MA2001 Linear Algebra I

Ryan Joo

AY25/26 Sem 1

Contents

1 Linear Systems and Gaussian Elimination

1.1 Linear Systems and Their Solutions

Definition 1.1. A **linear equation** in n variables $x_1, \dots, x_n$ is an equation of the form

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

Definition 1.2. A **linear system** of m linear equations in n variables $x_1, \dots, x_n$ is

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1}$$

Definition 1.3. Given $s_1, \dots, s_n$, we say $x_1 = s_1, \dots, x_n = s_n$ is a **solution** to the system if $x_1 = s_1, \dots, x_n = s_n$ is a solution to every equation in the system.

The set of all solutions to the system is the **solution set**.
An expression that gives the entire solution set is a **general solution**.

Definition 1.4. A linear system is **consistent** if it has at least one solution, and **inconsistent** if it has none.

Every linear system has either

- no solution (inconsistent)
- unique solution (consistent)
- infinitely many solutions (consistent)

1.2 Elementary Row Operations

Definition 1.5. The **augmented matrix** of the linear system (1) is

$$\left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \vdots & \vdots & \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right).$$

Definition 1.6. **Elementary row operations** (EROs):

1. Multiply a row by non-zero constant (kR_i)
2. Swap two rows ($R_i \leftrightarrow R_j$)
3. Add a multiple of one row to another row ($R_j + cR_i$)

Definition 1.7. Two augmented matrices are **row equivalent** if one can be obtained from the other by a series of EROs.

Theorem 1.8. *If augmented matrices of two linear systems are row equivalent, then the two systems have the same solution sets.*

Remark. The converse is not true. The two linear systems may consist of different number of equations. Then the augmented matrices have different number of rows and hence not row equivalent.

1.3 Row-Echelon Form

Definition 1.9. An augmented matrix is in **row-echelon form** (REF) if

(i) Zero rows are at the bottom of the matrix.

(ii) For any two successive non-zero rows, the leading entry of the lower row occurs further to the right than that of the higher row.

Example.

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

In an REF, the leading entry of a non-zero row is a **pivot points**.

A column is a **pivot column** if it contains a pivot point, and **non-pivot column** if not.

Definition 1.10. An augmented matrix is in **reduced row-echelon form** (RREF) if it is in REF, and

(iii) The leading entry of every non-zero row is 1.

(iv) In each pivot column, all non-pivot entries are zero.

Example.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right) \qquad \left(\begin{array}{cccc|c} 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

If the augmented matrix of a linear system is in REF or RREF, then we can get the solution(s) to the system easily.

1.4 Gaussian Elimination

Algorithm 1.11 (Gaussian elimination)**.** *Convert augmented matrix to REF*

1. *Find the leftmost non-zero column.*

2. *Check the top entry of the column. If it is 0, make it non-zero by swapping rows.*

3. *To rows underneath, add a multiple of the top row to make the rest of the column 0.*

4. *Cover the top row and repeat until done.*

Algorithm 1.12 (Gauss–Jordan elimination)**.** *Convert REF to RREF*

5. *Multiply rows by constants to make all leading entries 1.*

6. *Starting from the last non-zero row and working upwards, add multiples of it to rows above to make the rest of the pivot column 0.*

Consistency of a linear system:

- **No solution**: rightmost column is pivot column, i.e., leading entry occurs at the last column
- **Unique solution**: every column on the left is a pivot column
- **Infinitely many solutions**: at least one column on the left is not a pivot column

  Number of parameters = number of non-pivot columns on the left

Remark. Use a **branch diagram** to organise cases (for the values of unknowns) systematically.

1.5 Homogeneous Linear Systems

Definition 1.13. A linear system (1) is **homogeneous** if $b_1 = \cdots = b_n = 0$:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

A linear system is called **non-homogeneous** if it is not homogeneous.

The **trivial solution** is $x_1 = 0, \ldots, x_n = 0$; it is *always* a solution to any homogeneous system. All other solutions are **non-trivial solutions**.

Therefore a homogeneous linear system is always consistent. It can have

- only the trivial solution

$$\text{RREF} = \left(\begin{array}{ccc|c} 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{array}\right)$$

- infinitely many solutions in addition to the trivial solution.

Theorem 1.14. *A homogeneous linear system with more unknowns than equations has infinitely many solutions.*

2 Matrices

2.1 Introduction to Matrices

An $m \times n$ **matrix** is a rectangular array of numbers

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

or simply $A = (a_{ij})_{m\times n}$, where a_{ij} is the **(i, j)-entry** of A.

Notation. If the size of the matrix is already known, we simply write $A = (a_{ij})$.

- **Row matrix**: matrix with only one row

  Column matrix: matrix with only one column

- **Square matrix**: matrix with the same number of rows and columns.

  In particular, an $n \times n$ square matrix is called a square matrix of **order** n.

- The **diagonal** of square matrix $A = (a_{ij})_{n\times n}$ is the sequence of entries $a_{11}, \ldots, a_{nn}$.

  $$a_{ij} \text{ is a } \begin{cases} \textbf{diagonal entry} & (i = j) \\ \textbf{non-diagonal entry} & (i \neq j) \end{cases}$$

  Diagonal matrix: square matrix where all non-diagonal entries are 0

- **Scalar matrix**: diagonal matrix where all equal diagonal entries

- **Identity matrix** I: scalar matrix with all diagonal entries 1

- **Zero matrix** $\mathbf{0}$: matrix with all entries 0

- **Symmetric matrix**: square matrix with $a_{ij} = a_{ji}$ for all i, j (symmetric wrt diagonal)

- **Upper-triangular matrix**: square matrix where if $a_{ij} = 0$ whenever $i > j$.

  Lower-triangular matrix: square matrix where $a_{ij} = 0$ whenever $i < j$.

  Triangular matrix: upper/lower-triangular

2.2 Matrix Operations

Let $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{m\times n}$.

- Equality: $A = B$ iff $a_{ij} = b_{ij}$
- Addition: $(A + B)_{ij} = a_{ij} + b_{ij}$
- Scalar multiplication: $(cA)_{ij} = ca_{ij}$

The following properties follow:

- $A + B = B + A$
- $A + (B + C) = (A + B) + C$

- $c(\boldsymbol{A}+\boldsymbol{B}) = c\boldsymbol{A}+c\boldsymbol{B}$
- $(c+d)\boldsymbol{A} = c\boldsymbol{A}+d\boldsymbol{A}$
- $c(d\boldsymbol{A}) = (cd)\boldsymbol{A} = d(c\boldsymbol{A})$
- $\boldsymbol{A}+\boldsymbol{0} = \boldsymbol{0}+\boldsymbol{A} = \boldsymbol{A}$
- $\boldsymbol{A}-\boldsymbol{A} = \boldsymbol{0}$
- $0\boldsymbol{A} = \boldsymbol{0}$

Definition 2.1. Let $\boldsymbol{A} = (a_{ij})_{m\times p}$, $\boldsymbol{B} = (b_{ij})_{p\times n}$. The **product** $\boldsymbol{AB}$ is the $m\times n$ matrix whose (i,j)-entry is

$$\sum_{k=1}^{p} a_{ik}b_{kj}.$$

Remark. We can only multiply two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ (in the manner $\boldsymbol{AB}$) when # columns of $\boldsymbol{A}$ = # rows of $\boldsymbol{B}$.

The following properties follow:

- $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$
- $\boldsymbol{A}(\boldsymbol{B}_1+\boldsymbol{B}_2) = \boldsymbol{AB}_1+\boldsymbol{AB}_2$
  $(\boldsymbol{A}_1+\boldsymbol{A}_2)\boldsymbol{B} = \boldsymbol{A}_1\boldsymbol{B}+\boldsymbol{A}_2\boldsymbol{B}$
- $c(\boldsymbol{AB}) = (c\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(c\boldsymbol{B})$
- $\boldsymbol{A0} = \boldsymbol{0} = \boldsymbol{0A}$
- $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$

The **powers** of a square matrix $\boldsymbol{A}$ are

$$\boldsymbol{A}^n = \begin{cases} \boldsymbol{I} & (n=0) \\ \underbrace{\boldsymbol{A}\cdots\boldsymbol{A}}_{n \text{ times}} & (n\geq 1) \end{cases}$$

Block multiplication: Let $\boldsymbol{A} = (a_{ij})_{m\times n}$ and $a_i = \begin{pmatrix} a_{i1} & \cdots & a_{in} \end{pmatrix}$ denote the i-th row. Then $\boldsymbol{A} = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$.

Similarly, if $b_j = \begin{pmatrix} b_{1j} \\ \vdots \\ b_{pj} \end{pmatrix}$ is the j-th column, then $\boldsymbol{A} = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix}$.

If $\boldsymbol{A} = (a_{ij})_{m\times p}$ with i-th row a_i and $\boldsymbol{B} = (a_{ij})_{p\times n}$ with j-th column b_j, then

$$\boldsymbol{AB} = \begin{pmatrix} a_1b_1 & \cdots & a_1b_n \\ \vdots & & \vdots \\ a_mb_1 & \cdots & a_mb_n \end{pmatrix} = \begin{pmatrix} a_1\boldsymbol{B} \\ \vdots \\ a_m\boldsymbol{B} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}b_1 & \cdots & \boldsymbol{A}b_n \end{pmatrix}.$$

The linear system (1) can be written as $\boldsymbol{Ax} = \boldsymbol{b}$:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

$\boldsymbol{A}$ is the **coefficient matrix**, $\boldsymbol{x}$ is the **variable matrix**, $\boldsymbol{b}$ is the **constant matrix** of the linear system. A solution $x_1 = u_1, \ldots, x_n = u_n$ to the linear system can be represented by an $n \times 1$ column matrix

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

i.e. $\boldsymbol{u}$ is a **solution** to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ if $\boldsymbol{Au} = \boldsymbol{b}$.

Theorem 2.2. *Linear system with > 1 solution $\Rightarrow$ infinite solutions*

Proof. If $\boldsymbol{Ax} = \boldsymbol{b}$ has two distinct solutions $\boldsymbol{u_1}, \boldsymbol{u_2}$, then $\boldsymbol{u_2} + t(\boldsymbol{u_1} - \boldsymbol{u_2})$ is a solution $\forall t \in \mathbb{R}$. $\square$

Definition 2.3. The **transpose** of $\boldsymbol{A} = (a_{ij})_{m \times n}$ is the $n \times m$ matrix $\boldsymbol{A}^\top$ whose (i, j)-entry is a_{ji}.

Properties:

- $(\boldsymbol{A}^\top)^\top = \boldsymbol{A}$
- $(\boldsymbol{A} + \boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$
- $(c\boldsymbol{A})^\top = c\boldsymbol{A}^\top$
- $(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$
- $\boldsymbol{A}$ is symmetric $\iff \boldsymbol{A} = \boldsymbol{A}^\top$

2.3 Inverses of Square Matrices

Definition 2.4. A square matrix $\boldsymbol{A}$ of order n is **invertible** if there exists a square matrix $\boldsymbol{B}$ of order n such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$. Such a matrix $\boldsymbol{B}$ is called an **inverse** of $\boldsymbol{A}$.
A square matrix is called **singular** if it has no inverse.

Remark. To show a matrix is singular, use proof by contradiction.

The inverse of an invertible matrix $\boldsymbol{A}$ is unique, denoted by $\boldsymbol{A}^{-1}$.
Properties:

- $(c\boldsymbol{A})^{-1} = \frac{1}{c}\boldsymbol{A}^{-1}$
- $(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$
- $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$
- $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

- $A^{-n} = (A^{-1})^n = (A^n)^{-1}$

For positive integers n, **negative powers** of invertible matrix A are $A^{-n} := (A^{-1})^n = \underbrace{A^{-1} \cdots A^{-1}}_{n \text{ times}}$.

2.4 Elementary Matrices

Definition 2.5. A square matrix is an **elementary matrix** if it can be obtained from the identity matrix by performing a single ERO.

Theorem 2.6. *If E is the elementary matrix obtained by performing an ERO to I_m, then for any $m \times n$ matrix A, EA can be obtained by performing the same ERO to A.*

Theorem 2.7. *Every elementary matrix has an inverse that is also elementary.*

Hence A and B are row equivalent $\iff$ $E_k \cdots E_1 A = B$ for some elementary matrices $E_1, \ldots, E_k$.

Theorem 2.8 (Invertibility equivalences)**.** *Let A be a square matrix. Then the following are equivalent:*

(i) A is invertible.

(ii) $Ax = 0$ has only the trivial solution.

(iii) RREF of A is identity matrix.

(iv) A can be expressed as a product of elementary matrices.

Proof.
$\boxed{\text{(i)} \Rightarrow \text{(ii)}}$ Suppose A is invertible. If $Ax = 0$, then multiply A^{-1} on both sides to get $x = 0$.
$\boxed{\text{(ii)} \Rightarrow \text{(iii)}}$ Suppose $Ax = 0$ has only the trivial solution. Then RREF of A does not have zero rows.
$\boxed{\text{(iii)} \Rightarrow \text{(iv)}}$ There exist elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A = I$. Then $A = E_1^{-1} \cdots E_k^{-1}$.
$\boxed{\text{(iv)} \Rightarrow \text{(i)}}$ A product of invertible matrices is invertible. □

The inverse of a matrix can be found using Gaussian elimination:

Theorem 2.9. *If A is invertible, then*

$$\text{RREF}(A \mid I) = (I \mid A^{-1}).$$

Proof. Write $E_k \cdots E_1 A = I$ for some elementary matrices $E_1, \ldots, E_k$. Then $A^{-1} = E_k \cdots E_1$. □

Theorem 2.10 (Half-price theorem)**.** *Let A and B be square matrices of same order. If $AB = I$, then A and B are invertible and $A^{-1} = B$, $B^{-1} = A$.*

Proof. Consider $Bx = 0$. Multiply A on both sides to get $x = 0$, so the trivial solution is the only solution. Hence B invertible. □

Theorem 2.11. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two square matrices of same order. If $\boldsymbol{A}$ is singular, then $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are singular.*

Definition 2.12. **Elementary column operations** (ECOs) are EROs but on columns.

Theorem 2.13. *If $\boldsymbol{E}$ is obtained from $\boldsymbol{I}_n$ by a single elementary column operation, then $\boldsymbol{E}$ is an elementary matrix.*

Theorem 2.14. *If $\boldsymbol{E}$ is the elementary matrix obtained by performing an ECO to $\boldsymbol{I}_n$, then for any $m \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{AE}$ can be obtained by performing the same ECO to $\boldsymbol{A}$.*

Remark. We *post-multiply* E to $\boldsymbol{A}$, instead of pre-multiplying it.

2.5 Determinant

Definition 2.15. Let $\boldsymbol{M}_{ij}$ be the submatrix of $\boldsymbol{A}$ obtained by deleting the i-th row and j-th column of $\boldsymbol{A}$. The **(i, j)-cofactor** of $\boldsymbol{A}$ is

$$A_{ij} := (-1)^{i+j} \det\left(\boldsymbol{M}_{ij}\right).$$

Definition 2.16. Let $\boldsymbol{A} = (a_{ij})_{n\times n}$. The **determinant** of $\boldsymbol{A}$ is

$$\det(\boldsymbol{A}) := \sum_{k=1}^{n} a_{1k}A_{1k} = a_{11}A_{11} + \cdots + a_{1n}A_{1n}$$

if $n > 1$, and $\det(\boldsymbol{A}) := a_{11}$ if $n = 1$.

For 2×2 and 3×3 matrices,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Theorem 2.17 (Cofactor expansion). *Let $\boldsymbol{A} = (a_{ij})_{n\times n}$. Then for any $i, j \in \{1, \ldots, n\}$,*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} a_{ik}A_{ik} = a_{i1}A_{i1} + \cdots + a_{in}A_{in} \qquad [\textit{along i-th row}]$$

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} a_{kj}A_{kj} = a_{1j}A_{1j} + \cdots + a_{nj}A_{nj} \qquad [\textit{along j-th column}]$$

To make the computation easier, perform cofactor expansion along the row/column that has many 0s.

Theorem 2.18. *If $A = (a_{ij})_{n\times n}$ is triangular, then* $\det(A) = a_{11}\cdots a_{nn}$.

Proof. Induct on the order of A. □

Theorem 2.19. *Let A be a square matrix. Then* $\det(A) = \det\left(A^\top\right)$.

Proof. Induct on the order of A. □

Theorem 2.20. *The determinant of any square matrix with two identical rows/columns is zero.*

Proof. If two identical rows, reduce square matrix to row-echelon form with zero row. Then transpose to get two identical columns. □

Lemma 2.21. *If two square matrices of order n differ at the i-th row only, then their $(i,1),\dots,(i,n)$ cofactors are the same.*

Theorem 2.22 (Determinants under EROs)**.**

(i) $A \xrightarrow{cR_i} B \implies \det(B) = c\det(A)$

(ii) $A \xrightarrow{R_i \leftrightarrow R_j} B \implies \det(B) = -\det(A)$

(iii) $A \xrightarrow{R_i + cR_j} B \implies \det(B) = \det(A)$

Theorem 2.23. *For any elementary matrix E,* $\det(EA) = \det(E)\det(A)$.

Hence, to find $\det(A)$:

1. Perform Gaussian elimination on A, reduce it to REF (upper-triangular)
2. $\det(R)$ = product of diagonal entries
3. $E_k \cdots E_1 A = B \implies \det(E_k)\cdots\det(E_1)\det(A) = \det(R)$

Theorem 2.24 (Invertibility equivalences)**.**

(v) $\det(A) \neq 0$.

Proof. Let $B = E_k \cdots E_1 A$ be RREF of A. Then $\det(B) = \det(E_k)\cdots\det(E_1)\det(A)$.
$\boxed{\Rightarrow}$ If A is invertible, then B is an identity matrix, so $\det(B) \neq 0$. Thus $\det(A) \neq 0$.
$\boxed{\Leftarrow}$ We prove the contrapositive. If A is singular, then B contains at least one zero row, so $\det(B) = 0$ (by cofactor expansion along the zero row). But each $\det(E_i) \neq 0$, so $\det(A) = 0$. □

Theorem 2.25. *Let A and B be $n \times n$ matrices.*

(i) $\det(cA) = c^n \det(A)$

(ii) $\det(AB) = \det(A)\det(B)$

(iii) If $\boldsymbol{A}$ is invertible, then $\det\left(\boldsymbol{A}^{-1}\right) = \dfrac{1}{\det(\boldsymbol{A})}$.

Definition 2.26. Let $\boldsymbol{A}$ be an $n \times n$ matrix. The **adjoint** of $\boldsymbol{A}$ is the transpose of cofactor matrix:

$$\operatorname{adj}(\boldsymbol{A}) = (A_{ji})_{n\times n} = (A_{ij})^{\top}.$$

Theorem 2.27 (Method of adjoints)**.** *If $\boldsymbol{A}$ is invertible, then*

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}\operatorname{adj}(\boldsymbol{A}). \tag{2}$$

Proof. We need to show $\boldsymbol{A}[\operatorname{adj}(\boldsymbol{A})] = \det(\boldsymbol{A})\boldsymbol{I}$. Let $\boldsymbol{A}[\operatorname{adj}(\boldsymbol{A})] = (b_{ij})$. Then $b_{ij} = \sum_{k=1}^{n} a_{ik}A_{jk}$.

- Diagonal terms: $b_{ii} = \det(\boldsymbol{A})$ by cofactor expansion along i-th row.
- Non-diagonal terms: $b_{ij} = \det\left(\boldsymbol{A}'\right) = 0$, where $\boldsymbol{A}'$ is obtained from $\boldsymbol{A}$ replacing j-th row with i-th row.

□

Theorem 2.28 (Cramer's rule)**.** *Let $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ be a linear system, where $\boldsymbol{A}$ is an $n \times n$ matrix.*
Let $\boldsymbol{A}_i$ be the matrix obtained from $\boldsymbol{A}$ by replacing the i-th column of $\boldsymbol{A}$ by $\boldsymbol{b}$.
If $\boldsymbol{A}$ is invertible, then the linear system has a unique solution $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$:

$$\boldsymbol{x} = \frac{1}{\det(\boldsymbol{A})}\begin{pmatrix} \det(A_1) \\ \vdots \\ \det(A_n) \end{pmatrix}. \tag{3}$$

Proof. Let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$. By Method of adjoints,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \iff \boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} = \frac{1}{\det(\boldsymbol{A})}[\operatorname{adj}\boldsymbol{A}]\boldsymbol{b}.$$

Performing cofactor expansion along i-th column of $\boldsymbol{A}_i$, $\det(\boldsymbol{A}_i) = b_1A_{1i} + \cdots + b_nA_{ni}$. Thus

$$x_i = \frac{b_1A_{1i} + \cdots + b_nA_{ni}}{\det(\boldsymbol{A})} = \frac{\det(\boldsymbol{A}_i)}{\det(\boldsymbol{A})}.$$

□

3 Vector Spaces

3.1 Euclidean n-Spaces

Definition 3.1. An **n-vector** of real numbers is $\boldsymbol{v} = (v_1, \ldots, v_n)$.
v_i is the **i-th component** or **i-th coordinate** of $\boldsymbol{v}$.

Let $\boldsymbol{u} = (u_1, \ldots, u_n)$, $\boldsymbol{v} = (v_1, \ldots, v_n)$. Define

- $\boldsymbol{u} = \boldsymbol{v} \Leftrightarrow u_i = v_i \ \forall i$
- $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, \ldots, u_n + v_n)$
- $c\boldsymbol{u} = (cu_1, \ldots, cu_n)$

We can identify an n-vector $(u_1, \ldots, u_n)$ with a row matrix $\begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}$ or column matrix $\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$.

The **zero vector** is $\mathbf{0} = (0, \ldots, 0)$.

Definition 3.2. The **Euclidean n-space** $\mathbb{R}^n$ is the set of all n-vectors of real numbers.

Properties:

- $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
- $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$
- $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u} = \mathbf{0} + \boldsymbol{u}$
- $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$
- $c(d\boldsymbol{u}) = (cd)\boldsymbol{u}$
- $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$
  $(c + d)\boldsymbol{u} + c\boldsymbol{u} + d\boldsymbol{u}$
- $1\boldsymbol{u} = \boldsymbol{u}$

3.2 Linear Combinations and Linear Spans

Definition 3.3. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$. For any $c_1, \ldots, c_k \in \mathbb{R}$, the vector $\sum_{i=1}^{k} c_i \boldsymbol{v}_i$ is a **linear combination** of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$.

Definition 3.4. Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$. The **span** of S is the set of all linear combinations of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$:

$$\operatorname{span}(S) := \{c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k \mid c_1, \ldots, c_k \in \mathbb{R}\}.$$

Let $S = \{v_1, \dots, v_k\} \subseteq \mathbb{R}^n$. To check if $w \in \operatorname{span}(S)$, check if

$$\begin{aligned}
&\exists c_1, \dots, c_k \in \mathbb{R} \text{ s.t. } c_1 v_1 + \dots + c_k v_k = w \\
\iff& \begin{pmatrix} v_1 & \cdots & v_k \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} = w \text{ is consistent} \\
\iff& \left(\begin{array}{ccc|c} v_1 & \cdots & v_k & w \end{array} \right) \text{ is consistent}
\end{aligned}$$

- Inconsistent $\Rightarrow w \notin \operatorname{span}\{v_1, \dots, v_k\}$
- Unique solution $\Rightarrow$ only one way to express w as a linear combination of $v_1, \dots, v_k$
- Infinitely many solutions $\Rightarrow$ multiple ways to express w as a linear combination of $v_1, \dots, v_k$

More generally, to check if $w_1, \dots, w_m \in \operatorname{span}\{v_1, \dots, v_k\}$, check

$$\left(\begin{array}{ccc|ccc} v_1 & \cdots & v_k & w_1 & \cdots & w_m \end{array} \right) \text{ is consistent.}$$

To check if $\operatorname{span}\{v_1, \dots, v_k\} = \mathbb{R}^n$, it suffices to check $\mathbb{R}^n \subseteq \operatorname{span}\{v_1, \dots, v_k\}$, i.e., $x \in \operatorname{span}\{v_1, \dots, v_k\}$ for all $x \in \mathbb{R}^n$:

$$\begin{aligned}
&\left(\begin{array}{ccc|c} v_1 & \cdots & v_k & x \end{array} \right) \text{ is consistent for } \textbf{all } x \in \mathbb{R}^n \\
\iff& \text{REF}(A) \text{ has no zero rows} \\
\iff& A \text{ is invertible}
\end{aligned}$$

Theorem 3.5. *Let $S = \{v_1, \dots, v_k\} \subseteq \mathbb{R}^n$. If $k < n$, then S cannot span $\mathbb{R}^n$.*

Proof. If $k < n$, REF of $\begin{pmatrix} v_1 & \cdots & v_k \end{pmatrix}$ must have at least one zero row. □

Theorem 3.6. *Let $S = \{u_1, \dots, u_k\} \subseteq \mathbb{R}^n$.*

(i) $\mathbf{0} \in \operatorname{span}(S)$.

(ii) $c_1 v_1 + \dots + c_r v_r \in \operatorname{span}(S)$. *(closed under addition and scalar multiplication)*

Theorem 3.7. *Given $S_1 = \{u_1, \dots, u_k\}$, $S_2 = \{v_1, \dots, v_m\} \subseteq \mathbb{R}^n$,*

$$\operatorname{span}(S_1) \subseteq \operatorname{span}(S_2) \iff \textit{every } u_i \textit{ is a linear combination of } v_1, \dots, v_m.$$

Proof.
$\boxed{\Rightarrow}$ $S_1 \subseteq \operatorname{span}(S_1) \subseteq \operatorname{span}(S_2)$ implies every $u_i \in \operatorname{span}\{v_1, \dots, v_m\}$.
$\boxed{\Leftarrow}$ Any linear combination of $u_1, \dots, u_k$ is a linear combination of $v_1, \dots, v_m$. □

Remark. To show $\operatorname{span}(S_1) = \operatorname{span}(S_2)$, need to show $\operatorname{span}(S_1) \subseteq \operatorname{span}(S_2)$ and $\operatorname{span}(S_2) \subseteq \operatorname{span}(S_1)$; use Gaussian elimination above.

Theorem 3.8 (Redundancy). *If v_k is a linear combination of $v_1, \ldots, v_{k-1}$, then*

$$\operatorname{span}\{v_1, \ldots, v_{k-1}\} = \operatorname{span}\{v_1, \ldots, v_{k-1}, v_k\}.$$

Proof. Let $S_1 = \{v_1, \ldots, v_{k-1}\}$ and $S_2 = \{v_1, \ldots, v_{k-1}, v_k\}$.
$\boxed{\subseteq}$ By the previous theorem, $\operatorname{span}(S_1) \subseteq \operatorname{span}(S_2)$.
$\boxed{\supseteq}$ Since $v_1, \ldots, v_{k-1}, v_k$ are each linear combinations of $v_1, \ldots, v_{k-1}$, $\operatorname{span}(S_2) \subseteq \operatorname{span}(S_1)$. □

3.3 Subspaces

Definition 3.9. $V \subseteq \mathbb{R}^n$ is a **subspace** of $\mathbb{R}^n$ if $\exists v_1, \ldots, v_k \in \mathbb{R}^n$ s.t. $V = \operatorname{span}\{v_1, \ldots, v_k\}$.
V is the **subspace spanned** by $S = \{v_1, \ldots, v_k\}$.
$\{v_1, \ldots, v_k\}$ **spans** the subspace V.

$\{\mathbf{0}\} = \operatorname{span}\{\mathbf{0}\}$ is the **zero space**.

Lemma 3.10. *Since a subspace V is of the form $\operatorname{span}(S)$, we have*

- $\mathbf{0} \in V$
- $c\boldsymbol{v} \in V$ *for all* $\boldsymbol{v} \in V, c \in \mathbb{R}$
- $\boldsymbol{u} + \boldsymbol{v} \in V$ *for all* $\boldsymbol{u}, \boldsymbol{v} \in V$

By contrapositive, if any of the above fails, then V is not a subspace (of $\mathbb{R}^n$).

Theorem 3.11 (Subspace criterion). *Let $V \subseteq \mathbb{R}^n$ be non-empty. Then V is a subspace of $\mathbb{R}^n$ if and only if*

$$c\boldsymbol{u} + d\boldsymbol{v} \in V \qquad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in V,\ c, d \in \mathbb{R}.$$

Theorem 3.12. *The solution set of a homogeneous linear system of n variables is a subspace of $\mathbb{R}^n$.*

The solution set of a homogeneous linear system is called the **solution space** of the system.

Proof. If it only has the trivial solution, then $\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$.
Otherwise the homogeneous system has infinitely many solutions. Express $x_1, \ldots, x_n$ in terms of arbitrary parameters $t_1, \ldots, t_k$. □

3.4 Linear Independence

Definition 3.13. Let $S = \{v_1, \ldots, v_k\} \subseteq \mathbb{R}^n$. Consider the equation

$$c_1 v_1 + \cdots + c_k v_k = \mathbf{0}. \tag{4}$$

- If (4) only has trivial solution, we say that S is a **linearly independent** set; $v_1, \ldots, v_k$ are **linearly independent**:

$$c_1 v_1 + \cdots + c_k v_k = \mathbf{0} \implies c_1 = \cdots = c_k = 0.$$

- If (4) has non-trivial solutions, we say that S is a **linearly dependent** set; $v_1, \ldots, v_k$ are **linearly dependent**:

$$\exists c_1, \ldots, c_k \in \mathbb{R}, \text{ not all zero, such that } c_1 v_1 + \cdots + c_k v_k = \mathbf{0}.$$

To check if $\{v_1, \ldots, v_k\}$ is linearly independent, $c_1 v_1 + \cdots + c_k v_k = \mathbf{0}$ only has trivial solution, so solving

$$\left(\begin{array}{ccc|c} & & & 0 \\ v_1 & \cdots & v_k & \vdots \\ & & & 0 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \end{array}\right)$$

has unique solution (only the trivial solution).

Lemma 3.14. *If $\mathbf{0} \in S$, then S is linearly dependent.*

Proof. Suppose $c_0 \mathbf{0} + c_1 v_1 + \cdots + c_k v_k = \mathbf{0}$. A non-trivial solution is $c_0 = 1$, $c_1 = \cdots = c_k = 0$. ☐

Lemma 3.15. *Let S_1, S_2 be finite subsets of $\mathbb{R}^n$ such that $S_1 \subseteq S_2$*

- *S_1 linearly dependent $\Rightarrow$ S_2 linearly dependent*
- *S_2 linearly independent $\Rightarrow$ S_1 linearly independent*

Theorem 3.16. *Let $S = \{v_1, \ldots, v_k\} \subseteq \mathbb{R}^n$, $k \geq 2$. Then*

- *S is linearly dependent $\iff$ $\exists v_i$ such that it is a linear combination of the other vectors.*
- *S is linearly independent $\iff$ no vector in S is the linear combination of the other vectors.*

Proof.
$\boxed{\Rightarrow}$ Suppose S is linearly dependent. Then not all of $c_1, \ldots, c_k$ are 0. If $c_i \neq 0$, then

$$v_i = -\frac{c_1}{c_i} v_1 - \cdots - \frac{c_{i-1}}{c_i} v_{i-1} - \frac{c_{i+1}}{c_i} v_{i+1} - \cdots - \frac{c_k}{c_i} v_k.$$

$\boxed{\Leftarrow}$ Suppose there exists v_i such that

$$v_i = c_1 v_1 + \cdots + c_{i-1} v_{i-1} + c_{i+1} v_{i+1} + \cdots + c_k v_k.$$

Then bring v_i to RHS, so LHS becomes $\mathbf{0}$. Note that the coefficients of $v_1, \ldots, v_k$ are not all zero. ☐

Remark. If a set of vectors is linearly dependent, then one of the vectors is "redundant"; else, there is no "redundant" vector in the set.

Theorem 3.17. *Let $S = \{v_1, \ldots, v_k\} \subseteq \mathbb{R}^n$. If $k > n$, then S is linearly dependent.*

Proof. Suppose there exist $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$c_1 v_1 + \cdots + c_k v_k = \mathbf{0}.$$

Write the linear system as an augmented matrix

$$\left(\begin{array}{ccc|c} & & & 0 \\ v_1 & \cdots & v_k & \vdots \\ & & & 0 \end{array}\right)$$

which has k variables and n equations. Since $k > n$, the linear system has non-trivial solutions. □

The next result states that we can extend a linearly independent set.

Theorem 3.18. *Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$ be linearly independent.*
If $\boldsymbol{v}_{k+1} \notin \operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{v}_{k+1}\}$ is linearly independent.

Proof. Suppose there exist $c_1, \ldots, c_k, c_{k+1} \in \mathbb{R}$ such that

$$c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k + c_{k+1}\boldsymbol{v}_{k+1} = \mathbf{0}.$$

Suppose, for a contradiction, that $c_{k+1} \neq 0$. Then

$$v_{k+1} = -\frac{c_1}{c_{k+1}}v_1 - \cdots - \frac{c_k}{c_{k+1}}v_k$$

implies $\boldsymbol{v}_{k+1} \in \operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, a contradiction. Hence $c_{k+1} = 0$. □

3.5 Bases

Definition 3.19. V is a **vector space** if $V = \mathbb{R}^n$ or V is a subspace of $\mathbb{R}^n$ for some n.
If W and V are vector spaces and $W \subseteq V$, then W is a **subspace** of V.

Unless mentioned otherwise, V denotes a vector space.

Definition 3.20. S is a **basis** of V if S is linearly independent and spans V.

For convenience, $\varnothing$ is the basis for $\{\mathbf{0}\}$.
A basis for a vector space is the **smallest** spanning set, the **largest** linearly independent set.
Any basis for $\mathbb{R}^n$ has exactly n vectors (if $< n$ then not spanning set, if $> n$ then linearly dependent).

Theorem 3.21. *Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq V$. If S is a basis for V, then every $\boldsymbol{v} \in V$ is a unique linear combination of $v_1, \ldots, v_k$.*

Proof. Let $v \in V$. Since S spans V, v is a linear combination of $v_1, \ldots, v_k$. Suppose

$$v = c_1v_1 + \cdots + c_kv_k = d_1v_1 + \cdots + d_kv_k.$$

Then $(c_1 - d_1)v_1 + \cdots + (c_k - d_k)v_k = \mathbf{0}$. Since S is linearly dependent, $c_1 = d_1, \ldots, c_k = d_k$. □

Definition 3.22. Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq V$ be a basis for V. Let $\boldsymbol{v} \in V$, then $\boldsymbol{v} = \sum_{i=1}^{k} c_i\boldsymbol{v}_i$ uniquely. The coefficients $c_1, \ldots, c_k$ are *the* **coordinates** of $\boldsymbol{v}$ relative to the basis S.

The **coordinate vector** of $\boldsymbol{v}$ relative to the basis S is

$$(\boldsymbol{v})_S = (c_1, \ldots, c_k) \in \mathbb{R}^k.$$

Example. Let $E = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\} \subseteq \mathbb{R}^n$, where $\boldsymbol{e}_1 = (1, 0, \ldots, 0), \ldots, \boldsymbol{e}_n = (0, 0, \ldots, 1)$. Then E is a basis, called the **standard basis** for $\mathbb{R}^n$.

* For $m \times n$ matrix, $\boldsymbol{A}\boldsymbol{e}_j$ is j-th column of $\boldsymbol{A}$.

Theorem 3.23. *Let S be a basis of V.*

(i) $\boldsymbol{u} = \boldsymbol{v} \iff (\boldsymbol{u})_S = (\boldsymbol{v})_S$

(ii) $(c_1\boldsymbol{v}_1 + \cdots + c_r\boldsymbol{v}_r)_S = c_1(\boldsymbol{v}_1)_S + \cdots + c_r(\boldsymbol{v}_r)_S$

Theorem 3.24. *Let S be a basis of V, where $|S| = k$.*

(i) $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ *are linearly independent in* $V \iff (\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_r)_S$ *are linearly independent in* $\mathbb{R}^k$

(ii) $\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} = V \iff \operatorname{span}\{(\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_r)_S\} = \mathbb{R}^k$

Proof.

(i)

$$\begin{aligned} & c_1\boldsymbol{v}_1 + \cdots + c_r\boldsymbol{v}_r = \boldsymbol{0} \\ \iff & (c_1\boldsymbol{v}_1 + \cdots + c_r\boldsymbol{v}_r)_S = (\boldsymbol{0})_S \\ \iff & c_1(\boldsymbol{v}_1)_S + \cdots + c_r(\boldsymbol{v}_r)_S = (\boldsymbol{0})_S \end{aligned}$$

(ii) Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ be a basis of V.

$\boxed{\Rightarrow}$ Suppose $\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\} = V$. It suffices to show $\mathbb{R}^k \subseteq \operatorname{span}\{(\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_r)_S\}$.

Let $(a_1, \ldots, a_k) \in \mathbb{R}^k$. Let $\boldsymbol{w} = a_1\boldsymbol{u}_1 + \cdots + a_k\boldsymbol{u}_k \in V$. Since $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ span V, write $\boldsymbol{w} = c_1\boldsymbol{v}_1 + \cdots + c_r\boldsymbol{v}_r$. Then

$$\begin{aligned} (a_1, \ldots, a_k) = (\boldsymbol{w})_S &= (c_1\boldsymbol{v}_1 + \cdots + c_r\boldsymbol{v}_r)_S \\ &= c_1(\boldsymbol{v}_1)_S + \cdots + c_r(\boldsymbol{v}_r)_S. \end{aligned}$$

Thus $(a_1, \ldots, a_k) \in \operatorname{span}\{(\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_r)_S\}$.

$\boxed{\Leftarrow}$ Suppose $\operatorname{span}\{(\boldsymbol{v}_1)_S, \ldots, (\boldsymbol{v}_r)_S\} = \mathbb{R}^k$. It suffices to show $V \subseteq \operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$.

Let $\boldsymbol{w} \in V$. Since $(w)_S \in \mathbb{R}^k$,

$$(w)_S = c_1(v_1)_S + \cdots + c_r(v_r)_S = (c_1v_1 + \cdots + c_rv_r)_S$$

so $w = c_1v_1 + \cdots + c_rv_r \in \operatorname{span}\{v_1, \ldots, v_r\}$.

□

3.6 Dimensions

Theorem 3.25. *Let V have a basis of k vectors.*

(i) Any subset of V with $> k$ vectors is linearly dependent;

(ii) Any subset of V with $< k$ vectors cannot span V.

Proof. Consider a basis S, express all vectors in terms of coordinates with respect to the basis S. $\square$

As a consequence, all bases of a vector space have the same length.

Definition 3.26. Let S be a basis for V. The **dimension** of V is $\dim(V) := |S|$.
Define $\dim\{\mathbf{0}\} := 0$.

Given a homogeneous linear system $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$, to find dimension of solution space:

1. Use Gauss-Jordan elimination to convert augmented matrix to RREF
2. Set arbitrary variables, equate the other variables accordingly
3. Write general solution in the form $\sum_{i=1}^{k} t_i \boldsymbol{u}_i$
4. $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is a basis for the solution space
5. No. of non-pivot col in REF($\boldsymbol{A}$) = no. of arbitrary parameters in sol = $\dim(V)$

Theorem 3.27. *Let $S \subseteq V$ and $\dim(V) = k$. Then the following are equivalent:*

(i) S is a basis for V.

(ii) S is linearly independent and $|S| = k$.

(iii) S spans V and $|S| = k$.

Proof.
$\boxed{\text{(i)} \Rightarrow \text{(ii)}}$ This follows from definition.
$\boxed{\text{(i)} \Rightarrow \text{(iii)}}$ This follow from definition.
$\boxed{\text{(ii)} \Rightarrow \text{(i)}}$ Suppose S is linearly independent and $|S| = k$. Suppose, for a contradiction, that S is not a basis for V. Then $\operatorname{span}(S) \neq V$; fix $\boldsymbol{v} \in V \setminus \operatorname{span}(S)$. Then $S \cup \{bmv\}$ is a linearly independent set of length $k+1$, a contradiction.
$\boxed{\text{(iii)} \Rightarrow \text{(i)}}$ Suppose S spans V and $|S| = k$. Suppose, for a contradiction, that S is not a basis for V. Then S is linearly dependent. Fix $\boldsymbol{v} \in S$ which is a linear combination of other vectors in S. Then $S \setminus \{\boldsymbol{v}\}$ is a spanning set of $k-1$ vectors, a contradiction. $\square$

$\therefore$ If $|S| = k$, to show S is a basis, it suffices to show either S is linearly independent *or* spans V.
The next result shows that if a subspace has dimension equal the whole space, then the subspace is the whole space.

Theorem 3.28. *Let U be a subspace of V. Then $U = V \iff \dim(U) = \dim(V)$.*

Proof.

$\boxed{\Rightarrow}$ Trivial.

$\boxed{\Leftarrow}$ Suppose $\dim(U) = \dim(V)$. Let S be a basis for U.

Since S is linearly independent and $|S| = \dim(U) = \dim(V)$, S is a basis for V. But then $U = \operatorname{span}(S) = V$. ☐

Theorem 3.29. *Let U be a subspace of V. Then $\dim(U) \leq \dim(V)$.*
Furthermore, if $U \neq V$, then $\dim(U) < \dim(V)$.

Proof. Let S be a basis for U. Since $U \subseteq V$, S is a linearly independent subset of V, so $\dim(U) = |S| \leq \dim(V)$. ☐

Theorem 3.30 (Invertibility equivalences).

(vi) *The rows of A form a basis for $\mathbb{R}^n$.*

(vii) *The columns of A form a basis for $\mathbb{R}^n$.*

Proof.

$\boxed{\text{(vii)} \Leftrightarrow \text{(i)}}$ Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{pmatrix}$. Then

$$\begin{aligned} \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\} \text{ is a basis for } \mathbb{R}^n &\iff \operatorname{span}\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\} = \mathbb{R}^n \\ &\iff \operatorname{REF}(\boldsymbol{A}) \text{ has no zero row} \\ &\iff \boldsymbol{A} \text{ is invertible.} \end{aligned}$$

$\boxed{\text{(iv)} \Leftrightarrow \text{(i)}}$ Since $\boldsymbol{A}$ is invertible $\Leftrightarrow$ $\boldsymbol{A}^\top$ is invertible, and columns of $\boldsymbol{A}$ are rows of $\boldsymbol{A}^\top$, (vii) $\Leftrightarrow$ (vi). ☐

To check if S is a basis, form vectors into columns/rows, compute det of matrix.

3.7 Transition Matrices

Let $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ be a basis for V, and $(\boldsymbol{v})_S = (c_1, \dots, c_n)$. Write it as a column vector

$$[\boldsymbol{v}]_S = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

Let $S = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$, $T = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ be bases for V.

Let $\boldsymbol{w} \in V$. In terms of old basis, $\boldsymbol{w} = c_1\boldsymbol{u}_1 + \cdots + c_n\boldsymbol{u}_n$, i.e., $[\boldsymbol{w}]_S = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$.

Write each old basis vector in terms of new basis:

$$\begin{aligned} \boldsymbol{u}_1 &= a_{11}\boldsymbol{v}_1 + \cdots + a_{n1}\boldsymbol{v}_n \\ &\vdots \\ \boldsymbol{u}_n &= a_{1n}\boldsymbol{v}_1 + \cdots + a_{nn}\boldsymbol{v}_n \end{aligned} \qquad \text{i.e., } [\boldsymbol{u}_1]_T = \begin{pmatrix} a_{11} \\ \vdots \\ a_{k1} \end{pmatrix}, \quad \cdots, \quad [\boldsymbol{u}_n]_T = \begin{pmatrix} a_{k1} \\ \vdots \\ a_{kk} \end{pmatrix}.$$

In terms of new basis, $\boldsymbol{w} = (c_1a_{11} + \cdots + c_na_{1n})\boldsymbol{v}_1 + \cdots + (c_1a_{n1} + \cdots + c_na_{nn})\boldsymbol{v}_n$, i.e.,

$$[\boldsymbol{w}]_T = \begin{pmatrix} c_1a_{11} + \cdots + c_ka_{1n} \\ \vdots \\ c_1a_{n1} + \cdots + c_na_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$
$$= \begin{pmatrix} [\boldsymbol{u}_1]_T & \cdots & [\boldsymbol{u}_n]_T \end{pmatrix} [w]_S.$$

Definition 3.31. The **transition matrix** from S to T is

$$P := \begin{pmatrix} [\boldsymbol{u}_1]_T & \cdots & [\boldsymbol{u}_n]_T \end{pmatrix}.$$

By above, for all $\boldsymbol{w} \in V$,

$$[\boldsymbol{w}]_T = \boldsymbol{P}[\boldsymbol{w}]_S.$$

To find transition matrix from bases $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ to $T = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$:

1. Perform Gauss-Jordan elimination on $\left(\begin{array}{ccc|ccc} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n & \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{array} \right)$
2. Then we obtain coefficients to express $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ as linear combinations of $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$

Theorem 3.32. *Let S and T be bases for a vector space V, and $\boldsymbol{P}$ be the transition matrix from S to T. Then $\boldsymbol{P}$ is invertible, and*

$$\boldsymbol{P}^{-1} \text{ is the transition matrix from } T \text{ to } S.$$

Proof. Let $\boldsymbol{Q}$ be the transition matrix from T to S. By half-price theorem, it suffices to show that $\boldsymbol{QP} = \boldsymbol{I}$.

Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$. Note that $[\boldsymbol{u}_i]_S = \boldsymbol{e}_i$; recall that $\boldsymbol{A}\boldsymbol{e}_i$ is the i-th column of a matrix $\boldsymbol{A}$. Hence

$$i\text{-th column of } \boldsymbol{QP} = (\boldsymbol{QP})[\boldsymbol{u}_i]_S = \boldsymbol{Q}(\boldsymbol{P}[\boldsymbol{u}_i]_S) = \boldsymbol{Q}[\boldsymbol{u}_i]_T = [\boldsymbol{u}_i]_S = \boldsymbol{e}_i.$$

□

Exercises

Exercise 1. Let A be an $n \times n$ invertible matrix. Let $\mathrm{col}(A)$ be the set of column vectors of A, and $\mathrm{Col}(A)$ be the column space of A. What can you say about $\mathrm{col}(A)$ and $\mathrm{Col}(A)$?

Solution. A is the transition matrix from $\mathrm{col}(A)$ to the standard basis for $\mathbb{R}^n$.
$\mathrm{Col}(A) = \mathbb{R}^n$ and $\mathrm{col}(A)$ is a basis for $\mathbb{R}^n$. □

4 Vector Spaces Associated with Matrices

Let $\boldsymbol{R}$ denote REF of matrix $\boldsymbol{A}$.

4.1 Row Spaces and Column Spaces

Let $\boldsymbol{A} = (a_{ij})$. Consider the rows/columns as vectors. Denote the i-th row of $\boldsymbol{A}$ as $\boldsymbol{r}_i = \begin{pmatrix} a_{i1} & \cdots & a_{in} \end{pmatrix}$, and j-th column of $\boldsymbol{A}$ as $\boldsymbol{c}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = \begin{pmatrix} a_{1j} & \cdots & a_{mj} \end{pmatrix}^\top$

Definition 4.1. The **row space** of $\boldsymbol{A}$ is

$$\text{Row}(\boldsymbol{A}) := \text{span}\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m\}.$$

The **column space** of $\boldsymbol{A}$ is

$$\text{Col}(\boldsymbol{A}) := \text{span}\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}.$$

Note that $\text{Row}(\boldsymbol{A}) = \text{Col}(\boldsymbol{A}^\top)$ and $\text{Col}(\boldsymbol{A}) = \text{Row}(\boldsymbol{A}^\top)$.

Theorem 4.2. *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are row equivalent, then* $\text{Row}(\boldsymbol{A}) = \text{Row}(\boldsymbol{B})$.
(EROs preserve row space.)

Proof. Show that all 3 EROs don't change row space (same span). □

Remark. Row equivalence $\not\Rightarrow$ same column space; for instance, consider $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

Corollary 4.3.

- *Basis for* $\text{Row}(\boldsymbol{A})$ = *basis for* $\text{Row}(\boldsymbol{R})$ = $\{$*non-zero rows of* $\boldsymbol{R}\}$.
- *Basis for* $\text{Col}(\boldsymbol{A})$ = *basis for* $\text{Row}(\boldsymbol{A}^\top)$.

Theorem 4.4. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be row equivalent. Then*

- *A set of columns of $\boldsymbol{A}$ is linearly independent $\iff$ corresponding set of columns of $\boldsymbol{B}$ is linearly independent*
- *A set of columns of $\boldsymbol{A}$ is a basis* $\text{Col}(\boldsymbol{A})$ $\iff$ *corresponding set of columns of $\boldsymbol{B}$ is a basis for* $\text{Col}(\boldsymbol{B})$

Note that basis for $\text{Col}(\boldsymbol{R})$ = $\{$pivot columns of $\boldsymbol{R}\}$. Therefore:

Corollary 4.5. *Basis for* $\text{Col}(\boldsymbol{A})$ = $\{$*columns of $\boldsymbol{A}$ that correspond to pivot columns of* $\boldsymbol{R}\}$.

Two methods to **find a basis** of a subspace of $\mathbb{R}^n$ spanned by $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$:

1. View $v_1, \dots, v_m$ as row vectors, find REF of $A = \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$, and take **non-zero rows** of R.

2. View $v_1, \dots, v_m$ as column vectors, find REF of $A = \begin{pmatrix} v_1 & \cdots & v_m \end{pmatrix}$, and take columns of A **corresponding to pivot columns** of R.

Theorem 4.6. *Let A be $m \times n$ matrix. Then*

$$\mathrm{Col}(A) = \left\{ Au \mid u \in \mathbb{R}^n \right\}.$$

Consequently,

$$\begin{aligned} Ax = b \text{ is consistent } &\iff \exists u \in \mathbb{R}^n,\ Au = b \\ &\iff b \in \{Au \mid u \in \mathbb{R}^n\} = \mathrm{Col}(A). \end{aligned}$$

Proof. Let $A = \begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix}$.

$\boxed{\supseteq}$ Let $b \in \left\{ Au \mid u \in \mathbb{R}^n \right\}$. Then $b = Au$ for some $u \in \mathbb{R}^n$. Thus

$$b = Au = \begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = u_1 c_1 + \cdots + u_n c_n$$

is a linear combination of columns of A, so $b \in \mathrm{Col}(A)$.

$\boxed{\subseteq}$ Let $b \in \mathrm{Col}(A)$. Then there exist $u_1, \dots, u_n \in \mathbb{R}$ such that

$$b = u_1 c_1 + \cdots + u_n c_n = \begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = Au.$$

□

4.2 Ranks

Theorem 4.7. $\dim \mathrm{Row}(A) = \dim \mathrm{Col}(A)$.

We sometimes say row rank = $\dim \mathrm{Row}(A)$, column rank = $\dim \mathrm{Col}(A)$.

Proof. We have

$$\begin{aligned} \dim \mathrm{Row}(A) &= \dim \mathrm{Row}(R) = \text{no. of non-zero rows in } R \\ &= \text{no. of pivot columns of } R \\ \dim \mathrm{Col}(A) &= \dim \mathrm{span}\{\text{columns of } A \text{ corresponding to pivot columns of } R\} \\ &= \text{no. of pivot columns of } R. \end{aligned}$$

□

Definition 4.8. The **rank** of A is dimension of row/column space of A, denoted $\operatorname{rank}(A)$.

Note that

$$\operatorname{rank}(A) = \underbrace{\text{no. of non-zero rows of } R}_{\text{row rank}} = \underbrace{\text{no. of pivot columns in } R}_{\text{column rank}}.$$

Properties - Let A be $m \times n$ matrix

- $\operatorname{rank}(A) = \operatorname{rank}(A^\top)$ $\quad (\because \operatorname{Row}(A) = \operatorname{Col}(A^\top))$
- $\operatorname{rank}(A) = 0 \Leftrightarrow A = 0$
- $\operatorname{rank}(A) \leq \min\{m, n\}$ $\quad$ (largest no. of pivot columns in R)

* A is called **full rank** if $\operatorname{rank}(A) = \min\{m, n\}$

- A square matrix A is of full rank $\iff$ A is invertible

Given $Ax = b$, let $\{c_1, \ldots, c_n\}$ be columns of A. Then

$$\begin{aligned} Ax = b \text{ is consistent } &\iff b \in \operatorname{span}\{c_1, \ldots, c_n\} \\ &\iff \operatorname{span}\{c_1, \ldots, c_n\} = \operatorname{span}\{c_1, \ldots, c_n, b\} \\ &\iff \dim \operatorname{span}\{c_1, \ldots, c_n\} = \dim \operatorname{span}\{c_1, \ldots, c_n, b\} \\ &\iff \operatorname{rank}(A) = \operatorname{rank}(A \mid b) \end{aligned}$$

Remark. $\operatorname{rank}(A) \leq \operatorname{rank}(A \mid b) \leq \operatorname{rank}(A) + 1$

Theorem 4.9. *Let A and B be $m \times n$ and $n \times p$ matrices. Then*

$$\operatorname{rank}(AB) \leq \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}.$$

Proof. We need to show (i) $\operatorname{rank}(AB) \leq \operatorname{rank}(A)$, and (ii) $\operatorname{rank}(AB) \leq \operatorname{rank}(B)$.

(i) Let $A = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$ and $B = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix}$. By block multiplication,

$$AB = \begin{pmatrix} Ab_1 & \cdots & Ab_n \end{pmatrix}.$$

Note that each $Ab_i \in \operatorname{Col}(A)$. Hence $\operatorname{Col}(AB) \subseteq \operatorname{Col}(A)$.

(ii) Applying (i) to $B^\top A^\top$, we have $\operatorname{rank}(B^\top A^\top) \leq \operatorname{rank}(B^\top)$. Then

$$\operatorname{rank}(AB) = \operatorname{rank}((AB)^\top) = \operatorname{rank}(B^\top A^\top) \leq \operatorname{rank}(B^\top) = \operatorname{rank}(B).$$

□

4.3 Nullspaces and Nullities

Definition 4.10. Let A be a $m \times n$ matrix. The **nullspace** of A is the solution space of $Ax = \mathbf{0}$:

$$\operatorname{NS}(A) := \{u \in \mathbb{R}^n \mid Au = \mathbf{0}\}.$$

Definition 4.11. The **nullity** of $\boldsymbol{A}$ is dimension of nullspace, denoted $\operatorname{nullity}(\boldsymbol{A})$.

If $\boldsymbol{A}$ is $m \times n$, then $\operatorname{nullity}(\boldsymbol{A}) \leq n$ ($\because$ nullspace is subspace of $\mathbb{R}^n$)
Nullity = no. of arbitrary parameters = no. of non-pivot columns in REF
From now on, unless otherwise stated, vectors in the nullspace are viewed as column vectors.

Theorem 4.12 (Rank-nullity theorem)**.** *Let $\boldsymbol{A}$ be $m \times n$ matrix. Then*

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{nullity}(\boldsymbol{A}) = n. \tag{5}$$

Proof. $\operatorname{rank}(\boldsymbol{A})$ = no. of pivot columns in $\boldsymbol{R}$, $\operatorname{nullity}(\boldsymbol{A})$ = no. of non-pivot columns in $\boldsymbol{R}$ □

Theorem 4.13. *Suppose $\boldsymbol{Ax} = \boldsymbol{b}$ has a solution $\boldsymbol{u}$. Then solution set of $\boldsymbol{Ax} = \boldsymbol{b}$ is*

$$\{\boldsymbol{u} + \boldsymbol{v} \mid \boldsymbol{v} \in \textit{nullspace of } \boldsymbol{A}\}.$$

That is, $\boldsymbol{Ax} = \boldsymbol{b}$ has general solution

$$\boldsymbol{x} = (\text{one particular solution to } \boldsymbol{Ax} = \boldsymbol{b}) + (\text{general solution for } \boldsymbol{Ax} = \boldsymbol{0}).$$

As a consequence, a consistent linear system $Ax = b$ has only one solution if and only if $\mathrm{NS}(\boldsymbol{A}) = \{\boldsymbol{0}\}$.

Exercises

Exercise 2. Find the rank of the matrix below:

$$\begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{pmatrix}$$

where a, b, c are real numbers such that $a \neq b$, $a \neq c$ and $b \neq c$.

Solution. By Gaussian elimination,

$$\begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & b-a & c-a \\ 0 & b^2-a^2 & c^2-a^2 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & b-a & c-a \\ 0 & 0 & (c-a)(c-b) \end{pmatrix}$$

Hence the rank of the matrix is 3. □

5 Orthogonality

5.1 The Dot Product

Definition 5.1. Let $\boldsymbol{u} = (u_1, \ldots, u_n)$, $\boldsymbol{v} = (v_1, \ldots, v_n)$ be two vectors in $\mathbb{R}^n$.

- The **dot product** (**inner product**) of $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\boldsymbol{u} \cdot \boldsymbol{v} := \sum_{i=1}^{n} u_i v_i.$$

- The **norm** (**length**) of $\boldsymbol{v}$ is

$$\|\boldsymbol{v}\| := \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{\sum_{i=1}^{n} {v_i}^2}.$$

  We say $\boldsymbol{v}$ is a **unit vector** if $\|\boldsymbol{v}\| = 1$.

- The **distance** between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$d(\boldsymbol{u}, \boldsymbol{v}) := \|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}.$$

- The **angle** between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\theta = \cos^{-1}\left(\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right), \qquad 0^\circ \leq \theta \leq 180^\circ.$$

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are viewed as column vectors, then

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \boldsymbol{u}^\top \boldsymbol{v}.$$

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are viewed as row vectors, then

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \boldsymbol{u}\boldsymbol{v}^\top.$$

Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, $c \in \mathbb{R}$. The following properties hold:

- $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$
- $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$
  $\boldsymbol{w} \cdot (\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{w} \cdot \boldsymbol{u} + \boldsymbol{w} \cdot \boldsymbol{v}$
- $(c\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (c\boldsymbol{v}) = c(\boldsymbol{u} \cdot \boldsymbol{v})$
- $\|c\boldsymbol{u}\| = |c|\|\boldsymbol{u}\|$
- $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$ and $\boldsymbol{u} \cdot \boldsymbol{u} = 0 \iff \boldsymbol{u} = \boldsymbol{0}$

Theorem 5.2 (Inequalities). *Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$*

- $\|\boldsymbol{u} \cdot \boldsymbol{v}\| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|$ *(Cauchy-Schwarz inequality)*
- $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ *(triangle inequality)*
- $d(\boldsymbol{u}, \boldsymbol{w}) \leq d(\boldsymbol{u}, \boldsymbol{v}) + d(\boldsymbol{v}, \boldsymbol{w})$ *(triangle inequality)*

5.2 Orthogonal and Orthonormal Bases

Definition 5.3. We say $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ are **orthogonal** if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$, denoted $\boldsymbol{u} \perp \boldsymbol{v}$.
Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$.

- S is **orthogonal** if $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0 \ \forall i \neq j$.
- S is **orthonormal** if S is orthogonal and every vector in S is a unit vector.

Normalising: convert orthogonal vectors into orthonormal ones, $\boldsymbol{u}_i \to \dfrac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$

Theorem 5.4. *An orthogonal set of non-zero vectors is linearly independent.*

Proof. Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$. Suppose there exist $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$c_1\boldsymbol{u}_1 + \cdots + c_k\boldsymbol{u}_k = \mathbf{0}.$$

Dot product this linear combination with $\boldsymbol{u}_i$ over all $i = 1, \ldots, k$ to get $c_i = 0$. □

Definition 5.5. Let S be a basis for a vector space.

- S is an **orthogonal basis** if it is orthogonal.
- S is an **orthonormal basis** if it is orthonormal.

Theorem 5.6. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ be an orthogonal basis for a vector space V. Then for any $\boldsymbol{w} \in V$,*

$$\boldsymbol{w} = \frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\boldsymbol{u}_k. \tag{6}$$

Proof. Let $\boldsymbol{w} = c_1\boldsymbol{u}_1 + \cdots + c_k\boldsymbol{u}_k$. Take dot product of $\boldsymbol{w}$ and each $\boldsymbol{u}_i$ to get $c_i = \frac{\boldsymbol{w} \cdot \boldsymbol{u}_i}{\boldsymbol{u}_i \cdot \boldsymbol{u}_i}$. □

Corollary 5.7. *Let $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be an orthonormal basis for a vector space V. Then for any $\boldsymbol{w} \in V$,*

$$\boldsymbol{w} = (\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k) \cdot \boldsymbol{v}_k. \tag{7}$$

Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \subseteq \mathbb{R}^n$ and $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k \end{pmatrix}$. For any $\boldsymbol{w} \in V$,

- $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is orthogonal $\iff$ $\boldsymbol{A}^\top \boldsymbol{A}$ is diagonal
- $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is orthonormal $\iff$ $\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{I}_k$

Definition 5.8. Let V be a subspace of $\mathbb{R}^n$. A vector $\boldsymbol{u} \in \mathbb{R}^n$ is **orthogonal** to V if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ for all $\boldsymbol{v} \in V$.

Let V be a plane in $\mathbb{R}^3$ defined by the equation $ax + by + cz = 0$. Then

$$V = \left\{(x, y, z) \in \mathbb{R}^3 \mid ax + by + cz = 0\right\} = \{\boldsymbol{u} \in \mathbb{R}^n \mid (a, b, c) \cdot \boldsymbol{u} = 0\}.$$

The vector $\boldsymbol{n}$ is a **normal vector** of V.

Theorem 5.9. *Let $V = \operatorname{span}\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ be a subspace of $\mathbb{R}^n$. For $\boldsymbol{v} \in \mathbb{R}^n$,*

$$\boldsymbol{v} \text{ is orthogonal to } V \iff \boldsymbol{v} \cdot \boldsymbol{u}_i = 0 \qquad (i = 1, \dots, k).$$

Proof.
$\boxed{\Rightarrow}$ This follows from definition.
$\boxed{\Leftarrow}$ Consider the coordinate vector of any $\boldsymbol{v} \in V$. □

(Exericse) If W is a subspace of $\mathbb{R}^n$, then $W^{\perp} = \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v} \text{ is orthogonal to } W\}$ is subspace of $\mathbb{R}^n$

Definition 5.10. Let V be a subspace of $\mathbb{R}^n$. Every $\boldsymbol{u} \in \mathbb{R}^n$ can be uniquely written as

$$\boldsymbol{u} = \boldsymbol{n} + \boldsymbol{p}$$

where $\boldsymbol{n}$ is orthogonal to V, $\boldsymbol{p} \in V$. We call $\boldsymbol{p}$ the **orthogonal projection** of $\boldsymbol{u}$ onto V, denote $\boldsymbol{p} = \operatorname{proj}_V(\boldsymbol{u})$.

This is analogous to "dropping the perpendicular".

Theorem 5.11. *Let V be subspace of $\mathbb{R}^n$, and $\boldsymbol{w} \in \mathbb{R}^n$. If $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ is orthogonal basis for V, then the projection of $\boldsymbol{w}$ onto V is*

$$\operatorname{proj}_V(\boldsymbol{w}) = \frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\boldsymbol{u}_k. \tag{8}$$

That is, we sum the projections of $\boldsymbol{w}$ onto $\boldsymbol{u}_1, \dots, \boldsymbol{u}_k$.

Proof. Write $\boldsymbol{w} = \boldsymbol{n} + \boldsymbol{p}$, where $\boldsymbol{p}$ is defined as in (8).
Obviously $\boldsymbol{p} \in V$. Then show $\boldsymbol{n}$ is orthogonal to V, i.e., $\boldsymbol{n} \cdot \boldsymbol{u}_i = 0$:

$$\begin{aligned}
\boldsymbol{n} \cdot \boldsymbol{u}_i &= \boldsymbol{w} \cdot \boldsymbol{u}_i - \boldsymbol{p} \cdot \boldsymbol{u}_i \\
&= \boldsymbol{w} \cdot \boldsymbol{u}_i - \frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}(\boldsymbol{u}_1 \cdot \boldsymbol{u}_i) - \cdots - \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}(\boldsymbol{u}_k \cdot \boldsymbol{u}_i) \\
&= \boldsymbol{w} \cdot \boldsymbol{u}_i - \frac{\boldsymbol{w} \cdot \boldsymbol{u}_i}{\boldsymbol{u}_i \cdot \boldsymbol{u}_i}(\boldsymbol{u}_i \cdot \boldsymbol{u}_i) = 0.
\end{aligned}$$

□

The **Gram-Schmidt process** allows us to convert a **basis** to an **orthogonal basis**.

Theorem 5.12 (Gram-Schmidt process). *Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ be a basis for vector space V. Define*

$$\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{u}_1 \\
\boldsymbol{v}_2 &= \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 \\
\boldsymbol{v}_3 &= \boldsymbol{u}_3 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 \\
&\vdots \\
\boldsymbol{v}_k &= \boldsymbol{u}_k - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_1 - \cdots - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_{k-1}}{\boldsymbol{v}_{k-1} \cdot \boldsymbol{v}_{k-1}}\boldsymbol{v}_{k-1}
\end{aligned}$$

Then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is an orthogonal basis for V.

That is, we subtract away the components of $\boldsymbol{v}_i$ that are parallel to the previous vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}$, so that $\boldsymbol{v}_i$ is orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}$.

Normalising with $\boldsymbol{w}_i = \dfrac{\boldsymbol{v}_i}{\|\boldsymbol{v}_i\|}$, then $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\}$ is an orthonormal basis for V.

5.3 Best Approximations

Theorem 5.13. *Let V be a subspace of $\mathbb{R}^n$. For $\boldsymbol{u} \in \mathbb{R}^n$, let $\boldsymbol{p}$ be the projection of $\boldsymbol{u}$ onto V. Then*

$$d(\boldsymbol{u}, \boldsymbol{p}) \leq d(\boldsymbol{u}, \boldsymbol{v}) \qquad (\boldsymbol{v} \in V) \tag{9}$$

*i.e., $\boldsymbol{p}$ is the **best approximation** of $\boldsymbol{u}$ in V.*

Proof. Let $\boldsymbol{v} \in V$. Define

$$\boldsymbol{n} = \boldsymbol{u} - \boldsymbol{p}, \qquad \boldsymbol{w} = \boldsymbol{p} - \boldsymbol{v}, \qquad \boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v}.$$

We need to show $\|\boldsymbol{n}\| \leq \|\boldsymbol{x}\|$. Note that $\boldsymbol{x} = \boldsymbol{n} + \boldsymbol{w}$. Then since $\boldsymbol{n} \cdot \boldsymbol{w} = 0$,

$$\begin{aligned}
\|\boldsymbol{x}\|^2 = \boldsymbol{x} \cdot \boldsymbol{x} &= (\boldsymbol{n} + \boldsymbol{w}) \cdot (\boldsymbol{n} + \boldsymbol{w}) \\
&= \|\boldsymbol{n}\|^2 + \|\boldsymbol{w}\|^2 \geq \|\boldsymbol{n}\|^2.
\end{aligned}$$

□

Least squares method: Suppose we want to fit points (x_i, y_i) on a straight line $y = c + dx$. For each i, observed value is y_i, predicted value is $c + dx_i$; error is $y_i - (c + dx_i)$. Since error terms may be positive or negative, and to prevent the error terms from cancelling out each other, we look at the square of error terms instead; we want to minimise the *sum of squares of errors* (SSE). Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \boldsymbol{x} = \begin{pmatrix} c \\ d \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then

$$\sum_{i=1}^{n} (y_i - (c + dx_i))^2 = \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|^2.$$

Definition 5.14. Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system, where $\boldsymbol{A}$ is an $m \times n$ matrix. We say $\boldsymbol{u} \in \mathbb{R}^n$ is a **least squares solution** to the linear system if

$$\|\boldsymbol{b} - \boldsymbol{Au}\| \leq \|\boldsymbol{b} - \boldsymbol{Av}\| \qquad (\boldsymbol{v} \in \mathbb{R}^n).$$

A least squares solution solves the equation $\boldsymbol{Ax} = \boldsymbol{b}$ as closely as possible, in the sense that the sum of the squares of the difference $\boldsymbol{b} - \boldsymbol{Ax}$ is minimised.

- If $\boldsymbol{Ax} = \boldsymbol{b}$ is consistent, then there is an exact solution.
- If $\boldsymbol{Ax} = \boldsymbol{b}$ is inconsistent, recall that $\mathrm{Col}(\boldsymbol{A}) = \{\boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n\}$. Then the projection of $\boldsymbol{b}$ onto $\mathrm{Col}(\boldsymbol{A})$ is the best approximation of $\boldsymbol{b}$ in $\mathrm{Col}(\boldsymbol{A})$, i.e., $\|\boldsymbol{b} - \boldsymbol{Au}\|$ is minimised.

Theorem 5.15. *Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system, where $\boldsymbol{A}$ is an $m \times n$ matrix. Let $\boldsymbol{p}$ be the projection of $\boldsymbol{b}$ onto $\mathrm{Col}(\boldsymbol{A})$. Then*

$$\left\|\boldsymbol{b} - \boldsymbol{p}\right\| \leq \|\boldsymbol{b} - \boldsymbol{Av}\| \qquad (\boldsymbol{v} \in \mathbb{R}^n)$$

i.e., $\boldsymbol{u}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b} \iff \boldsymbol{u}$ is a solution to $\boldsymbol{Ax} = \boldsymbol{p}$.

Proof. Let $V = \mathrm{Col}(\boldsymbol{A})$. By the preceding theorem, for any $\boldsymbol{w} \in V$,

$$\left\|\boldsymbol{b} - \boldsymbol{p}\right\| = d(\boldsymbol{b}, \boldsymbol{p}) \leq d(\boldsymbol{b}, \boldsymbol{w}) = \|\boldsymbol{b} - \boldsymbol{w}\|.$$

Since $V = \{\boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n\}$, it follows that $\left\|\boldsymbol{b} - \boldsymbol{p}\right\| \leq \|\boldsymbol{b} - \boldsymbol{Av}\|$ for all $\boldsymbol{v} \in \mathbb{R}^n$. $\square$

To find least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$:

1. Find an orthogonal basis for $\mathrm{Col}(\boldsymbol{A})$.
2. Find $\boldsymbol{p}$, the orthogonal projection of $\boldsymbol{b}$ onto $\mathrm{Col}(\boldsymbol{A})$.
3. Solve $\boldsymbol{Ax} = \boldsymbol{p}$.

Alternatively, there is a relatively straightforward method to find a least squares solution, without having to first find an orthogonal basis for $\mathrm{Col}(\boldsymbol{A})$.

Theorem 5.16. *Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system. Then $\boldsymbol{u}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b} \iff u$ is a solution to $\boldsymbol{A}^\top \boldsymbol{Ax} = \boldsymbol{A}^\top \boldsymbol{b}$.*

$\boldsymbol{A}^\top \boldsymbol{Ax} = \boldsymbol{A}^\top \boldsymbol{b}$ is called the **normal equation** to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$.

Proof. Let $A = \begin{pmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{pmatrix}$, where $\boldsymbol{a}_i$ is the i-th column of $\boldsymbol{A}$. Let $V = \mathrm{Col}(\boldsymbol{A}) = \mathrm{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} =$

$\{\boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n\}$. Then

$$\begin{aligned}
& \boldsymbol{u} \text{ is a least squares solution to } \boldsymbol{Ax} = \boldsymbol{b} \\
\iff & \boldsymbol{Au} \text{ is the projection of } \boldsymbol{b} \text{ onto } V \\
\iff & \boldsymbol{b} - \boldsymbol{Au} \text{ is orthogonal to } V \\
\iff & \boldsymbol{b} - \boldsymbol{Au} \text{ is orthogonal to } \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \\
\iff & \boldsymbol{a}_1 \cdot (\boldsymbol{b} - \boldsymbol{Au}) = 0, \quad \ldots, \quad \boldsymbol{a}_n \cdot (\boldsymbol{b} - \boldsymbol{Au}) = 0 \\
\iff & \begin{pmatrix} \boldsymbol{a}_1 \cdot (\boldsymbol{b} - \boldsymbol{Au}) \\ \vdots \\ \boldsymbol{a}_n \cdot (\boldsymbol{b} - \boldsymbol{Au}) \end{pmatrix} = \boldsymbol{0} \\
\iff & \boldsymbol{A}^\top (\boldsymbol{b} - \boldsymbol{Au}) = \boldsymbol{0} \\
\iff & \boldsymbol{A}^\top \boldsymbol{Au} = \boldsymbol{A}^\top \boldsymbol{b}.
\end{aligned}$$

□

5.4 Orthogonal Matrices

Definition 5.17. A square matrix $\boldsymbol{A}$ is **orthogonal** if $\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$.

Equivalently, $\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}$.

Properties:

- If $\boldsymbol{A}$ is orthogonal matrix, so is $\boldsymbol{A}^\top$
- If $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal, then $\boldsymbol{AB}$ is orthogonal:

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} = \boldsymbol{B}^\top \boldsymbol{A}^\top = (\boldsymbol{AB})^\top.$$

Theorem 5.18. *Let $\boldsymbol{A}$ be a square matrix of order n. Then*

$$\begin{aligned}
\boldsymbol{A} \textit{ is orthogonal} \iff & \textit{rows of } \boldsymbol{A} \textit{ form orthonormal basis for } \mathbb{R}^n \\
\iff & \textit{columns of } \boldsymbol{A} \textit{ form orthonormal basis for } \mathbb{R}^n
\end{aligned}$$

Proof. Let $A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$, where a_i is the i-th row of A. Note that

$$\begin{aligned}
AA^\top = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} a_1{}^\top & \cdots & a_n{}^\top \end{pmatrix} &= \begin{pmatrix} a_1 a_1{}^\top & \cdots & a_1 a_n{}^\top \\ \vdots & & \vdots \\ a_n a_1{}^\top & \cdots & a_n a_n{}^\top \end{pmatrix} \\
&= \begin{pmatrix} a_1 \cdot a_1 & \cdots & a_1 \cdot a_n \\ \vdots & & \vdots \\ a_n \cdot a_1 & \cdots & a_n \cdot a_n \end{pmatrix}
\end{aligned}$$

Then

$$
\begin{aligned}
\boldsymbol{A} \text{ is orthogonal} &\iff \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I} \\
&\iff \text{for all } i, j, \quad \boldsymbol{a}_i \cdot \boldsymbol{a}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \\
&\iff \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \text{ are orthonormal.}
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\boldsymbol{A} \text{ is orthogonal} &\iff \boldsymbol{A}^\top \text{ is orthogonal} \\
&\iff \text{rows of } \boldsymbol{A}^\top \text{ form orthonormal basis for } \mathbb{R}^n \\
&\iff \text{columns of } \boldsymbol{A} \text{ form orthonormal basis for } \mathbb{R}^n.
\end{aligned}
$$

□

Theorem 5.19. *Let S and T be orthonormal bases for a vector space. Let $\boldsymbol{P}$ be transition matrix from S to T. Then $\boldsymbol{P}$ is orthogonal, so $\boldsymbol{P}^\top$ is transition matrix from T to S.*

Proof. Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$, $T = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$. Since T is orthonormal,

$$
\begin{aligned}
\boldsymbol{u}_1 &= (\boldsymbol{u}_1 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_k)\boldsymbol{v}_k, \\
&\vdots \\
\boldsymbol{u}_k &= (\boldsymbol{u}_k \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (u_k \cdot v_k)v_k.
\end{aligned}
$$

Thus the transition matrix from S to T is

$$
\boldsymbol{P} = \begin{pmatrix} u_1 \cdot v_1 & \cdots & u_k \cdot v_1 \\ \vdots & & \vdots \\ u_1 \cdot v_k & \cdots & u_k \cdot v_k \end{pmatrix}.
$$

Similarly, the transition matrix from S to T is

$$
\boldsymbol{Q} = \begin{pmatrix} v_1 \cdot u_1 & \cdots & v_k \cdot u_1 \\ \vdots & & \vdots \\ v_1 \cdot u_k & \cdots & v_k \cdot u_k \end{pmatrix}.
$$

Hence $\boldsymbol{Q} = \boldsymbol{P}^\top$. But we know $\boldsymbol{Q} = \boldsymbol{P}^{-1}$. Therefore $\boldsymbol{P}^{-1} = \boldsymbol{P}^\top$. □

Example. Let $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ be standard basis for $\mathbb{R}^2$: $\boldsymbol{e}_1 = (1, 0)$, $\boldsymbol{e}_2 = (0, 1)$. Rotate the original xy-coordinates anticlockwise about the origin through angle θ. The transformed basis vectors are

$$
\begin{aligned}
u_1 &= (\cos\theta, \sin\theta) = (\cos\theta)\boldsymbol{e}_1 + (\sin\theta)\boldsymbol{e}_2, \\
u_2 &= (-\sin\theta, \cos\theta) = (-\sin\theta)\boldsymbol{e}_1 + (\cos\theta)\boldsymbol{e}_2.
\end{aligned}
$$

$S = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is an orthonormal basis for $\mathbb{R}^2$. Thus the transition matrix from S to E is $\boldsymbol{P} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Hence the transition matrix from E to S is $\boldsymbol{P}^\top = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

Exercises

1. Let $S = \{u_1, u_2, \ldots, u_k\}$ be an orthonormal basis for a subspace V of $\mathbb{R}^n$.

   (a) For any vector $v \in V$, show that $\|(v)_S\| = \|v\|$.

   (b) For any vectors $v, w \in V$, show that $d((v)_S, (w)_S) = d(v, w)$.

   Hints:

   - Writing $\boldsymbol{v}$ as a column vector, $\boldsymbol{v} = \boldsymbol{U}(\boldsymbol{v})_S$ where $\boldsymbol{U} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k \end{pmatrix}$.
   - S is an orthonormal set $\iff$ $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_n$.
   - $\|\boldsymbol{v}\|^2 = \boldsymbol{v} \cdot \boldsymbol{v} = \boldsymbol{v}^\top \boldsymbol{v}$.

   Solution.

   (a) It is easier to work with norm squared:

   $$\begin{aligned} \|\boldsymbol{v}\|^2 = \boldsymbol{v}^\top \boldsymbol{v} &= (\boldsymbol{U}(\boldsymbol{v})_S)^\top (\boldsymbol{U}(\boldsymbol{v})_S) \\ &= {(\boldsymbol{v})_S}^\top \boldsymbol{U}^\top \boldsymbol{U}(\boldsymbol{v})_S \\ &= {(\boldsymbol{v})_S}^\top (\boldsymbol{v})_S = \|(\boldsymbol{v})_S\|^2. \end{aligned}$$

   Hence $\|v\| = \|(v)_S\|$.

   (b) We proceed similarly as above:

   $$\begin{aligned} d(\boldsymbol{v}, \boldsymbol{w})^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2 &= (\boldsymbol{v} - \boldsymbol{w})^\top (\boldsymbol{v} - \boldsymbol{w}) \\ &= (\boldsymbol{U}(\boldsymbol{v})_S - \boldsymbol{U}(\boldsymbol{w})_S)^\top (\boldsymbol{U}(\boldsymbol{v})_S - \boldsymbol{U}(\boldsymbol{w})_S) \\ &= (\boldsymbol{U}((\boldsymbol{v})_S - (\boldsymbol{w})_S))^\top \boldsymbol{U}((\boldsymbol{v})_S - (\boldsymbol{w})_S) \\ &= ((\boldsymbol{v})_S - (\boldsymbol{w})_S)^\top \boldsymbol{U}^\top \boldsymbol{U}((\boldsymbol{v})_S - (\boldsymbol{w})_S) \\ &= ((\boldsymbol{v})_S - (\boldsymbol{w})_S)^\top ((\boldsymbol{v})_S - (\boldsymbol{w})_S) \\ &= \|(\boldsymbol{v})_S - (\boldsymbol{w})_S\|^2 = d((\boldsymbol{v})_S, (\boldsymbol{w})_S) \end{aligned}$$

   $\square$

   Alternative solution.

   (a) Let $(\boldsymbol{v})_S = (a_1, \ldots, a_k)$, i.e., $\boldsymbol{v} = a_1\boldsymbol{u}_1 + \cdots + a_k\boldsymbol{u}_k$. Then

   $$\begin{aligned} \|\boldsymbol{v}\|^2 &= \boldsymbol{v} \cdot \boldsymbol{v} \\ &= (a_1\boldsymbol{u}_1 + \cdots + a_k\boldsymbol{u}_k) \cdot (a_1\boldsymbol{u}_1 + \cdots + a_k\boldsymbol{u}_k) \\ &= {a_1}^2 + \cdots + {a_k}^2 \\ &= \|(\boldsymbol{v})_S\|. \end{aligned}$$

   (b) Using (a),

   $$\begin{aligned} d((\boldsymbol{v})_S, (\boldsymbol{w})_S) &= \|(\boldsymbol{v})_S - (\boldsymbol{w})_S\| \\ &= \|(\boldsymbol{v} - \boldsymbol{w})_S\| \\ &= \|\boldsymbol{v} - \boldsymbol{w}\| \\ &= d(\boldsymbol{v}, \boldsymbol{w}) \end{aligned}$$

□

2. Let V be a subspace of $\mathbb{R}^n$, and $u \in \mathbb{R}^n$. Show that $\boldsymbol{u}$ can be uniquely written as

$$\boldsymbol{u} = \boldsymbol{n} + \boldsymbol{p}$$

such that $\boldsymbol{n}$ is a vector orthogonal to V, and $\boldsymbol{p}$ is a vector in V.

Solution.

Lemma 1: Orthogonal subspaces intersect only in the zero subspace.

Proof. Suppose $V \perp W$. Let $v \in V \cap W$. Then $v \in V$ and $v \in W$, so $\|v\|^2 = v \cdot v = 0 \implies v = 0$. □

Lemma 2: Suppose $\{w_1, \ldots, w_r\}$ is a basis for W, and $\{w_{r+1}, \ldots, w_n\}$ is a basis for $W^\perp$. Then $\{w_1, \ldots, w_n\}$ is a basis for $\mathbb{R}^n$.

Proof. It suffices to check linear independence. Suppose $c_1 w_1 + \cdots + c_n w_n = 0$. Then

$$c_1 w_1 + \cdots + c_r w_r = -(c_{r+1} w_{r+1} + \cdots + c_n w_n).$$

Note that the LHS lies in W, and the RHS lies in $W^\perp$. By Lemma 1, we have $c_1 = \cdots = c_r = 0$ and $c_{r+1} = \cdots = c_n = 0$. □

$\boxed{\text{Existence}}$ Write $u = c_1 w_1 + \cdots + c_n w_n$. Then

$$\begin{aligned} p &= c_1 w_1 + \cdots + c_r w_r, \\ n &= c_{r+1} w_{r+1} + \cdots + c_n w_n. \end{aligned}$$

$\boxed{\text{Uniqueness}}$ If have 2 distinct n, p, contradicts $\{w_1, \ldots, w_n\}$ being a basis.

□

3. Let W be a subspace of $\mathbb{R}^n$. Define the *orthogonal complement* of W as

$$W^\perp := \left\{ \boldsymbol{u} \in \mathbb{R}^n \mid \boldsymbol{u} \text{ is orthogonal to } W \right\}.$$

(a) Show that $W^\perp$ is a subspace of $\mathbb{R}^n$.

(b) Prove that $\dim(W) + \dim(W^\perp) = n$.

(c) If $W = \left\{ (a, a+b, b, b+c, c) \mid a, b, c \in \mathbb{R} \right\} \subseteq \mathbb{R}^5$, find $W^\perp$.

Solution.

(a) Check that $W^\perp$ is closed under addition and scalar multiplication.

(b) Let $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\}$ be an orthogonal basis for W (find a basis for W, then use Gram-Schmidt process). Extend $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r\}$ to an orthogonal basis $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ for $\mathbb{R}^n$.
Claim: $W^\perp = \operatorname{span}\{\boldsymbol{w}_{r+1}, \ldots, \boldsymbol{w}_n\}$.
$\boxed{\supseteq}$ Each w_k, $k = r+1, \ldots, n$ is orthogonal to each w_j, $j = 1, \ldots, r$. Thus each w_k is orthogonal to W, so $w_k \in W^\perp$. Hence any linear combination of $w_{r+1}, \ldots, w_n \in W^\perp$.

$\boxed{\subseteq}$ Let $w = c_1 w_1 + \cdots + c_n w_n \in W^\perp$. Then

$$(c_1 w_1 + \cdots + c_n w_n) \cdot \boldsymbol{w}_j = 0 \qquad (j = 1, \ldots, r).$$

Thus $c_j(w_j \cdot w_j) = 0$. Since $w_j \cdot w_j \neq 0$, it follows that $c_j = 0$.
Hence $w = c_{r+1} w_{r+1} + \cdots + c_n w_n$, so $w \in \operatorname{span}\{w_{r+1}, \ldots, w_n\}$.

(c) Use $(\text{Row}\, A)^\perp = \text{NS}(A)$:
A basis for W is $\{(1,1,0,0,0), (0,1,1,1,0), (0,0,0,1,1)\}$, so

$$W = \text{Row}\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Hence

$$W^\perp = \text{NS}\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{pmatrix} \right\}.$$

□

Solution. see Dr Ng's solution for alternative □

6 Diagonalisation

All vectors are written as column vectors.

6.1 Eigenvalues and Eigenvectors

Definition 6.1. Let $\boldsymbol{A}$ be a square matrix of order n. We say $\boldsymbol{u} \in \mathbb{R}^n$, $\boldsymbol{u} \neq \mathbf{0}$ is an **eigenvector** of $\boldsymbol{A}$ if there exists $\lambda \in \mathbb{R}$ such that

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}.$$

We call λ an **eigenvalue** of $\boldsymbol{A}$, and $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ *associated* with eigenvalue λ.

Note:

$$\begin{aligned}
& \lambda \text{ is eigenvalue of } \boldsymbol{A} \\
\iff & \boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u} \text{ for some } \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{u} \neq \mathbf{0} \\
\iff & (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{u} = \mathbf{0} \text{ for some } \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{u} \neq \mathbf{0} \\
\iff & (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \mathbf{0} \text{ has non-trivial solutions} && \text{(do Gaussian elimination)} \\
\iff & (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} \text{ is singular} \\
\iff & \det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0. && \text{(find det using Gaussian elimination)}
\end{aligned}$$

If expanded, $\det(\lambda\boldsymbol{I} - \boldsymbol{A})$ is a polynomial of λ of degree n.

Definition 6.2. Let $\boldsymbol{A}$ be a square matrix of order n. The characteristic polynomial of $\boldsymbol{A}$ is

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}).$$

The characteristic equation of $\boldsymbol{A}$ is

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0.$$

Hence the roots of the characteristic equation of $\boldsymbol{A}$ are the eigenvalues of $\boldsymbol{A}$.
If two matrices have the same characteristic equations, then they have the same set of eigenvalues:

- $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ have the characteristic polynomial, since

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = \det\left(\lambda\boldsymbol{I} - \boldsymbol{A}^\top\right).$$

- If $\boldsymbol{P}$ is invertible, A and $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ have the same characteristic polynomial, since

$$\det\left(\lambda\boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\right) = \det\left(\boldsymbol{P}^{-1}(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{P}\right) = \det(\lambda\boldsymbol{I} - \boldsymbol{A}).$$

- If the characteristic polynomial of $\boldsymbol{A}$ is $p(\lambda) = a_0 + a_1\lambda + \cdots + a_{n-1}\lambda^{n-1} + \lambda^n$, then

$$a_0 = p(0) = \det(0\boldsymbol{I} - \boldsymbol{A}) = \det(-\boldsymbol{A}) = (-1)^n \det(\boldsymbol{A}).$$

Theorem 6.3 (Invertibility equivalences).

(viii) $\operatorname{rank}(\boldsymbol{A}) = n$.

(ix) *0 is not an eigenvalue of* $\boldsymbol{A}$.

Proof.

(viii) $\operatorname{rank}(\boldsymbol{A}) = n \iff \boldsymbol{A}$ has full rank $\iff \boldsymbol{A}$ is invertible.

(ix) Since $\det(0\boldsymbol{I} - A) = \det(-\boldsymbol{A}) = (-1)^n \det(\boldsymbol{A})$,

$$\boldsymbol{A} \text{ is invertible} \iff \det(\boldsymbol{A}) \neq 0 \iff 0 \text{ is not eigenvalue of } \boldsymbol{A}.$$

□

Theorem 6.4. *If* $\boldsymbol{A}$ *is triangular, then eigenvalues of* $\boldsymbol{A}$ *are diagonal entries of* $\boldsymbol{A}$.

Proof. Let $\boldsymbol{A} = (a_{ij})_{n\times n}$ be triangular. Then $\lambda\boldsymbol{I} - \boldsymbol{A}$ is triangular, with diagonal entries $\lambda - a_{11}, \ldots, \lambda - a_{nn}$. Then

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = (\lambda - a_{11}) \cdots (\lambda - a_{nn}).$$

Hence the diagonal entries $a_{11}, \ldots, a_{nn}$ are the eigenvalues of $\boldsymbol{A}$. □

Definition 6.5. Let $\boldsymbol{A}$ be a square matrix of order n. Let λ be an eigenvalue of $\boldsymbol{A}$. The **eigenspace** of $\boldsymbol{A}$ associated with λ is the solution space of $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$:

$$E_\lambda := \left\{\boldsymbol{x} \in \mathbb{R}^n \mid (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}\right\}.$$

Thus if $\boldsymbol{u} \in E_\lambda$ is non-zero, then $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ associated with eigenvalue λ.
Solve $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ by Gaussian elimination to get general solution, to write E_λ explicitly.

6.2 Diagonalisation

Definition 6.6. A square matrix $\boldsymbol{A}$ is **diagonalisable** if

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$$

for some invertible $\boldsymbol{P}$, diagonal $\boldsymbol{D}$; we say $\boldsymbol{P}$ diagonalises $\boldsymbol{A}$.

To compute using $\boldsymbol{A}$, it is convenient to write

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D} \iff \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}.$$

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be diagonalisable matrices of the same order.

- $\boldsymbol{A}^\top$ is diagonalisable:
$$\boldsymbol{A}^\top = (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1})^\top = (\boldsymbol{P}^\top)^{-1}\boldsymbol{D}^\top\boldsymbol{P}^\top.$$
- $\boldsymbol{A}^2$ is diagonalisable:
$$\boldsymbol{A}^2 = (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1})(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}) = \boldsymbol{P}\boldsymbol{D}^2\boldsymbol{P}^{-1}.$$

This makes it easy to compute powers of $\boldsymbol{A}$.

- If the same invertible $\boldsymbol{P}$ diagonalises both $\boldsymbol{A}$ and $\boldsymbol{B}$, then $\boldsymbol{AB}$ is diagonalisable:

$$\boldsymbol{AB} = (\boldsymbol{P}\boldsymbol{D}_A\boldsymbol{P}^{-1})(\boldsymbol{P}\boldsymbol{D}_B\boldsymbol{P}^{-1}) = \boldsymbol{P}(\boldsymbol{D}_A\boldsymbol{D}_B)\boldsymbol{P}^{-1}.$$

- $\boldsymbol{I} + \boldsymbol{A}$ is diagonalisable:

$$\boldsymbol{I} + \boldsymbol{A} = \boldsymbol{I} + \boldsymbol{PDP}^{-1} = \boldsymbol{PIP}^{-1} + \boldsymbol{PDP}^{-1} = \boldsymbol{P}(\boldsymbol{I} + \boldsymbol{D})\boldsymbol{P}^{-1}.$$

The next result provides a necessary and sufficient condition for a square matrix to be diagonalisable.

Theorem 6.7. *Let A be a square matrix of order n. Then A is diagonalisable if and only if A has n linearly independent eigenvectors.*

$\mathbb{R}^n$ needs to have a basis of eigenvectors

Proof.
$\boxed{\Rightarrow}$ Suppose $\boldsymbol{A}$ is diagonalisable. Then there exists an invertible matrix $\boldsymbol{P}$ such that

$$\boldsymbol{P}^{-1}\boldsymbol{AP} = \boldsymbol{D} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Let $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix}$, where u_i is the i-th column of $\boldsymbol{P}$. Since $\boldsymbol{AP} = \boldsymbol{PD}$, we have

$$\boldsymbol{A}\begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix}\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

so

$$\begin{pmatrix} \boldsymbol{Au}_1 & \cdots & \boldsymbol{Au}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\boldsymbol{u}_1 & \cdots & \lambda_n\boldsymbol{u}_n \end{pmatrix}.$$

Thus $\boldsymbol{Au}_i = \lambda_i\boldsymbol{u}_i$, i.e., $\boldsymbol{u}_i$ are eigenvectors of $\boldsymbol{A}$.
Since $\boldsymbol{P}$ is invertible, its columns $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ form a basis for $\mathbb{R}^n$, and thus are linearly independent.
$\boxed{\Leftarrow}$ Suppose $\boldsymbol{A}$ has n linearly independent eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ associated with eigenvalues $\lambda_1, \ldots, \lambda_n$ respectively. Then $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ is a basis for $\mathbb{R}^n$.
Define $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix}$, which is invertible, since its columns form a basis for $\mathbb{R}^n$. Then

$$\boldsymbol{AP} = \begin{pmatrix} \boldsymbol{Au}_1 & \cdots & \boldsymbol{Au}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\boldsymbol{u}_1 & \cdots & \lambda_n\boldsymbol{u}_n \end{pmatrix} = \boldsymbol{PD}$$

where $\boldsymbol{D} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$. Hence $\boldsymbol{P}^{-1}\boldsymbol{AP} = \boldsymbol{D}$, i.e., $\boldsymbol{A}$ is diagonalisable. $\square$

Algorithm 6.8. *Determine whether $\boldsymbol{A}$ is diagonalisable, and how to find $\boldsymbol{P}$.*

1. *Find all distinct eigenvalues of* $\boldsymbol{A}$ *by solving characteristic equation:*

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{r_1} \cdots (\lambda - \lambda_k)^{r_k}$$

   where $\lambda_1, \ldots, \lambda_k$ *are distinct eigenvalues of* $\boldsymbol{A}$.

2. $\boldsymbol{A}$ *is symmetric* $\iff$ *algebraic and geometric multiplicities of all eigenvalues coincide, i.e.,* $\dim(E_{\lambda_i}) = r_i$.
   - *The algebraic multiplicity of* λ_i *is the degree of the root* λ_i *in the polynomial* $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$.
   - *The geometric multiplicity of* λ_i *is the dimension of the eigenspace associated with the eigenvalue* λ_i.

3. *For each eigenvalue, find a basis for the eigenspace.*

   Put the basis vectors together to form $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{pmatrix}$.

Remark. Geometric multiplicity is bounded by algebraic multiplicity: why?

Theorem 6.9. *Let* $\boldsymbol{A}$ *be a square matrix of order n. If* $\boldsymbol{A}$ *has n distinct eigenvalues, then* $\boldsymbol{A}$ *is diagonalisable.*

Proof. If the characteristic polynomial of a square matrix $\boldsymbol{A}$ of order n can be factorised into exactly n factors

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1) \cdots (\lambda - \lambda_n),$$

then each basis S_{λ_i} has exactly one eigenvector, so $S = S_{\lambda_1} \cup \cdots \cup S_{\lambda_n}$ has n linearly independent eigenvectors that are required to form the matrix $\boldsymbol{P}$. Thus $\boldsymbol{A}$ is diagonalisable.
if algebraic multiplicity = 1, then geometric multiplicity = 1 □

Remark. The converse not true: $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ is diagonalisable but has eigenvalues 3 and 0.

6.3 Orthogonal Diagonalisation

"Orthogonally diagonalisable" is a stronger condition than "diagonalisable", where we impose an additional requirement on $\boldsymbol{P}$. Note that we can write $\boldsymbol{P}^\top$ in place of $\boldsymbol{P}^{-1}$.

Definition 6.10. A square matrix A is **orthogonally diagonalisable** if

$$\boldsymbol{P}^\top \boldsymbol{A} \boldsymbol{P} = \boldsymbol{D}$$

for some orthogonal $\boldsymbol{P}$, diagonal $\boldsymbol{D}$; we say $\boldsymbol{P}$ *orthogonally diagonalises* $\boldsymbol{A}$.

For orthogonal diagonalisation, we have a very simple, necessary and sufficient condition.

Theorem 6.11 (Spectral theorem)**.** *A square matrix is orthogonally diagonalisable if and only if it is symmetric.*

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be orthogonally diagonalisable matrices of the same order.

- $\boldsymbol{A}+\boldsymbol{B}$ is orthogonally diagonalisable
- If $\boldsymbol{AB}=\boldsymbol{BA}$, then $\boldsymbol{AB}$ is orthogonally diagonalisable
- If $\boldsymbol{u}$ and $\boldsymbol{v}$ are eigenvectors of $\boldsymbol{A}$ associated with eigenvalues λ and μ, then $\boldsymbol{u}\cdot\boldsymbol{v}=0$:

$$\lambda(\boldsymbol{u}\cdot\boldsymbol{v})=(\lambda\boldsymbol{u})^\top\boldsymbol{v}=(\boldsymbol{A}\boldsymbol{u})^\top\boldsymbol{v}=\boldsymbol{u}^\top\boldsymbol{A}^\top\boldsymbol{v}=\boldsymbol{u}^\top\boldsymbol{A}\boldsymbol{v}=\boldsymbol{u}^\top(\mu\boldsymbol{v})=\mu(\boldsymbol{u}\cdot\boldsymbol{v}).$$

Algorithm 6.12. *Given symmetric matrix $\boldsymbol{A}$, find orthogonal matrix $\boldsymbol{P}$*

1. *Find all distinct eigenvalues of $\boldsymbol{A}$.*
2. *For each eigenvalue λ_i, find a basis for the eigenspace.*

   Apply Gram-Schmidt process to convert the basis into an orthonormal basis.
3. *Put the basis vectors together to form $\boldsymbol{P}=\begin{pmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n\end{pmatrix}$.*

6.4 Quadratic Forms and Conic Sections

Definition 6.13. A **quadratic form** in n variables $x_1,\dots,x_n$ is

$$Q(x_1,\dots,x_n):=\sum_{i=1}^n\sum_{j=1}^n q_{ij}x_ix_j.$$

Expanding the summation out gives

$$\begin{aligned} q_{11}{x_1}^2+q_{12}x_1x_2+\cdots+q_{1n}x_1x_n \\ +q_{22}{x_2}^2+\cdots+q_{2n}x_2x_n \\ +\cdots\cdots \\ +q_{nn}{x_n}^2 \end{aligned}$$

Let $\boldsymbol{x}=(x_1,\dots,x_n)^\top$. Then

$$Q(x_1,\dots,x_n)=\begin{pmatrix}x_1 & x_2 & \cdots & x_n\end{pmatrix}\begin{pmatrix} q_{11} & \frac{1}{2}q_{12} & \cdots & \frac{1}{2}q_{1n} \\ \frac{1}{2}q_{12} & q_{22} & \cdots & \frac{1}{2}q_{2n} \\ \vdots & \vdots & & \vdots \\ \frac{1}{2}q_{1n} & \frac{1}{2}q_{2n} & \cdots & q_{nn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{x}.$$

Thus the quadratic form can be regarded as a mapping $Q\colon\mathbb{R}^n\to\mathbb{R}$ defined by

$$Q(\boldsymbol{x})=\boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{x}\qquad(\boldsymbol{x}\in\mathbb{R}^n).$$

Remark. Diagonal terms of $\boldsymbol{A}$ are coefficients for square terms; non-diagonal terms are half of coefficients for cross terms.

Definition 6.14. A **quadratic equation** in two variables x and y has the form

$$ax^2 + bxy + cy^2 + dx + ey = f$$

where a, b, c are not all zero.

We can rewrite the equation as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d & e \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = f.$$

Let $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\boldsymbol{A} = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} d & e \end{pmatrix}$. Then the equation becomes

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} = f.$$

The term $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ is the **quadratic form** associated with the quadratic equation.

The equation of a non-degenerated conic section is in **standard form** if it belongs to one of the following cases:

1. **Circle/ellipse**: $\dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} = 1$

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$

2. **Hyperbola**: $\dfrac{x^2}{\alpha^2} - \dfrac{y^2}{\beta^2} = 1$ or $-\dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} = 1$

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & -\frac{1}{\beta^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1 \quad \text{or} \quad \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} -\frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$

3. **Parabola**: $x^2 = \alpha y$ or $y^2 = \alpha x$

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 & -\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0 \quad \text{or} \quad \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -\alpha & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

Algorithm 6.15. *If $\boldsymbol{A}$ is not diagonal, we need to simplify the quadratic form:*

1. *Since $\boldsymbol{A}$ is symmetric, find an orthogonal matrix $\boldsymbol{P}$ that orthogonally diagonalises $\boldsymbol{A}$:*

$$\boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

2. *Let $\boldsymbol{y} = \boldsymbol{P}^\top \boldsymbol{x}$. Then the quadratic form becomes*

$$\begin{aligned}
\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{x}^\top (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top)\boldsymbol{x} \\
&= (\boldsymbol{P}^\top \boldsymbol{x})^\top \boldsymbol{D}(\boldsymbol{P}^\top \boldsymbol{x}) \\
&= \boldsymbol{y}^\top \boldsymbol{D}\boldsymbol{y} \\
&= \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\
&= \lambda_1 {y_1}^2 + \cdots + \lambda_n {y_n}^2.
\end{aligned}$$

In the case of two variables, let $\begin{pmatrix} x' \\ y' \end{pmatrix} = \boldsymbol{P}^\top \begin{pmatrix} x \\ y \end{pmatrix}$.

7 Linear Transformations

All vectors are written as column vectors.

7.1 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

Definition 7.1. A **linear transformation** is a mapping $T\colon \mathbb{R}^n \to \mathbb{R}^m$ of the form

$$T\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}.$$

The matrix $(a_{ij})_{m\times n}$ above is called the **standard matrix** for T.
If $n = m$, we call T a **linear operator** on $\mathbb{R}^n$.

Let $\boldsymbol{A}$ be the standard matrix for T, i.e. $T(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$ for all $\boldsymbol{u} \in \mathbb{R}^n$.

Theorem 7.2 (Linearity)**.** *Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.*

(i) $T(\mathbf{0}) = \mathbf{0}$;

(ii) $T(c_1\boldsymbol{u}_1 + \cdots + c_k\boldsymbol{u}_k) = c_1T(\boldsymbol{u}_1) + \cdots + c_kT(\boldsymbol{u}_k)$.

Proof. Write $T(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$, then use properties of matrix multiplication. □

Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a basis for $\mathbb{R}^n$. Any vector $\boldsymbol{v} \in \mathbb{R}^n$ can be written as

$$\boldsymbol{v} = c_1\boldsymbol{u}_1 + \cdots + c_n\boldsymbol{u}_n$$

for some $c_1, \ldots, c_n \in \mathbb{R}$. Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then

$$T(\boldsymbol{v}) = c_1T(\boldsymbol{u}_1) + \cdots + c_nT(\boldsymbol{u}_n).$$

Hence the image $T(\boldsymbol{v})$ of $\boldsymbol{v}$ is **completely determined** by the images $T(\boldsymbol{u}_1), \ldots, T(\boldsymbol{u}_n)$ of the basis vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$.
If T has standard matrix $\boldsymbol{A}$, since $T(\boldsymbol{e}_i) = \boldsymbol{A}\boldsymbol{e}_i = i$-th column of $\boldsymbol{A}$, we have

$$\boldsymbol{A} = \begin{pmatrix} T(\boldsymbol{e}_1) & \cdots & T(\boldsymbol{e}_n) \end{pmatrix}.$$

Definition 7.3. Let $S\colon \mathbb{R}^n \to \mathbb{R}^m$ and $T\colon \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations. The **composition** of T with S is a mapping from $\mathbb{R}^n \to \mathbb{R}^k$ such that

$$(T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u})) \qquad (\boldsymbol{u} \in \mathbb{R}^n).$$

Theorem 7.4. *Let $S\colon \mathbb{R}^n \to \mathbb{R}^m$ and $T\colon \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations. Then $T \circ S\colon \mathbb{R}^n \to \mathbb{R}^k$ is a linear transformation.*
Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be the standard matrices for S and T. Then the standard matrix for $T \circ S$ is $\boldsymbol{BA}$.

Proof. For all $\boldsymbol{u} \in \mathbb{R}^n$, $(T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u})) = T(\boldsymbol{Au}) = \boldsymbol{BAu}$. □

7.2 Ranges and Kernels

Definition 7.5. Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **range** of T is

$$R(T) = \left\{T(\boldsymbol{u}) \mid \boldsymbol{u} \in \mathbb{R}^n\right\}.$$

Theorem 7.6. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and $\{u_1, \ldots, u_n\}$ a basis for $\mathbb{R}^n$. Then*

$$R(T) = \operatorname{span}\{T(\boldsymbol{u}_1), \ldots, T(\boldsymbol{u}_n)\}.$$

Hence $R(T)$ is a subspace of $\mathbb{R}^n$.

Proof.
$\boxed{\subseteq}$ The image of every $\boldsymbol{v} \in \mathbb{R}^n$ under T is a linear combination of $T(\boldsymbol{u}_1)$, $\ldots$, $T(\boldsymbol{u}_n)$, so $T(\boldsymbol{v}) \in \operatorname{span}\{T(\boldsymbol{u}_1), \ldots, T(\boldsymbol{u}_n)\}$.
$\boxed{\supseteq}$ Let $\boldsymbol{v} \in \operatorname{span}\{T(\boldsymbol{u}_1), \ldots, T(\boldsymbol{u}_n)\}$. Then there exist $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$\begin{aligned}\boldsymbol{v} &= c_1 T(\boldsymbol{u}_1) + \cdots + c_n T(\boldsymbol{u}_n)\\ &= T(c_1\boldsymbol{u}_1 + \cdots + c_n\boldsymbol{u}_n)\\ &= T(\boldsymbol{w})\end{aligned}$$

where $\boldsymbol{w} \in \mathbb{R}^n$. Thus $\boldsymbol{v} \in R(T)$. □

In particular, if we choose the basis for $\mathbb{R}^n$ as the standard basis for $\mathbb{R}^n$, then

$$R(T) = \operatorname{span}\{T(\boldsymbol{e}_1), \ldots, T(\boldsymbol{e}_n)\} = \operatorname{Col}(\boldsymbol{A})$$

since $T(\boldsymbol{e}_i)$ is the i-th column of $\boldsymbol{A}$.

Definition 7.7. Let T be a linear transformation. The **rank** of T is

$$\operatorname{rank}(T) := \dim R(T).$$

If $\boldsymbol{A}$ is the standard matrix for T, then

$$\operatorname{rank}(T) = \dim R(T) = \dim \operatorname{Col}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}).$$

Definition 7.8. Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The **kernel** of T is

$$\operatorname{Ker}(T) := \left\{u \in \mathbb{R}^n \mid T(\boldsymbol{u}) = \boldsymbol{0}\right\}.$$

Theorem 7.9. *Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with standard matrix $\boldsymbol{A}$. Then*

$$\mathrm{Ker}(T) = \mathrm{NS}(\boldsymbol{A})$$

which is a subspace of $\mathbb{R}^n$.

Proof.

$$\boldsymbol{u} \in \mathrm{Ker}(T) \iff T(\boldsymbol{u}) = \boldsymbol{0} \iff \boldsymbol{A}\boldsymbol{u} = \boldsymbol{0} \iff \boldsymbol{u} \in \mathrm{NS}(\boldsymbol{A}).$$

□

Definition 7.10. Let T be a linear transformation. The **nullity** of T is

$$\mathrm{nullity}(T) := \dim \mathrm{Ker}(T).$$

If $\boldsymbol{A}$ is the standard matrix for T, then

$$\mathrm{nullity}(T) = \dim \mathrm{Ker}(T) = \dim \mathrm{NS}(\boldsymbol{A}) = \mathrm{nullity}(\boldsymbol{A}).$$

Theorem 7.11 (Rank-nullity theorem)**.** *Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then*

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = n. \tag{10}$$

Proof. Let $\boldsymbol{A}$ be the standard matrix for T. Using rank-nullity theorem for matrices,

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \mathrm{rank}(\boldsymbol{A}) + \mathrm{nullity}(\boldsymbol{A}) = n.$$

□

7.3 Geometric Linear Transformations

Given a geometric transformation $T: \mathbb{R}^n \to \mathbb{R}^n$, consider the images of standard basis vectors under the transformation: $T(\boldsymbol{e}_1), \ldots, T(\boldsymbol{e}_n)$. Then form the standard matrix

$$\begin{pmatrix} T(\boldsymbol{e}_1) & \cdots & T(\boldsymbol{e}_n) \end{pmatrix}.$$

Theorem 7.12. *In $\mathbb{R}^2$, a **scaling** $S: \mathbb{R}^2 \to \mathbb{R}^2$ along the x and y-axes by factors of λ_1 and λ_2 respectively has standard matrix*

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Note that

$$S\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix}, \quad S\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ \lambda_2 \end{pmatrix}.$$

When $\lambda_1 = \lambda_2 = \lambda$, the scaling is a **dilation** if $\lambda > 1$, and a **contraction** if $\lambda < 1$.

Suppose a linear transformation $T: \mathbb{R}^2 \to \mathbb{R}^2$ has standard matrix $\boldsymbol{A}$ that is **not diagonal, but**

diagonalisable. Then there exists an invertible $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{pmatrix}$ such that

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

for some $\lambda_1, \lambda_2 > 0$. Note that $\boldsymbol{u}_1$ is an eigenvector of $\boldsymbol{A}$ associated with eigenvalue λ_1, and $\boldsymbol{u}_2$ is an eigenvector of $\boldsymbol{A}$ associated with eigenvalue λ_2. Then

$$\begin{aligned} T(\boldsymbol{u}_1) &= \boldsymbol{A}\boldsymbol{u}_1 = \lambda_1 \boldsymbol{u}_1, \\ T(\boldsymbol{u}_2) &= \boldsymbol{A}\boldsymbol{u}_2 = \lambda_2 \boldsymbol{u}_2. \end{aligned}$$

Thus T can be regarded as a scaling that scales along the axes **in the directions of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$** by factors of λ_1 and λ_2 respectively.
Similarly, in $\mathbb{R}^3$, the standard matrices for scaling along the x, y and z-axes by factors of λ_1, λ_2 and λ_3 respectively are

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

When $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$, the scaling is a **dilation** if $\lambda > 1$, and a **contraction** if $\lambda < 1$.

Theorem 7.13. *A* ***reflection*** *$F\colon \mathbb{R}^2 \to \mathbb{R}^2$ about a line through the origin making an angle θ with x-axis has standard matrix*

$$\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$$

In particular, reflections about x-axis, y-axis, and the line $y = x$ have standard matrices

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

The formula for a reflection F can also be written as

$$F(\boldsymbol{u}) = \boldsymbol{u} - 2(\boldsymbol{u} \cdot \boldsymbol{n})\boldsymbol{n}$$

where $\boldsymbol{n} = (\sin\theta, -\cos\theta)^\top$.
Similarly, in $\mathbb{R}^3$, the standard matrices for reflections about xy-plane, xz-plane and yz-plane are

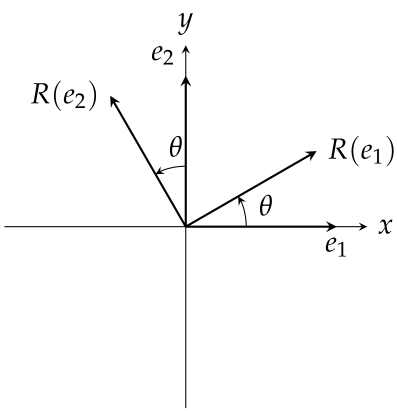
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Theorem 7.14. *An anti-clockwise* ***rotation*** *$R\colon \mathbb{R}^2 \to \mathbb{R}^2$ in $\mathbb{R}^2$ about the origin and through angle θ has standard matrix*

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Note that

$$S\left(\begin{pmatrix}1\\0\end{pmatrix}\right)=\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix},\qquad S\left(\begin{pmatrix}0\\1\end{pmatrix}\right)=\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}.$$



Similarly, in $\mathbb{R}^3$, rotation about z-axis means that $\boldsymbol{e}_3$ is fixed under rotation, $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are rotated in $\mathbb{R}^2$:

$$\begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}.$$

The standard matrices for rotations about x-axis and y-axis are

$$\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\theta & -\sin\theta\\ 0 & \sin\theta & \cos\theta\end{pmatrix},\qquad \begin{pmatrix}\cos\theta & 0 & \sin\theta\\ 0 & 1 & 0\\ -\sin\theta & 0 & \cos\theta\end{pmatrix}.$$

Theorem 7.15. *A **translation** $T\colon \mathbb{R}^2\to\mathbb{R}^2$ in $\mathbb{R}^2$ is*

$$T\left(\begin{pmatrix}x\\y\end{pmatrix}\right)=\begin{pmatrix}x+a\\y+b\end{pmatrix}.$$

Remark. Translation is not a linear transformation, since $T(\mathbf{0})\neq\mathbf{0}$.

Theorem 7.16. *A **shear** $H\colon \mathbb{R}^2\to\mathbb{R}^2$ in the x-direction by a factor of k is*

$$H\left(\begin{pmatrix}x\\y\end{pmatrix}\right)=\begin{pmatrix}x+ky\\y\end{pmatrix}=\begin{pmatrix}1 & k\\0 & 1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

Similarly, the standard matrix for the shear in the y-direction by a factor of k is

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

In $\mathbb{R}^3$, a **shear** $H' \colon \mathbb{R}^3 \to \mathbb{R}^3$ in x-direction by a factor of k_1 and in y-direction by a factor of k_2 is

$$H'\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x + k_1 z \\ y + k_2 z \\ z \end{pmatrix}.$$

Exercises

Exercise 3. Let $S \colon \mathbb{R}^n \to \mathbb{R}^m$ and $T \colon \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations. Show that

(i) $\operatorname{Ker}(S) \subseteq \operatorname{Ker}(T \circ S)$;

(ii) $R(T \circ S) \subseteq R(T)$.

Solution.

(i) Let $\boldsymbol{u} \in \operatorname{Ker}(S)$. Then $S(\boldsymbol{u}) = \boldsymbol{0}$, so $(T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u})) = T(\boldsymbol{0}) = \boldsymbol{0}$. Thus $\boldsymbol{u} \in \operatorname{Ker}(T \circ S)$.

(ii) Let $\boldsymbol{v} \in R(T \circ S)$. Then there exists $\boldsymbol{u} \in \mathbb{R}^n$ such that $\boldsymbol{v} = (T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u}))$.
Take $\boldsymbol{w} = S(\boldsymbol{u}) \in \mathbb{R}^m$; then $\boldsymbol{v} = T(\boldsymbol{w})$, so $\boldsymbol{v} \in R(T)$.

□

Exercise 4. Let $T \colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Show that

(i) if $m > n$, then $\operatorname{Ker}(T) \neq \{\boldsymbol{0}\}$;

(ii) if $m < n$, then $R(T) \neq \mathbb{R}^n$.

Solution. Use rank-nullity theorem.

(i) If $\operatorname{Ker}(T) = \{\boldsymbol{0}\}$, then $\operatorname{nullity}(T) = 0$, so $\operatorname{rank}(T) = n$.

(ii)

□

4/11 IS THE LAST LECTURE

A Selected Problems

A.1 Linear Systems and Gaussian Elimination

Exercise 5 (AY24/25)**.** Consider the following linear system, where a and b are constants.

$$\begin{cases} x - 7y + 4z &= 4 \\ -2x - 3y + 9z &= 9 \\ 7x - 15y + az &= b. \end{cases}$$

(a) Find all possible values for a such that the linear system has a unique solution for all $b \in \mathbb{R}$.

(b) Find a pair of values for (a, b) such that the linear system has infinitely many solutions. For your choice of (a, b), write down a general solution of the linear system.

(c) Find a pair of values for (a, b) such that the linear system is inconsistent.

Solution.

(a) Reduce augmented matrix to REF:

$$\left(\begin{array}{ccc|c} 1 & -7 & 4 & 4 \\ 0 & -17 & 17 & 17 \\ 0 & 0 & a+6 & b+6 \end{array}\right).$$

When $a \neq -6$, the linear system has a unique solution for all $b \in \mathbb{R}$.

(b) The linear system has infinitely many solutions when $(a, b) = (-6, -6)$. Then the RREF is

$$\left(\begin{array}{ccc|c} 1 & 0 & -3 & -3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

A general solution for the linear system is

$$\begin{cases} x = -3 + 3t \\ y = -1 + t \\ z = t \end{cases} \qquad t \in \mathbb{R}$$

(c) The linear system is inconsistent when $(a, b) = (-6, 0)$ (or any $b \neq -6$).

□

A.2 Matrices

Exercise 6 (AY24/25)**.** For $n \geq 2$, let $X_n = (x_{ij})_{n \times n}$ be a square matrix of order n defined as follows

$$x_{ij} = \begin{cases} -1 & (i \neq j) \\ n & (i = j) \end{cases}$$

(i) Write down the matrix X_2. Show that X_2 is invertible and find ${X_2}^{-1}$ using Gaussian Elimination.

(ii) Write down the matrix X_3. Show that X_3 is invertible and find X_3^{-1} using the method of adjoints.

(iii) Make use of your findings in (i) and (ii), show that X_n is always invertible for all integers $n \geq 2$. (Hint: Guess how the inverse of X_n looks like and test it.)

A.3 Vector Spaces

Exercise 7 (AY24/25)**.** Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 3 & 4 \end{pmatrix}, \qquad \boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \qquad \boldsymbol{u}_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \qquad \boldsymbol{u}_3 = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix}.$$

Note that $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$ are the columns of $\boldsymbol{A}$.

(i) Show that $\boldsymbol{b}$ belongs to $\text{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$.

(ii) Is it true that for any $\boldsymbol{c} \in \mathbb{R}^3$, if $\boldsymbol{Ax} = \boldsymbol{c}$ is consistent, then $\boldsymbol{b} + \boldsymbol{c}$ belongs to $\text{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$? Justify your answer.

Solution.

(i) To show that $\boldsymbol{b} \in \text{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, we show that the following vector equation is consistent.

$$xu_1 + yu_2 + zu_3 = b \iff \boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix} \iff Ax = b.$$

Solving the linear system

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 4 & 5 \end{array}\right).$$

(ii)

□