begin{aligned} \begin{cases} T \begin{bmatrix} 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ T \begin{bmatrix} 1 \\ -1 \end{bmatrix} &= \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{cases} \implies [T]_{\beta} = \begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix} \\ \implies \begin{cases} I_V(\beta'_1) = \beta'_1 = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ I_V(\beta'_2) = \beta'_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{cases} \end{aligned}$$

$$\implies Q = [I_V]_{\beta'}^\beta = \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} \implies Q^{-1} = \frac{1}{5} \begin{bmatrix} 1 & -2 \\ 1 & 3 \end{bmatrix}.$$

Then we get by [Theorem 14.2](#) that

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q = \frac{1}{5} \begin{bmatrix} 1 & -2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ -2 & 2 \end{bmatrix}$$

Now suppose $\beta = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. Here $T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$, hence

$$\begin{aligned} [T]_\beta &= \begin{bmatrix} 3 & -1 \\ 1 & 3 \end{bmatrix} \implies Q = [I_V]_{\beta'}^\beta = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix} \implies Q^{-1} = \begin{bmatrix} -1/10 & 3/10 \\ 2/5 & -1/5 \end{bmatrix} \\ \implies [T]_{\beta'} &= Q^{-1}[T]_\beta Q = \begin{bmatrix} 4 & 1 \\ -2 & 2 \end{bmatrix}. \quad \square \end{aligned}$$

Corollary 14.3. Let $A \in \text{Mat}_{n \times n}(\mathbb{F})$ and α be an ordered basis for $\mathbb{F}^n$. Then $[L_A]_\alpha = Q^{-1}AQ$, where Q is the $n \times n$ matrix whose j -th column is given by the j -th vector in the basis α .

Proof. Let β be the standard basis for $\mathbb{F}^n$. Then we have $[L_A]_\alpha = Q^{-1}[L_A]_\beta Q$, where $Q = [I_V]_{\beta'}^\alpha$. But we also have $[L_A]_\beta = A$, since the action of L_A on the standard basis vectors is just matrix multiplication by A . Thus, $[L_A]_\alpha = Q^{-1}AQ$. $\square$

Remark 8. 1. $[L_A]_\beta = A$ where β is the standard ordered basis of $\mathbb{F}^n$.

2. $Q = [I_{\mathbb{F}^n}]_{\beta'}^\alpha$ is the change of basis matrix from β to α , so $I_V(\alpha_j) = \alpha_j = \sum_{i=1}^n Q_{ij}\beta_i$, i.e., the j -th column of Q is given by the coordinates of α_j with respect to the basis β .

Example 14.4. Take $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 3 \\ 0 & -1 & 0 \end{bmatrix}$, and $\alpha = \left\{ \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ be an ordered basis for $\mathbb{R}^3$. Find $[L_A]_\alpha$.

Solution. By definition,

$$\begin{aligned} L_A(\alpha_1) &= A\alpha_1 = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} = -\alpha_2, \\ L_A(\alpha_2) &= A\alpha_2 = \begin{bmatrix} 5 \\ 3 \\ -1 \end{bmatrix} = 2\alpha_1 + 4\alpha_2 - \alpha_3, \\ L_A(\alpha_3) &= A\alpha_3 = \begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix} = 8\alpha_1 + 6\alpha_2 - \alpha_3. \end{aligned}$$

Thus, $[L_A]_\alpha = \begin{bmatrix} 0 & 2 & 8 \\ -1 & 4 & 6 \\ 0 & -1 & -1 \end{bmatrix}$. By the corollary, $Q = \begin{bmatrix} -1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ and $Q^{-1} = \begin{bmatrix} -1 & 2 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$,

hence

$$[L_A]_\alpha = Q^{-1}AQ = \begin{bmatrix} 0 & 2 & 8 \\ -1 & 4 & 6 \\ 0 & -1 & -1 \end{bmatrix}. \quad \square$$

Definition 14.5. Let $A, B \in \text{Mat}_{n \times n}(\mathbb{F})$. We say that A and B are similar if there exists an invertible matrix $Q \in \text{Mat}_{n \times n}(\mathbb{F})$ such that $B = Q^{-1}AQ$.

§15 Lecture 15—08th March, 2022

§15.1 The similar matrices and eigendecompositions

Recall these from last time:

Theorem 15.1. Let T be a linear operator on a finite-dimensional vector space V , and let β and β' be two ordered bases for V . Suppose that Q is the change of coordinate matrix that changes β' -coordinates to β -coordinates. Then

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q.$$

Definition 15.1. Let $A, B \in \text{Mat}_{n \times n}(\mathbb{F})$. We say that A and B are similar if there exists an invertible matrix $P \in \text{Mat}_{n \times n}(\mathbb{F})$ such that $B = P^{-1}AP$.

We will now go over the following important properties of similar matrices:

Proposition 15.2. A is similar to itself.

Proof. Let $P = I_n$ be the identity matrix. Then we have $P^{-1}AP = I_n^{-1}AI_n = A$. $\square$

Proposition 15.3. If B is similar to A , then A is similar to B .

Proof. If B is similar to A , then $B = P^{-1}AP$ for some invertible matrix P , hence $A = PBP^{-1}$. Since $P^{-1} := C$ is also invertible, we have $A = C^{-1}BC$, so A is similar to B . $\square$

Proposition 15.4. If A is similar to B and B is similar to C , then A is similar to C .

Proof. Take $A = P^{-1}BP$ and $B = Q^{-1}CQ$ for some invertible matrices P, Q . Then

$$A = P^{-1}(Q^{-1}CQ)P = (QP)^{-1}C(QP) \quad \text{since } (QP)^{-1} = P^{-1}Q^{-1}$$

and since the product of invertible matrices is invertible, we conclude that A is similar to C . $\square$

With the notion of similarity, we can now discuss eigendecompositions of a matrix.

Definition 15.5. We say that the matrix A is diagonalisable if A is similar to a diagonal matrix constructed by the eigenvalues of A .

Recall the eigenvalue problem $AX = \lambda X$, where λ is an eigenvalue and X is the corresponding eigenvector. Let

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \dots & x^{(n)} \\ | & | & & | \end{bmatrix}.$$

Λ and X are not unique due to the order of λ_i and $x^{(i)}$. In fact these eigenvectors are not unique. Indeed:

$$\begin{aligned} AX &= A \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \dots & x^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ax^{(1)} & Ax^{(2)} & \dots & Ax^{(n)} \\ | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} | & | & & | \\ \lambda_1 x^{(1)} & \lambda_2 x^{(2)} & \dots & \lambda_n x^{(n)} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \dots & x^{(n)} \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \\ &= X\Lambda. \end{aligned}$$

Clearly we have an immediate condition for the diagonalisability of A : it is sufficient that X^{-1} exists and is well-defined. If this condition is satisfied, that is,

$$\text{if } X \text{ exists} \implies \begin{cases} A = X\Lambda X^{-1} & A \text{ is similar to } \Lambda \\ \Lambda = X^{-1}AX & A \text{ is diagonalisable.} \end{cases}$$

This framing immediately gives the following result:

Proposition 15.6. An $n \times n$ matrix A is diagonalisable if and only if A has n linearly independent eigenvectors.

Proof. The inverse X^{-1} exists if and only if the columns of X are linearly independent, which is true if and only if A has n linearly independent eigenvectors. $\square$

Proposition 15.7. Similar matrices have the same eigenvalues.

Proof. We claim that $\det(B - \lambda I) = \det(A - \lambda I)$ for eigenvalue λ and $B = P^{-1}AP$. Indeed,

$$\det(B - \lambda I) = \det(P^{-1}AP - \lambda I) = \det(P^{-1}(A - \lambda I)P) = \det(P^{-1}) \det(A - \lambda I) \det(P) = \det(A - \lambda I).$$

Thus A and B have the same characteristic polynomial, and hence the same eigenvalues. $\square$

Proposition 15.8. For $k \in [1, \infty)$, if A is diagonalisable, then A^k has the same eigenvectors in X and eigenvalues λ_i^k . In particular, as $k \rightarrow \infty$, $A^k \rightarrow 0$ if $|\lambda_i| < 1$ for all i .

The proof is straightforward, just take $A^k = \underbrace{(X\Lambda X^{-1})(X\Lambda X^{-1}) \cdots (X\Lambda X^{-1})}_{k \text{ times}} = X\Lambda^k X^{-1}$, since in the middle all the $X^{-1}X$ cancel out.

Definition 15.9. We say that a matrix A is nilpotent if there exists a positive integer m such that $A^m = 0$.

Definition 15.10. A matrix $S \in \mathbb{R}^{n \times n}$ is symmetric if $S = S^\top$.

From the above discussion, it is clear that every real symmetric matrix has real eigenvalues and orthonormal eigenvectors, that is, $S = Q\Lambda Q^{-1} = Q\Lambda Q^\top = \sum_{i=1}^n \lambda_i q_i q_i^\top$, hence we can write

$$\begin{aligned} S = Q\Lambda Q^\top &= \begin{bmatrix} | & | & \cdots & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} - & q_1^\top & - \\ - & q_2^\top & - \\ & \vdots & \\ - & q_n^\top & - \end{bmatrix} \\ &= \begin{bmatrix} | & | & \cdots & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} - & \lambda_1 q_1^\top & - \\ - & \lambda_2 q_2^\top & - \\ & \vdots & \\ - & \lambda_n q_n^\top & - \end{bmatrix} = \sum_{i=1}^n \lambda_i q_i q_i^\top, \end{aligned}$$

which is the same as before so I'm not sure why I frantically typed this again. Haha.

§16 Lecture 16—10th March, 2022

§16.1 Diagonalisation

Now we will discuss the diagonalisation problem. Take T a linear transform on a finite-dimensional vector space V . We want to answer the following questions:

1. Does there exist an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix?
2. If such a basis exists, how can we find it?

Definition 16.1. A linear operator T on a finite-dimensional vector space V is called diagonalisable if there exists an ordered basis β for V such that $[T]_\beta$ is a diagonal matrix.

If we have diagonal $D = [T]_\beta = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$, then for each vector in β , we have

$$T(v_j) = \sum_{i=1}^n D_{ij} v_i = D_{jj} v_j = \lambda_j v_j, \quad \text{where } 1 \leq j \leq n.$$

Conversely, if $\beta = \{v_1, v_2, \dots, v_n\}$ is an ordered basis for V such that $T(v_j) = \lambda_j v_j$ for some scalars $\lambda_1, \dots, \lambda_n$, then

$$[T]_\beta = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

Remark 9. 1. Since v_j lies in a basis, v_j is nonzero.

2. If $\dim V = n$, then T has n eigenvalues (counting multiplicities).

Definition 16.2. Let T be a linear operator on a vector space V . A nonzero vector $v \in V$ is called an eigenvector of T if there exists a scalar $\lambda \in \mathbb{F}$ such that $T(v) = \lambda v$. The scalar λ is called the eigenvalue of T corresponding to the eigenvector v .

Remark 10. 1. For each eigenvalue λ , there are infinitely many corresponding eigenvectors, since αv is an eigenvector if v is an eigenvector and $\alpha \neq 0$. In particular, for nonzero $\alpha \in \mathbb{F}$,

$$T \text{ is linear} \implies T(\alpha v) = \alpha T(v) = \alpha \lambda v = \lambda(\alpha v).$$

2. (λ, v) is an eigenpair (read: “eigenvalue-eigenvector pair”) of an $n \times n$ matrix A if and only if λ is an eigenvalue of the linear operator L_A and v is a corresponding eigenvector of L_A .

Given $A \in \text{Mat}_{n \times n}(\mathbb{F})$, a nonzero vector $v \in \mathbb{F}^n$ is called an eigenvector of A if v is an eigenvector of L_A , i.e. if there exists a scalar $\lambda \in \mathbb{F}$ such that $Av = \lambda v$. In this case λ is called an eigenvalue of A corresponding to the eigenvector v . These eigenvectors (resp. values) are also called the characteristic vectors (resp. values) or proper vectors (resp. values), and they span the eigenspace corresponding to the eigenvalue λ .

Definition 16.3. The eigenspace corresponding to an eigenvalue λ of a linear operator T is the set of all eigenvectors of T corresponding to λ , together with the zero vector. It is denoted by $E_\lambda(T)$ and is a subspace of the vector space V .

Example 16.4. Let $T: \mathbb{P}(\mathbb{R}) \rightarrow \mathbb{P}(\mathbb{R})$ and $T[f(t)] = f'(t)$. Find the eigenvalues and eigenvectors of T .

Solution. Immediately we know there are infinitely many eigenvalues since $\dim(\mathbb{P}(\mathbb{R})) = \infty$. Now with $f \neq 0$,

$$T(f) = \lambda f \implies f' = \lambda f \implies f = Ce^{\lambda t}, \quad C \neq 0.$$

Every real number λ is an eigenvalue of T corresponding to $f = e^{\lambda t}$. For $\lambda = 0$, the eigenvector is 1. □

Example 16.5. Take $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_2(\mathbb{R})$ and $T(f(x)) = f(x) + (x+1)f'(x)$. Find the eigenvalues and eigenvectors of T .

This is an office-hours problem, we’ll go over it then.

Theorem 16.1. Let $A \in \text{Mat}_{n \times n}(\mathbb{F})$. A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$.

Proof. For $v \neq 0$,

$$\begin{aligned} Av = \lambda v &\iff (A - \lambda I)v = 0 \iff v \in \text{Nul}(A - \lambda I) \setminus \{0\} \\ &\iff \text{Nul}(A - \lambda I) \neq \{0\} \iff A - \lambda I \text{ is not invertible} \\ &\iff \det(A - \lambda I) = 0. \end{aligned}$$

□

Let $A \in \text{Mat}_{n \times n}(\mathbb{F})$. The polynomial $f(\lambda) = \det(A - \lambda I)$ is called the characteristic polynomial of A , and the eigenvalues of A are precisely the roots of $f(\lambda)$. This characteristic polynomial of A is a polynomial of degree n in λ with leading coefficient $(-1)^n$ times the determinant of A , and A has at most n distinct eigenvalues.

Example 16.6. Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$.

Solution. The characteristic polynomial of A is

$$\det(A - \lambda I_2) = 0 \implies \lambda^2 - 3\lambda - 10 = (\lambda + 2)(\lambda - 5) \implies \lambda = -2, 5.$$

The eigenvectors corresponding to $\lambda = -2$ and $\lambda = 5$ can be found by solving

$$(A + 2I_2)v_1 = 0 \quad \text{and} \quad (A - 5I_2)v_2 = 0,$$

and indeed we get

$$v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \text{ or } \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 4/3 \end{bmatrix}.$$

□

§17 Lecture 17—22nd March, 2022

§17.1 Eigenvalues and eigenvectors (II)

Definition 17.1. Let T be a linear operator on an n -dimensional vector space V with ordered basis α . We define the characteristic polynomial $f(t)$ of T to be the characteristic polynomial of $A = [T]_\alpha$, i.e. $f(t) = \det(A - tI) = \det([T]_\alpha - tI)$.

Note that this definition is independent of the choice of the ordered basis α , and that λ is an eigenvalue of T if and only if λ is an eigenvalue of $[T]_\alpha$. We often denote the characteristic polynomial of an operator T by $\det(T - tI)$.

Example 17.2. Take $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_2(\mathbb{R})$ and $T(f(x)) = f(x) + (x+1)f'(x)$. Find the eigenvalues of T , using the standard basis $\alpha = \{1, x, x^2\}$ and $A = [T]_\alpha$.

Solution. We will get the eigenvalues of T using the eigenvalues of $[T]_\alpha$. Check that

$$\begin{cases} T(1) = 1 \\ T(x) = 2x + 1 \\ T(x^2) = 3x^2 + 2x \end{cases} \implies A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 2 & 3 \end{bmatrix} \implies \det(A - \lambda I_3) = (1 - \lambda)(2 - \lambda)(3 - \lambda) = 0,$$

thus $\lambda = 1, 2, 3$. For fun, suppose we used the basis $\gamma = \{1, 1 + x, x^2\}$:

$$\begin{cases} T(1) = 1 \\ T(x) = 2(1 + x) \\ T(x^2) = 2x + 3x^2 \end{cases} \implies A = \begin{bmatrix} 1 & 0 & -2 \\ 1 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} \implies \det(A - \lambda I_3) = (1 - \lambda)(2 - \lambda)(3 - \lambda) = 0,$$

and so again we find $\lambda = 1, 2, 3$. □

This is an important lesson by the way: the eigenvalues of a linear operator are independent of the choice of basis.

Theorem 17.1. *Let T be a linear operator on a vector space V and let λ be an eigenvalue of T . A vector $v \in V$ is an eigenvector of T corresponding to the eigenvalue λ if and only if $v \neq 0$ and $v \in \mathcal{N}(T - \lambda I)$.*

The proof is left as a homework problem.

Remark 11. 1. Sometimes it is not easy to get eigenvectors of T by definition, or by the above theorem.
2. Since the eigenvalues of T are the same as the eigenvalues of $[T]_\alpha$, are the eigenvalues of T related to eigenvectors of $[T]_\alpha$ corresponding to the same eigenvalue?

Theorem 17.2. *Let T be a linear operator on a finite-dimensional vector space V and $\alpha = \{v_1, \dots, v_n\}$ be an ordered basis for V . Then the coordinate vector of any eigenvector of T is the eigenvector of $[T]_\alpha$ corresponding to the same eigenvalue.*

Proof. Starting with $T(v) = \lambda v$ for nonzero v , we want $[T]_\alpha[v]_\alpha = \lambda[v]_\alpha$. Now

$$[T(v)]_\alpha = [\lambda v]_\alpha = \lambda[v]_\alpha \iff T[v]_\alpha = \left[T \left(\sum_{j=1}^n ([v]_\alpha)_j v_j \right) \right]_\alpha = \sum_{j=1}^n ([v]_\alpha)_j [T(v_j)]_\alpha.$$

Now,

$$T(v_j) = A_{1j}v_1 + A_{2j}v_2 + \dots + A_{nj}v_n \implies [T(v_j)]_\alpha = \begin{bmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{nj} \end{bmatrix} = a_j,$$

so we get $\sum_{j=1}^n ([v]_\alpha)_j [T(v_j)]_\alpha \iff \sum_{j=1}^n ([v]_\alpha)_j a_j = A[v]_\alpha = [T]_\alpha[v]_\alpha$. Thus $[T]_\alpha[v]_\alpha = \lambda[v]_\alpha$, as required. □

Example 17.3. As before, take $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_2(\mathbb{R})$ and $T(f(x)) = f(x) + (x+1)f'(x)$, under the standard basis $\alpha = \{1, x, x^2\}$. Find the eigenvectors of T .

Solution. We already have $[T]_\alpha = A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$, and $\lambda = 1, 2, 3$. So first we get the eigenvectors of $A = [T]_\alpha$.

$$\begin{aligned} \lambda_1 = 1 &\implies (A - \lambda_1 I_3)v_1 = 0 \implies \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \\ v_{13} \end{bmatrix} = 0 \implies v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \\ \lambda_2 = 2 &\implies (A - \lambda_2 I_3)v_2 = 0 \implies \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{21} \\ v_{22} \\ v_{23} \end{bmatrix} = 0 \implies v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \\ \lambda_3 = 3 &\implies (A - \lambda_3 I_3)v_3 = 0 \implies \begin{bmatrix} -2 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_{31} \\ v_{32} \\ v_{33} \end{bmatrix} = 0 \implies v_3 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}. \end{aligned}$$

Thus the eigenvectors of T correspond to $f_1 = 1$, $f_2 = 1 + x$, and $f_3 = 1 + 2x + x^2$.

If we take the basis $\gamma = \{1, 1 + x, x^2\}$ instead, we have $f_1 = 1$, $f_2 = 1 + x$, and $f_3 = -1 + 2(1 + x) + x^2 = 1 + 2x + x^2$. So again, the eigenvectors of T are independent of the choice of basis.

To diagonalise, from a basis by the eigenvectors $\beta = \{f_1, f_2, f_3\}$, i.e. compute $[T]_\beta$. We know that

$$\begin{aligned} T(f_1) &= 1 \\ T(f_2) &= 1 + x + (1 + x) \cdot 1 = 2 + 2x = 2 \cdot (1 + x) \\ T(f_3) &= 1 + 2x + x^2 + (1 + x)(2 + 2x) = 3 \cdot (1 + 2x + x^2). \end{aligned}$$

Hence $[T]_\beta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

□

§18 Lecture 18—24th March, 2022

§18.1 Diagonalisation of linear operators (I): distinct and complex eigenvalues

Theorem 18.1. Let $T: V \rightarrow V$ be a linear operator. Assume that $\{\lambda_i\}_{i=1}^n$ are distinct eigenvalues of T and $\{v_i\}_{i=1}^n$ are corresponding eigenvectors. Then $\{[v_i]_\alpha\}_{i=1}^n$ is a linearly independent set in the coordinate space $\mathbb{R}^n$.

Proof. We proceed by induction. Clearly it holds for $n = 1$, since eigenvectors are necessarily nonzero. So suppose the statement holds for $n = k$, that is, $\sum_{i=1}^k a_i v_i = 0$ implies $a_1 = 0$. We want

to show it holds for $n = k + 1$. Let $\sum_{i=0}^{k+1} a_i v_i = 0$. Then we have

$$\begin{aligned}
(T - \lambda_{k+1}I) \left(\sum_{i=0}^{k+1} a_i v_i \right) &= (T - \lambda_{k+1}I)(0) \implies \sum_{i=0}^{k+1} (T(a_i v_i) - \lambda_{k+1}I(a_i v_i)) = 0 \\
&\implies \sum_{i=0}^{k+1} a_i T(v_i) - \lambda_{k+1} a_i v_i = 0 \\
&\implies \sum_{i=0}^{k+1} a_i \lambda_i v_i - \lambda_{k+1} a_i v_i = 0 \\
&\implies \sum_{i=0}^{k+1} a_i (\lambda_i - \lambda_{k+1}) v_i = 0 \\
&\implies \sum_{i=0}^k a_i (\lambda_i - \lambda_{k+1}) v_i + a_{k+1} (\lambda_{k+1} - \lambda_{k+1}) v_{k+1} = 0 \\
&\implies \sum_{i=0}^k a_i (\lambda_i - \lambda_{k+1}) v_i = 0.
\end{aligned}$$

Since $\{v_1, \dots, v_k\}$ are linearly independent, it follows that

$$a_1(\lambda_1 - \lambda_{k+1}) = a_2(\lambda_2 - \lambda_{k+1}) = \dots = a_k(\lambda_k - \lambda_{k+1}) = 0 \implies a_1 = a_2 = \dots = a_k = 0,$$

and so we only need prove that $a_{k+1} = 0$. But this follows from the original equation, since $\sum_{i=1}^{k+1} a_i v_i = 0$ and v_{k+1} is not in the span of $\{v_1, \dots, v_k\}$. Thus $a_{k+1} = 0$ and we are done. $\square$

Corollary 18.1. *Let T be a linear operator on V and $\dim(V) = n < \infty$. If T has n distinct eigenvalues, then T is diagonalisable.*

Proof. Let $\beta = \{v_1, \dots, v_n\}$ be the set of eigenvectors corresponding to the n distinct eigenvalues $\{\lambda_1, \dots, \lambda_n\}$. By the previous theorem, the set $\{[v_i]_\alpha\}_{i=1}^n$ is linearly independent in $\mathbb{R}^n$. Since $\dim(V) = n$, it follows that β spans V . Therefore, β is a basis for V and we can conclude that T is diagonalisable. $\square$

Example 18.2. 1. Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be defined as $Tx = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} x$ for $x \in \mathbb{R}^2$. Is T diagonalisable?

Solution. The characteristic polynomial is $(1 - \lambda)(3 - \lambda) = 0 \implies \lambda_1 = 1, \lambda_2 = 3$. Thus $\lambda_1 \neq \lambda_2$, so T is diagonalisable. $\square$

2. Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be defined as $Tx = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} x$ for $x \in \mathbb{R}^2$. Is T diagonalisable?

Solution. The characteristic polynomial is $(1 - \lambda)^2 = 0 \implies \lambda_1 = 1$. Thus T has only one eigenvalue, so we cannot conclude whether T is diagonalisable or not. $\square$

3. Let $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_2(\mathbb{R})$ be defined as $T(f(x)) = f(x) + (x+1)f'(x)$. Is T diagonalisable?

Solution. We already have $[T]_{\alpha} = A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$, and $\lambda = 1, 2, 3$. Since these eigenvalues are distinct, we conclude that T is diagonalisable. $\square$

(I left early, but there were basically more examples like this.)

§19 Lecture 19—29th March, 2022

§19.1 Diagonalisation of linear operators (II): repeated eigenvalues

Definition 19.1. Let $f(t)$ be the characteristic polynomial of a linear operator T on a vector space V . The algebraic multiplicity of an eigenvalue λ is the largest positive integer k for which $(t - \lambda)^k$ is a factor of $f(t) = \det(T - \lambda I)$.

Example 19.2. 1. Let $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be defined as $T(x) = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix} x$. Then we get $f(t) = \det(A - tI) = -(t - 3)^2(t - 4)$, so the eigenvalue 3 has algebraic multiplicity 2 and the eigenvalue 4 has algebraic multiplicity 1.

2. Let $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be defined as $T(x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x$. Then we get $f(t) = \det(A - tI) = -(t - 1)^3$, so the eigenvalue 1 has algebraic multiplicity 3.

Definition 19.3. Let $T: V \rightarrow V$ be a linear operator and λ an eigenvalue of T . Define

$$E_{\lambda} = \{x \in V : T(x) = \lambda x\} \equiv \mathcal{N}(T - \lambda I_V).$$

Then E_{λ} is called the eigenspace of T corresponding to λ .

Theorem 19.1. Let T be a linear operator on an n -dimensional ($n < \infty$) vector space V with distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Then T is diagonalizable if and only if $\dim(E_{\lambda_i})$ equals the algebraic multiplicity of λ_i for each $i = 1, \dots, k$.

Proof. Let m_i be the algebraic multiplicity of λ_i and $d_i = \dim(E_{\lambda_i})$ be its geometric multiplicity. We know that for each eigenvalue λ_i , we have $1 \leq d_i \leq m_i$. The sum of the algebraic multiplicities is the dimension of the vector space, $\sum_{i=1}^k m_i = n$.

($\implies$) Suppose T is diagonalizable. This means there exists a basis of V consisting of eigenvectors of T . Let $\mathcal{B}_i$ be a basis for the eigenspace E_{λ_i} . The vectors in $\mathcal{B}_i$ are eigenvectors corresponding to λ_i , and there are $d_i = \dim(E_{\lambda_i})$ of them. Since eigenvectors corresponding to distinct eigenvalues are linearly independent, the set $\mathcal{B} = \bigcup_{i=1}^k \mathcal{B}_i$ is a linearly independent set of vectors in V . The total number of vectors in $\mathcal{B}$ is $\sum_{i=1}^k d_i$. Since T is diagonalizable, the sum of the dimensions of its

eigenspaces must be equal to the dimension of V . That is, $n = \sum_{i=1}^k d_i$. We have $\sum_{i=1}^k d_i = n$ and $\sum_{i=1}^k m_i = n$. Combined with the fact that $d_i \leq m_i$ for all i , the only way for the sums to be equal is if $d_i = m_i$ for all $i = 1, \dots, k$.

($\Leftarrow$) Suppose that $d_i = m_i$ for all $i = 1, \dots, k$. We know that $\sum_{i=1}^k m_i = n$, so our assumption implies $\sum_{i=1}^k d_i = n$. Let $\mathcal{B}_i$ be a basis for each eigenspace E_{λ_i} . The size of $\mathcal{B}_i$ is d_i . Let $\mathcal{B} = \bigcup_{i=1}^k \mathcal{B}_i$. This is a set of eigenvectors. Since eigenvectors corresponding to distinct eigenvalues are linearly independent, $\mathcal{B}$ is a linearly independent set of vectors. The total number of vectors in $\mathcal{B}$ is $\sum_{i=1}^k d_i = n$. So, $\mathcal{B}$ is a set of n linearly independent vectors in an n -dimensional vector space V . Therefore, $\mathcal{B}$ is a basis for V . Since we have found a basis for V that consists entirely of eigenvectors of T , the operator T is diagonalizable. $\square$

Note:

- the *algebraic multiplicity* counts the repetition of λ among the eigenvalues; we find it by counting the factors of $(t - \lambda)$ in the characteristic polynomial $f(t)$.
- the *geometric multiplicity* counts the independent eigenvectors for λ ; we find it by finding the dimension of the eigenspace E_λ .

Recall: a diagonalisable linear operator needs n linearly independent eigenvectors.

Example 19.4. Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be defined as $T(x) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x$. Then we get $f(t) = \det(A - tI) = (1 - t)^2$, so the eigenvalue $\lambda = 1$ has algebraic multiplicity 2. However, the eigenspace is given by

$$\begin{aligned} E_\lambda &= \{x \in \mathbb{R}^2 : (T - \lambda I)x = 0\} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies x_2 = 0 \\ \implies E_\lambda &= \left\{ x \in \mathbb{R}^2 : x = c \begin{bmatrix} 1 \\ 0 \end{bmatrix}, c \in \mathbb{R} \right\} \implies \dim(E_\lambda) = 1. \end{aligned}$$

Thus $\text{AM}(\lambda = 1) = 2$ and $\text{GM}(\lambda = 1) = 1$. Therefore T is not diagonalisable.

Example 19.5. Let $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_2(\mathbb{R})$ be defined as $T(f) = f(1) + f'(0)x + (f'(0) + f''(0))x^2$. We want to tell if T is diagonalisable. First, let $\beta = \{1, x, x^2\}$ be the standard basis of $V = \mathbb{P}_2(\mathbb{R})$. Then

$$\begin{cases} T(1) = 1, \\ T(x) = 1 + x + x^2, \\ T(x^2) = 1 + 2x^2 \end{cases} \implies [T]_\beta = A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \implies f(t) = \begin{vmatrix} 1-t & 1 & 1 \\ 0 & 1-t & 0 \\ 0 & 1 & 2-t \end{vmatrix},$$

so $f(t) = -(t - 1)^2(t - 2)$, and therefore $\text{AM}(\lambda = 1) = 2$ and $\text{AM}(\lambda = 2) = 1$. Now, $E_{\lambda=1} = \{f \in V : (T - \lambda_1 I)f = 0\}$, so

$$(A - \lambda I)x = 0 \implies \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \implies x_2 = -x_3 \implies x = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \\ \implies f = a - bx + bx^2 = a + b(-x + x^2).$$

Therefore $E_{\lambda=1} = \{f = a + b(-x + x^2) : a, b \in \mathbb{R}\}$, so $\text{GM}(\lambda = 1) = 2$. Thus T is diagonalisable since T has 3 linearly independent eigenvectors.

In the same way we can actually also show that $E_{\lambda=2} = \{f = c(1 + x^2) : c \in \mathbb{R}\}$, so $\text{GM}(\lambda = 2) = 1$.

Example 19.6. Recall the Fibonacci numbers 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, ... generated according to the rule $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$ with initial conditions $F_0 = 0$ and $F_1 = 1$. Let's say we want to find F_{100} . We set

$$U_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}, \\ \implies U_{k+1} = \begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} F_{k+1} + F_k \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = AU_k,$$

which then would imply

$$U_{k+1} = AU_k = A(AU_{k-1}) = A^2(AU_{k-2}) = \dots = A^{k+1}U_0 = A^{k+1} \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = A^{k+1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Therefore we want

$$F_{100} \implies U_{100} = \begin{bmatrix} F_{101} \\ F_{100} \end{bmatrix} = A^{100}U_0 \implies U_{100} = BD^{100}B^{-1}U_0,$$

where $A = BDB^{-1}$ is the diagonalisation of A .

§20 Lecture 20—31st March, 2022

§20.1 Applications: homogeneous linear system with constant coefficients

Consider the system of the first-order linear differential equations with constant coefficients:

$$\begin{cases} x'_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n, \\ x'_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n, \\ \vdots \\ x'_n = a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n. \end{cases} \iff x' = Ax,$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

(At this point my laptop shut down and remained shut down, so there isn't more to add. But we covered approx. Sec. 5.2 of Friedberg-Insel-Spence, 5th edition.)

§21 Lecture 21—05th April, 2022

§21.1 Inner product spaces

In our discussions so far we have only worked on abstract vector spaces for which there is no canonical structure. We will now try to look at vector spaces upon which we may introduce some structure.

Definition 21.1 (Inner product). *Let V be a vector space over the field $\mathbb{C}$. An inner product on V is a function*

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$$

satisfying the following properties for all $u, v, w \in V$ and $\alpha \in \mathbb{C}$:

1. Conjugate symmetry: $\langle u, v \rangle = \overline{\langle v, u \rangle}$
2. Linearity in the first argument: $\langle \alpha u + v, w \rangle = \alpha \langle u, w \rangle + \langle v, w \rangle$ and $\langle cx, y \rangle = c \langle x, y \rangle$ for all $c \in \mathbb{C}$.
3. Positive-definiteness: $\langle v, v \rangle \geq 0$ with equality if and only if $v = 0$.

A vector space V over $\mathbb{C}$ endowed with a specific inner product is called an *inner product space*.

Example 21.2. Merely by checking the properties above, we can show that:

1. Let $V = \mathbb{C}^n$. Define $\langle x, y \rangle = x^\top \bar{y} = y^* x = \sum_{i=1}^n x_i \bar{y}_i$, $x, y \in V$. Then V is an inner product space.
2. Let $V = C([0, 1])$, and define $\langle f, g \rangle = \int_0^1 f(x) \overline{g(x)} dx$. Then V is an inner product space.
3. Let $V \in \mathbb{C}^{n \times n}$. Then $\langle A, B \rangle = \text{trace}(AB^*)$ defines an inner product on V .

Definition 21.3. *Let V be an inner product space. Then for all $x, y, z \in V$ and $c \in \mathbb{C}$, we have*

1. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$
2. $\langle x, cy \rangle = \bar{c} \langle x, y \rangle$
3. $\langle x, 0 \rangle = \langle 0, x \rangle = 0$
4. $\langle x, x \rangle = 0$ if and only if $x = 0$

5. if $\langle x, y \rangle = \langle x, z \rangle$ for all $x \in V$, then $y = z$.

Definition 21.4. Let V be an inner product space. For $x \in V$, we define the norm of x as $\|x\| = \sqrt{\langle x, x \rangle}$.

Remark 12. 1. If $V = \mathbb{R}^n$, then $\|x\| = \sqrt{x^\top x}$ is the length of the vector x .

2. The norm is exactly a generalisation of the length of a vector in an inner product space.

Theorem 21.1. Let V be an inner product space over $\mathbb{C}$. Then for all $x, y \in V$ and $c \in \mathbb{C}$, we have:

1. $\|cx\| = |c|\|x\|$
2. $\|x\| = 0$ if and only if $x = 0$
3. (Cauchy-Schwarz inequality) $|\langle x, y \rangle| \leq \|x\|\|y\|$
4. (Triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$
5. (Parallelogram law) $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$
6. (Polarization identity for reals) $\|x + y\|^2 - \|x - y\|^2 = 4\langle x, y \rangle$

Proof. 1. We have $\|cx\|^2 = \langle cx, cx \rangle = c\langle x, cx \rangle = c\bar{c}\langle x, x \rangle = |c|^2\|x\|^2$. Taking the square root of both sides gives $\|cx\| = |c|\|x\|$.

2. This follows directly from the positive-definiteness property of the inner product. $\|x\| = 0 \iff \|x\|^2 = 0 \iff \langle x, x \rangle = 0 \iff x = 0$.

3. If $y = 0$, the inequality is trivial. Assume $y \neq 0$. Let $z = x - \frac{\langle x, y \rangle}{\|y\|^2}y$. Then

$$\begin{aligned}
 0 \leq \|z\|^2 &= \left\langle x - \frac{\langle x, y \rangle}{\|y\|^2}y, x - \frac{\langle x, y \rangle}{\|y\|^2}y \right\rangle \\
 &= \langle x, x \rangle - \frac{\overline{\langle x, y \rangle}}{\|y\|^2} \langle x, y \rangle - \frac{\langle x, y \rangle}{\|y\|^2} \langle y, x \rangle + \frac{|\langle x, y \rangle|^2}{\|y\|^4} \langle y, y \rangle \\
 &= \|x\|^2 - \frac{|\langle x, y \rangle|^2}{\|y\|^2} - \frac{|\langle x, y \rangle|^2}{\|y\|^2} + \frac{|\langle x, y \rangle|^2}{\|y\|^2} \\
 &= \|x\|^2 - \frac{|\langle x, y \rangle|^2}{\|y\|^2}.
 \end{aligned}$$

Rearranging gives $|\langle x, y \rangle|^2 \leq \|x\|^2\|y\|^2$, and taking the square root gives the desired inequality.

4. We have

$$\begin{aligned}
 \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\
 &= \|x\|^2 + \langle x, y \rangle + \overline{\langle x, y \rangle} + \|y\|^2 \\
 &= \|x\|^2 + 2\Re(\langle x, y \rangle) + \|y\|^2.
 \end{aligned}$$

By the Cauchy-Schwarz inequality, $\Re(\langle x, y \rangle) \leq |\langle x, y \rangle| \leq \|x\|\|y\|$. Thus, $\|x + y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$. Taking the square root gives the result.

5. We have $\|x + y\|^2 = \|x\|^2 + 2\Re(\langle x, y \rangle) + \|y\|^2$. Similarly,

$$\begin{aligned}\|x - y\|^2 &= \langle x - y, x - y \rangle = \langle x, x \rangle - \langle x, y \rangle - \langle y, x \rangle + \langle y, y \rangle \\ &= \|x\|^2 - 2\Re(\langle x, y \rangle) + \|y\|^2.\end{aligned}$$

Adding these two equations gives $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

6. Subtracting the equation for $\|x - y\|^2$ from the one for $\|x + y\|^2$ gives

$$\|x + y\|^2 - \|x - y\|^2 = 4\Re(\langle x, y \rangle).$$

For a real inner product space, this is the result. For a complex inner product space, we have

$$\langle x, y \rangle = \Re(\langle x, y \rangle) + i\Im(\langle x, y \rangle).$$

Replacing y with iy , we get

$$\|x + iy\|^2 - \|x - iy\|^2 = 4\Re(\langle x, iy \rangle) = 4\Re(-i\langle x, y \rangle) = 4\Im(\langle x, y \rangle).$$

The identity as stated in the theorem holds only for real inner product spaces. A more general form is

$$\langle x, y \rangle = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2). \quad \square$$

Definition 21.5. Let V be an inner product space. Two vectors $x, y \in V$ are said to be orthogonal if $\langle x, y \rangle = 0$. A subset $S \subseteq V$ is said to be orthogonal if every pair of distinct vectors in S is orthogonal.

§22 Lecture 22—07th April, 2022

§22.1 The Gram-Schmidt orthogonalisation process

Definition 22.1. Let V be an inner product space. A subset S of V is an orthonormal basis for V if it is an ordered basis for V and is orthonormal (i.e., each vector in S has norm 1 and is mutually orthogonal to the other vectors in S).

Example 22.2. Let $V = \mathbb{P}([-1, 1])$ and $\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx$. Show that the standard basis $\beta = \{1, x, x^2\}$ is not an orthonormal basis for $\mathbb{P}_2([-1, 1])$.

Solution. We check the conditions for an orthonormal basis. First, we check if the vectors have norm 1. The norm of the vector $f(x) = 1$ is

$$\|1\|^2 = \langle 1, 1 \rangle = \int_{-1}^1 1 \cdot 1 dx = \int_{-1}^1 1 dx = [x]_{-1}^1 = 2.$$

Since $\|1\| = \sqrt{2} \neq 1$, the basis is not orthonormal.

Alternatively, we can check for orthogonality. Let's check the inner product of $f(x) = 1$ and

$$g(x) = x^2.$$

$$\langle 1, x^2 \rangle = \int_{-1}^1 1 \cdot x^2 \, dx = \left[\frac{x^3}{3} \right]_{-1}^1 = \frac{1}{3} - \left(-\frac{1}{3} \right) = \frac{2}{3}.$$

Since $\langle 1, x^2 \rangle \neq 0$, the vectors 1 and x^2 are not orthogonal. Therefore, the basis is not orthonormal. $\square$

Theorem 22.1. *Let V be an inner product space and $S = \{v_1, \dots, v_n\}$ be a subset of non-zero vectors.*

1. *If S is orthogonal, then for all $y \in \text{span}(S)$, we have*

$$y = \sum_{i=1}^n \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i.$$

2. *If S is orthogonal, then S is linearly independent.*

Proof. 1. Let $y \in \text{span}(S)$. Then there exist scalars $c_1, \dots, c_n$ such that $y = \sum_{j=1}^n c_j v_j$. For any $i \in \{1, \dots, n\}$, we take the inner product with v_i as $\langle y, v_i \rangle = \left\langle \sum_{j=1}^n c_j v_j, v_i \right\rangle = \sum_{j=1}^n c_j \langle v_j, v_i \rangle$. Since S is orthogonal, $\langle v_j, v_i \rangle = 0$ for $j \neq i$. Thus the sum becomes $\langle y, v_i \rangle = c_i \langle v_i, v_i \rangle = c_i \|v_i\|^2$. Since v_i is a non-zero vector, $\|v_i\| \neq 0$, so we can write $c_i = \langle y, v_i \rangle / \|v_i\|^2$. Substituting this back gives the desired result.

2. Suppose that $\sum_{i=1}^n c_i v_i = 0$ for some scalars $c_1, \dots, c_n$. We take the inner product with v_j for any $j \in \{1, \dots, n\}$ as $\langle \sum_{i=1}^n c_i v_i, v_j \rangle = \langle 0, v_j \rangle = 0$. By linearity, this is $\sum_{i=1}^n c_i \langle v_i, v_j \rangle = 0$. Since S is orthogonal, this simplifies to $c_j \langle v_j, v_j \rangle = c_j \|v_j\|^2 = 0$. As v_j is non-zero, $\|v_j\| \neq 0$, so we must have $c_j = 0$. Since this holds for all j , the set S is linearly independent. $\square$

Theorem 22.2 (Gram-Schmidt orthogonalisation process). *Let V be an inner product space and let $\{u_1, \dots, u_n\}$ be a basis for V . Define $v_1 = u_1$ and for $2 \leq k \leq n$, define*

$$v_k = u_k - \sum_{j=1}^{k-1} \frac{\langle u_k, v_j \rangle}{\|v_j\|^2} v_j.$$

Then $\{v_1, \dots, v_n\}$ is an orthogonal basis for V , and $\text{span}\{v_1, \dots, v_n\} = \text{span}(u_1, \dots, u_n)$.

Proof. We proceed by induction on k . For $k = 1$, we have $v_1 = u_1$. Then $\{v_1\}$ is an orthogonal set and $\text{span}(v_1) = \text{span}(u_1)$. Assume that for some $1 \leq m < n$, the set $\{v_1, \dots, v_m\}$ is orthogonal and $\text{span}(v_1, \dots, v_m) = \text{span}(u_1, \dots, u_m)$. By definition,

$$v_{m+1} = u_{m+1} - \sum_{j=1}^m \frac{\langle u_{m+1}, v_j \rangle}{\|v_j\|^2} v_j.$$

For any $1 \leq i \leq m$, we have

$$\langle v_{m+1}, v_i \rangle = \left\langle u_{m+1} - \sum_{j=1}^m \frac{\langle u_{m+1}, v_j \rangle}{\|v_j\|^2} v_j, v_i \right\rangle = \langle u_{m+1}, v_i \rangle - \sum_{j=1}^m \frac{\langle u_{m+1}, v_j \rangle}{\|v_j\|^2} \langle v_j, v_i \rangle.$$

By the induction hypothesis, $\{v_1, \dots, v_m\}$ is an orthogonal set, so $\langle v_j, v_i \rangle = 0$ for $j \neq i$. The sum simplifies to the term where $j = i$:

$$\langle v_{m+1}, v_i \rangle = \langle u_{m+1}, v_i \rangle - \frac{\langle u_{m+1}, v_i \rangle}{\|v_i\|^2} \langle v_i, v_i \rangle = \langle u_{m+1}, v_i \rangle - \langle u_{m+1}, v_i \rangle = 0.$$

Thus, v_{m+1} is orthogonal to v_i for all $1 \leq i \leq m$, hence $\{v_1, \dots, v_{m+1}\}$ is orthogonal.

Also, v_{m+1} is a linear combination of u_{m+1} and $\{v_1, \dots, v_m\}$. By the induction hypothesis, $\{v_1, \dots, v_m\} \subset \text{span}(u_1, \dots, u_m)$. Thus $v_{m+1} \in \text{span}(u_1, \dots, u_{m+1})$, and so $\text{span}(v_1, \dots, v_{m+1}) \subseteq \text{span}(u_1, \dots, u_{m+1})$. From the definition of v_{m+1} , we can write

$$u_{m+1} = v_{m+1} + \sum_{j=1}^m \frac{\langle u_{m+1}, v_j \rangle}{\|v_j\|^2} v_j.$$

This shows $u_{m+1} \in \text{span}(v_1, \dots, v_{m+1})$. Since by induction $\text{span}(u_1, \dots, u_m) = \text{span}(v_1, \dots, v_m)$, we have $\text{span}(u_1, \dots, u_{m+1}) \subseteq \text{span}(v_1, \dots, v_{m+1})$. Thus the spans are equal.

Finally, if $v_k = 0$ for some k , then $u_k = \sum_{j=1}^{k-1} \frac{\langle u_k, v_j \rangle}{\|v_j\|^2} v_j$. This means $u_k \in \text{span}(v_1, \dots, v_{k-1}) = \text{span}(u_1, \dots, u_{k-1})$, which contradicts the linear independence of $\{u_1, \dots, u_n\}$. So all v_k are non-zero. Since $\{v_1, \dots, v_n\}$ is an orthogonal set of non-zero vectors in an n -dimensional space, it is an orthogonal basis. $\square$

Example 22.3. Take $V = \mathbb{P}([-1, 1])$ and $\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx$. Consider the subspace $\mathbb{P}_2([-1, 1])$ and the standard basis $\beta = \{1, x, x^2\}$. Find an orthonormal basis for $\mathbb{P}_2([-1, 1])$.

Solution. Let the standard basis be $\{u_1, u_2, u_3\} = \{1, x, x^2\}$. We apply the Gram-Schmidt process to obtain an orthogonal basis $\{v_1, v_2, v_3\}$.

Step 1: Find v_1 . Let $v_1 = u_1 = 1$. We have $\|v_1\|^2 = \langle 1, 1 \rangle = \int_{-1}^1 1 dx = 2$.

Step 2: Find v_2 .

$$v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 = x - \frac{\langle x, 1 \rangle}{2} \cdot 1.$$

The inner product is $\langle x, 1 \rangle = \int_{-1}^1 x dx = 0$. So, $v_2 = x$. We have $\|v_2\|^2 = \langle x, x \rangle = \int_{-1}^1 x^2 dx = \frac{2}{3}$.

Step 3: Find v_3 .

$$v_3 = u_3 - \frac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2 = x^2 - \frac{\langle x^2, 1 \rangle}{2} \cdot 1 - \frac{\langle x^2, x \rangle}{2/3} \cdot x.$$

The inner products are:

$$\langle x^2, 1 \rangle = \int_{-1}^1 x^2 dx = \frac{2}{3}.$$

$$\langle x^2, x \rangle = \int_{-1}^1 x^3 dx = 0.$$

So, $v_3 = x^2 - \frac{2/3}{2} \cdot 1 - 0 = x^2 - \frac{1}{3}$. We have $\|v_3\|^2 = \langle x^2 - \frac{1}{3}, x^2 - \frac{1}{3} \rangle = \int_{-1}^1 (x^2 - \frac{1}{3})^2 dx = \int_{-1}^1 (x^4 - \frac{2}{3}x^2 + \frac{1}{9}) dx = \left[\frac{x^5}{5} - \frac{2x^3}{9} + \frac{x}{9} \right]_{-1}^1 = 2 \left(\frac{1}{5} - \frac{2}{9} + \frac{1}{9} \right) = 2 \left(\frac{1}{5} - \frac{1}{9} \right) = \frac{8}{45}$.

The orthogonal basis is $\{1, x, x^2 - \frac{1}{3}\}$. To get an orthonormal basis, we normalise these vectors:

$$w_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{2}}$$

$$w_2 = \frac{v_2}{\|v_2\|} = \frac{x}{\sqrt{2/3}} = \sqrt{\frac{3}{2}}x$$

$$w_3 = \frac{v_3}{\|v_3\|} = \frac{x^2 - 1/3}{\sqrt{8/45}} = \sqrt{\frac{45}{8}}(x^2 - \frac{1}{3}) = \frac{3\sqrt{5}}{2\sqrt{2}}(x^2 - \frac{1}{3}) = \sqrt{\frac{5}{8}}(3x^2 - 1).$$

The orthonormal basis is $\{\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1)\}$. These are the first three normalised Legendre polynomials. $\square$

Theorem 22.3. Let V be an n -dimensional inner product space with orthonormal basis $\beta = \{v_1, \dots, v_n\}$. Then for any $x \in V$, we have $x = \sum_{i=1}^n \langle x, v_i \rangle v_i$.

Proof. Let $x \in V$. Since β is a basis for V , there exist unique $c_1, \dots, c_n$ such that $x = \sum_{j=1}^n c_j v_j$. Taking the inner product with v_i for $1 \leq i \leq n$, we have $\langle x, v_i \rangle = \langle \sum_{j=1}^n c_j v_j, v_i \rangle = \sum_{j=1}^n c_j \langle v_j, v_i \rangle = \sum_{j=1}^n c_j \delta_{ij} = c_i$. Thus $c_i = \langle x, v_i \rangle$ for each i , and the result follows. $\square$

Example 22.4. Let $\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx$ for $\mathbb{P}_2([-1, 1])$. Represent $f(x) = 1 + 2x + 3x^2$ as a linear combination of the vectors in the orthonormal basis $\beta = \{\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1)\}$.

Solution. Let $f(x) = 1 + 2x + 3x^2$. Let the orthonormal basis be $\beta = \{v_1, v_2, v_3\}$, where

$$v_1(x) = \frac{1}{\sqrt{2}}, \quad v_2(x) = \sqrt{\frac{3}{2}}x, \quad v_3(x) = \sqrt{\frac{5}{8}}(3x^2 - 1).$$

By Theorem 22.3, we have $f = \langle f, v_1 \rangle v_1 + \langle f, v_2 \rangle v_2 + \langle f, v_3 \rangle v_3$. We compute the inner products:

$$\langle f, v_1 \rangle = \int_{-1}^1 (1 + 2x + 3x^2) \frac{1}{\sqrt{2}} dx = \frac{1}{\sqrt{2}} [x + x^2 + x^3]_{-1}^1 = \frac{4}{\sqrt{2}} = 2\sqrt{2}.$$

$$\begin{aligned} \langle f, v_2 \rangle &= \int_{-1}^1 (1 + 2x + 3x^2) \sqrt{\frac{3}{2}}x dx = \sqrt{\frac{3}{2}} \int_{-1}^1 (x + 2x^2 + 3x^3) dx \\ &= \sqrt{\frac{3}{2}} \left[\frac{x^2}{2} + \frac{2x^3}{3} + \frac{3x^4}{4} \right]_{-1}^1 = \sqrt{\frac{3}{2}} \left(\left(\frac{1}{2} + \frac{2}{3} + \frac{3}{4} \right) - \left(\frac{1}{2} - \frac{2}{3} + \frac{3}{4} \right) \right) = \frac{2\sqrt{6}}{3}. \end{aligned}$$

$$\begin{aligned} \langle f, v_3 \rangle &= \int_{-1}^1 (1 + 2x + 3x^2) \sqrt{\frac{5}{8}}(3x^2 - 1) dx = \sqrt{\frac{5}{8}} \int_{-1}^1 (9x^4 + 6x^3 - 2x - 1) dx \\ &= \sqrt{\frac{5}{8}} \left[\frac{9x^5}{5} + \frac{3x^4}{2} - x^2 - x \right]_{-1}^1 = \sqrt{\frac{5}{8}} \left(\left(\frac{9}{5} - 1 \right) - \left(-\frac{9}{5} + 1 \right) \right) = \frac{2\sqrt{10}}{5}. \end{aligned}$$

Thus,

$$f(x) = 2\sqrt{2}v_1(x) + \frac{2\sqrt{6}}{3}v_2(x) + \frac{2\sqrt{10}}{5}v_3(x).$$

As a check,

$$\begin{aligned}
 f(x) &= 2\sqrt{2} \left(\frac{1}{\sqrt{2}} \right) + \frac{2\sqrt{6}}{3} \left(\sqrt{\frac{3}{2}}x \right) + \frac{2\sqrt{10}}{5} \left(\sqrt{\frac{5}{8}}(3x^2 - 1) \right) \\
 &= 2 + \frac{2\sqrt{18}}{3\sqrt{2}}x + \frac{2\sqrt{50}}{5\sqrt{8}}(3x^2 - 1) = 2 + \frac{2 \cdot 3\sqrt{2}}{3\sqrt{2}}x + \frac{2 \cdot 5\sqrt{2}}{5 \cdot 2\sqrt{2}}(3x^2 - 1) \\
 &= 2 + 2x + (3x^2 - 1) = 1 + 2x + 3x^2.
 \end{aligned}$$

□

Theorem 22.4. Let V be an n -dimensional inner product space with an orthonormal basis $\beta = \{v_1, \dots, v_n\}$. Let T be a linear operator on V , and let $A = [T]_\beta$. Then $A_{ij} = \langle Tv_j, v_i \rangle$ for $1 \leq i, j \leq n$.

Proof. By definition of the matrix of a linear operator, the j -th column of A is the coordinate vector of $T(v_j)$ with respect to the basis β . That is, $T(v_j) = \sum_{i=1}^n A_{ij}v_i$. By Theorem 22.3, we also have that for any vector $x \in V$, $x = \sum_{i=1}^n \langle x, v_i \rangle v_i$. Applying this to $x = T(v_j)$, we get $T(v_j) = \sum_{i=1}^n \langle T(v_j), v_i \rangle v_i$. Since the representation of a vector in a basis is unique, we can equate the coefficients to get $A_{ij} = \langle T(v_j), v_i \rangle$. □

§23 Lecture 23—12th April, 2022

§23.1 Adjoint operators

Definition 23.1. Let V and W be inner product spaces over $\mathbb{C}$. Let $T \in \mathcal{L}(V, W)$. The adjoint of T is the unique operator $T^* \in \mathcal{L}(W, V)$ such that

$$\langle Tv, w \rangle = \langle v, T^*w \rangle$$

for all $v \in V$ and $w \in W$.

It is clear from the definition that the form of the adjoint operator of a linear operator depends on the inner product used in the definition.

Let $V = \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ be a real-valued matrix. Define a linear operator on V and let $T = \mathcal{L}_A$. Then for $x, y \in V$,

$$\langle T(x), y \rangle = \langle Ax, y \rangle = y^\top Ax = (A^\top y)^\top x = \langle x, A^\top y \rangle.$$

Thus, we have

$$\langle T(x), y \rangle = \langle x, A^\top y \rangle.$$

This shows that the adjoint operator T^* is given by $T^* = \mathcal{L}_{A^\top}$.

If instead A is a complex matrix, we have

$$\langle T(x), y \rangle = \langle Ax, y \rangle = y^* Ax = (A^* y)^* x = \langle x, A^* y \rangle.$$

Thus, we have

$$\langle T(x), y \rangle = \langle x, A^* y \rangle.$$

This shows that the adjoint operator T^* is given by

$$T^* = \mathcal{L}_{A^*}.$$

If the matrix is symmetric, then $\mathcal{L}_S^* = \mathcal{L}_S$, i.e. S is self-adjoint. If the matrix is Hermitian ($A = A^*$), then $\mathcal{L}_A^* = \mathcal{L}_A$, i.e. A is self-adjoint.

Example 23.2. Let $V = \mathbb{R}^2$ and T an operator defined by

$$T(x) = \begin{bmatrix} 2x_1 + x_2 \\ -3x_1 + 2x_2 \end{bmatrix}.$$

Find the adjoint operator T^* .

Solution. The operator T can be represented by the matrix

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

The adjoint operator T^* is represented by the transpose of this matrix, $A^\top$.

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

Thus, the adjoint operator T^* is given by

$$T^*(y) = A^\top y = \begin{bmatrix} 2 & -3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2y_1 - 3y_2 \\ y_1 + 2y_2 \end{bmatrix}. \quad \square$$

It turns out that every linear operator on a finite-dimensional inner product space has a unique adjoint operator. To prove this, we will need the following representation theorem.

Theorem 23.1 (Finite-dimensional Riesz representation theorem). *Let V be a finite-dimensional inner product space over $\mathbb{C}$ and $g: V \rightarrow \mathbb{C}$ a linear transformation. Then there exists a unique vector $v_g \in V$ such that*

$$g(w) = \langle w, v_g \rangle$$

for all $w \in V$.

Proof. Let $\{e_1, \dots, e_n\}$ be an orthonormal basis for V .

Existence: We construct the vector v_g . Let

$$v_g = \sum_{i=1}^n \overline{g(e_i)} e_i.$$

Then for any $w \in V$, we can write $w = \sum_{i=1}^n \langle w, e_i \rangle e_i$. Since g is a linear transformation,

$$g(w) = g\left(\sum_{i=1}^n \langle w, e_i \rangle e_i\right) = \sum_{i=1}^n \langle w, e_i \rangle g(e_i).$$

On the other hand, using our constructed v_g :

$$\begin{aligned}\langle w, v_g \rangle &= \left\langle w, \sum_{j=1}^n \overline{g(e_j)} e_j \right\rangle \\ &= \sum_{j=1}^n \overline{g(e_j)} \langle w, e_j \rangle \quad (\text{conjugate linearity in the second argument}) \\ &= \sum_{j=1}^n g(e_j) \langle w, e_j \rangle.\end{aligned}$$

Comparing the two expressions, we see that $g(w) = \langle w, v_g \rangle$ for all $w \in V$. This proves the existence of such a vector v_g .

Uniqueness: Suppose there exists another vector $v'_g \in V$ such that $g(w) = \langle w, v'_g \rangle$ for all $w \in V$. Then we have $\langle w, v_g \rangle = \langle w, v'_g \rangle$ for all $w \in V$. This implies $\langle w, v_g \rangle - \langle w, v'_g \rangle = 0$, and by linearity, $\langle w, v_g - v'_g \rangle = 0$ for all $w \in V$. In particular, we can choose $w = v_g - v'_g$. This gives

$$\langle v_g - v'_g, v_g - v'_g \rangle = 0.$$

This is equivalent to $\|v_g - v'_g\|^2 = 0$. By the positive-definiteness property of the inner product, this implies $v_g - v'_g = 0$, so $v_g = v'_g$. Thus, the vector v_g is unique. $\square$

Theorem 23.2 (Existence and uniqueness of adjoint operators). *Let V be a finite-dimensional inner product space and T a linear operator on V . Then T has a unique adjoint that is also a linear operator on V .*

Proof. Let w be an arbitrary vector in V . Define a function $g_w : V \rightarrow \mathbb{C}$ by $g_w(v) = \langle T(v), w \rangle$. We first show that g_w is a linear functional on V . For any $v_1, v_2 \in V$ and $c \in \mathbb{C}$:

$$\begin{aligned}g_w(cv_1 + v_2) &= \langle T(cv_1 + v_2), w \rangle \\ &= \langle cT(v_1) + T(v_2), w \rangle \quad (\text{since } T \text{ is linear}) \\ &= c\langle T(v_1), w \rangle + \langle T(v_2), w \rangle \quad (\text{by linearity of inner product}) \\ &= cg_w(v_1) + g_w(v_2).\end{aligned}$$

Thus, g_w is a linear functional. By the Riesz representation theorem, for each $w \in V$, there exists a unique vector in V , which we will denote by $T^*(w)$, such that $g_w(v) = \langle v, T^*(w) \rangle$ for all $v \in V$. Therefore, we have defined an operator $T^* : V \rightarrow V$ that satisfies $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$ for all $v, w \in V$.

Next, we must show that this operator T^* is linear. Let $w_1, w_2 \in V$ and $c \in \mathbb{C}$. For any $v \in V$:

$$\begin{aligned}\langle v, T^*(cw_1 + w_2) \rangle &= \langle T(v), cw_1 + w_2 \rangle \quad (\text{by definition of } T^*) \\ &= \overline{c}\langle T(v), w_1 \rangle + \langle T(v), w_2 \rangle \quad (\text{by conjugate linearity in the second argument}) \\ &= \overline{c}\langle v, T^*(w_1) \rangle + \langle v, T^*(w_2) \rangle \\ &= \langle v, cT^*(w_1) \rangle + \langle v, T^*(w_2) \rangle \\ &= \langle v, cT^*(w_1) + T^*(w_2) \rangle.\end{aligned}$$

Since this holds for all $v \in V$, we must have $T^*(cw_1 + w_2) = cT^*(w_1) + T^*(w_2)$. Thus, T^* is a linear operator. This proves the existence of an adjoint operator.

Finally, we prove uniqueness. Suppose T_1^* and T_2^* are two operators that both satisfy the condition for being an adjoint of T . Then for all $v, w \in V$:

$$\langle v, T_1^*(w) \rangle = \langle T(v), w \rangle = \langle v, T_2^*(w) \rangle.$$

This implies $\langle v, T_1^*(w) - T_2^*(w) \rangle = 0$ for all $v, w \in V$. Let's choose $v = T_1^*(w) - T_2^*(w)$. Then we have

$$\langle T_1^*(w) - T_2^*(w), T_1^*(w) - T_2^*(w) \rangle = 0,$$

which means $\|T_1^*(w) - T_2^*(w)\|^2 = 0$. By the properties of the norm, this implies $T_1^*(w) - T_2^*(w) = 0$, so $T_1^*(w) = T_2^*(w)$. Since this is true for all $w \in V$, the operators are the same: $T_1^* = T_2^*$. This proves uniqueness. $\square$

Example 23.3. Let $V = \{f : f \in C^1([-1, 1]) \text{ and } f(1) = f(-1)\}$. Define the linear operator T by $T(f) = f' + 3f$ and the inner product by $\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx$. Find the adjoint operator T^* .

Solution. We want to find an operator T^* such that $\langle T(f), g \rangle = \langle f, T^*(g) \rangle$ for all $f, g \in V$. Let's start with the left-hand side:

$$\begin{aligned} \langle T(f), g \rangle &= \langle f' + 3f, g \rangle \\ &= \int_{-1}^1 (f'(x) + 3f(x))g(x) dx \\ &= \int_{-1}^1 f'(x)g(x) dx + \int_{-1}^1 3f(x)g(x) dx. \end{aligned}$$

We use integration by parts on the first term, $\int f'g dx$. Let $u = g(x)$ and $dv = f'(x)dx$. Then $du = g'(x)dx$ and $v = f(x)$.

$$\int_{-1}^1 f'(x)g(x) dx = [f(x)g(x)]_{-1}^1 - \int_{-1}^1 f(x)g'(x) dx.$$

The boundary term is $f(1)g(1) - f(-1)g(-1)$. Since $f, g \in V$, we have the boundary conditions $f(1) = f(-1)$ and $g(1) = g(-1)$. Therefore, the boundary term becomes $f(1)g(1) - f(1)g(1) = 0$.

Substituting this back into the expression for the inner product:

$$\begin{aligned} \langle T(f), g \rangle &= - \int_{-1}^1 f(x)g'(x) dx + \int_{-1}^1 3f(x)g(x) dx \\ &= \int_{-1}^1 f(x)(-g'(x) + 3g(x)) dx \\ &= \langle f, -g' + 3g \rangle. \end{aligned}$$

By definition of the adjoint, we must have $\langle T(f), g \rangle = \langle f, T^*(g) \rangle$. Comparing the two expressions, we can identify the adjoint operator as

$$T^*(g) = -g' + 3g.$$

We must also ensure that T^* maps from V to V . If $g \in V$, then $g \in C^1([-1, 1])$ and $g(1) = g(-1)$. Then $T^*(g) = -g' + 3g$ is also in $C^1([-1, 1])$. We check the boundary condition:

$$(T^*g)(1) = -g'(1) + 3g(1) \quad \text{and} \quad (T^*g)(-1) = -g'(-1) + 3g(-1).$$

The problem statement does not impose conditions on the derivatives at the boundary, so we cannot assume $(T^*g)(1) = (T^*g)(-1)$. The domain of T^* is V , but the codomain is not necessarily V . However, the operator is well-defined. $\square$

Theorem 23.3 (Matrix representation of the adjoint operator). *Let V be a finite-dimensional inner product space and $\beta = \{v_1, \dots, v_n\}$ an orthonormal basis for V . Let T be a linear operator on V and $[T]_\beta$ the matrix representation of T with respect to β . Then*

$$[T^*]_\beta = [T]_\beta^*.$$

Proof. Let $A = [T]_\beta$ and $B = [T^*]_\beta$. We want to show that $B = A^*$.

Since β is an orthonormal basis, the entries of the matrix representation of a linear operator L are given by $([L]_\beta)_{ij} = \langle L(v_j), v_i \rangle$.

Applying this formula to T , we get the entries of matrix A :

$$A_{ij} = \langle T(v_j), v_i \rangle.$$

Applying the same formula to the adjoint operator T^* , we get the entries of matrix B :

$$B_{ij} = \langle T^*(v_j), v_i \rangle.$$

Now, we relate the entries of B to the entries of A . By the definition of the adjoint operator, we have $\langle u, T^*(w) \rangle = \langle T(u), w \rangle$. Let $u = v_i$ and $w = v_j$. Then

$$\langle v_i, T^*(v_j) \rangle = \langle T(v_i), v_j \rangle.$$

Let's look at the entry B_{ij} :

$$\begin{aligned} B_{ij} &= \langle T^*(v_j), v_i \rangle \\ &= \overline{\langle v_i, T^*(v_j) \rangle} \quad (\text{by conjugate symmetry of the inner product}) \\ &= \overline{\langle T(v_i), v_j \rangle} \quad (\text{using the relation derived above}). \end{aligned}$$

The term $\langle T(v_i), v_j \rangle$ is the entry A_{ji} of the matrix A . So, we have shown that

$$B_{ij} = \overline{A_{ji}}.$$

This is precisely the definition of the conjugate transpose of the matrix A . The (i, j) -th entry of A^* is the complex conjugate of the (j, i) -th entry of A . Therefore, $B = A^*$, which means $[T^*]_\beta = [T]_\beta^*$. $\square$

Theorem 23.4 (Properties of the adjoint operator). *Let V be an inner product space, and let T and U be two linear operators on V . Then:*

1. $(T + U)^* = T^* + U^*$.
2. $(cT)^* = \bar{c}T^*$ for any scalar c .
3. $(TU)^* = U^*T^*$.
4. $(T^*)^* = T$.
5. $I^* = I$, where I is the identity operator on V .

Proof. We prove each property using the definition of the adjoint operator, $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$, and its uniqueness.

1. For any $v, w \in V$:

$$\begin{aligned}
\langle (T + U)(v), w \rangle &= \langle T(v) + U(v), w \rangle \quad (\text{by definition of operator sum}) \\
&= \langle T(v), w \rangle + \langle U(v), w \rangle \quad (\text{by linearity of inner product}) \\
&= \langle v, T^*(w) \rangle + \langle v, U^*(w) \rangle \quad (\text{by definition of adjoints}) \\
&= \langle v, T^*(w) + U^*(w) \rangle \quad (\text{by linearity of inner product}) \\
&= \langle v, (T^* + U^*)(w) \rangle \quad (\text{by definition of operator sum}).
\end{aligned}$$

By the uniqueness of the adjoint, we conclude that $(T + U)^* = T^* + U^*$.

2. For any $v, w \in V$ and scalar c :

$$\begin{aligned}
\langle (cT)(v), w \rangle &= \langle c(T(v)), w \rangle \quad (\text{by definition of scalar-operator product}) \\
&= c\langle T(v), w \rangle \quad (\text{by linearity of inner product}) \\
&= c\langle v, T^*(w) \rangle \quad (\text{by definition of adjoint}) \\
&= \langle v, \bar{c}T^*(w) \rangle \quad (\text{by conjugate linearity in second argument}) \\
&= \langle v, (\bar{c}T^*)(w) \rangle \quad (\text{by definition of scalar-operator product}).
\end{aligned}$$

By the uniqueness of the adjoint, we conclude that $(cT)^* = \bar{c}T^*$.

3. For any $v, w \in V$:

$$\begin{aligned}
\langle (TU)(v), w \rangle &= \langle T(U(v)), w \rangle \quad (\text{by definition of operator composition}) \\
&= \langle U(v), T^*(w) \rangle \quad (\text{by definition of } T^*) \\
&= \langle v, U^*(T^*(w)) \rangle \quad (\text{by definition of } U^*) \\
&= \langle v, (U^*T^*)(w) \rangle \quad (\text{by definition of operator composition}).
\end{aligned}$$

By the uniqueness of the adjoint, we conclude that $(TU)^* = U^*T^*$.

4. For any $v, w \in V$:

$$\begin{aligned}
\langle v, (T^*)^*(w) \rangle &= \langle T^*(v), w \rangle \quad (\text{by definition of the adjoint of } T^*) \\
&= \overline{\langle w, T^*(v) \rangle} \quad (\text{by conjugate symmetry}) \\
&= \overline{\langle T(w), v \rangle} \quad (\text{by definition of } T^*) \\
&= \langle v, T(w) \rangle \quad (\text{by conjugate symmetry}).
\end{aligned}$$

Since this holds for all $v \in V$, we must have $(T^*)^*(w) = T(w)$ for all $w \in V$. Thus, $(T^*)^* = T$.

5. For any $v, w \in V$:

$$\langle I(v), w \rangle = \langle v, w \rangle.$$

We are looking for an operator I^* such that $\langle I(v), w \rangle = \langle v, I^*(w) \rangle$. Comparing the two equations, we have $\langle v, w \rangle = \langle v, I^*(w) \rangle$. This implies $\langle v, w - I^*(w) \rangle = 0$ for all $v \in V$. Choosing $v = w - I^*(w)$, we get $\|w - I^*(w)\|^2 = 0$, which implies $w - I^*(w) = 0$, so $I^*(w) = w = I(w)$. Thus, $I^* = I$. $\square$

§24 Lecture 24—14th April, 2022

§24.1 The singular value decomposition (SVD)

In our study of linear operators, the spectral theorem has been a powerful tool. It tells us that for any self-adjoint operator on an inner product space, there exists an orthonormal basis of eigenvectors. In this basis, the operator is represented by a simple diagonal matrix. This is a very satisfying result, as it completely characterizes the operator's action as simple scaling along orthogonal directions.

However, what about operators that are not self-adjoint? Or linear transformations between different vector spaces, $T : V \rightarrow W$, which are represented by non-square matrices? We can no longer hope for a single basis of eigenvectors, as the domain and codomain are different.

The central question is: can we still find a pair of orthonormal bases, one for the domain V and one for the codomain W , such that T has a simple, diagonal-like structure with respect to these bases? The Singular Value Decomposition (SVD) provides a resounding "yes" to this question. It is arguably one of the most important and useful matrix factorizations in all of linear algebra.

Geometrically, the SVD states that any linear transformation $T : V \rightarrow W$ maps an orthonormal basis of V to an orthogonal (but not necessarily orthonormal) set of vectors in W . This means that the action of any linear map can be understood as:

1. A rotation/reflection in the domain space V .
2. A scaling along the new coordinate axes.
3. A rotation/reflection in the codomain space W .

This implies that a linear map always transforms a hypersphere in its domain into a hyperellipsoid (possibly a degenerate one, like a line or a point) in its codomain. The singular values are precisely the lengths of the semi-axes of this resulting hyperellipsoid.

Connecting SVD to Eigen-decomposition To find the special bases for SVD, we can't use T directly, as it maps between different spaces. The key insight is to construct a related operator that is self-adjoint and lives in one space. Given $T : V \rightarrow W$, we can form two important self-adjoint operators:

- $T^*T : V \rightarrow V$
- $TT^* : W \rightarrow W$

These operators are self-adjoint (since $(T^*T)^* = T^*(T^*)^* = T^*T$) and positive semi-definite (since $\langle T^*T(v), v \rangle = \langle T(v), T(v) \rangle = \|T(v)\|^2 \geq 0$).

By the spectral theorem, we can find an orthonormal basis for V consisting of eigenvectors of T^*T . Let these eigenvectors be $\{v_1, \dots, v_n\}$ with corresponding non-negative eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$. These eigenvectors will be our special basis for the domain V .

The SVD for Matrices Let's formalize this by translating it into the language of matrices. Let $A \in \mathbb{C}^{m \times n}$ represent the linear transform $T : \mathbb{C}^n \rightarrow \mathbb{C}^m$. The adjoint T^* is represented by the conjugate transpose A^* . The operators T^*T and TT^* correspond to the matrices A^*A (an $n \times n$ Hermitian matrix) and AA^* (an $m \times m$ Hermitian matrix).

Theorem 24.1 (Singular Value Decomposition). *Let $A \in \mathbb{C}^{m \times n}$ be a matrix with rank r . Then A can be factored as*

$$A = U\Sigma V^*$$

where:

- $U \in \mathbb{C}^{m \times m}$ is a unitary matrix. Its columns $\{u_1, \dots, u_m\}$ are the left singular vectors.
- $V \in \mathbb{C}^{n \times n}$ is a unitary matrix. Its columns $\{v_1, \dots, v_n\}$ are the right singular vectors.
- $\Sigma \in \mathbb{R}^{m \times n}$ is a rectangular diagonal matrix. The diagonal entries $\sigma_i = \Sigma_{ii}$ are the singular values of A , ordered such that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ and $\sigma_i = 0$ for $i > r$.

Proof. The matrix A^*A is an $n \times n$ Hermitian, positive semi-definite matrix. By the Spectral Theorem, there exists a unitary matrix $V \in \mathbb{C}^{n \times n}$ whose columns are orthonormal eigenvectors of A^*A , such that

$$V^*(A^*A)V = D = \text{diag}(\lambda_1, \dots, \lambda_n)$$

where λ_i are the real, non-negative eigenvalues of A^*A , ordered such that $\lambda_1 \geq \dots \geq \lambda_n \geq 0$. The rank of A^*A is equal to the rank of A , which is r . Thus, $\lambda_1, \dots, \lambda_r > 0$ and $\lambda_{r+1}, \dots, \lambda_n = 0$.

We define the **singular values** as the square roots of these eigenvalues: $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \dots, n$. Thus, $\sigma_1 \geq \dots \geq \sigma_r > 0$ and $\sigma_{r+1}, \dots, \sigma_n = 0$.

Now, we define the first r left singular vectors $\{u_1, \dots, u_r\}$ as:

$$u_i = \frac{1}{\sigma_i} Av_i \quad \text{for } i = 1, \dots, r.$$

Let's verify that these vectors are orthonormal. For $1 \leq i, j \leq r$:

$$\langle u_i, u_j \rangle = u_j^* u_i = \left(\frac{1}{\sigma_j} Av_j \right)^* \left(\frac{1}{\sigma_i} Av_i \right) = \frac{1}{\sigma_i \sigma_j} v_j^* A^* A v_i.$$

Since v_i is an eigenvector of A^*A with eigenvalue $\lambda_i = \sigma_i^2$, we have $A^* A v_i = \sigma_i^2 v_i$.

$$\langle u_i, u_j \rangle = \frac{1}{\sigma_i \sigma_j} v_j^* (\sigma_i^2 v_i) = \frac{\sigma_i^2}{\sigma_i \sigma_j} (v_j^* v_i).$$

Since the columns of V form an orthonormal set, $v_j^* v_i = \delta_{ij}$.

$$\langle u_i, u_j \rangle = \frac{\sigma_i^2}{\sigma_i \sigma_j} \delta_{ij} = \delta_{ij}.$$

So, $\{u_1, \dots, u_r\}$ is an orthonormal set in $\mathbb{C}^m$. This set can be extended to a full orthonormal basis $\{u_1, \dots, u_m\}$ for $\mathbb{C}^m$ using, for example, the Gram-Schmidt process on a basis for the orthogonal complement of $\text{span}(u_1, \dots, u_r)$. Let U be the matrix with these vectors as columns.

By construction, we have $Av_i = \sigma_i u_i$ for $i = 1, \dots, r$. For $i > r$, we have $\lambda_i = 0$. Then $\|Av_i\|^2 = (Av_i)^*(Av_i) = v_i^* A^* Av_i = v_i^* (\lambda_i v_i) = 0$. This implies $Av_i = 0$ for $i > r$. We can write all these relations in a single matrix equation:

$$AV = U\Sigma$$

where Σ is the $m \times n$ rectangular matrix with the singular values $\sigma_1, \dots, \sigma_r$ on its main diagonal and zeros everywhere else. Since V is unitary, $V^{-1} = V^*$. Multiplying on the right by V^* gives the desired result:

$$A = U\Sigma V^*.$$

□

Relationship with Eigenvectors of AA^* We constructed the right singular vectors v_i from the eigenvectors of A^*A . What about the left singular vectors u_i ? They are eigenvectors of AA^* . From $A = U\Sigma V^*$, we can derive the eigendecomposition of AA^* :

$$AA^* = (U\Sigma V^*)(V\Sigma^T U^*) = U(\Sigma\Sigma^T)U^*.$$

Since U is unitary, this is the eigendecomposition of AA^* . The columns of U are the eigenvectors of AA^* , and the eigenvalues are the diagonal entries of $\Sigma\Sigma^T$, which are $\sigma_1^2, \dots, \sigma_r^2$ and $m - r$ zeros. So, the left singular vectors $\{u_i\}$ are the orthonormal eigenvectors of AA^* . This gives us a symmetric way to think about the SVD:

- The right singular vectors V diagonalize A^*A .
- The left singular vectors U diagonalize AA^* .
- Both matrices A^*A and AA^* have the same non-zero eigenvalues, $\sigma_1^2, \dots, \sigma_r^2$.

§25 Lecture 25—19th April, 2022

§25.1 The singular value decomposition: forms and subspaces

The SVD can be expressed in several forms depending on the application. Let $T: V \rightarrow W$ be a linear transform between finite-dimensional inner product spaces, with $\dim(V) = n$ and $\dim(W) = m$. Let $r = \text{rank}(T)$. The singular value theorem gives us orthonormal bases $\beta = \{v_1, \dots, v_n\}$ for V and $\gamma = \{u_1, \dots, u_m\}$ for W , and singular values $\sigma_1 \geq \dots \geq \sigma_r > 0$ such that

$$T(v_i) = \begin{cases} \sigma_i u_i & \text{if } 1 \leq i \leq r \\ 0 & \text{if } i > r. \end{cases}$$

This representation of the operator T is fundamental. When we represent T as a matrix $A = [T]_\beta^\gamma \in \mathbb{C}^{m \times n}$, this leads to the matrix decomposition $A = U\Sigma V^*$.

- **Full SVD:** The matrices are $U \in \mathbb{C}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{C}^{n \times n}$. This is the most general form. The matrix Σ has the singular values on its main diagonal, and is zero otherwise.
- **Thin SVD:** If $m \geq n$, the last $m - n$ rows of Σ are zero. The product $U\Sigma$ only uses the first n columns of U . We can write $A = U_1 \hat{\Sigma} V^*$, where $U_1 \in \mathbb{C}^{m \times n}$ consists of the first n columns of U and $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ is a square matrix with the singular values on its diagonal. This is the “thin” SVD. Similarly, if $n > m$, we can take the first m columns of V .
- **Truncated SVD:** The most compact and widely used form expresses T (or A) as a sum of rank-one operators (or matrices).

$$T(v) = \sum_{i=1}^r \sigma_i \langle v, v_i \rangle u_i \quad \text{for any } v \in V.$$

This is the outer product form. The corresponding matrix form is $A = \sum_{i=1}^r \sigma_i u_i v_i^*$. This is also called the reduced SVD.

SVD and the Four Fundamental Subspaces The SVD provides orthonormal bases for the four fundamental subspaces associated with a linear operator $T: V \rightarrow W$.

1. *Range*, $\text{range}(T) = \{T(v) \mid v \in V\} \subseteq W$.
2. *Kernel (Null Space)*, $\ker(T) = \{v \in V \mid T(v) = 0\} \subseteq V$.
3. *Range of Adjoint*, $\text{range}(T^*) = \{T^*(w) \mid w \in W\} \subseteq V$. This corresponds to the row space for matrices.
4. *Kernel of Adjoint*, $\ker(T^*) = \{w \in W \mid T^*(w) = 0\} \subseteq W$. This corresponds to the left null space for matrices.

Let $\{v_1, \dots, v_n\}$ and $\{u_1, \dots, u_m\}$ be the orthonormal bases from the singular value theorem for T , and let $r = \text{rank}(T)$.

Theorem 25.1 (SVD and Fundamental Subspaces). *With the notation above:*

1. The set $\{u_1, \dots, u_r\}$ is an orthonormal basis for $\text{range}(T)$.
2. The set $\{v_{r+1}, \dots, v_n\}$ is an orthonormal basis for $\ker(T)$.
3. The set $\{v_1, \dots, v_r\}$ is an orthonormal basis for $\text{range}(T^*)$.
4. The set $\{u_{r+1}, \dots, u_m\}$ is an orthonormal basis for $\ker(T^*)$.

Proof. From the singular value theorem, we have the relations $T(v_i) = \sigma_i u_i$ for $i \leq r$ and $T(v_i) = 0$ for $i > r$. The adjoint operator T^* satisfies $T^*(u_i) = \sigma_i v_i$ for $i \leq r$ and $T^*(u_i) = 0$ for $i > r$.

1. **Range** $\text{range}(T)$: Any vector in the range is of the form $T(v)$ for some $v \in V$. We can express v in the basis $\{v_j\}$: $v = \sum_{j=1}^n c_j v_j$.

$$T(v) = T\left(\sum_{j=1}^n c_j v_j\right) = \sum_{j=1}^n c_j T(v_j) = \sum_{j=1}^r c_j (\sigma_j u_j) + \sum_{j=r+1}^n c_j (0) = \sum_{j=1}^r (c_j \sigma_j) u_j.$$

This shows that any vector in $\text{range}(T)$ is a linear combination of $\{u_1, \dots, u_r\}$. Since these vectors are orthonormal, they are linearly independent and thus form an orthonormal basis for $\text{range}(T)$.

2. **Kernel** $\ker(T)$: We seek vectors v such that $T(v) = 0$. Let $v = \sum_{j=1}^n c_j v_j$.

$$T(v) = \sum_{j=1}^r (c_j \sigma_j) u_j = 0.$$

Since $\{u_1, \dots, u_r\}$ are linearly independent and $\sigma_j > 0$ for $j \leq r$, we must have $c_j = 0$ for $j = 1, \dots, r$. So, any vector v in the kernel must be of the form $v = \sum_{j=r+1}^n c_j v_j$. This means $\{v_{r+1}, \dots, v_n\}$ spans the kernel. Being orthonormal, they form an orthonormal basis for $\ker(T)$.

3. **Range** $\text{range}(T^*)$: The proof is symmetric to the range of T . Any vector in $\text{range}(T^*)$ is of the form $T^*(w)$ for $w \in W$. Express $w = \sum_{i=1}^m d_i u_i$.

$$T^*(w) = T^*\left(\sum_{i=1}^m d_i u_i\right) = \sum_{i=1}^m d_i T^*(u_i) = \sum_{i=1}^r d_i (\sigma_i v_i).$$

This shows that $\{v_1, \dots, v_r\}$ is an orthonormal basis for $\text{range}(T^*)$.

4. **Kernel** $\ker(T^*)$: Symmetric to the kernel of T . We seek w such that $T^*(w) = 0$. Let $w = \sum_{i=1}^m d_i u_i$.

$$T^*(w) = \sum_{i=1}^r (d_i \sigma_i) v_i = 0.$$

Since $\{v_1, \dots, v_r\}$ are linearly independent, we must have $d_i = 0$ for $i = 1, \dots, r$. So, any vector w in the kernel of T^* must be of the form $w = \sum_{i=r+1}^m d_i u_i$. Thus, $\{u_{r+1}, \dots, u_m\}$ is an orthonormal basis for $\ker(T^*)$.

□

This gives the orthogonal decompositions of the domain and codomain:

- $V = \text{range}(T^*) \oplus \ker(T)$
- $W = \text{range}(T) \oplus \ker(T^*)$

The operator T is a bijection from $\text{range}(T^*)$ to $\text{range}(T)$. This provides a complete, geometric understanding of the operator's action.

§26 Lecture 26—21st April, 2022

§26.1 Low-rank approximation and geometric interpretation

The Eckart-Young-Mirsky Theorem One of the most powerful applications of the SVD is finding the best low-rank approximation to a given linear operator. This is the theoretical foundation for applications like data compression and noise reduction.

The problem is as follows: Given a linear operator $T: V \rightarrow W$ of rank r , and an integer $k < r$, find an operator $T_k: V \rightarrow W$ of rank k that is closest to T . The notion of “closest” is measured by an operator norm.

The SVD provides a direct answer. The best rank- k approximation is formed by truncating the singular value expansion of T . Let the SVD of T be given by $T(v) = \sum_{i=1}^r \sigma_i \langle v, v_i \rangle u_i$. The optimal rank- k approximation is the operator T_k defined by:

$$T_k(v) = \sum_{i=1}^k \sigma_i \langle v, v_i \rangle u_i$$

Theorem 26.1 (Eckart-Young-Mirsky). *Let $T: V \rightarrow W$ be a linear operator of rank r , with singular values $\sigma_1 \geq \dots \geq \sigma_r > 0$. For any k with $0 \leq k < r$, let $T_k = \sum_{i=1}^k \sigma_i \langle \cdot, v_i \rangle u_i$. Then, T_k is the best rank- k approximation to T in the operator norm:*

$$\|T - T_k\| = \min_{\text{rank}(S)=k} \|T - S\| = \sigma_{k+1}$$

where $\|T\|$ denotes the operator norm induced by the vector norms on V and W .

Proof. First, let's compute the norm of the error operator $T - T_k$.

$$(T - T_k)(v) = \sum_{i=k+1}^r \sigma_i \langle v, v_i \rangle u_i.$$

This is the singular value expansion for the operator $T - T_k$. The basis vectors are the same, but the singular values are now $\{\sigma_{k+1}, \dots, \sigma_r\}$. The largest singular value of $T - T_k$ is σ_{k+1} . The operator norm of an operator is its largest singular value. Therefore,

$$\|T - T_k\| = \sigma_{k+1}.$$

Now, we must show that for any other rank- k operator $S: V \rightarrow W$, we have $\|T - S\| \geq \sigma_{k+1}$. Let S be any operator with $\text{rank}(S) = k$. The kernel of S , $\ker(S)$, has dimension $\dim(\ker(S)) = \dim(V) - \text{rank}(S) = n - k$. Consider the subspace $V_{k+1} = \text{span}\{v_1, v_2, \dots, v_{k+1}\}$. The dimension of V_{k+1} is $k + 1$. Since $\dim(\ker(S)) + \dim(V_{k+1}) = (n - k) + (k + 1) = n + 1 > n = \dim(V)$, the two subspaces must have a non-trivial intersection. So, there exists a non-zero vector $z \in \ker(S) \cap V_{k+1}$. We can normalize this vector such that $\|z\| = 1$.

Since $z \in V_{k+1}$, we can write $z = \sum_{i=1}^{k+1} c_i v_i$, where $c_i = \langle z, v_i \rangle$. Because $\{v_i\}$ are orthonormal, $\|z\|^2 = \sum_{i=1}^{k+1} |c_i|^2 = 1$.

Now, let's examine the action of the operator $(T - S)$ on z . Since $z \in \ker(S)$, we have $S(z) = 0$. So, $\|(T - S)(z)\| = \|T(z)\|$. Let's compute $T(z)$:

$$T(z) = T\left(\sum_{i=1}^{k+1} c_i v_i\right) = \sum_{i=1}^{k+1} c_i T(v_i) = \sum_{i=1}^{k+1} c_i \sigma_i u_i.$$

Now we compute the squared norm:

$$\|T(z)\|^2 = \left\| \sum_{i=1}^{k+1} c_i \sigma_i u_i \right\|^2$$

Since $\{u_i\}$ are also orthonormal, this is:

$$\|T(z)\|^2 = \sum_{i=1}^{k+1} |c_i \sigma_i|^2 = \sum_{i=1}^{k+1} |c_i|^2 \sigma_i^2.$$

We know that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{k+1}$. Therefore, $\sigma_i^2 \geq \sigma_{k+1}^2$ for all $i \leq k+1$.

$$\|T(z)\|^2 \geq \sum_{i=1}^{k+1} |c_i|^2 \sigma_{k+1}^2 = \sigma_{k+1}^2 \sum_{i=1}^{k+1} |c_i|^2 = \sigma_{k+1}^2 (1) = \sigma_{k+1}^2.$$

So, we have shown that $\|(T - S)(z)\| \geq \sigma_{k+1}$ for this specific unit vector z .

The operator norm of an operator L is defined as $\|L\| = \sup_{\|v\|=1} \|L(v)\|$. Therefore,

$$\|T - S\| = \sup_{\|v\|=1} \|(T - S)(v)\| \geq \|(T - S)(z)\| \geq \sigma_{k+1}.$$

This completes the proof. A similar argument holds for the Frobenius norm when working with matrices. $\square$

Geometric Interpretation The SVD provides a powerful geometric interpretation of the action of a linear operator $T: V \rightarrow W$. The transformation can be seen as a sequence of three fundamental geometric actions:

1. **A rotation in the domain V :** The basis $\{v_i\}$ is an orthonormal basis for V . A vector $v \in V$ is expressed in this basis via the coordinates $\langle v, v_i \rangle$.
2. **A scaling along coordinate axes:** The operator acts by scaling the components in the v_i basis. The component $\langle v, v_i \rangle$ is scaled by the singular value σ_i . For $i > r$, the scaling is by zero, which is a projection onto the subspace $\text{span}\{v_1, \dots, v_r\} = \text{range}(T^*)$.
3. **A rotation in the codomain W :** The scaled components are then expressed in the orthonormal basis $\{u_i\}$ of W . The i -th scaled component becomes the coefficient of the basis vector u_i .

A common way to visualize this is to consider the image of the unit sphere in V under the operator T . Let $S = \{v \in V : \|v\| = 1\}$ be the unit sphere in V . The image of S under T , i.e., the set $\{T(v) \mid v \in S\}$, is an ellipsoid in W . The SVD reveals the geometry of this ellipsoid. The vectors $\{u_i\}$ define the principal axes of the ellipsoid. The lengths of the semi-axes are the singular values $\sigma_1, \sigma_2, \dots, \sigma_r$. The remaining $m - r$ axes have length 0, meaning the ellipsoid is contained within the r -dimensional subspace $\text{range}(T)$. The longest semi-axis is in the direction of u_1 with length $\sigma_1 = \|T(v_1)\| = \|T\|$, the operator norm of T . The shortest non-zero semi-axis is in the direction of u_r with length $\sigma_r = \|T(v_r)\|$.

§27 Lecture 27—26th April, 2022

§27.1 The singular value theorem for linear transforms

Theorem 27.1 (Singular value theorem for linear transforms). *Let V and W be finite-dimensional inner product spaces, and let $T: V \rightarrow W$ be a linear transform of rank r . Then there exist orthonormal*

bases $\{v_1, \dots, v_n\}$ for V and $\{u_1, \dots, u_m\}$ for W as well as positive scalars $\sigma_1 \geq \dots \geq \sigma_r$ such that

$$T(v_i) = \begin{cases} \sigma_i u_i & \text{if } 1 \leq i \leq r \\ 0 & \text{if } i > r. \end{cases}$$

Conversely, suppose that the preceding conditions are satisfied. Then for $1 \leq i \leq n$, v_i is an eigenvector of T^*T with eigenvalue σ_i^2 if $1 \leq i \leq r$, and 0 if $i > r$. In this way the scalars σ_i are uniquely determined by the linear transform T .

Proof. The operator T^*T is self-adjoint and positive semi-definite. By the spectral theorem, there is an orthonormal basis $\{v_1, \dots, v_n\}$ for V consisting of eigenvectors of T^*T . Let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. Since T^*T is positive semi-definite, these eigenvalues are non-negative. We may reorder them so that $\lambda_1 \geq \dots \geq \lambda_n \geq 0$. The rank of T^*T is equal to the rank of T , which is r . Thus $\lambda_1, \dots, \lambda_r$ are positive, and $\lambda_i = 0$ for $i > r$.

For $1 \leq i \leq r$, define $\sigma_i = \sqrt{\lambda_i}$. Then $\sigma_1 \geq \dots \geq \sigma_r > 0$. For $1 \leq i \leq r$, define $u_i = \frac{1}{\sigma_i} T(v_i)$. Then

$$\langle u_i, u_j \rangle = \frac{1}{\sigma_i \sigma_j} \langle T(v_i), T(v_j) \rangle = \frac{1}{\sigma_i \sigma_j} \langle T^*T(v_i), v_j \rangle = \frac{\lambda_i}{\sigma_i \sigma_j} \langle v_i, v_j \rangle = \frac{\sigma_i^2}{\sigma_i \sigma_j} \delta_{ij} = \delta_{ij}.$$

Thus $\{u_1, \dots, u_r\}$ is an orthonormal set. We may extend it to an orthonormal basis $\{u_1, \dots, u_m\}$ for W . By construction, $T(v_i) = \sigma_i u_i$ for $1 \leq i \leq r$. For $i > r$, we have $T^*T(v_i) = 0$, so $\langle T(v_i), T(v_i) \rangle = \langle T^*T(v_i), v_i \rangle = 0$, which implies that $T(v_i) = 0$.

Conversely, suppose the conditions are satisfied. Then

$$T^*T(v_i) = T^*(\sigma_i u_i) = \sigma_i T^*(u_i).$$

We have that $T^*(u_j) = \sum_{k=1}^n \langle T^*(u_j), v_k \rangle v_k = \sum_{k=1}^n \langle u_j, T(v_k) \rangle v_k$. For $k \leq r$, $\langle u_j, T(v_k) \rangle = \langle u_j, \sigma_k u_k \rangle = \sigma_k \delta_{jk}$. For $k > r$, $\langle u_j, T(v_k) \rangle = \langle u_j, 0 \rangle = 0$. Thus $T^*(u_j) = \sigma_j v_j$ for $j \leq r$, and $T^*(u_j) = 0$ for $j > r$. It follows that for $i \leq r$, $T^*T(v_i) = \sigma_i T^*(u_i) = \sigma_i^2 v_i$. For $i > r$, $T(v_i) = 0$, so $T^*T(v_i) = 0$. Thus v_i is an eigenvector of T^*T with eigenvalue σ_i^2 for $i \leq r$ and 0 for $i > r$. Since the eigenvalues of T^*T are uniquely determined, so are the scalars σ_i . $\square$

Definition 27.1. The unique scalars $\sigma_1, \dots, \sigma_r$ in [Theorem 27.1](#) are called the singular values of the linear transform T .

Remark 13. If $r < k = \min\{m, n\}$, then the term “singular value” is extended to $\sigma_{r+1} = \dots = \sigma_k = 0$.

Example 27.2. Let $\mathbb{P}_2(\mathbb{R})$ and $\mathbb{P}_1(\mathbb{R})$ be the polynomial spaces with inner products defined by

$$\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx.$$

Let $T: \mathbb{P}_2(\mathbb{R}) \rightarrow \mathbb{P}_1(\mathbb{R})$ and $T(f(x)) = f'(x)$. Find orthonormal bases $\beta = \{v_1, v_2, v_3\}$ for $\mathbb{P}_2(\mathbb{R})$ and $\gamma = \{u_1, u_2\}$ for $\mathbb{P}_1(\mathbb{R})$ such that $T(v_i) = \sigma_i u_i$ for $i = 1, 2$ and $T(v_3) = 0$ where $\sigma_1 \geq \sigma_2 > 0$ are the non-zero singular values of T .

Solution. Let $\alpha' = \{1, x, x^2\}$ and $\gamma' = \{1, x\}$ be the standard bases for $\mathbb{P}_2(\mathbb{R})$ and $\mathbb{P}_1(\mathbb{R})$ respectively. An orthonormal basis for $\mathbb{P}_2(\mathbb{R})$ is given by the normalised Legendre polynomials

$$p_0(x) = \frac{1}{\sqrt{2}}, \quad p_1(x) = \sqrt{\frac{3}{2}}x, \quad p_2(x) = \sqrt{\frac{5}{8}}(3x^2 - 1).$$

An orthonormal basis for $\mathbb{P}_1(\mathbb{R})$ is $\{p_0(x), p_1(x)\}$. Let $\alpha = \{p_0, p_1, p_2\}$ and $\gamma' = \{p_0, p_1\}$. The operator is $T(f(x)) = f'(x)$. We have

$$T(p_0(x)) = 0$$

$$T(p_1(x)) = \sqrt{\frac{3}{2}} = \sqrt{3}p_0(x)$$

$$T(p_2(x)) = \sqrt{\frac{5}{8}}(6x) = \sqrt{\frac{5}{8}} \cdot 6 \cdot \sqrt{\frac{2}{3}}p_1(x) = \sqrt{\frac{5 \cdot 36 \cdot 2}{8 \cdot 3}}p_1(x) = \sqrt{15}p_1(x).$$

The matrix of T with respect to these bases is

$$A = \begin{bmatrix} 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{15} \end{bmatrix}.$$

Then the matrix of T^*T is

$$A^*A = \begin{bmatrix} 0 & 0 \\ \sqrt{3} & 0 \\ 0 & \sqrt{15} \end{bmatrix} \begin{bmatrix} 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{15} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 15 \end{bmatrix}.$$

The eigenvalues of T^*T are $\lambda_1 = 15, \lambda_2 = 3, \lambda_3 = 0$. The singular values are $\sigma_1 = \sqrt{15}, \sigma_2 = \sqrt{3}$. The eigenvectors of A^*A are e_3, e_2, e_1 . Thus the basis for $\mathbb{P}_2(\mathbb{R})$ is $\{v_1, v_2, v_3\}$ where $v_1 = p_2, v_2 = p_1, v_3 = p_0$.

$$v_1(x) = \sqrt{\frac{5}{8}}(3x^2 - 1), \quad v_2(x) = \sqrt{\frac{3}{2}}x, \quad v_3(x) = \frac{1}{\sqrt{2}}.$$

The basis for $\mathbb{P}_1(\mathbb{R})$ is $\{u_1, u_2\}$ where

$$u_1 = \frac{1}{\sigma_1}T(v_1) = \frac{1}{\sqrt{15}}T(p_2) = p_1(x) = \sqrt{\frac{3}{2}}x$$

$$u_2 = \frac{1}{\sigma_2}T(v_2) = \frac{1}{\sqrt{3}}T(p_1) = p_0(x) = \frac{1}{\sqrt{2}}.$$

We have $T(v_1) = \sqrt{15}u_1$, $T(v_2) = \sqrt{3}u_2$, and $T(v_3) = 0$. □

§27.2 Applications of the Singular Value Decomposition

The Moore-Penrose Pseudoinverse For a linear operator $T: V \rightarrow W$ which is not invertible, we can define a pseudoinverse $T^+: W \rightarrow V$. The SVD provides a clear definition.

Definition 27.3 (Moore-Penrose Pseudoinverse). Let $T: V \rightarrow W$ be a linear operator with singular value expansion $T(v) = \sum_{i=1}^r \sigma_i \langle v, v_i \rangle u_i$. The **pseudoinverse** of T is the linear operator $T^+: W \rightarrow V$ defined by

$$T^+(w) = \sum_{i=1}^r \frac{1}{\sigma_i} \langle w, u_i \rangle v_i.$$

If $A = U\Sigma V^*$ is the matrix of T , the matrix of T^+ is $A^+ = V\Sigma^+U^*$, where Σ^+ is formed by taking the transpose of Σ and reciprocating the non-zero singular values.

The pseudoinverse is the unique operator satisfying the four Moore-Penrose conditions:

1. $TT^+T = T$
2. $T^+TT^+ = T^+$
3. $(TT^+)^* = TT^+$ (TT^+ is self-adjoint)
4. $(T^+T)^* = T^+T$ (T^+T is self-adjoint)

Note that T^+T is the orthogonal projector onto $\text{range}(T^*)$, and TT^+ is the orthogonal projector onto $\text{range}(T)$.

Least Squares Problem The pseudoinverse provides a complete solution to the linear least squares problem. Given $T: V \rightarrow W$ and a vector $b \in W$, we want to find $x \in V$ that minimizes $\|T(x) - b\|$.

Theorem 27.2. Consider the least squares problem $\min_{x \in V} \|T(x) - b\|$.

1. A solution always exists. The set of all solutions is the affine subspace $T^+(b) + \ker(T)$.
2. The vector $x^+ = T^+(b)$ is the unique solution of minimum norm.

Proof. The vector $T(x)$ is a vector in $\text{range}(T)$. The vector in $\text{range}(T)$ closest to b is the orthogonal projection of b onto $\text{range}(T)$, which we call b_T . So we need to solve $T(x) = b_T$. The operator TT^+ is the orthogonal projector onto $\text{range}(T)$, so $b_T = TT^+(b)$. The equation becomes $T(x) = TT^+(b)$. One particular solution is $x = T^+(b)$. The general solution is this particular solution plus any vector in the kernel of T , giving $T^+(b) + \ker(T)$.

For the minimum norm property, recall that $x^+ = T^+(b) = \sum_{i=1}^r \frac{\langle b, u_i \rangle}{\sigma_i} v_i$. This shows x^+ is in $\text{range}(T^*)$. Any other solution is $x = x^+ + z$ where $z \in \ker(T)$. Since $\text{range}(T^*) \perp \ker(T)$, we have $\|x\|^2 = \|x^+\|^2 + \|z\|^2 \geq \|x^+\|^2$. The norm is minimized when $z = 0$. $\square$

Principal Component Analysis (PCA) PCA is a technique to reduce the dimensionality of data while preserving as much variance as possible. Given a set of data vectors $\{x_1, \dots, x_N\}$ in a high-dimensional space $\mathbb{R}^d$, we can form a data matrix $X \in \mathbb{R}^{N \times d}$ where each row is a data point. Assume the data is centered (the mean of each column is zero).

The goal is to find a low-dimensional representation of the data. This is achieved by finding a new orthogonal basis for the data space, called principal components. The first principal component is the direction of maximum variance.

The sample covariance matrix is $C = \frac{1}{N-1}X^\top X$. The variance of the data projected onto a direction w (a unit vector) is $w^\top C w$. To maximize this, w must be the eigenvector of C with the largest eigenvalue. The principal components are the eigenvectors of the covariance matrix, ordered by their corresponding eigenvalues.

The SVD of the data matrix $X = U\Sigma V^\top$ reveals the principal components directly.

$$C = \frac{1}{N-1}X^\top X = \frac{1}{N-1}(V\Sigma^\top U^\top)(U\Sigma V^\top) = \frac{1}{N-1}V(\Sigma^\top \Sigma)V^\top.$$

This is the eigendecomposition of C .

- The columns of V (the right singular vectors of X) are the principal components.
- The eigenvalues of C are $\lambda_i = \frac{\sigma_i^2}{N-1}$, where σ_i are the singular values of X .

Using SVD to find the principal components is numerically more stable than forming the covariance matrix and computing its eigendecomposition. The matrix $U\Sigma$ contains the coordinates of the data in the new principal component basis, often called the scores.

References

- [Axl15] Sheldon Axler. *Linear algebra done right*. Springer, 2015. [3](#)
- [FIS14] Stephen H Friedberg, Arnold J Insel, and Lawrence E Spence. *Linear algebra*, vol. 4, 2014. [3](#)
- [Hal13] Paul R Halmos. *Finite-dimensional vector spaces*. Princeton University Press, 2013. [3](#)
- [Lan87] Serge Lang. *Linear algebra*. Springer Science & Business Media, 1987. [3](#)
- [Str06] Gilbert Strang. *Linear algebra and its applications*. Thomson Learning, 2006. [3](#)