MA51100: Linear Algebra And Its Applications (Notes)

Summer 2021

Ziyue "Alan" Xiang

July 31, 2021

Preface

It was not until summer of 2021 that I started to attend lectures in person frequently. The long absence of classroom experience and my shaky foundations in linear algebra both contributed to the unusually strong interest in the course. Therefore, I decided to devote a serious amount of my time in compiling the notes for MA511 (Linear Algebra with Applications) at Purdue University. I wish the notes to be of high quality that it would be easy to understand whenever I need to come back for some forgotten knowledge about linear algebra. However, I did not have enough time to prove all the theorems listed in the note, especially towards the final topics of the lecture.

I was really fortunate to have prof. Rongqing Ye teaching this course. He adapted the materials aptly for students of diverse backgrounds. He was one of the best instructors I have ever seen at presenting lectures using new technologies for both on-campus and remote students. He had tremendous passion in teaching and patience for students. I wish he could continue his career in the academia, for he must be able to influence many students for years to come.

As the Delta strain emerges, it is unclear when the pandemic will be over. Purdue may cancel residential classes again in the next semester. The inability to show up in a classroom somehow reduces my efficiency to study. However, health is the number one priority, and both us and the Purdue University are making sacrifices to get through this together.

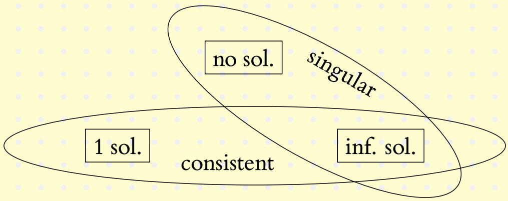
Pray for all the innocent souls that suffered in this tragic event.

Alan Xiang
July 31, 2021

Contents

1 The Basics

- A solution set of a linear system is the set of all solutions (expressed using set notation: $\{s_1, s_2, ...\}$).
- A linear system is consistent if it has at least 1 solution.
- A linear system is inconsistent if it has no solution.
- A linear system is singular if it has no solution or infinitely many solutions.
- The number of solutions to a linear system: a linear system can only have no solution, one solution or infinitely many solutions.



- Identity matrix of order n, denoted by $\mathbf{I}_n$, is an $n \times n$ matrix with 1's on the diagonal and 0's elsewhere.
- Elementary row operations:
  - Replacement: add to one row the multiple of another: $R_i \to R_i + cR_j$.
    Corresponding matrix: setting the i, j entry to c in $\mathbf{I}$ ($j < i$).
  - Interchange: interchange two rows: $R_i \leftrightarrow R_j$.
    Corresponding matrix: swapping i-th and j-th column in $\mathbf{I}$.
  - Scaling: scale one row by a nonzero scalar: $R_i \to cR_i$, $c \neq 0$.
    Corresponding matrix: setting the i, i entry to c in $\mathbf{I}$.
- Augmented matrix: a matrix where the coefficient matrix and biases is juxtaposed together. For linear system $\mathbf{Ax} = \mathbf{b}$, the augmented matrix is

$$\left[\begin{array}{c:c} \mathbf{A} & \mathbf{b} \end{array}\right]. \tag{1.1}$$

- Gaussian elimination: using row operations to transform the matrix into upper triangular form.

Example.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 1 & 2 & 2 & 11 \\ 2 & 3 & -4 & 3 \end{array}\right] \xrightarrow{R_2 \to R_2 - R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 1 & 5 \\ 2 & 3 & -4 & 3 \end{array}\right] \xrightarrow{R_3 \to R_3 - 2R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 1 & 5 \\ 0 & 1 & -6 & -9 \end{array}\right] \tag{1.2, 1.3, 1.4}$$

$$\xrightarrow{R_3 \to R_3 - 2R_1} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 1 & 5 \\ 0 & 1 & -6 & -9 \end{array}\right] \xrightarrow{R_3 \to R_2 - R_3} \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 1 & 5 \\ 0 & 0 & 7 & 14 \end{array}\right] \tag{1.5, 1.6}$$

- Pivot positions: the first nonzero entries in each row of an *upper triangular system*.
- Pivots: The nonzero numbers at pivot positions of an upper triangular system.
- ***Theorem 1.1 (Number of Solutions to a Linear System)***
  - A linear system is consistent iff. the last column of its augmented matrix does not have a pivot position.
  - If a Linear system is consistent:
    - It has exactly one solution if all columns in its coefficient matrix have pivot positions.
    - It has infinitely many solutions if some columns in its coefficient matrix have no pivot position.

  Remark.

  For the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$:
    - If each row of $\mathbf{A}$ has a pivot position, then $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b}$.
    - If each column of $\mathbf{A}$ has a pivot position, then $\mathbf{A}\mathbf{x} = \mathbf{b}$ has at most 1 solution.
- Matrix multiplication
  - Matrix-vector multiplication: let matrix $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_l] \in \mathbb{R}^{m\times l}$, where $\mathbf{A}_i$ is the i-th column of $\mathbf{A}$. Let vector $\mathbf{v} = [v_1, v_2, \ldots, v_l] \in \mathbb{R}^l$.Then the matrix-vector multiplication $\mathbf{A}\mathbf{v}$ is given by

$$\mathbf{A}\mathbf{v} = v_1\mathbf{A}_1 + v_2\mathbf{A}_2 + \cdots + v_l\mathbf{A}_l. \tag{1.7}$$

– Matrix-matrix multiplication: let matrix $\mathbf{A} \in \mathbb{R}^{m\times l}$, let matrix $\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, \dots,$ $\mathbb{R}^{l\times p}$. Then the matrix-matrix multiplication $\mathbf{AB}$ is given by

$$\mathbf{AB} = [\mathbf{AB}_1, \mathbf{AB}_2, \dots, \mathbf{AB}_p] \in \mathbb{R}^{m\times p}. \tag{1.8}$$

– Remark.

  * The a_{ij} entry of $\mathbf{AB}$ is the inner product of i-th row of $\mathbf{A}$ and j-th column of $\mathbf{B}$.
  * The j-th column of $\mathbf{AB}$ is the product of $\mathbf{A}$ and j-th column of $\mathbf{B}$. Each column of $\mathbf{AB}$ is a linear combination of columns in $\mathbf{A}$.
  * The i-th row of $\mathbf{AB}$ is the product of i-th row of $\mathbf{A}$ and $\mathbf{B}$. Each row of $\mathbf{AB}$ is a linear combination of rows in $\mathbf{B}$.

– Properties

  * (Associativity) $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) = \mathbf{ABC}$
  * (Distributivity) $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$, $(\mathbf{B}+\mathbf{C})\mathbf{D} = \mathbf{BD}+\mathbf{CD}$
  * In general, $\mathbf{AB} \neq \mathbf{BA}$
  * In general, cancellation law does not work. That is, $\mathbf{AB} = \mathbf{AC} \not\Rightarrow \mathbf{B} = \mathbf{C}$.
  * If $\mathbf{AB} = \mathbf{0}$, we cannot conclude $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.

- Triangular factorization

  – Any matrix $\mathbf{A}$ can be written as $\mathbf{PA} = \mathbf{LU}$, where $\mathbf{P}$ is a permutation matrix, $\mathbf{L}$ is a lower triangular matrix, and $\mathbf{U}$ is an upper triangular matrix. This factorization can be acquired from the Gaussian elimination process.

Example.

$$\underset{(1.9)}{\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & 3 & 4 \end{bmatrix}} \xrightarrow{R_1 \leftrightarrow R_2} \underset{(1.10)}{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 2 & 3 & 4 \end{bmatrix}} \xrightarrow{R_3 \to R_3 - 2R_1 - 3R_2} \underset{(1.11)}{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}}$$

Therefore, the **P**, **L**, **U** matrices are given by

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 1 \end{bmatrix}, \mathbf{U} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}. \tag{1.12}$$

Notice how the coefficients of **L** can be acquired directly from the Gaussian elimination process. Notice that this is only possible when each row is replaced by the rows above it.

- This factorization can also be written as $\mathbf{PA} = \mathbf{LDV}$, where $\mathbf{DV} = \mathbf{U}$. **D** is a diagonal matrix whose diagonal values are taken from **U**'s. **V** is an upper triangular matrix with each pivot standardized to 1.
- If **A** is invertible, then **LDV** is uniquely determined by **A**.
- For linear system $\mathbf{Ax} = \mathbf{b}$, we know that it is equivalent to $\mathbf{LUx} = \mathbf{b}$. Therefore, we can first solve $\mathbf{Lc} = \mathbf{b}$ and then solve $\mathbf{Ux} = \mathbf{c}$.

• Inverse

- The inverse of **A**, denoted by $\mathbf{A}^{-1}$, is a matrix such that $\mathbf{AA}^{-1} = \mathbf{I}$.
- If **A** is invertible, then $\mathbf{A}^{-1}$ is unique.
- If **A** is invertible, then $\mathbf{Ax} = \mathbf{b}$ has a solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
- If $\mathbf{Ax} = \mathbf{0}$ has a nonzero solution, then it is not invertible.
- Properties
  * $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
  * $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
  * In general, $(\mathbf{A} + \mathbf{B})^{-1} \neq \mathbf{A}^{-1} + \mathbf{B}^{-1}$
- The inverse of a matrix can be found with the Gaussian-Jordan method.

  Example.

$$\left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right] \tag{1.13}$$

$$\xrightarrow{R_2 \to R_2 + \frac{1}{2}R_1} \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right] \tag{1.14}$$

$$\xrightarrow{R_3 \to R_3 + \frac{2}{3}R_2} \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & -1 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right] \tag{1.15}$$

$$\xrightarrow{R_2 \to R_2 + \frac{3}{4}R_3} \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right] \tag{1.16}$$

$$\xrightarrow{R_1 \to R_1 + \frac{2}{3}R_2} \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & \frac{3}{2} & 1 & \frac{1}{2} \\ 0 & \frac{3}{2} & 0 & \frac{3}{4} & \frac{3}{2} & \frac{3}{4} \\ 0 & 0 & \frac{4}{3} & \frac{1}{3} & \frac{2}{3} & 1 \end{array}\right] \tag{1.17}$$

$$\xrightarrow{\text{divide by pivot}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array}\right] \tag{1.18}$$

- Transpose
  - The transpose of $\mathbf{A}$, denoted by $\mathbf{A}^T$, is a matrix such that $(\mathbf{A}^T)_{ij} = (\mathbf{A})_{ji}$.
  - Properties
    * $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
    * $(\mathbf{A}^T)^T = \mathbf{A}$
    * $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
    * $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
- Symmetric matrix
  - A matrix $\mathbf{A}$ is symmetric if $\mathbf{A}^T = \mathbf{A}$.
  - Symmetric matrices are square matrices.
  - For any matrix $\mathbf{A}$, $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$ are symmetric.
  - If a symmetric matrix $\mathbf{A}$ is factorized into $\mathbf{A} = \mathbf{LDV}$, then $\mathbf{V} = \mathbf{L}^T$.

2 *Vector Spaces*

- A vector space is an *nonempty* set V of objects called vectors, on which two operations are defined: addition and multiplication by scalars. The set V subjects to the ten axioms below, which must hold true for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and for all scalars $c, d \in \mathbb{R}$:

1. $\mathbf{u} + \mathbf{v} \in V$
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
4. There is a zero vector $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$
5. For each $\mathbf{u} \in V$, there is a vector $-\mathbf{u} \in V$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
6. $\forall c, c\mathbf{u} \in V$
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
8. $c(d\mathbf{u}) = (cd)\mathbf{u}$
9. $1\mathbf{u} = \mathbf{u}$

- A subspace of a vector space is a nonempty set H satisfying:
  - $\forall \mathbf{x}, \mathbf{y} \in H, \mathbf{x} + \mathbf{y} \in H$ (closed under addition)
  - $\forall \mathbf{x} \in H$, $c \in \mathbb{R}$, $c\mathbf{x} \in H$ (closed under scalar multiplication)

  Remark.

  - If we can prove a subspace satisfies the two conditions, it is automatically a vector space.
  - There are two trivial subspaces for any vector space V: $H = V$ and $H = \{\mathbf{0}\}$.

- Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ be vectors in V. The spanning set of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ is the set of all linear combinations of them. That is,

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1, c_2\mathbf{v}_2, \dots, c_k\mathbf{v}_k, \forall c_i \in \mathbb{R}\}. \tag{2.1}$$

- A spanning set is a subspace.
- The column space of $\mathbf{A}$, denoted by $C(\mathbf{A})$, is defined to be the spanning set of its columns.

  Remark.

  - $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent iff. $\mathbf{b} \in C(\mathbf{A})$.

- The nullspace of $\mathbf{A}$, denoted by $N(\mathbf{A})$, is the solution set of $\mathbf{A}\mathbf{x} = \mathbf{0}$.

- A linear system is homogeneous if $\mathbf{b} = \mathbf{0}$. It is inhomogeneous if $\mathbf{b} \neq \mathbf{0}$.

  Remark.

  – The solution set of $\mathbf{A}\mathbf{x} = \mathbf{0}$ is $N(\mathbf{A})$.

  – A homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$ is always consistent.

- An upper triangular matrix $\mathbf{U}$ is an echelon matrix if:

  1. The pivots are the first nonzero entries in their rows.
  2. Below each pivot are all zeros.
  3. Each pivot lies to the right of the pivot in the row above.

  It is further called a row reduced echelon matrix (form) (RREF) if it further satisfies:

  1. Each pivot is 1.
  2. Each pivot is the only nonzero entry in its column.

- ***Theorem 2.1 (Transformation to Echelon Matrix)***

  Any matrix $\mathbf{A}$ can be transformed into an echelon matrix $\mathbf{U}$ by a sequence of elementary row operations.

  Remark.

  – Such echelon matrix $\mathbf{U}$ is called an echelon form of $\mathbf{A}$.

  – There are infinitely many echelon forms of a *nonzero* matrix, but there is only a unique reduced echelon form.

  – Variables on the pivot columns are called pivot variables. Variables on non-pivot columns are called free variables.

- A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are linearly independent if the linear system $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_k\mathbf{v}_k = \mathbf{0}$ only has the trivial solution $\mathbf{v}_1 = \mathbf{v}_2 = \cdots = \mathbf{v}_k = \mathbf{0}$.

- ***Theorem 2.2 (Linear Independence)***

  The following statements are equivalent. That is, they are either all true or all false.

  – The columns of $\mathbf{A}$ are linearly independent.

- $N(\mathbf{A}) = \{\mathbf{0}\}$
- Each column of $\mathbf{A}$ has a pivot position. If $\mathbf{A}$ is a square matrix, then also $\mathbf{A}$ is invertible.

- A basis of a vector space V is a set of vectors $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ such that:
  1. B is linearly independent
  2. $V = \text{span}(B)$

  Remark.

  - In general, the columns of $\mathbf{I}_n$ forms the standard basis of $\mathbb{R}^n$.
  - A basis is a maximum independent set. That is, any linearly independent set in V can be extended to a basis, by adding more vectors if necessary.
  - A basis is a minimal spanning set. That is, any spanning set in V can be reduced to a basis, by discarding vectors if necessary.

- ***Theorem 2.3 (Basis and Invertibility)***

  If an $n \times n$ matrix $\mathbf{A}$ is intertible, then its columns form a basis for $\mathbb{R}^n$.

- Any two basis of a vector space V must have the same number of vectors. This number is called the dimension of V.

- The four fundamental subspaces

  Let $\mathbf{A}$ be an $m \times l$ matrix. The four fundamental subspaces associated to $\mathbf{A}$ are:

  1. The column space of $\mathbf{A}$, $C(\mathbf{A})$, which is the spanning set of the columns of $\mathbf{A}$. The rank of $\mathbf{A}$ is $\dim C(\mathbf{A})$.
  2. The nullspace of $\mathbf{A}$, $N(\mathbf{A})$, which is the solution set of $\mathbf{A}\mathbf{x} = \mathbf{0}$. The nullity of $\mathbf{A}$ is $\dim N(\mathbf{A})$.
  3. The row space of $\mathbf{A}$, which is the column space of $\mathbf{A}^T$. It is the spanning set of rows of $\mathbf{A}$.
  4. The left nullspace of $\mathbf{A}$, which is the nullspace of $\mathbf{A}^T$. It is the solution set of $\mathbf{x}\mathbf{A} = \mathbf{0}$.

- ***Theorem 2.4 (Finding Basis of Column Space)***

  The pivot columns of $\mathbf{A}$ form a basis for $C(\mathbf{A})$.

Example.

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 4 & 6 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{R_2 \to R_2 - R_1} \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{R_3 \to R_3 - R_2} \begin{bmatrix} 0 & \boxed{1} & 2 & 3 & 4 \\ 0 & 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2.2)\ (2.3)\ (2.4)$$

Therefore, a basis of the column space is

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \right\}. \tag{2.5}$$

Notice that the vectors must be taken from the original matrix, instead of the upper triangular matrix after row operations.

- **Theorem 2.5 (Finding Basis of Nullspace)**

The vectors associated to the free variables of a parametric form of solutions to $\mathbf{Ax} = \mathbf{0}$ form a basis of $N(\mathbf{A})$.

Example.

First of all, devise the parametric form of solutions, express the pivot variables in terms of free variables. Then, find the vectors associated with the free variables.

$$\begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \end{bmatrix} \xrightarrow{R_2 \to R_2 + 2R_1} \begin{bmatrix} \boxed{1} & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.6)\ (2.7)$$

It can be seen that x_1 is a pivot variable; x_2, x_3 are free variables. Therefore, we have

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2x_2 + -3x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}. \tag{2.8}$$

It can be seen that a basis for $N(\mathbf{A})$ is

$$\left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}. \right\} \tag{2.9}$$

Example.

Now, we have a inhomogeneous system.

$$\overset{(2.10)}{\left[\begin{array}{cccc|c} 1 & 0 & -1 & 2 & -5 \\ 3 & 1 & 2 & 4 & -2 \\ 1 & 1 & 4 & 0 & 8 \end{array}\right]} \xrightarrow[R_3 \to R_3 - R_1]{R_2 \to R_2 - 3R_1} \overset{(2.11)}{\left[\begin{array}{cccc|c} 1 & 0 & -1 & 2 & -5 \\ 0 & 1 & 5 & -2 & 13 \\ 0 & 1 & 5 & -2 & 13 \end{array}\right]}$$

$$\xrightarrow{R_3 \to R_3 - R_2} \overset{(2.12)}{\left[\begin{array}{cccc|c} \boxed{1} & 0 & -1 & 2 & -5 \\ 0 & \boxed{1} & 5 & -2 & 13 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]}$$

It can be seen that x_1, x_2 are pivot variables, x_3, x_4 are free variables. Therefore, we can write

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -5 + x_3 - 2x_4 \\ 13 - 5x_3 + 2x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -5 \\ 13 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -5 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 2 \\ 0 \\ 1 \end{bmatrix}. \tag{2.13}$$

It can be seen that the solution can be written as $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$, where $\mathbf{x}_n \in N(\mathbf{A})$. $\mathbf{x}_p$ is known as the particular solution.

- Dimensionality, nullity, rank

  Suppose $\mathbf{A} \in \mathbb{R}^{m \times l}$.

  - $\operatorname{rank} \mathbf{A} = \dim C(\mathbf{A})$ = number of pivot columns
  - $\operatorname{null} \mathbf{A} = \dim N(\mathbf{A})$ = number of non-pivot columns
  - $\operatorname{rank} \mathbf{A} \leq \min(m, l)$
  - $\operatorname{rank} \mathbf{A} + \operatorname{null} \mathbf{A} = l$
  - $\operatorname{rank} \mathbf{A} = \operatorname{rank} \mathbf{A}^T$
  - If $\mathbf{A}$ is square $(m = l)$, then $\operatorname{rank} \mathbf{A} = m \Leftrightarrow \mathbf{A}$ is invertible

Remark.

– In general, null $\mathbf{A} \neq$ null $\mathbf{A}^T$. The equality holds only when $\mathbf{A}$ is a square matrix.

- ***Theorem 2.6 (Finding Basis of Row space)***
  The elementary row operations do not alter the row space.Therefore, the nonzero rows of an echelon form can serve as a basis for the row space.

  Example.

$$\begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -4 \end{bmatrix} \tag{2.14} \xrightarrow{R_2 \to R_2 + 2R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \tag{2.15}$$

  Therefore, a basis for the row space is

$$\left\{ \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 2 \end{bmatrix} \right\}. \tag{2.16}$$

- The geometry of four fundamental subspaces

  – $N(\mathbf{A})$ is perpendicular to $C(\mathbf{A}^T)$

  Proof. Suppose $\mathbf{A}^T\mathbf{y} = \mathbf{b}$ for some $\mathbf{y}$, that is, $\mathbf{b} \in C(\mathbf{A}^T)$; let $\mathbf{v} \in N(\mathbf{A})$, that is, $\mathbf{A}\mathbf{v} = \mathbf{0}$. It can be seen that $\mathbf{b}^T\mathbf{v} = (\mathbf{A}^T\mathbf{y})^T\mathbf{v} = \mathbf{y}^T\mathbf{A}\mathbf{v} = \mathbf{0}$.

  – $N(\mathbf{A}^T)$ is perpendicular to $C(\mathbf{A})$

  Proof. Suppose $\mathbf{A}\mathbf{y} = \mathbf{b}$ for some $\mathbf{y}$, that is, $\mathbf{b} \in C(\mathbf{A})$; let $\mathbf{v} \in N(\mathbf{A}^T)$, that is, $\mathbf{A}^T\mathbf{v} = \mathbf{0}$. It can be seen that $\mathbf{b}^T\mathbf{v} = (\mathbf{A}\mathbf{y})^T\mathbf{v} = \mathbf{y}^T\mathbf{A}^T\mathbf{v} = \mathbf{0}$.

- A transformation T from $\mathbb{R}^l$ to $\mathbb{R}^m$ is a rule that assigns to each $\mathbf{x} \in \mathbb{R}^l$ a vector $T(\mathbf{x}) \in \mathbb{R}^m$.

  Remark. Let $\mathbf{A} \in \mathbb{R}^{m \times l}$. Then $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is a transformation.

- A transformation $T : \mathbb{R}^l \to \mathbb{R}^m$ is linear if:

  1. $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$
  2. $T(c\mathbf{x}) = cT(\mathbf{x})$

  Remark. Matrix transformations are linear.

- The standard basis for $\mathbb{R}^l$ is $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}^l\}$, which are columns of $\mathbf{I}_l$. If $\mathbf{x} = (x_1, x_2, \dots, x_l)$, then $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_k\mathbf{e}_k$

- **Theorem 2.7 (Linear Transformation Matrix)**
  Let $T : \mathbb{R}^l \to \mathbb{R}^m$ be a linear transformation. Then there is a matrix $\mathbf{A}$, called the standard matrix of T, such that $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Moreover,

$$\mathbf{A} = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_l) \end{bmatrix}. \tag{2.17}$$

Example. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation given by

$$T\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, T\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \tag{2.18}$$

Derive the standard matrix $\mathbf{A}$ of T.
Because T is a linear transformation, we can write $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Therefore, we have

$$\mathbf{A}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}. \tag{2.19}$$

It can be seen that

$$\mathbf{A} = \mathbf{A}\mathbf{I} = \mathbf{A}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}^{-1}. \tag{2.20}$$

- **Theorem 2.8 (Composition of Linear Transformations)**
  Suppose $T : \mathbb{R}^l \to \mathbb{R}^p$ is linear with standard matrix $\mathbf{A}$, and $G : \mathbb{R}^p \to \mathbb{R}^m$ is linear with standard matrix $\mathbf{B}$. Then the standard matrix for $G \circ T : \mathbb{R}^l \xrightarrow{T} \mathbb{R}^p \xrightarrow{G} \mathbb{R}^m$ is $\mathbf{BA}$.

- Let $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Then any vector $\mathbf{v}$ in V can be uniquely written as $\mathbf{v} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n$. $[\mathbf{v}]_B = [x_1, x_2, \ldots, x_n]^T$ is called the B-coordinate of $\mathbf{v}$.

Example. Let $\mathbb{P}_2 = \{\text{all polynomials of degree} \le 2\}$. Let $L = \{t+1, t-1, t^2-1\}$. Prove that L is a basis for $\mathbb{P}_2$. Derive the L-coordinate for $1 + 2t + 3t^2$.
A polynomial in $\mathbb{P}_2$ can be written as $c + bt + at^2$. Its coordinate under the

standard basis is $[c, b, a]^T$. Putting the three coordinate vectors of L in one matrix, we have

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_2 \to R_2 - R_1} \begin{bmatrix} \boxed{1} & -1 & 0 \\ 0 & \boxed{2} & 1 \\ 0 & 0 & \boxed{1} \end{bmatrix}$$

(2.21) (2.22)

Because there are three pivots, we know that the vectors in L are linearly independent. Because the dimension of $\mathbb{P}_2$ is 3, we know that L is a basis of $\mathbb{P}_2$. Denote the L-coordinate of $1 + 2t + 3t^2$ by $\mathbf{x}$. We know that

$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}. \tag{2.23}$$

Therefore, we have

$$\mathbf{x} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}. \tag{2.24}$$

Example. Let $\mathbb{P}_2 = \{\text{all polynomials of degree} \le 2\}$. Let $W = \{p \in \mathbb{P}_2 : p(1) = p'(1) = 0\}$. Show that W is a subspace of $\mathbb{P}_2$ and find a basis for W.

Suppose $p_1, p_2 \in W$. It can be seen that $(p_1 + p_2)(1) = p_1(1) + p_2(1) = 0$, $(p_1 + p_2)'(1) = p_1'(1) + p_2'(1) = 0$ that is, $p_1 + p_2 \in W$; $\forall c \in \mathbb{R}$, $cp_1(1) = 0$, $cp_1'(1) = 0$, that is, $cp_1 \in W$. As a result, W is a subspace of $\mathbb{P}_2$. A polynomial in $\mathbb{P}_2$ can be written as $c + bt + at^2$. Now, $p(1) = p'(1)$ is equivalent to the condition

$$\begin{cases} a + b + c = 0 \\ 2a + b = 0 \end{cases}. \tag{2.25}$$

The corresponding matrix of this homogeneous system is

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 0 \end{bmatrix}. \tag{2.26}$$

Finding a basis of the homogeneous system is essentially finding a basis for the nullspace of the matrix. Therefore, we have

$$\underset{(2.27)}{\begin{bmatrix}1 & 1 & 1\\ 2 & 1 & 0\end{bmatrix}} \xrightarrow{R_2 \to R_2 - 2R_1} \underset{(2.28)}{\begin{bmatrix}1 & 1 & 1\\ 0 & -1 & -2\end{bmatrix}} \xrightarrow{R_1 \to R_1 + R_2} \underset{(2.29)}{\begin{bmatrix}1 & 1 & 1\\ 0 & -1 & -2\end{bmatrix}} \xrightarrow{R_1 \to R_1 + R_2} \underset{(2.30)}{\begin{bmatrix}\boxed{1} & 0 & -1\\ 0 & \boxed{-1} & -2\end{bmatrix}}$$

There is only one free variable c. It can be seen that

$$\begin{bmatrix}a\\ b\\ c\end{bmatrix} = \begin{bmatrix}c\\ -2c\\ c\end{bmatrix} = c\begin{bmatrix}1\\ -2\\ 1\end{bmatrix}. \tag{2.31}$$

Therefore, a basis for W is $1 - 2t + t^2$. (*Remember to write the basis in polynomial form.*)

Example. Let $T : \mathbb{P}_2 \to \mathbb{P}_2$. Defined by $T(p) = p(0)t + p(1)t^2$. Let B be the standard basis of $\mathbb{P}_2$. Find a B-matrix $\mathbf{A}$ to represent T. Let L be $\{t - 1, t + 1, t^2 - 1\}$. Find an L-matrix representation of $\mathbf{A}$.

A polynomial in $\mathbb{P}_2$ can be written as $c + bt + at^2$. Therefore, we have $T(p) = ct + (a + b + c)t^2$. As a result, the B-matrix of $\mathbf{A}$ is given by

$$\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ 1 & 1 & 1\end{bmatrix}. \tag{2.32}$$

The basis of L is given by

$$\mathbf{C} = \begin{bmatrix}-1 & 1 & -1\\ 1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}. \tag{2.33}$$

Suppose vector $\mathbf{y}$ is in L-coordinate. In order to apply $\mathbf{A}$ to $\mathbf{y}$ in L-coordinate, we need to transform $\mathbf{y}$ back to B-coordinate first, apply $\mathbf{A}$, and then transform the result back to L-coordinate. Therefore, the corresponding transform matrix is

$$\mathbf{C}^{-1}\mathbf{A}\mathbf{C} = \begin{bmatrix}-1 & 1 & -1\\ 1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}^{-1} \begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ 1 & 1 & 1\end{bmatrix} \begin{bmatrix}-1 & 1 & -1\\ 1 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} \tag{2.34}$$

$$= \begin{bmatrix} 0 & -1 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 2 & 0 \end{bmatrix}. \tag{2.35}$$

3 Orthogonality

- An inner product on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, associates a real number $< \mathbf{u}, \mathbf{v} >$ that satisfies:

  1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
  2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
  3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
  4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$

  Remark. Only vector spaces with inner products have geometry.

- Geometry on the inner product space
  - The length of $\mathbf{v} \in V$ is $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$
  - Let the angle between $\mathbf{v}$ and $\mathbf{w}$ be θ, then $\langle \mathbf{v}, \mathbf{w} \rangle = ||\mathbf{v}||\ ||\mathbf{w}|| \cos\theta$

- Two vectors $\mathbf{v}$ and $\mathbf{w}$ are orthogonal if $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{0}$.

- **Theorem 3.1 (Orthogonality and Linear Independence)**

  Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ be mutually orthogonal nonzero vectors, then they are linearly independent.

  Proof. Apply the dot product of $\mathbf{v}_i$ to both sides of $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots, c_k\mathbf{v}_k = \mathbf{0}$, we get $c_i\mathbf{v}_i^T\mathbf{v}_i = 0$. Because $\mathbf{v}_i$'s are nonzero, it must be that $c_i = 0$.

- If the vectors in a basis of V are mutually orthogonal, it is called an orthogonal basis. If each vector in the orthogonal matrix has unit length, it is called an orthonormal basis.

- Let V be a vector space with an inner product. Two subspaces G and H of V are orthogonal (written as $G \perp V$) if $\forall\, \mathbf{g} \in G$, $\forall\, \mathbf{h} \in H$, $\langle \mathbf{g}, \mathbf{h} \rangle = 0$.

- **Theorem 3.2 (Orthogonality of Matrix Subspaces)**

Let $\mathbf{A} \in \mathbb{R}^{m \times l}$

– $C(\mathbf{A}^T) \perp N(\mathbf{A})$

– $C(\mathbf{A}) \perp N(\mathbf{A}^T)$

- Let V be a vector with an inner product. Given a subspace H of V, the space of all vectors orthogonal to H is called orthogonal complement of H. It is denoted by $H^\perp$.

Remark.

– $\dim H + \dim H^\perp = \dim V$

– $(H^\perp)^\perp = H$

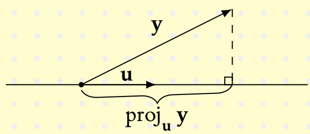
Example. Let W be a subspace in $\mathbb{R}^3$ defined by $x + 2y - 3z = 0$. Find a basis for $W^\perp$.

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, then $W = N(\mathbf{A})$. $W^\perp = N(\mathbf{A})^\perp = C(\mathbf{A}^T)$. A basis for the row space is $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$, which is the basis for $W^\perp$.

Example. $\mathbf{Ax} = \mathbf{b}$ is consistent when $\mathbf{b} \in C(\mathbf{A})$. That is, $\mathbf{b} \perp N(\mathbf{A}^T)$.

- Projection onto a vector: the projection of vector $\mathbf{y}$ onto a vector $\mathbf{u}$ is given by

$$\operatorname{proj}_{\mathbf{u}} \mathbf{y} = \frac{\langle \mathbf{u}, \mathbf{y} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u} = \frac{\mathbf{u}\mathbf{u}^T \mathbf{y}}{\mathbf{u}^T \mathbf{u}}. \tag{3.1}$$



- Projection formula: suppose $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is an orthogonal basis for W. The orthogonal projection of vector $\mathbf{y}$ onto W is given by

$$\operatorname{proj}_W \mathbf{y} = \operatorname{proj}_{\mathbf{u}_1} \mathbf{y} + \operatorname{proj}_{\mathbf{u}_2} \mathbf{y} + \cdots + \operatorname{proj}_{\mathbf{u}_k} \mathbf{y}. \tag{3.2}$$

The standard matrix of the projection operation $\mathbf{P}$ is given by

$$\mathbf{P} = \frac{\mathbf{u}_1\mathbf{u}_1^T}{\mathbf{u}_1^T\mathbf{u}_1} + \frac{\mathbf{u}_2\mathbf{u}_2^T}{\mathbf{u}_2^T\mathbf{u}_2} + \cdots + \frac{\mathbf{u}_k\mathbf{u}_k^T}{\mathbf{u}_k^T\mathbf{u}_k}. \tag{3.3}$$

Remark. It works only when B is an orthogonal basis.

- ***Theorem 3.3 (Projection, Generic)***

  Let W be a subspace of $\mathbb{R}^n$ with a basis given by columns of a matrix $\mathbf{A}$. Then the orthogonal projection matrix onto W is $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

  Proof. Suppose $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is a basis for W. Let $\mathbf{A} = \left[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\right]$, we know $W = C(\mathbf{A})$. We can write $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, where $\hat{\mathbf{y}} \in C(\mathbf{A})$, $\mathbf{z} \in C(\mathbf{A})^\perp = N(\mathbf{A}^T)$. It can be seen that $\exists \mathbf{x}, \hat{\mathbf{y}} = \mathbf{A}\mathbf{x}$; $\mathbf{A}^T\mathbf{z} = \mathbf{0}$. We need to find $\mathbf{x}$.

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z} \tag{3.4}$$

$$\Rightarrow \quad \mathbf{A}^T\mathbf{y} = \mathbf{A}^T\mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{z} \tag{3.5}$$

$$\Rightarrow \quad \mathbf{A}^T\mathbf{y} = \mathbf{A}^T\mathbf{A}\mathbf{x} \tag{3.6}$$

$$\Rightarrow \quad \mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} \tag{3.7}$$

  That is,

$$\hat{\mathbf{y}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}. \tag{3.8}$$

  Remark. If the columns of $\mathbf{A}$ are linearly independent, then $\mathbf{A}^T\mathbf{A}$ is invertible. Proof: $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = 0 \Rightarrow (\mathbf{A}\mathbf{x})^T\mathbf{A}\mathbf{x} = 0 \Rightarrow \|\mathbf{A}\mathbf{x}\|^2 = 0 \Rightarrow \mathbf{A}\mathbf{x} = \mathbf{0}$. That is to say, the nullspace of $\mathbf{A}$ and $\mathbf{A}^T\mathbf{A}$ are the same.

- The orthogonal complement of $\mathbf{y}$ onto W, denoted by $\operatorname{orth}_W \mathbf{y}$, is given by

$$\operatorname{orth}_W \mathbf{y} = \mathbf{y} - \operatorname{proj}_W \mathbf{y}. \tag{3.9}$$

  Remark.

  - Proof of $\operatorname{orth}_W \mathbf{y} \perp \mathbf{W}$: let $\mathbf{z} = \mathbf{y} - \operatorname{proj}_W \mathbf{y} = \mathbf{y} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. Multiply $\mathbf{A}^T$ to both sides, we have $\mathbf{A}^T\mathbf{z} = \mathbf{A}^T\mathbf{y} - \mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} = \mathbf{0}$. That is, $\mathbf{z} \in N(\mathbf{A}^T) \Rightarrow \mathbf{z} \perp C(\mathbf{A}) \Rightarrow \mathbf{z} \in C(\mathbf{A})^\perp$.
  - The distance between $\mathbf{y}$ and W is given by $\|\operatorname{orth}_W \mathbf{y}\|$.

Example. Let $\mathbf{A}$ be given by

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} \tag{3.10}$$

and $W = C(\mathbf{A})$. Find the projection of $\mathbf{y} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$ on W. Determine the distance between $\mathbf{y}$ and W.

Step 1: find a basis for W

$$\underset{(3.11)}{\begin{bmatrix} 1 & 1 & 2 \\ -1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix}} \xrightarrow{R_2 \to R_2 + R_1} \underset{(3.12)}{\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}} \xrightarrow{R_3 \to R_3 + R_2} \underset{(3.13)}{\begin{bmatrix} \boxed{1} & 1 & 2 \\ 0 & \boxed{1} & 1 \\ 0 & 0 & 0 \end{bmatrix}}$$

Therefore, a basis for W is given by

$$\mathbf{V} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3.14}$$

Step 2: Find the projection matrix

$$\mathbf{P} = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T \tag{3.15}$$

$$= \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \tag{3.16}$$

$$= \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}. \tag{3.17}$$

Step 3: Compute projection

$$\hat{\mathbf{y}} = \mathbf{P}\mathbf{y} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \tag{3.18}$$

Step 4: Compute orthogonal complement and distance

$$\text{orth}_W\, y = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}. \tag{3.19}$$

The distance is $\|\text{orth}_W\, \mathbf{y}\| = 2\sqrt{3}$.

- ***Theorem 3.4 (Gram-Schmidt Process)***

  Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_1 = \mathbf{x}_1 \tag{3.20}$$
$$\mathbf{v}_2 = \mathbf{x}_2 - \text{proj}_{\mathbf{v}_1} \mathbf{x}_2 \tag{3.21}$$
$$\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{\mathbf{v}_1} \mathbf{x}_3 - \text{proj}_{\mathbf{v}_2} \mathbf{x}_3 \tag{3.22}$$
$$\vdots \tag{3.23}$$
$$\mathbf{v}_p = \mathbf{x}_p - \text{proj}_{\mathbf{v}_1} \mathbf{x}_p - \text{proj}_{\mathbf{v}_2} \mathbf{x}_p - \cdots - \text{proj}_{\mathbf{v}_{p-1}} \mathbf{x}_p, \tag{3.24}$$

  then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W.

  Remark.

  - $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ for $1 \le k \le p$.
  - $\left\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \ldots, \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}$ is an orthonormal basis for W.

- If a matrix $\mathbf{Q}$ has orthonormal columns, it is called an orthogonal matrix. It can be seen that $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$, that is, $\mathbf{Q}^{-1} = \mathbf{Q}^T$.

  Remark. When $\mathbf{Q}$ is a square matrix, we have $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ as well.

- ***Theorem 3.5 (QR Factorization)***

Let $\mathbf{A}$ be an $m \times l$ matrix of rank l. That is, $l \leq m$, and columns of $\mathbf{A}$ are linearly independent. Then, there exists an $m \times l$ matrix $\mathbf{Q}$ and an $l \times l$ matrix $\mathbf{R}$ such that $\mathbf{A} = \mathbf{QR}$ satisfying:

1. $\mathbf{Q} \in \mathbb{R}^{m \times l}$ is an orthogonal matrix
2. $\mathbf{R} = \mathbf{Q}^T\mathbf{A} \in \mathbb{R}^{l \times l}$ is an upper triangular matrix with positive diagonal entries

Remark.

- We can find $\mathbf{Q}$ by applying Gram-Schmidt process to the columns of $\mathbf{A}$.
- $\mathbf{Q}$ and $\mathbf{R}$ are unique.

- A least-square solution of $\mathbf{Ax} = \mathbf{b}$ is $\hat{\mathbf{x}}$ such that

$$||\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}|| \leq ||\mathbf{b} - \mathbf{Ax}|| \tag{3.25}$$

 for all $\mathbf{x}$. The length $||\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}||$ is called the least square error of the approximation of $\mathbf{Ax} = \mathbf{b}$.

- Finding a least square solution for $\mathbf{Ax} = \mathbf{b}$

  - Idea: find the projection of $\mathbf{b}$ in $C(\mathbf{A})$, which is denoted by $\hat{\mathbf{b}}$. Solve $\mathbf{Ax} = \hat{\mathbf{b}}$ instead.

$$\mathbf{Ax} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{3.26}$$

$$\Rightarrow \mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{3.27}$$

$$\Rightarrow \mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b} \tag{3.28}$$

  - **Theorem 3.6 (Normal Equation)**
    $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ is the normal equation of $\mathbf{Ax} = \mathbf{b}$, which is *always consistent*. The solutions to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ are the least-square solutions to $\mathbf{Ax} = \mathbf{b}$.

- **Theorem 3.7 (Properties of Least Squares)**
  Let $\mathbf{A}$ be an $m \times l$ matrix. The following statements are equivalent:

  - $\mathbf{Ax} = \mathbf{b}$ has an unique least-square solution for each $\mathbf{b} \in \mathbb{R}^n$.
  - The columns of $\mathbf{A}$ are linearly independent.
  - The matrix $\mathbf{A}^T\mathbf{A}$ is invertible.

When these statements are true, the least-square solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \tag{3.29}$$

- Least squares and QR factorization: let $\mathbf{A}$ be an $m \times l$ matrix with linearly independent columns. $\mathbf{A}$ has QR factorization $\mathbf{A} = \mathbf{QR}$. The least square solution of $\mathbf{Ax} = \mathbf{b}$ is given by $\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b}$.

  Remark. In practice, an easier way to get the solution is to solve the upper triangular system $\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^T\mathbf{b}$.

- Least squares and polynomial approximation

  Example. Let $V = C[0,1]$ and $\langle f,g\rangle = \int_0^1 f(t)g(t)dt$ be an inner product on V. Let $W = \text{span}\{p_1,p_2,p_3\}$, where

$$p_1 = 1 \tag{3.30}$$
$$p_2 = 2t - 1 \tag{3.31}$$
$$p_3 = 12t^2. \tag{3.32}$$

  Find an orthogonal basis of W.

  It can be seen that:

$$q_1 = p_1 = 1; \tag{3.33}$$

$$\begin{aligned} q_2 &= p_2 - \text{proj}_{q_1} p_2 &(3.34)\\ &= p_2 - \frac{\langle q_1,p_2\rangle}{\langle q_1,q_1\rangle} q_1 &(3.35)\\ &= (2t-1) - \frac{\int_0^1 2t - 1dt}{\int_0^1 1dt}\cdot 1 &(3.36)\\ &= 2t-1. &(3.37)\end{aligned}$$

$$\begin{aligned} q_3 &= q_3 - \text{proj}_{q_1} p_3 - \text{proj}_{q_2} p_3 &(3.38)\\ &= (12t^2) - \frac{\langle q_1,p_3\rangle}{\langle q_1,q_1\rangle} q_1 - \frac{\langle q_2,p_3\rangle}{\langle q_2,q_2\rangle} q_2 &(3.39)\end{aligned}$$

$$= (12t^2) - \frac{\int_0^1 12t^2 dt}{\int_0^1 1 dt} \cdot 1 - \frac{\int_0^1 (2t-1)12t^2 dt}{\int_0^1 (2t-1)(2t-1)dt} \cdot (2t-1) \tag{3.40}$$

$$= (12t^2) - 4 - 6(2t-1) = 12t^2 - 12t + 2. \tag{3.41}$$

Example. Let $V = \mathbb{P}_4$ with inner product given by evaluation at -2, 1, 0, 1, 2. Let $W = \mathbb{P}_2$ be a subspace of V. Find a, b, c such that $a + bt + ct^2$ is closest to $5 - \frac{1}{2}t^4$.

We can write our problem as

$$\underbrace{\begin{bmatrix} 1 & t & t^2 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\mathbf{x}} = \underbrace{5 - \frac{1}{2}t^4}_{\mathbf{b}}. \tag{3.42}$$

The normal equation is $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, which is given by

$$\begin{bmatrix} \langle 1,1\rangle & \langle 1,t\rangle & \langle 1,t^2\rangle \\ \langle t,1\rangle & \langle t,t\rangle & \langle t,t^2\rangle \\ \langle t^2,1\rangle & \langle t^2,t\rangle & \langle t^2,t^2\rangle \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \langle 1, 5 - \frac{1}{2}t^4\rangle \\ \langle t, 5 - \frac{1}{2}t^4\rangle \\ \langle t^2, 5 - \frac{1}{2}t^4\rangle \end{bmatrix}, \tag{3.43}$$

where

$$1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, t = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, t^2 = \begin{bmatrix} 4 \\ 1 \\ 0 \\ 1 \\ 4 \end{bmatrix}, 5 - \frac{1}{2}t^4 = \begin{bmatrix} -3 \\ \frac{9}{2} \\ 5 \\ \frac{9}{2} \\ -3 \end{bmatrix}. \tag{3.44}$$

After the computation of inner product, we get

$$\begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 8 \\ 0 \\ -15 \end{bmatrix}. \tag{3.45}$$

It can be seen that a least square solution is

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}^{-1} \begin{bmatrix} 8 \\ 0 \\ -15 \end{bmatrix} \tag{3.46}$$

$$= \begin{bmatrix} \frac{17}{35} & 0 & -\frac{1}{7} \\ 0 & \frac{1}{10} & 0 \\ -\frac{1}{7} & 0 & \frac{1}{14} \end{bmatrix} \begin{bmatrix} 8 \\ 0 \\ -15 \end{bmatrix} = \begin{bmatrix} \frac{211}{35} \\ 0 \\ -\frac{31}{14} \end{bmatrix}. \tag{3.47}$$

Example. Let $V = C[0,1]$ and $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ be an inner product on V. Find $a + bx$ that is closest to x^5 in V.

We can write our problem as

$$\underbrace{\begin{bmatrix} 1 & t \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\mathbf{x}} = \underbrace{x^5}_{\mathbf{b}}. \tag{3.48}$$

The normal equation is $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, which is given by

$$\begin{bmatrix} \langle 1,1 \rangle & \langle 1,x \rangle \\ \langle x,1 \rangle & \langle x,x \rangle \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \langle 1,x^5 \rangle \\ \langle t,x^5 \rangle \end{bmatrix}. \tag{3.49}$$

That is,

$$\begin{bmatrix} \int_0^1 dx & \int_0^1 xdx \\ \int_0^1 xdx & \int_0^1 x^2dx \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \int_0^1 x^5dx \\ \int_0^1 x^6dx \end{bmatrix} \tag{3.50}$$

$$\Rightarrow \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{1}{6} \\ \frac{1}{7} \end{bmatrix}. \tag{3.51}$$

Therefore, we have

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{6} \\ \frac{1}{7} \end{bmatrix} = \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \begin{bmatrix} \frac{1}{6} \\ \frac{1}{7} \end{bmatrix} = \begin{bmatrix} -\frac{4}{21} \\ \frac{5}{7} \end{bmatrix}. \tag{3.52}$$

- Complex number

- A complex number can be written as $z = a + bi$, where $a, b \in \mathbb{R}$, and i is the imaginary number unit.
- Notations: suppose $z = a + bi$
  * $\Re(z) = a$, the real part
  * $\Im(z) = b$, the imaginary part
  * $\bar{z} = a - bi$, the complex conjugate of z
  * $\mathbb{C}$: the set of all complex numbers
  * $|z| = \sqrt{a^2 + b^2}$, the absolute value; $|z|^2 = z\bar{z}$
  * $\arg(z) = \arctan \frac{b}{a}$
- Polar form: a complex number can be written as $z = a + bi = r\cos\phi + ir\sin\phi$, where $r = |z|$, $\phi = \arg(z)$.

• Given a sequence of numbers $y_0, y_1, \ldots, y_{n-1}$. Suppose these numbers are sampled from a function $f(t)$ at $t = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}$. The Fourier transform of $f(t)$ is given by

$$F(t) = c_0 + c_1 e^{2\pi i \cdot t} + c_2 e^{2\pi i \cdot 2t} + \cdots + c_{n-1} e^{2\pi i \cdot (n-1)t} \tag{3.53}$$

Remark. Only n terms are enough. Consider the $kn + l$ term, where $k \geq 1, 0 \leq l < n$. We have

$$e^{2\pi i \cdot (kn+l)t} = e^{2\pi i \cdot knt} e^{2\pi i \cdot lt} = e^{2\pi i \cdot lt}. \tag{3.54}$$

That is, terms beyond n will be equal to the first n terms.

- Derivation of Discrete Fourier Transform

  In order to acquire the coefficients c_k, consider $F\left(\frac{1}{n}\right)$, which is given by

$$F\left(\frac{1}{n}\right) = c_0 + c_1 \underbrace{e^{2\pi i \cdot \frac{1}{n}}}_{w} + c_2 \underbrace{e^{2\pi i \cdot \frac{2}{n}}}_{w^2} + \cdots + c_{n-1} \underbrace{e^{2\pi i \cdot \frac{n-1}{n}}}_{w^{n-1}}. \tag{3.55}$$

Putting $t = 0, \frac{1}{n}, \dots, \frac{n-1}{n}$ into the above equation:

$$\begin{aligned} c_0 + c_1 \quad &+ c_2 &&+ \cdots + c_{n-1} &&= y_0 \\ c_0 + wc_1 \quad &+ w^2c_2 &&+ \cdots + w^{n-1}c_{n-1} &&= y_1 \\ c_0 + w^2c_1 \quad &+ w^4c_2 &&+ \cdots + w^{2(n-1)}c_{n-1} &&= y_2 \\ &\vdots \\ c_0 + w^{n-1}c_1 &+ w^{2(n-1)}c_2 &&+ \cdots + w^{(n-1)^2}c_{n-1} &&= y_{n-1}, \end{aligned} \tag{3.56}$$

where $w = e^{\frac{2\pi i}{n}}$. The system above can be written in matrix form as:

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{n-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \cdots & w^{(n-1)^2} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{n-1} \end{bmatrix}}_{\mathbf{c}} = \underbrace{\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}}_{\mathbf{y}}. \tag{3.57}$$

That is, the Fourier coefficients $\mathbf{c}$ can be found by multiplying the inverse of Fourier matrix $\mathbf{F}$ and the function values $\mathbf{y}$ ($\mathbf{c} = \mathbf{F}^{-1}\mathbf{y}$).

– ***Theorem 3.8 (Properties of the Fourier Matrix)***

Let $w = e^{\frac{2\pi i}{n}}$ and $\mathbf{F}$ be the $n \times n$ Fourier matrix. Then $\bar{w} = w^{-1}$ and $\mathbf{F}\bar{\mathbf{F}} = n\mathbf{I}$. That is,

$$\mathbf{F}^{-1} = \frac{1}{n}\bar{\mathbf{F}} = \frac{1}{n}\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w^{-1} & w^{-2} & \cdots & w^{-(n-1)} \\ 1 & w^{-2} & w^{-4} & \cdots & w^{-2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{-(n-1)} & w^{-2(n-1)} & \cdots & w^{-(n-1)^2} \end{bmatrix}. \tag{3.58}$$

Proof. It can be shown that for complex number z, $z^{-1} = \frac{\hat{z}}{|z|}$. Because $|w| = 1$, we know $\bar{w} = w^{-1}$. Here are some facts:

* $w^n = 1$: $w^n = \left(e^{\frac{2\pi i}{n}}\right)^n = e^{2\pi i} = 1$.

* For any non-negative integer j, $w^{nj} = 1$.
* Let $A_w = 1 + w^j + w^{2j} + \cdots + w^{(n-1)j}$, then $A_w = n$ if $j = 0$; $A_w = 0$ if j is an positive integer.

  It is clear that $A_w = n$ when $j = 0$. When $j > 0$, A_w is the sum of a geometric series with common ratio w^j. Therefore, we have $A_w = \frac{1-(w^j)^n}{1-w^j} = 0$.

Since $\mathbf{F}$ is a symmetric matrix, its rows and columns are the same. Therefore, the k-th row of $\mathbf{F}$ is given by

$$\begin{bmatrix}1 & w^k & w^{2k} & \cdots & w^{(n-1)k}\end{bmatrix}. \tag{3.59}$$

The l-th column of $\bar{\mathbf{F}}$ is given by

$$\begin{bmatrix}1 & w^{-l} & w^{-2l} & \cdots & w^{-(n-1)l}\end{bmatrix}^T. \tag{3.60}$$

Therefore, the (k, l) element of $\mathbf{F}\bar{\mathbf{F}}$ is

$$1 + w^k w^{-l} + w^{2k} w^{-2l} + \cdots + w^{(n-1)k} w^{-(n-1)l} \tag{3.61}$$

$$= 1 + w^{(k-l)} + w^{2(k-l)} + \cdots + w^{(n-1)(k-l)} \tag{3.62}$$

$$= \begin{cases} n, & k = l \\ 0, & k \neq l \end{cases}. \tag{3.63}$$

It can be seen that $\mathbf{F}\bar{\mathbf{F}} = n\mathbf{I}$.

- Fast Fourier Transform (FFT)

Assume the Fourier matrix is of $n \times n$, where $n = 2m$. Let $w_n = e^{\frac{2\pi i}{n}}$, $\mathbf{F} = [w_n^{jk}]$, $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1})$. It can be seen that $w_n^{2kj} = \left(e^{\frac{2\pi kj}{2m}}\right)^2 = e^{\frac{2\pi kj}{m}} = w_m^{kj}$. Our goal is to compute $\mathbf{y} = \mathbf{Fc}$ efficiently. Consider the j-th component of $\mathbf{y}$:

$$\mathbf{y}_j = c_0 + w_n^j c_1 + \cdots + w_n^{j(n-1)} c_{n-1} \tag{3.64}$$

$$= \sum_{k=0}^{n-1} w_n^{jk} c_k = \sum_{k=0}^{m-1} w_n^{2kj} c_{2k} + \sum_{k=0}^{m-1} w_n^{(2k+1)j} c_{2k+1} \tag{3.65}$$

$$= \underbrace{\sum_{k=0}^{m-1} w_m^{kj} c_{2k}}_{\mathbf{y}'_j} + w_n^j \underbrace{\sum_{k=0}^{m-1} w_m^{kj} c_{2k+1}}_{\mathbf{y}''_j} \tag{3.66}$$

One can see that $\mathbf{y}'_j$ comes from the smaller sub-problem. Therefore, we can derive an efficient algorithm as follows:

Step 1: Separate $\mathbf{c}$ into even and odd numbered components:

$$\mathbf{c}' = (c_0, c_2, \dots, c_{n-2}) \tag{3.67}$$

$$\mathbf{c}'' = (c_1, c_3, \dots, c_{n-1}) \tag{3.68}$$

Step 2: Let $m = \frac{n}{2}$, compute:

$$\mathbf{y}' = \mathbf{F}_m \mathbf{c}' \tag{3.69}$$

$$\mathbf{y}'' = \mathbf{F}_m \mathbf{c}'' \tag{3.70}$$

Step 3: Merge $\mathbf{y}'$ and $\mathbf{y}''$ to get $\mathbf{y}'$, for $j = 0, 1, \dots, m-1$:

$$\mathbf{y}_j = \mathbf{y}'_j + w_n^j \mathbf{y}''_j \tag{3.71}$$

Because $\mathbf{y}_{m+j} = \mathbf{y}'_j + w_n^{m+j} \mathbf{y}''_j$, also $w_n^m = \left(e^{\frac{2\pi kj}{2m}}\right)^m = e^{\pi i} = -1$, we have

$$\mathbf{y}_{m+j} = \mathbf{y}'_j - w_n^j \mathbf{y}''_j. \tag{3.72}$$

4 Determinants

- Let $\mathbf{A}$ be an $n \times n$ matrix. Let $\mathbf{A}_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the submatrix of $\mathbf{A}$, which is formed by deleting the i-th row and j-th column of $\mathbf{A}$. The (i, j)- cofactor of $\mathbf{A}$, denoted by c_{ij}, is given by

$$c_{ij} = (-1)^{i+j} \det \mathbf{A}_{ij}. \tag{4.1}$$

- Let $\mathbf{A} = [a_{ij}]$ be an $n \times n$ matrix with $n \geq 2$. The determinant of $\mathbf{A}$, denoted by $\det \mathbf{A}$, is defined by

$$\det \mathbf{A} = a_{11}c_{11} + a_{12}c_{12} + \cdots + a_{1n}c_{1n} \tag{4.2}$$

$$= a_{11} \det \mathbf{A}_{11} - a_{12} \det \mathbf{A}_{12} + \cdots + (-1)^{n+1} a_{1n} \det \mathbf{A}_{1n}. \tag{4.3}$$

This is also known as the cofactor expansion of $\mathbf{A}$ on the first row.

Remark.

- $\det \mathbf{A}$ can be computed by cofactor expansion on any arbitrary row/column.
- The time complexity of cofactor expansion is $O(n!)$.

- **Theorem 4.1 (Determinant of Triangular Matrix)**
  If $\mathbf{A}$ is a triangular matrix, then $\det \mathbf{A}$ is the product of the diagonal entries of $\mathbf{A}$.

- **Theorem 4.2 (Determinant and Row Operations)**
  Let $\mathbf{A}$ be an $n \times n$ matrix.

  1. Replacement does not change determinants. If $\mathbf{A} \xrightarrow{R_i \to R_i + cR_j} \mathbf{B}$, then $\det \mathbf{B} = \det \mathbf{A}$.

  2. Interchange changes the signs of determinants. If $\mathbf{A} \xrightarrow{R_i \leftrightarrow R_j} \mathbf{B}$, then $\det \mathbf{B} = -\det \mathbf{A}$.

  3. Scaling scales determinants. If $\mathbf{A} \xrightarrow{R_i \to cR_i} \mathbf{B}$, then $\det \mathbf{B} = c \det \mathbf{A}$.

Proof.

3. It can be easily proved using cofactor expansion.

2. Suppose we interchange i-th and $i+1$-th row of $\mathbf{A}$, which results in $\mathbf{B}$. If we expand on the i-th row of $\mathbf{B}$, we have

$$\det \mathbf{B} = \sum_{k=1}^{n} b_{ij}(-1)^{i+j} \det \mathbf{B}_{ij} \tag{4.4}$$

$$= \sum_{k=1}^{n} a_{ij}(-1)^{i+j+1} \det \mathbf{A}_{ij} = -\det \mathbf{A}. \tag{4.5}$$

Now we need to interchange i-th and j-th row. Suppose $i < j$, we can interchange $(i, i+1), (i+1, i+2), \dots, (j-1, j)$ rows until the i-th row becomes the j-th row. Now, we have done $j - i$ interchanges, and j-th row becomes the $j-1$-th row. We keep interchanging $(j-1, j-2), (j-2, j-3), \dots, (i+1, i)$ rows until the $j-1$-th row becomes the i-th row. In total, we did $2(j-i)-1$ interchanges. Therefore, $\det \mathbf{B} = -\det \mathbf{A}$.

1. The determinant of $\mathbf{B}$ is given by

$$\det \mathbf{B} = \sum_{k=1}^{n} (a_{ik} + ca_{jk})(-1)^{i+k} \det \mathbf{A}_{ik} \tag{4.6}$$

$$= \sum_{k=1}^{n} a_{ik}(-1)^{i+k} \det \mathbf{A}_{ik} + c \sum_{k=1}^{n} a_{jk}(-1)^{i+k} \det \mathbf{A}_{ik} \tag{4.7}$$

$$= \det \mathbf{A} + c \sum_{k=1}^{n} a_{jk}(-1)^{i+k} \det \mathbf{A}_{ik}. \tag{4.8}$$

To compute the second term in the sum, we can construct a matrix $\mathbf{C}$ as follows: let $\mathbf{C} = \mathbf{A}$, then let the i-th row of $\mathbf{C}$ be equal to the j-th row of $\mathbf{C}$. Obviously, $\mathbf{C}$ is a matrix with two identical rows. If we interchange the two identical rows of $\mathbf{C}$ to get $\mathbf{C}'$, we know that $\mathbf{C} = \mathbf{C}'$. Given that $\det \mathbf{C} = -\det \mathbf{C}' = -\det \mathbf{C}$, we know $\det \mathbf{C} = 0$. As a result, $\det \mathbf{B} = \det \mathbf{C}$.

Remark. $\det(c\mathbf{A}) = c^n \det \mathbf{A}$

Thoughts.

We can also use replacement and scaling to prove interchange. Use $\mathbf{B}$ to denote the matrix after row operations. Denote the initial row i and row j by R_i' and R_j', respectively. Interchanging is equivalent to the following sequence row operations:

Step	Row Operation	Value of R_i	Value of R_j	Value of $\det \mathbf{B}$
1	$R_i \to R_i + R_j$	$R_i' + R_j'$	R_j'	$\det \mathbf{A}$
2	$R_j \to -2R_j$	$R_i' + R_j'$	$-2R_j'$	$-2 \det \mathbf{A}$
3	$R_j \to R_j + R_i$	$R_i' + R_j'$	$R_i' - R_j'$	$-2 \det \mathbf{A}$
4	$R_i \to R_i - R_j$	$2R_j'$	$R_i' - R_j'$	$-2 \det \mathbf{A}$
5	$R_i \to \frac{1}{2} R_i$	R_j'	$R_i' - R_j'$	$-\det \mathbf{A}$
6	$R_j \to R_j + R_i$	R_j'	R_i'	$-\det \mathbf{A}$

- Properties of Determinant
  - A square matrix $\mathbf{A}$ is invertible iff. $\det \mathbf{A} \neq 0$.

Proof. Use row operations to reduce the matrix to triangular form; multiply the values on the diagonal.

– $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$

Proof. If $\mathbf{A}$ is not invertible, we need to prove $\mathbf{AB}$ is not invertible. Suppose $\mathbf{B}$ is not invertible, then $\exists \mathbf{y} \neq \mathbf{0}$, $\mathbf{By} = \mathbf{ABy} = 0$. Therefore, $\mathbf{AB}$ is not invertible. Suppose $\mathbf{B}$ is invertible, because $\mathbf{A}$ is not invertible, $\exists \mathbf{x} \neq \mathbf{0}$, $\mathbf{Ax} = 0$. We can find $\mathbf{y}$ such that $\mathbf{By} = \mathbf{x} \Rightarrow \mathbf{y} = \mathbf{B}^{-1}\mathbf{x}$. In this case, $\mathbf{ABy} = \mathbf{0}$, $\mathbf{y} \neq \mathbf{0}$. Therefore, we can conclude that $\mathbf{AB}$ is not invertible. In this case, it is easy to see that $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$.

If $\mathbf{A}$ is invertible, we can write $\mathbf{A}$ as the product of a series of elementary row operation matrices

$$\mathbf{A} = \mathbf{E}_k \mathbf{E}_{k-1} \cdots \mathbf{E}_1 \mathbf{I}. \tag{4.9}$$

Since each elementary row operation will introduce a constant scale factor to the result of the determinant, we can write

$$\det \mathbf{A} = \det(\mathbf{E}_k \mathbf{E}_{k-1} \cdots \mathbf{E}_1) = e_k e_{k-1} \cdots e_1. \tag{4.10}$$

It can be seen that

$$\det \mathbf{AB} = \det(\mathbf{E}_k \mathbf{E}_{k-1} \cdots \mathbf{E}_1 \mathbf{B}) = e_k e_{k-1} \cdots e_1 \det \mathbf{B}. \tag{4.11}$$

Therefore, we know that $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$.

– $\det \mathbf{A}^T = \det \mathbf{A}$

Proof. Suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$. We can use mathematical induction to prove that $\det \mathbf{A}^T = \det \mathbf{A}$ for $n \geq 1$.

Base Case: When $n = 1, 2$, it is obvious that $\det \mathbf{A}^T = \det \mathbf{A}$.

Induction Hypothesis: Suppose $\det \mathbf{A}^T = \det \mathbf{A}$ for $n \geq 1$.

Induction Step: We need to prove $\det \mathbf{B}^T = \det \mathbf{B}$, where $\mathbf{B} \in \mathbb{R}^{(n+1)\times(n+1)}$. The cofactor expansion of the first row of $\mathbf{B}^T$ is given by

$$\det \mathbf{B}^T = \sum_{j=1}^{n+1} b_{j1} (-1)^{1+j} \det \mathbf{B}_{1j}^T \tag{4.12}$$

$$= \sum_{j=1}^{n+1} b_{j1} (-1)^{1+j} \det \mathbf{B}_{1j} = \det \mathbf{B}. \tag{4.13}$$

By mathematical induction, we can conclude that $\det \mathbf{A}^T = \det \mathbf{A}$ for $n \geq 1$.

– $\det \mathbf{A}^{-1} = \dfrac{1}{\det \mathbf{A}}$

Proof. $\det \mathbf{I} = \det(\mathbf{A}\mathbf{A}^{-1}) = \det \mathbf{A} \det \mathbf{A}^{-1} = 1$.

– $\det(\mathbf{B}^{-1}\mathbf{A}\mathbf{B}) = \det \mathbf{A}$

- ***Theorem 4.3 (Cramer's Rule)***

  Let $\mathbf{A}$ be an $\mathbf{n} \times \mathbf{n}$ invertible matrix. For any $\mathbf{b} \in \mathbb{R}^n$, the unique solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ has entries given by

$$\mathbf{x}_i = \frac{\det \mathbf{A}_i(\mathbf{b})}{\det \mathbf{A}}, \tag{4.14}$$

  where $\mathbf{A}_i(\mathbf{b})$ denotes the matrix $\mathbf{A}$ with the i-th column replaced by $\mathbf{b}$.

  Proof. Denote the i-th column of $\mathbf{A}$ by $\mathbf{a}_i$, it can be seen that

$$\begin{aligned} \mathbf{I}_i(\mathbf{x}) &= [\ \mathbf{e}_1 \ \cdots \ \mathbf{e}_{i-1} \ \mathbf{x} \ \cdots \ \mathbf{e}_n\ ], \\ \mathbf{A}\mathbf{I}_i(\mathbf{x}) &= [\ \mathbf{A}\mathbf{e}_1 \ \cdots \ \mathbf{A}\mathbf{e}_{i-1} \ \mathbf{A}\mathbf{x} \ \cdots \ \mathbf{A}\mathbf{e}_n\ ] \\ &= [\ \mathbf{a}_1 \ \cdots \ \mathbf{a}_{i-1} \ \mathbf{b} \ \cdots \ \mathbf{a}_n\ ] \\ &= \mathbf{A}_i(\mathbf{b}). \end{aligned} \tag{4.15}$$

  That is, $\det(\mathbf{A}\mathbf{I}_i(\mathbf{x})) = \det \mathbf{A}_i(\mathbf{b}) \Rightarrow \det \mathbf{A} \det \mathbf{I}_i(\mathbf{x}) = \det \mathbf{A}_i(\mathbf{b})$. Consider the value $\det \mathbf{I}_i(\mathbf{x})$, if we use cofactor expansion across the i-th column, the cofactors corresponding to elements other than $\mathbf{x}_i$ will be 0, because there will be an all zero row in the submatrix; the cofactor corresponding to $\mathbf{x}_i$ will be $(-1)^{i+i} \det \mathbf{I}_{n-1} = 1$. As a result, we know that $\det(\mathbf{A}\mathbf{I}_i(\mathbf{x})) = \mathbf{x}_i$. That is, $(\det \mathbf{A})\mathbf{x}_i = \det \mathbf{A}_i(\mathbf{b})$.

- ***Theorem 4.4 (Determinant and Inverse)***

  Let $\mathbf{A}$ be an $n \times n$ matrix. The adjugate of $\mathbf{A}$ is given by $\operatorname{adj} \mathbf{A} = \mathbf{C}^T$, where the cofactor matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}. \tag{4.16}$$

If $\mathbf{A}$ is invertible, its inverse is given by

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \operatorname{adj} \mathbf{A} = \frac{1}{\det \mathbf{A}} \mathbf{C}^T \tag{4.17}$$

Proof. Consider the matrix $\mathbf{A} \operatorname{adj} \mathbf{A} = \mathbf{A}\mathbf{C}^T$, whose (i, j) entry is given by

$$\mathbf{A}\mathbf{C}^T_{i,j} = \sum_{k=1}^{n} a_{ik} c_{jk}. \tag{4.18}$$

When $i = j$, it is the cofactor expansion of $\det \mathbf{A}$. Therefore, the diagonal elements of $\mathbf{A}\mathbf{C}^T$ equal to $\det \mathbf{A}$. When $i \neq j$, it is equivalent to the cofactor expansion of $\mathbf{A}'$, which is matrix $\mathbf{A}$ with the i-th row equal to the j-th row. It can be seen that $\det \mathbf{A}' = 0$. That is,

$$\mathbf{A}\mathbf{C}^T = \begin{bmatrix} \det \mathbf{A} & & & \\ & \det \mathbf{A} & & \\ & & \ddots & \\ & & & \det \mathbf{A} \end{bmatrix} = (\det \mathbf{A})\mathbf{I}. \tag{4.19}$$

It can be seen that $\mathbf{A}\frac{\mathbf{C}^T}{\det \mathbf{A}} = \mathbf{I}$, which implies $\frac{\mathbf{C}^T}{\det \mathbf{A}} = \mathbf{A}^{-1}$.

- ***Theorem 4.5 (Volume After Linear Transformation)***

  Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix $\mathbf{A}$. Then $\operatorname{vol}(T(\mathbf{v})) = |\det \mathbf{A}| \operatorname{vol}(\mathbf{v})$.

- The parallelpiped determined by n vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is the subset

$$P = \{a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n \mid 0 \leq a_1, a_2, \dots a_n \leq 1\}. \tag{4.20}$$

- ***Theorem 4.6 (Determinant and Volume)***

  Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be n vectors in $\mathbb{R}^n$, let P be the parallelpiped determined by these vectors, and let $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$. The volume of P is given by $\operatorname{vol}(P) = |\det \mathbf{A}|$.

Proof. A proof for the two theorems above can be found in https://textbooks.math.gatech.edu/ila/determinants-volumes.html.

5 Eigenvalues and Eigenvectors

- Let $\mathbf{A}$ be an $n \times n$ matrix. An eigenvector of $\mathbf{A}$ is a vector $\mathbf{v}$ such that
  - $\mathbf{v} \neq \mathbf{0}$
  - $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ

  A scalar λ is called an eigenvalue if $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ has a nontrivial solution; such an $\mathbf{x}$ is called an eigenvector corresponding to λ (λ-eigenvector).

  Remark. Eigenvectors cannot be zero, eigenvalues can be zero.

- Let λ be an eigenvalue of $\mathbf{A}$. The eigenspace of $\mathbf{A}$ corresponding to λ is the set of all solutions to $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, or $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$. That is, the λ-eigenspace is $N(\mathbf{A} - \lambda\mathbf{I})$.

- Finding eigenvalues

  We know that

$$\begin{aligned} &\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \text{ has a nontrivial solution} \\ \Rightarrow\ &(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \text{ has a nontrivial solution} \\ \Rightarrow\ &\mathbf{A} - \lambda\mathbf{I} \text{ is not invertible} \\ \Rightarrow\ &\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}. \end{aligned}$$

  To find eigenvalues, we only need to solve $\det(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{0}$, which is known as the characteristic polynomial.

- The sum of eigenvalues of $\mathbf{A}$ equals to the sum of diagonal entries $\mathbf{A}$. The sum of diagonal entries is known as the trace of $\mathbf{A}$, which is denoted by $\operatorname{tr}\mathbf{A}$.

- The product of eigenvalues of $\mathbf{A}$ equals to the determinant of $\mathbf{A}$.

- ***Theorem 5.1 (Eigenvalues of Triangular Matrices)***

  The eigenvalues of a triangular matrix are the entries on the main diagonal.

  Example. Compute the eigenvalues and corresponding eigenvectors for

$$\mathbf{A} = \begin{bmatrix} -1 & 3 \\ 2 & 0 \end{bmatrix}. \tag{5.1}$$

  Step 1: Find all eigenvalues by the characteristic equation.

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} -1-\lambda & 3 \\ 2 & -\lambda \end{vmatrix} = \lambda^2 + \lambda - 6 = 0. \tag{5.2}$$

It can be seen that $\lambda_1 = -3$, $\lambda_2 = 2$.

Step 2: Find a basis for $N(\mathbf{A} - \lambda \mathbf{I})$ for each λ.

For $\lambda_1 = -3$, the matrix is

$$\begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix}. \tag{5.3}$$

Transforming it to RREF:

$$\overset{(5.4)}{\begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix}} \xrightarrow{R_2 \to R_2 - R_1} \overset{(5.5)}{\begin{bmatrix} 2 & 3 \\ 0 & 0 \end{bmatrix}} \xrightarrow{R_1 \to \frac{1}{2} R_1} \overset{(5.6)}{\begin{bmatrix} \boxed{1} & \frac{3}{2} \\ 0 & 0 \end{bmatrix}}$$

It can be seen that x_2 is a free variable. Therefore, we have

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -\frac{3}{2} \\ 1 \end{bmatrix}. \tag{5.7}$$

We can scale it to get the λ_1 eigenvector: $\begin{bmatrix} -3 & 2 \end{bmatrix}^T$.

For $\lambda_2 = 2$, the matrix is

$$\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}. \tag{5.8}$$

Transforming it to RREF:

$$\overset{(5.9)}{\begin{bmatrix} -3 & 3 \\ 2 & -2 \end{bmatrix}} \xrightarrow{R_2 \to R_2 + \frac{2}{3} R_1} \overset{(5.10)}{\begin{bmatrix} -3 & 3 \\ 0 & 0 \end{bmatrix}} \xrightarrow{R_1 \to -\frac{1}{3} R_1} \overset{(5.11)}{\begin{bmatrix} \boxed{1} & -1 \\ 0 & 0 \end{bmatrix}}$$

It can be seen that x_2 is a free variable. Therefore, we have

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{5.12}$$

The λ_1 eigenvector is $\begin{bmatrix} 1 & 1 \end{bmatrix}^T$.

- ***Theorem 5.2 (Linear Independence of Eigenvectors)***

  If $\mathbf{v}_1, \dots, \mathbf{v}_r$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of an $n \times n$ matrix $\mathbf{A}$. Then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly independent.

  Proof. We can prove it by contradiction. Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is linearly dependent. That is, there exists some nonzero $c_1, \dots, c_r$ such that $c_1\mathbf{v}_1 + \dots + c_r\mathbf{v}_r = \mathbf{0}$.

  Without loss of generality, assume $c_1 \neq 0$, and λ_1 is the biggest of all eigenvectors. Because $\mathbf{A}^k\mathbf{v} = \lambda^k\mathbf{v}$, if we multiply $\mathbf{A}^k$ to the equation above, we have

$$c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_r\lambda_r^k\mathbf{v}_r = \mathbf{0} \tag{5.13}$$

$$c_1\mathbf{v}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k \mathbf{v}_2 + \cdots + c_r\left(\frac{\lambda_r}{\lambda_1}\right)^k \mathbf{v}_r = \mathbf{0}. \tag{5.14}$$

  Because k is arbitrary, on the left hand side, we can get

$$\lim_{k\to\infty} c_1\mathbf{v}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k \mathbf{v}_2 + \cdots + c_r\left(\frac{\lambda_r}{\lambda_1}\right)^k \mathbf{v}_r = c_1\mathbf{v}_1. \tag{5.15}$$

  Because $c_1\mathbf{v}_1 \neq \mathbf{0}$, we have reached a contradiction. Therefore, $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ must be linearly independent.

- Properties of Eigenvalues/Eigenvectors

  Suppose λ is an eigenvalue of $\mathbf{A} \in \mathbb{R}^n$; $\mathbf{v}$ is the corresponding eigenvector of λ

  - $\mathbf{A}$ has at most n distinct eigenvalues
  - λ is an eigenvalue of $\mathbf{A}^T$
  - $\mathbf{v}$ is a λ^k-eigenvector of $\mathbf{A}^k$ ($\mathbf{A}^k\mathbf{v} = \lambda^k\mathbf{v}$)
  - $\mathbf{v}$ is a $c\lambda$-eigenvector of $c\mathbf{A}$ ($c\mathbf{A}\mathbf{v} = c(\mathbf{A}\mathbf{v}) = (c\lambda)\mathbf{v}$)
  - If $\mathbf{A}$ is invertible, $\mathbf{v}$ is a λ^{-1}-eigenvector of $\mathbf{A}^{-1}$ ($\mathbf{A}^{-1}\mathbf{A}\mathbf{v} = \mathbf{A}^{-1}\lambda\mathbf{v} \Rightarrow \mathbf{A}^{-1}\mathbf{v} = \lambda^{-1}\mathbf{v}$)

  Example. If $-1, 1, 2$ are eigenvalues of $\mathbf{A}$. Find eigenvalues of $\mathbf{A}^2 - \mathbf{A} + \mathbf{I}$.

  It can be seen that

$$(\mathbf{A}^2 - \mathbf{A} + \mathbf{I})\mathbf{v} = \mathbf{A}^2\mathbf{v} - \mathbf{A}\mathbf{v} + \mathbf{I}\mathbf{v} = \lambda^2\mathbf{v} - \lambda\mathbf{v} + \mathbf{v} = (\lambda^2 - \lambda + 1)\mathbf{v}. \tag{5.16}$$

For $\lambda = -1, 1, 2$, we know that the eigenvalues of $\mathbf{A}^2 - \mathbf{A} + \mathbf{I}$ are given by $3, 1, 3$.

- An $n \times n$ matrix $\mathbf{A}$ is similar to another $n \times n$ matrix $\mathbf{B}$ if there is an invertible $n \times n$ matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{PBP}^{-1}$.
- A square matrix $\mathbf{A}$ is said to be diagonalizable if $\mathbf{A}$ is similar to a diagonal matrix. That is, if $\mathbf{A} = \mathbf{PDP}^{-1}$ for some invertible matrix $\mathbf{P}$ and some diagonal matrix $\mathbf{D}$.

  Remark. If we can find $\mathbf{A} = \mathbf{PDP}^{-1}$, then $\mathbf{A}^k = (\mathbf{PDP}^{-1})^k = \mathbf{PD}^k\mathbf{P}^{-1}$.

- ***Theorem 5.3 (Diagonalization)***

  An $n \times n$ matrix $\mathbf{A}$ is diagonalizable iff. $\mathbf{A}$ has n linearly independent eigenvectors. Moreover, if $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a set of linearly independent eigenvectors with corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, then we can take

$$\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \tag{5.17}$$

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}. \tag{5.18}$$

  Proof. From $\mathbf{A} = \mathbf{PDP}^{-1}$ we know $\mathbf{AP} = \mathbf{PD}$. Looking at each matrix column by column, we have

$$\begin{bmatrix} \mathbf{Av}_1 & \mathbf{Av}_2 & \cdots & \mathbf{Av}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}. \tag{5.19}$$

  It can be seen that columns of $\mathbf{P}$ are eigenvectors; the diagonal of $\mathbf{D}$ is made up of corresponding eigenvalues.

  Remark. $\mathbf{P}$ and $\mathbf{D}$ are not unique.

- ***Theorem 5.4 (Algebraic and Geometric Multiplicity)***

  Let $\mathbf{A}$ be an $n \times n$ matrix whose distinct eigenvalues are:

$$\begin{gathered} \lambda_1 \text{ of multiplicity } m_1 \\ \vdots \\ \lambda_p \text{ of multiplicity } m_p \end{gathered}$$

  Then $\mathbf{A}$ is diagonalizable iff. both of the following statements are true:

1. $\dim(\lambda_i\text{-eigenspace}) = m_i$, $i = 1, 2, \dots, p$

2. $m_1 + m_2 + \cdots + m_p = n$

$\dim(\lambda_i$-eigenspace) is known as the geometric multiplicity of λ_i; m_i is known as the algebraic multiplicity of λ_i. It holds that

$$1 \leq \dim(\lambda_i\text{-eigenspace}) \leq m_i. \tag{5.20}$$

Remark. An $n \times n$ matrix is diagonalizable if it has n distinct eigenvalues.

- Differential Equations

Suppose we want to solve a linear system of differential equations

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}, \tag{5.21}$$

where

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_n(t) \end{bmatrix} \tag{5.22}$$

is a vector of smooth functions, and $\mathbf{A} \in \mathbb{R}^{n\times n}$. Consider the simple case $u'(t) = au(t)$, we can write

$$u'(t) = au(t) \tag{5.23}$$

$$\Rightarrow \frac{du}{dt} = au(t) \tag{5.24}$$

$$\Rightarrow \frac{du}{u} = a\,dt \tag{5.25}$$

$$\Rightarrow \int \frac{du}{u} = \int a\,dt \tag{5.26}$$

$$\Rightarrow \ln|u| = at + c \tag{5.27}$$

$$\Rightarrow |u| = e^c e^{at} \tag{5.28}$$

$$\Rightarrow u = c'e^{at}, \tag{5.29}$$

where c' is a constant.

– Let $\mathbf{A}$ be an $n \times n$ matrix. The matrix exponential of $\mathbf{A}$ is an $n \times n$ matrix defined by

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \cdots + \frac{\mathbf{A}^n}{n!}. \tag{5.30}$$

It can be seen that the definition is derived from Maclaurin series.

- If $\mathbf{D}$ is diagonal with $\mathbf{D} = \text{diag}(\lambda_1, \ldots, \lambda_n)$, then $e^{\mathbf{D}} = \text{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n})$.
- If $\mathbf{A}$ is diagonalizable, that is, $\mathbf{A} = \mathbf{PDP}^{-1}$, we know $\mathbf{A}^k = \mathbf{PA}^k\mathbf{P}^{-1}$. Then, we have $\mathbf{e}^A = \sum_{n=0}^{\infty} \frac{(\mathbf{PDP}^{-1})^n}{n!} = \mathbf{P}\left(\sum_{n=0}^{\infty} \frac{\mathbf{D}^n}{n!}\right)\mathbf{P}^{-1} = \mathbf{P}e^{\mathbf{D}}\mathbf{P}^{-1}$.
- Properties
  * $e^{\mathbf{A}s}e^{\mathbf{A}t} = e^{\mathbf{A}(s+t)}$
  * $e^{\mathbf{A}} \neq e^{\mathbf{B}}$ because $\mathbf{AB} \neq \mathbf{BA}$ in general
  * $e^{\mathbf{A}t}e^{-\mathbf{A}t} = \mathbf{I}$
  * $\frac{d}{dt}\left(e^{\mathbf{A}t}\right) = \mathbf{A}e^{\mathbf{A}t}$

  From 5.29 and the properties above, we know that the solution to $\frac{d\mathbf{u}}{dt} = \mathbf{Au}$ is $\mathbf{u}(t) = e^{\mathbf{A}t}\mathbf{c}$, where $\mathbf{c}$ is a vector of constants.
- ***Theorem 5.5 (Solution to First Order Differential Equations)***

  Let $\mathbf{A}$ be an $n \times n$ matrix. Then $\frac{d\mathbf{u}}{dt} = \mathbf{Au}$ has the solution

$$\mathbf{u}(t) = e^{\mathbf{A}t}\mathbf{u}(0). \tag{5.31}$$

  If $\mathbf{A}$ is diagonalizable and $\mathbf{A} = \mathbf{PDP}^{-1}$, then

$$\mathbf{u}(t) = \mathbf{P}e^{\mathbf{D}t}\mathbf{P}^{-1}\mathbf{u}(0), \tag{5.32}$$

  where columns of $\mathbf{P}$ are the eigenvectors of $\mathbf{A}$, and the diagonal entries of $\mathbf{D}$ are the corresponding eigenvalues.

  Remark. When $\mathbf{A}$ is diagonalizable, if we set $\mathbf{c} = \mathbf{P}^{-1}\mathbf{u}(0)$, we can write

$$\mathbf{u}(t) = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_1 t} & & \\ & & \ddots & \\ & & & e^{\lambda_n t} \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_n \end{bmatrix} \tag{5.33}$$

$$= \mathbf{c}_1 e^{\lambda_1 t}\mathbf{v}_1 + \mathbf{c}_2 e^{\lambda_2 t}\mathbf{v}_2 + \cdots + \mathbf{c}_n e^{\lambda_n t}\mathbf{v}_n. \tag{5.34}$$

Therefore, when $\mathbf{A}$ is diagonalizable, finding the general solution to $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ is essentially the same as finding the eigenvalues and eigenvectors of $\mathbf{A}$.

Example. Find the solutions to the differential equation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{5.35}$$

The eigenvalues and eigenvectors are given by $\lambda_1 = -1$, $\mathbf{v}_1 = (1, 1)$; $\lambda_2 = -3$, $\mathbf{v}_2 = (1, -1)$, respectively. The fundamental solutions are given by

$$e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{5.36}$$

A general solution is given by

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = c_1 e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 e^{-3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{5.37}$$

Suppose we are given $\mathbf{u}(0) = (2, 1)$, we need to solve

$$c_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{5.38}$$

for a particular solution. This is essentially solving the linear system

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \tag{5.39}$$

It can be seen that

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{5.40}$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{1}{2} \end{bmatrix}. \tag{5.41}$$

Remark. Steps of solving $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ when $\mathbf{A}$ is diagonalizable:

1. Find all eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$
2. Find a basis for each $N(\mathbf{A} - \lambda\mathbf{I})$ to get the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$

3. Write the general solution using 5.34; the $\mathbf{c}_i$'s are left in the result as unknowns

4. If given an initial condition (e.g., $\mathbf{u}(0) = \mathbf{l}$), solve the equation $\mathbf{c}_1\mathbf{v}_1 + \cdots + \mathbf{c}_n\mathbf{v}_n = \mathbf{u}(0)$ for $\mathbf{c}_1, \ldots, \mathbf{c}_n$

– The complex case: when real matrix $\mathbf{A}$ has complex eigenvalues

∗ If $\mathbf{v}$ is a λ-eigenvector of a real matrix $\mathbf{A}$, then $\bar{\mathbf{v}}$ is a $\bar{\lambda}$-eigenvector of $\mathbf{A}$.

Proof. Because $\mathbf{A}$ is real, we know $\mathbf{A}\bar{\mathbf{v}} = \overline{\mathbf{A}\mathbf{v}} = \overline{\lambda\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$.

∗ Consider a pair of complex eigenvalues and eigenvectors $\lambda = a + bi$, $\mathbf{v}$ and $\bar{\lambda} = a - bi$, $\bar{\mathbf{v}}$. For the fundamental solution $e^{\lambda t}\mathbf{v}$, suppose we can decompose $\mathbf{v} = \Re(\mathbf{v}) + i\Im(\mathbf{v})$, then we have

$$e^{\lambda t}\mathbf{v} = e^{(a+bi)t}[\Re(v) + i\Im(v)] \tag{5.42}$$
$$= e^a e^{ibt}[\Re(v) + i\Im(v)] \tag{5.43}$$
$$= e^a \{(\cos bt + i \sin bt)[\Re(v) + i\Im(v)]\} \tag{5.44}$$
$$= e^a \{\cos bt\Re(v) - \sin bt\Im(v) + i[\cos bt\Im(v) + \sin bt\Re(v)]\}\,. \tag{5.45}$$

Meanwhile, we can write

$$e^{\bar{\lambda} t}\bar{\mathbf{v}} = e^{(a-bi)t}[\Re(v) - i\Im(v)] \tag{5.46}$$
$$= e^a e^{-ibt}[\Re(v) - i\Im(v)] \tag{5.47}$$
$$= e^a \{(\cos bt - i \sin bt)[\Re(v) - i\Im(v)]\} \tag{5.48}$$
$$= e^a \{\cos bt\Re(v) - \sin bt\Im(v) - i[\cos bt\Im(v) + \sin bt\Re(v)]\}\,. \tag{5.49}$$

It can be seen that $e^{\lambda t}\mathbf{v}$ and $e^{\bar{\lambda} t}\bar{\mathbf{v}}$ are complex conjugates.

∗ For now, the solutions for $\mathbf{u}$ has complex numbers. If we split $\mathbf{u}(t) = \mathbf{f}(t) + i\mathbf{g}(t)$, where $\mathbf{f}$ and $\mathbf{g}$ are real functions. It can be seen that

$$\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u} \tag{5.50}$$
$$\Rightarrow \mathbf{f}'(t) + i\mathbf{g}'(t) = \mathbf{A}[\mathbf{f}(t) + i\mathbf{g}(t)] \tag{5.51}$$
$$\Rightarrow \mathbf{f}'(t) + i\mathbf{g}'(t) = \mathbf{A}\mathbf{f}(t) + i\mathbf{A}\mathbf{g}(t) \tag{5.52}$$

Since $\mathbf{f}$ and $\mathbf{g}$ are real, it can be seen that they are both solutions to $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$.

∗ Because $e^{\lambda t}\mathbf{v}$ and $e^{\bar{\lambda} t}\bar{\mathbf{v}}$ are complex conjugates, their real parts are the same; their absolute values of their imaginary parts are identical. Therefore, we only need the real part and imaginary part of $e^{\lambda t}\mathbf{v}$, and they are the two real solutions to the differential equations.

Example. Find a real solution to

$$\mathbf{x}'(t) = \begin{bmatrix} -2 & -2.5 \\ 10 & -2 \end{bmatrix} \mathbf{x}(t). \tag{5.53}$$

The eigenvalues are given by $\lambda_1 = -2 + 5i$, $\lambda_2 = -2 - 5i$. Because they are complex conjugates, we only need to compute the real solution for λ_1. The λ_1-eigenvector is $(i, 2)$. Decomposing $\mathbf{v}e^{\lambda t}$, we have

$$\begin{bmatrix} i \\ 2 \end{bmatrix} e^{(-2+5i)t} = \begin{bmatrix} i \\ 2 \end{bmatrix} e^{-2t} e^{i(5t)} \tag{5.54}$$

$$= \begin{bmatrix} i \\ 2 \end{bmatrix} e^{-2t} (\cos 5t + i \sin 5t) \tag{5.55}$$

$$= \begin{bmatrix} i(\cos 5t + i \sin 5t) \\ 2(\cos 5t + i \sin 5t) \end{bmatrix} e^{-2t} \tag{5.56}$$

$$= \begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix} e^{-2t} + i \begin{bmatrix} \cos 5t \\ 2 \sin 5t \end{bmatrix} e^{-2t}. \tag{5.57}$$

Therefore, the real solution is given by

$$\mathbf{x}(t) = c_1 \begin{bmatrix} -\sin 5t \\ 2\cos 5t \end{bmatrix} e^{-2t} + c_2 \begin{bmatrix} \cos 5t \\ 2 \sin 5t \end{bmatrix} e^{-2t}. \tag{5.58}$$

– **Theorem 5.6 (Stability of First Order Differential Equations)**

The differential equation $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ is

* stable if all $\Re(\lambda_i) < 0$
* neutrally stable if all $\Re(\lambda_i) \leq 0$ and at least one $\Re(\lambda_i) = 0$
* unstable if any $\Re(\lambda_i) > 0$

Here, λ_i's are the eigenvalues of $\mathbf{A}$.

– Second order equations

Consider $\frac{d^2\mathbf{u}}{dt^2} = \mathbf{A}\mathbf{u}$, where $\mathbf{A}$ only has *negative* eigenvalues. We take functions of the form $\mathbf{u}(t) = e^{iwt}\mathbf{v}$, it can be seen that

$$\frac{d^2}{dt^2}e^{iwt}\mathbf{v} = (iw)^2 e^{iwt}\mathbf{v} = -w^2 e^{iwt}\mathbf{v}. \tag{5.59}$$

If we let $\mathbf{A}e^{iwt}\mathbf{v} = -w^2 e^{iwt}\mathbf{v}$, we have

$$\mathbf{A}\mathbf{v} = -w^2\mathbf{v}. \tag{5.60}$$

If we get an eigenvalue λ and corresponding eigenvector $\mathbf{v}$ for $\mathbf{A}$, we have to solutions

$$e^{i\sqrt{-\lambda}t}\mathbf{v} \quad \text{and} \quad e^{-i\sqrt{-\lambda}t}\mathbf{v}. \tag{5.61}$$

In this case, $w = \sqrt{-\lambda}$ is called the frequency connected to the decay rate λ.

– **Theorem 5.7 (Solution to Second Order Differential Equations)**

Let $\mathbf{A}$ be an $n \times n$ matrix. If $\mathbf{A}$ has negative eigenvalues $\lambda_1, \dots, \lambda_n$, let $w_j = \sqrt{-\lambda_j}$, then $\frac{d^2\mathbf{u}}{dt^2} = \mathbf{A}\mathbf{u}$ has the solution

$$\mathbf{u}(t) = \left(c_1 e^{iw_1 t} + d_1 e^{-iw_1 t}\right)\mathbf{v}_1 + \cdots + \left(c_n e^{iw_n t} + d_n e^{-iw_n t}\right)\mathbf{v}_n, \tag{5.62}$$

where $\mathbf{v}_i$ is a λ_i-eigenvector. The general real solution is given by

$$\mathbf{u}(t) = (a_1 \cos w_1 t + b_1 \sin w_1 t)\mathbf{v}_1 + \cdots + (a_n \cos w_n t + b_n \sin w_n t)\mathbf{v}_n. \tag{5.63}$$

Example. Solve

$$\frac{d^2\mathbf{u}}{dt^2} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}\mathbf{u}. \tag{5.64}$$

The eigenvalues and eigenvectors are $\lambda_1 = -1$, $\mathbf{v}_1 = (1, 1)$; $\lambda_2 = -3$, $\mathbf{v}_2 = (1, -1)$. The frequencies are $w_1 = \sqrt{-\lambda_1} = 1$, $w_2 = \sqrt{-\lambda_2} = \sqrt{3}$.
The general solution has the form

$$\mathbf{u}(t) = (a_1 \cos t + b_1 \sin t)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + (a_2 \cos\sqrt{3}t + b_2 \sin\sqrt{3}t)\begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{5.65}$$

Suppose the initial conditions are given by $\mathbf{u}(0) = (1, 0)$, $\mathbf{u}'(0) = (0, 0)$. Here, we can interpret $\mathbf{u}$ as the position, $\mathbf{u}'$ as velocity, and $\mathbf{u}''$ as acceleration. It can be seen that

$$\mathbf{u}'(t) = (-a_1 \sin t + b_1 \cos t)\begin{bmatrix}1\\1\end{bmatrix} + (-\sqrt{3}a_2 \sin\sqrt{3}t + b_2\sqrt{3}\cos\sqrt{3}t)\begin{bmatrix}1\\-1\end{bmatrix}. \tag{5.66}$$

We can write the following equations:

$$a_1\begin{bmatrix}1\\1\end{bmatrix} + a_2\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix} \tag{5.67}$$

$$b_1\begin{bmatrix}1\\1\end{bmatrix} + \sqrt{3}b_2\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}. \tag{5.68}$$

It can be seen that $a_1 = a_2 = \frac{1}{2}$, $b_1 = b_2 = 0$.

- The conjugate transpose of $\mathbf{A}$, denoted by $\mathbf{A}^H$, is equal to $\bar{\mathbf{A}}^T$. $\mathbf{A}^H$ is called "A Hermitian".

  Remark. If $\mathbf{A}$ is a real matrix, then $\mathbf{A}^H = \mathbf{A}^T$.

- Inner product on $\mathbb{C}^n$: let $\mathbf{x}$ and $\mathbf{y}$ be two complex vectors, then their inner product is $\mathbf{x}^H\mathbf{y}$.

- An $n \times n$ complex matrix $\mathbf{A}$ is a Hermitian matrix if $\mathbf{A}^H = \mathbf{A}$; $\mathbf{A}$ is skew-Hermatian if $\mathbf{A}^H = -\mathbf{A}$.

  Remark. A real symmetric matrix is Hermitian.

- Properties of Hermitian matrix

  Let $\mathbf{A}$ be an $n \times n$ Hermitian matrix, that is, $\mathbf{A}^H = \mathbf{A}$.

  - For any $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x}^H\mathbf{A}\mathbf{x}$ is a real number.

    Proof. $(\mathbf{x}^H\mathbf{A}\mathbf{x})^H = \mathbf{x}^H\mathbf{A}^H(\mathbf{x}^H)^H = \mathbf{x}^H\mathbf{A}\mathbf{x}$. Therefore, $\mathbf{x}^H\mathbf{A}\mathbf{x}$ must be real.

  - Every eigenvalue of $\mathbf{A}$ is real.

    Proof. Suppose $\mathbf{v}$ is a λ-eigenvector of $\mathbf{A}$. We have

$$\mathbf{v}^H\mathbf{A}\mathbf{v} = \mathbf{v}^H\lambda\mathbf{v} = \lambda\mathbf{v}^H\mathbf{v}. \tag{5.69}$$

    That is,

$$\lambda = \frac{\mathbf{v}^H\mathbf{A}\mathbf{v}}{\mathbf{v}^H\mathbf{v}}. \tag{5.70}$$

Since both the numerator and denominator are real, we know λ is real.

– Let $\mathbf{v}_1$, $\mathbf{v}_2$ be two eigenvectors of $\mathbf{A}$ corresponding to distinct eigenvalues λ_1 and λ_2. Then $\mathbf{v}_1 \perp \mathbf{v}_2$.

Proof.

$$\bar{\lambda}_1 \mathbf{v}_1^H \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1)^H \mathbf{v}_2 = (\mathbf{A}\mathbf{v}_1)^H \mathbf{v}_2 = \mathbf{v}_1^H \mathbf{A}^H \mathbf{v}_2 \tag{5.71}$$

$$= \mathbf{v}_1^H \mathbf{A} \mathbf{v}_2 = \mathbf{v}_1^H (\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1^H \mathbf{v}_2. \tag{5.72}$$

Because the eigenvalues of $\mathbf{A}$ are real, we know $\bar{\lambda}_1 = \lambda_1$. Therefore, we can write $(\lambda_1 - \lambda_2)\mathbf{v}_1^H \mathbf{v}_2 = 0$. Since $\lambda_1 - \lambda_2 \neq 0$, it must be that $\mathbf{v}_1^H \mathbf{v}_2 = 0$.

Remark. A real symmetric matrix $\mathbf{A}$ can be orthogonally diagonalized. That is, $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ for some orthogonal matrix $\mathbf{Q}$ and diagonal matrix $\mathbf{\Lambda}$.

Example. Find an orthogonal diagonalization of

$$\mathbf{A} = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}. \tag{5.73}$$

It can be seen that for $\lambda_1 = 7$, $\mathbf{v}_1 = (-1, 2, 0)$, $\mathbf{v}_2 = (1, 0, 1)$; for $\lambda_2 = -2$, $\mathbf{v}_3 = (-2, -1, 2)$. Because $\mathbf{A}$ is symmetric, we know $\mathbf{v}_1 \perp \mathbf{v}_3$, $\mathbf{v}_2 \perp \mathbf{v}_3$. However, within the same eigenspace, $\mathbf{v}_1 \not\perp \mathbf{v}_2$. Therefore, we need to apply Gram-Schmidt process to find a diagonal basis for λ_1-eigenspace. It can be seen that

$$\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}; \tag{5.74}$$

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_1^T \mathbf{v}_2}{\mathbf{v}_1^T \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{1}{5} \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \tag{5.75}$$

$$= \begin{bmatrix} \frac{4}{5} \\ \frac{2}{5} \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 \\ 2 \\ 5 \end{bmatrix}. \tag{5.76}$$

Normalizing each vector:

$$\mathbf{u}_1 = \frac{\mathbf{w_1}}{\|\mathbf{w_1}\|} = \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix} \tag{5.77}$$

$$\mathbf{u}_2 = \frac{\mathbf{w_2}}{\|\mathbf{w_2}\|} = \begin{bmatrix} \frac{4}{\sqrt{45}} \\ \frac{2}{\sqrt{45}} \\ \frac{5}{\sqrt{45}} \end{bmatrix} \tag{5.78}$$

$$\mathbf{u}_3 = \frac{\mathbf{v_3}}{\|\mathbf{v_3}\|} = \begin{bmatrix} -\frac{2}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix}. \tag{5.79}$$

Therefore, we have

$$\mathbf{Q} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{4}{\sqrt{45}} & -\frac{2}{3} \\ \frac{2}{\sqrt{5}} & \frac{2}{\sqrt{45}} & -\frac{1}{3} \\ 0 & \frac{5}{\sqrt{45}} & \frac{2}{3} \end{bmatrix} \tag{5.80}$$

$$\mathbf{\Lambda} = \begin{bmatrix} 7 & & \\ & 7 & \\ & & -2 \end{bmatrix}. \tag{5.81}$$

- An $n \times n$ complex matrix with orthonormal columns is called a unitary matrix. That is, $\mathbf{U}^H\mathbf{U} = \mathbf{I}$.

  Remark. An orthogonal real matrix is also unitary.

  Example. If $\mathbf{K}$ is a skew-Hermitian matrix, then $e^{\mathbf{K}}$ is a unitary matrix. Because $\left(e^{\mathbf{K}}\right)^H e^{\mathbf{K}} = e^{\mathbf{K}^H} e^{\mathbf{K}} = e^{-\mathbf{K}} e^{\mathbf{K}} = \mathbf{I}$.

- Properties of Unitary Matrices

Let $\mathbf{U}$ be unitary, that is, $\mathbf{U}^H\mathbf{U} = \mathbf{I}$.

– $(\mathbf{U}\mathbf{x})^H(\mathbf{U}\mathbf{y}) = \mathbf{x}^H\mathbf{y}$

– The eigenvalues of $\mathbf{U}$ is of the form $e^{i\theta}$

Proof. Let $\mathbf{v}$ be a λ-eigenvector of $\mathbf{U}$. It can be seen that $\mathbf{v}^H\mathbf{v} = (\mathbf{U}\mathbf{v})^H(\mathbf{U}\mathbf{v}) = (\lambda\mathbf{v})^H(\lambda\mathbf{v}) = |\lambda|^2\mathbf{v}^H\mathbf{v}$. Therefore, we have $|\lambda| = 1$.

- Two $n \times n$ matrices are similar if there is an invertible matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.
- Let $\mathbf{P}$ be an $n \times n$ invertible matrix. The transformation $\mathbf{A} \to \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is called a similarity transformation.
- ***Theorem 5.8 (Eigenvalues of Similar Matrices)***

Suppose $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is a matrix similar to $\mathbf{A}$. Then $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues. Moreover, if $\mathbf{v}$ is a λ-eigenvector of $\mathbf{A}$, then $\mathbf{P}^{-1}\mathbf{v}$ is a λ-eigenvector of $\mathbf{B}$.

Proof. $\det(\mathbf{A} - \lambda\mathbf{I}) = \det[\mathbf{P}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{P}] = \det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P} - \lambda\mathbf{I})$.

$\mathbf{B}\mathbf{P}^{-1}\mathbf{v} = (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})\mathbf{P}^{-1}\mathbf{V} = \mathbf{P}^{-1}\mathbf{A}\mathbf{v} = \mathbf{P}^{-1}(\lambda\mathbf{v}) = \lambda(\mathbf{P}^{-1}\mathbf{v})$.

Remark.

– If $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues, they are not necessarily similar.

– If an $n \times n$ matrix $\mathbf{A}$ has n distinct eigenvalues, then any matrix $\mathbf{B}$ has the same eigenvalues is similar to $\mathbf{A}$.

Proof. If $\mathbf{A}$, $\mathbf{B}$ both has n distinct eigenvalues, they are diagonalizable. We can write $\mathbf{A} = \mathbf{P}_1^{-1}\mathbf{D}\mathbf{P}_1$, $\mathbf{B} = \mathbf{P}_2^{-1}\mathbf{D}\mathbf{P}_2$. By theorem 5.2, we know that both $\mathbf{P}_1$ and $\mathbf{P}_2$ are invertible. Let $\mathbf{X} = \mathbf{P}_2^{-1}\mathbf{P}_1$. Then $\mathbf{X}^{-1}\mathbf{B}\mathbf{X} = (\mathbf{P}_1^{-1}\mathbf{P}_2)\mathbf{P}_2^{-1}\mathbf{D}\mathbf{P}_2(\mathbf{P}_2^{-1}\mathbf{P}_1) = \mathbf{P}_1^{-1}\mathbf{D}\mathbf{P}_1 = \mathbf{A}$. Therefore, $\mathbf{A}$ and $\mathbf{B}$ are similar.

- ***Theorem 5.9 (Similarity and Linear Transformation)***

Let $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ be a basis of $\mathbb{R}^n$. The B-matrix of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is similar to its standard matrix $\mathbf{A}$. Moreover, if $\mathbf{P} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{bmatrix}$, then $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

- ***Theorem 5.10 (Similarity to Triangular Matrix)***

For any $n \times n$ matrix $\mathbf{A}$, there is a unitary matrix $\mathbf{U}$ such that $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{U}^H\mathbf{A}\mathbf{U} = \mathbf{T}$ is an upper triangular matrix. This is also known as Schur decomposition.

Example. Find a unitary matrix $\mathbf{U}$ so that $\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{T}$ is a triangular matrix.

(1) We need to find an eigenvector of $\mathbf{A}$. In this case, we know $\lambda_1 = 1$, and the corresponding $\mathbf{v}_1 = (1, -1, 1)$.

(2) Find a linearly independent set containing $\mathbf{v}$. In this case, we can take $\{\mathbf{v}, \mathbf{e}_1, \mathbf{e}_2\}$.

(3) Apply Gram-Schmidt process to get a orthogonal basis. It can be seen that

$$\mathbf{w}_1 = \mathbf{v}_1; \tag{5.82}$$

$$\mathbf{w}_2 = \mathbf{e}_1 - \frac{\mathbf{w}_1^T\mathbf{e}_1}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} \to \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}; \tag{5.83}$$

$$\mathbf{w}_3 = \mathbf{e}_2 - \frac{\mathbf{w}_1^T\mathbf{e}_2}{\mathbf{w}_1^T\mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{w}_2^T\mathbf{e}_2}{\mathbf{w}_2^T\mathbf{w}_2}\mathbf{w}_2 \tag{5.84}$$

$$= \begin{bmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \to \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}. \tag{5.85}$$

(4) Normalize vectors to get $\mathbf{U}$.

$$\mathbf{U}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix}. \tag{5.86}$$

(5) Compute $\mathbf{T}_1 = \mathbf{U}_1^{-1}\mathbf{A}\mathbf{U} = \mathbf{U}^T\mathbf{A}\mathbf{U}$. It can be seen that

$$\mathbf{T}_1 = \begin{bmatrix} 1 & \frac{24}{\sqrt{18}} & \frac{20}{\sqrt{6}} \\ 0 & -\frac{3}{2} & \frac{1}{\sqrt{12}} \\ 0 & -\frac{3}{\sqrt{12}} & -\frac{5}{2} \end{bmatrix}. \tag{5.87}$$

The lower block is not a triangular matrix yet. We need to repeat the process for this block.

(6) Let

$$\mathbf{B} = \begin{bmatrix} -\frac{3}{2} & \frac{1}{\sqrt{12}} \\ -\frac{3}{\sqrt{12}} & -\frac{5}{2}. \end{bmatrix} \tag{5.88}$$

Repeat the steps (1)-(5) to get $\mathbf{U}_2$ and $\mathbf{T}_2$. In this case, we have

$$\mathbf{U}_2 = \begin{bmatrix} -\frac{1}{2} & \frac{3}{\sqrt{12}} \\ \frac{3}{\sqrt{12}} & \frac{1}{2} \end{bmatrix} \tag{5.89}$$

$$\mathbf{T}_2 = \begin{bmatrix} -2 & -\frac{4}{\sqrt{12}} \\ 0 & -2 \end{bmatrix}. \tag{5.90}$$

It can be seen that $\mathbf{T}_2$ is triangular.

(7) We need to combine $\mathbf{U}_1$ and $\mathbf{U}_2$ to get the final $\mathbf{U}$.

$$\mathbf{A} = \mathbf{U}_1\mathbf{T}_1\mathbf{U}_1^{-1} \tag{5.91}$$

$$= \mathbf{U}_1 \begin{bmatrix} 1 & \frac{24}{\sqrt{18}} & \frac{20}{\sqrt{6}} \\ 0 & \multicolumn{2}{c}{\mathbf{B}} \\ 0 & & \end{bmatrix} \mathbf{U}_1^{-1} \tag{5.92}$$

$$= \mathbf{U}_1 \begin{bmatrix} 1 & \frac{24}{\sqrt{18}} & \frac{20}{\sqrt{6}} \\ 0 & \multicolumn{2}{c}{\mathbf{U}_2\mathbf{T}_2\mathbf{U}_2^{-1}} \\ 0 & & \end{bmatrix} \mathbf{U}_1^{-1} \tag{5.93}$$

$$= \mathbf{U}_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & & \\ 0 & \multicolumn{2}{c}{\mathbf{U}_2} \end{bmatrix} \begin{bmatrix} 1 & \frac{24}{\sqrt{18}} & \frac{20}{\sqrt{6}} \\ 0 & -2 & -\frac{4}{\sqrt{12}} \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & & \\ 0 & \multicolumn{2}{c}{\mathbf{U}_2^{-1}} \end{bmatrix} \mathbf{U}_1^{-1}. \tag{5.94}$$

Therefore, we have

$$\mathbf{U} = \mathbf{U}_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & & \\ 0 & \multicolumn{2}{c}{\mathbf{U}_2} \end{bmatrix} \tag{5.95}$$

$$= \begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \end{bmatrix}. \tag{5.96}$$

It can be seen that

$$\mathbf{T} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \begin{bmatrix} 1 & \frac{24}{\sqrt{18}} & \frac{20}{\sqrt{6}} \\ 0 & -2 & -\frac{4}{\sqrt{12}} \\ 0 & 0 & -2 \end{bmatrix}. \tag{5.97}$$

- ***Theorem 5.11 (Spectral Theorem)***

  Every real symmetric matrix $\mathbf{A}$ can be diagonalized by an orthogonal matrix $\mathbf{Q}$. Every Hermitian matrix can be diagonalized by a unitary matrix $\mathbf{U}$. That is,

$$\text{(real) } \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{\Lambda} \quad \text{or} \quad \mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \tag{5.98}$$

$$\text{(complex) } \mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{\Lambda} \quad \text{or} \quad \mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H. \tag{5.99}$$

  The columns of $\mathbf{Q}$ (or $\mathbf{U}$) are the orthonormal eigenvectors of $\mathbf{A}$.

- The matrix $\mathbf{N}$ is normal if $\mathbf{N}\mathbf{N}^H = \mathbf{N}^H\mathbf{N}$.

  Example. Hermitian matrices, skew-Hermitian matrices, and unitary matrices are normal.

- Properties of Normal Matrices

  – Normality is preserved under unitary similarity.

Proof. Suppose $\mathbf{N}$ is similar to $\mathbf{X}$, that is, $\mathbf{X} = \mathbf{U}^H\mathbf{N}\mathbf{U}$. It can be seen that $\mathbf{X}^H\mathbf{X} = \mathbf{U}^H\mathbf{N}^H\mathbf{U}\mathbf{U}^H\mathbf{N}\mathbf{U} = \mathbf{U}^H\mathbf{N}^H\mathbf{N}\mathbf{U} = \mathbf{U}^H\mathbf{N}\mathbf{N}^H\mathbf{U} = \mathbf{U}^H\mathbf{N}\mathbf{U}\mathbf{U}^H\mathbf{N}^H\mathbf{U} = \mathbf{X}\mathbf{X}^H$. Therefore, $\mathbf{X}$ is also normal.

– A triangular matrix is normal iff. it is diagonal

Proof. We can prove by induction. Denote the proposition "$n \times n$ triangular normal matrices are diagonal" by T_n.

Base case: When $n = 1$, it is obvious that T_1 is true.

Induction Hypothesis: Suppose T_n is true.

Induction Step: Suppose $\mathbf{A} \in \mathbb{C}^{(n+1)\times(n+1)}$ is an upper triangular matrix. We can write

$$\mathbf{A} = \left[\begin{array}{c|c} t_{11} & \mathbf{x}^H \\ \hline 0 & \mathbf{B} \end{array}\right], \tag{5.100}$$

where $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{B} \in \mathbb{C}^{n\times n}$. $\mathbf{B}$ is also an upper triangular normal matrix. It is easy to see

$$\mathbf{A}^H = \left[\begin{array}{c|c} \bar{t}_{11} & 0 \\ \hline \mathbf{x} & \mathbf{B}^H \end{array}\right]. \tag{5.101}$$

It can be seen that

$$\mathbf{A}\mathbf{A}^H = \left[\begin{array}{c|c} |t_{11}|^2 + \mathbf{x}^H\mathbf{x} & \mathbf{y}^H \\ \hline \mathbf{y} & \mathbf{B}\mathbf{B}^H \end{array}\right], \tag{5.102}$$

$$\mathbf{A}^H\mathbf{A} = \left[\begin{array}{c|c} |t_{11}|^2 & \mathbf{z}^H \\ \hline \mathbf{z} & \mathbf{x}\mathbf{x}^H + \mathbf{B}^H\mathbf{B} \end{array}\right], \tag{5.103}$$

Because $\mathbf{A}$ is normal, we know $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$. That implies $\mathbf{x}^H\mathbf{x} = 0 \Rightarrow \mathbf{x} = \mathbf{0}$. Therefore, $\mathbf{y} = \mathbf{z} = \mathbf{0}$. In the lower part, we have $\mathbf{x}\mathbf{x}^H + \mathbf{B}^H\mathbf{B} = \mathbf{B}^H\mathbf{B} = \mathbf{B}\mathbf{B}^H$. Because $\mathbf{B}$ is normal and upper triangular, by the induction hypothesis, it is diagonal. As a result, $\mathbf{A}$ is also diagonal.

By induction, we have proved T_n for $n \geq 1$.

- ***Theorem 5.12 (Normal Matrices)***
  - Normal matrices $\mathbf{N}$ are exactly those matrices with $\mathbf{T} = \mathbf{U}^{-1}\mathbf{N}\mathbf{U}$ being diagonal matrices for some unitary matrices $\mathbf{U}$.

    Proof. Suppose $\mathbf{T}$ is similar to $\mathbf{N}$. By Schur's decomposition, we know that it is similar to triangular matrix $\mathbf{T}$. Because $\mathbf{T}$ is also normal, $\mathbf{T}$ must be diagonal. Therefore, $\mathbf{N}$ is unitarily diagonalizable.
  - Normal matrices are exactly those that have a complete set of orthonormal eigenvectors.

    Proof. Because $\mathbf{N}$ is unitarily diagonalizable, it has a complete set of orthonormal eigenvectors.
- A Jordan block is of the form

$$\begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{bmatrix}, \tag{5.104}$$

where λ is any number. A Jordan block of $\mathbb{C}^{n\times n}$ is called a Jordan block of size n.

- A matrix $\mathbf{J}$ is in Jordan form if it has Jordan blocks on the diagonal. That is,

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \mathbf{J}_s \end{bmatrix}, \tag{5.105}$$

  where $\mathbf{J}_i$ are Jordan blocks.

- ***Theorem 5.13 (Jordan Normal Form)***

  If a matrix $\mathbf{A}$ has s linearly independent eigenvectors, then it is similar to a matrix $\mathbf{J}$ that is in Jordan form, with s Jordan blocks on the diagonal. That is, $\mathbf{J} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$, or equivalently, $\mathbf{A} = \mathbf{M}\mathbf{J}\mathbf{M}^{-1}$. Each block $\mathbf{J}_i$ is

$$\begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}, \tag{5.106}$$

  where λ_i is an eigenvalue.

  Remark.

  - A square matrix $\mathbf{A}$ is diagonalizable iff. all Jordan blocks for $\mathbf{A}$ are of size 1.
  - $\mathbf{J}$ is unique up to a permutation of $\mathbf{J}_1, \dots, \mathbf{J}_s$.
  - Finding $\mathbf{M}$ and $\mathbf{J}$ for $\mathbf{A}$

    For each λ-eigenvector of $\mathbf{v}$, start with $\mathbf{x}_1 = \mathbf{v}$, then:

    * Get an $\mathbf{x}_2$ such that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}_2 = \mathbf{x}_1$
    * Get an $\mathbf{x}_3$ such that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}_3 = \mathbf{x}_2$
    * ...

    Until some k such that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}_{k+1} = \mathbf{x}_k$ has no solution. Then we get a Jordan block of size k. All these $\mathbf{x}_1, \dots, \mathbf{x}_k$ would be columns of $\mathbf{M}$.
  - A quick way to verify $\mathbf{J}$ and $\mathbf{M}$ is to check whether $\mathbf{AM} = \mathbf{MJ}$.

Example. Find the JCF of

$$\mathbf{A} = \begin{bmatrix} 8 & 0 & 0 & 8 & 8 \\ 0 & 0 & 0 & 8 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}. \tag{5.107}$$

The eigenvalues and eigenvectors of $\mathbf{A}$ are given by $\lambda_1 = 8$ (of multiplicity 2), $\mathbf{v}_1 = (1,0,0,0,0)$; $\lambda_2 = 0$ (of multiplicity 3), $\mathbf{v}_2 = (0,1,0,0,0)$ and $\mathbf{v}_3 = (0,0,1,0,0)$.

For $\lambda_1 = 8$, let $\mathbf{x}_1 = \mathbf{v}_1$. Then, we need to solve $(\mathbf{A}-8\mathbf{I})\mathbf{x}_2 = \mathbf{x}_1$. A solution is given by $\mathbf{x}_2 = (0, \frac{1}{8}, 0, 0, \frac{1}{8})$. Next, we solve $(\mathbf{A}-8\mathbf{I})\mathbf{x}_3 = \mathbf{x}_2$, which has no solution. Therefore, the size of this Jordan block is 2.

For $\lambda_2 = 0$, let $\mathbf{x}_4 = \mathbf{v}_2$. Then, we need to solve $\mathbf{A}\mathbf{x}_5 = \mathbf{x}_4$. A solution is given by $\mathbf{x}_5 = (0,0,0,\frac{1}{8},0)$. Next, we solve $\mathbf{A}\mathbf{x}_6 = \mathbf{x}_5$, which has no solution. The size of this Jordan block is 2. Let $\mathbf{x}_7 = \mathbf{v}_3$, we need to solve $\mathbf{A}\mathbf{x}_8 = \mathbf{x}_7$, which has no solution. Therefore, the size of this Jordan block is 1.

As a result, we can write

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & \frac{1}{8} & 0 & 0 & 0 \end{bmatrix}, \tag{5.108}$$

$$\mathbf{J} = \begin{bmatrix} 8 & 1 & & & \\ & 8 & & & \\ & & 0 & 1 & \\ & & & 0 & \\ & & & & 0 \end{bmatrix}. \tag{5.109}$$

Remark. One can scale the columns of $\mathbf{J}$. However, the columns corresponding to the same Jordan block must be scaled by the same factor.

6 Positive Definitive Matrices

- A point (x_0, y_0) is a stationary point or a critical point of a differentiable function $F(x, y)$ if

$$\frac{\partial F}{\partial x}(x_0, y_0) = 0 \quad \text{and} \quad \frac{\partial F}{\partial y}(x_0, y_0). \tag{6.1}$$

  If (x_0, y_0) is a stationary point of F, then it can be

  – A local minimum

  – A local maximum

  – A saddle point

- **Theorem 6.1 (Stationary Point Type of Quadratic Forms)**

  For a quadratic form $f(x, y) = ax^2 + 2bxy + cy^2$, $(0, 0)$ is a stationary point and

  1. It is a minimum if $a > 0$ and $ac > b^2$. In this case, f is said to be positive definite.
  2. It is a minimum if $a < 0$ and $ac > b^2$. In this case, f is said to be negative definite.
  3. It is a saddle point if $ac < b^2$.
  4. If $ac = b^2$, then
     (a) f is said to be positive semidefinite if $a > 0$.
     (b) f is said to be negative semidefinite if $a < 0$.

  Proof. $f(x, y) = ax^2 + 2bxy + cy^2 = a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2.$

  Remark. If F is not a quadratic form, suppose (α, β) is a stationary point, we can use Taylor series at this point:

$$\begin{aligned} F(x, y) = F(\alpha, \beta) &+ F_x(\alpha, \beta)(x - \alpha) + F_y(\alpha, \beta)(y - \beta) \\ &+ \frac{1}{2}F_{xx}(\alpha, \beta)(x - \alpha)^2 + \frac{1}{2}F_{yy}(\alpha, \beta)(y - \alpha)^2 \\ &+ F_{xy}(\alpha, \beta)(x - \alpha)(y - \beta) + \text{higher order terms.} \end{aligned} \tag{6.2}$$

Since (α, β) is stationary, it is equivalent to

$$\begin{aligned} F(x,y) &= F(\alpha,\beta) \\ &\quad + \frac{1}{2}F_{xx}(\alpha,\beta)(x-\alpha)^2 + \frac{1}{2}F_{yy}(\alpha,\beta)(y-\alpha)^2 \\ &\quad + F_{xy}(\alpha,\beta)(x-\alpha)(y-\beta) + \text{higher order terms}. \end{aligned} \tag{6.3}$$

If at least one of the $F_{xx}(\alpha,\beta)$, $F_{yy}(\alpha,\beta)$ and $F_{xy}(\alpha,\beta)$ is not zero, then the type of stationary point (α,β) is the same as the type of stationary point $(0,0)$ of the quadratic form

$$f(x,y) = F_{xx}(\alpha,\beta)x^2 + 2F_{xy}(\alpha,\beta)xy + F_{yy}(\alpha,\beta)y^2. \tag{6.4}$$

$\frac{1}{2}f$ is called the quadratic part of F.

- The quadratic form $f(x_1,x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2$ can be expressed in terms of multiplication with a symmetric matrix

$$ax_1^2 + 2bx_1x_2 + cx_2^2 = \underbrace{\begin{bmatrix} x_1 & x_2 \end{bmatrix}}_{\mathbf{x}^T} \underbrace{\begin{bmatrix} a & b \\ b & c \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}}. \tag{6.5}$$

Remark. In matrix $\mathbf{A}$, the value used is b, which is half of the coefficient of the cross term in $f(x_1,x_2)$.

- For any $n \times n$ *symmetric* matrix $\mathbf{A}$, $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ is a pure quadratic form on $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$.
  - f is positive definite if $f(x) > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
  - f is negative definite if $f(x) < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
  - f is positive semidefinite if $f(x) \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$.
  - f is negative semidefinite if $f(x) \leq 0$ for all $\mathbf{x} \neq \mathbf{0}$.

  If f is positive definite, $\mathbf{A}$ is positive definite.

- ***Theorem 6.2 (Test for Positive Definiteness)***

  Each of the following tests is a necessary and sufficient condition for the real symmetric matrix $\mathbf{A}$ to be positive definite.

  (I) $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all nonzero real vector $\mathbf{x}$.

(II) All eigenvalues of $\mathbf{A}$ are greater than zero.

(III) All upper left submatrices $\mathbf{A}_k$ have determinants greater than zero.

(IV) All diagonal entries of $\mathbf{D}$ in $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T$ are positive.

(V) There is a matrix $\mathbf{R}$ with independent columns such that $\mathbf{A} = \mathbf{R}^T\mathbf{R}$.

Remark.

- (Law of inertia) The number of positive, negative and zero entries in $\mathbf{\Lambda}$ and $\mathbf{D}$ are the same.
- Three ways to find $\mathbf{R}$:
  1. If $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T$, then $\mathbf{R} = \sqrt{\mathbf{D}}\mathbf{L}^T$.
  2. If $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, then $\mathbf{R} = \sqrt{\mathbf{\Lambda}}\mathbf{Q}^T$.
  3. If $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, then $\mathbf{R} = \mathbf{Q}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^T$. Notice that in this case, $\mathbf{R}$ is symmetric.

Proof.

- (I) This is the definition of positive definiteness.
- (V)→(I) If $\mathbf{A} = \mathbf{R}^T\mathbf{R}$, where $\mathbf{R}$ is real and invertible, we have

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = (\mathbf{R}\mathbf{x})^T(\mathbf{R}\mathbf{x}) \geq 0. \tag{6.6}$$

  Since $\mathbf{R}$ has linearly independent columns, $\mathbf{R}\mathbf{x} = \mathbf{0}$ only has zero solution. Therefore, for nonzero $\mathbf{x}$, we have $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$.
- (II)→(V) By Theorem 5.11, we can write $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, where $\mathbf{Q}$ is orthogonal. Because the eigenvalues of $\mathbf{A}$ are greater than zero, the diagonal entries of λ are greater than zero. Therefore, we can take $\mathbf{R} = \sqrt{\mathbf{\Lambda}}\mathbf{Q}$.
- (IV)→(V) We can take $\mathbf{R} = \sqrt{\mathbf{D}}\mathbf{L}^T$. Because $\mathbf{L}$ is lower triangular, it must be invertible.
- (III)→(I) Consider $\mathbf{x}_k \in \mathbb{R}^k$, where $k < n$. Construct a vector $\mathbf{x} =$

$\begin{bmatrix}\mathbf{x}_k^T & \mathbf{0}\end{bmatrix}^T \in \mathbb{R}^n$. Then,

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \begin{bmatrix}\mathbf{x}_k^T & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{A}_k & * \\ * & *\end{bmatrix}\begin{bmatrix}\mathbf{x}_k \\ \mathbf{0}\end{bmatrix} = \mathbf{x}_k^T\mathbf{A}_k\mathbf{x}_k. \tag{6.7}$$

That is to say, $\mathbf{A}$ is positive definite iff. the first k submatrices are also positive definite. Because determinant equals the product of eigenvalues, we can prove (III)→(I) using (II).

- ***Theorem 6.3 (Test for Positive Semidefiniteness)***

  Each of the following tests is a necessary and sufficient condition for the real symmetric matrix $\mathbf{A}$ to be positive semidefinite.

  (I) $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all nonzero real vector $\mathbf{x}$.

  (II) All eigenvalues of $\mathbf{A}$ are greater than or equal to zero.

  (III) All upper left submatrices $\mathbf{A}_k$ have determinants greater than or equal to zero.

  (IV) All diagonal entries of $\mathbf{D}$ in $\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T$ are nonnegative.

  (V) There is a matrix $\mathbf{R}$ such that $\mathbf{A} = \mathbf{R}^T\mathbf{R}$.

- If $\mathbf{A}$ is an $n \times n$ symmetric positive definite matrix, then $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is an ellipsoid in $\mathbb{R}^n$. It is an ellipse when $n = 2$.

- ***Theorem 6.4 (Shape of Ellipsoid)***

  Suppose $\mathbf{A}$ is positive definite with spectral decomposition $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$. Then $\mathbf{y} = \mathbf{Q}^T\mathbf{x}$ simplifies the ellipsoid $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$. More specifically, $\mathbf{y}^T\mathbf{\Lambda}\mathbf{y}$ is the equation of the simplified ellipsoid. Its axes have lengths $\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_n}}$ from the center, where λ_i are eigenvalues. In the original $\mathbf{x}$-space, they point along the eigenvectors of $\mathbf{A}$.

- ***Theorem 6.5 (Singular Value Decomposition)***

  Let $\mathbf{A}$ be an $m \times l$ matrix with rank r. There exist an $m \times m$ orthogonal matrix $\mathbf{U}$, an $m \times l$ diagonal matrix $\mathbf{\Sigma}$ and an $l \times l$ orthogonal matrix $\mathbf{V}$ such that

  $$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{6.8}$$

  – The columns of $\mathbf{U}$ are eigenvectors of $\mathbf{A}\mathbf{A}^T$.

  – The columns of $\mathbf{V}$ are eigenvectors of $\mathbf{A}^T\mathbf{A}$.

– There are r nonzero values in the diagonal of $\mathbf{\Sigma}$, and they are the square roots of the common nonzero eigenvectors of both $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$.

Remark.

– $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ share distinct eigenvectors. More concretely, suppose $\mathbf{x} \neq \mathbf{0}$ in an eigenvector of $\mathbf{A}^T\mathbf{A}$, that is, $\mathbf{A}^T\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Then, we have $\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{x}) = \lambda(\mathbf{A}\mathbf{x})$. That is, $\mathbf{A}\mathbf{x}$ is an eigenvector of $\mathbf{A}\mathbf{A}^T$ with the same eigenvalue.

– Let $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \cdots \mathbf{u}_m\end{bmatrix}$, $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \cdots \mathbf{v}_l\end{bmatrix}$, and

$$\mathbf{\Sigma} = \begin{bmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}. \tag{6.9}$$

From $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, we have $\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}$. That is,

$$\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i \tag{6.10}$$

for $i = 1, 2, \dots, r$. We have the freedom to choose $\mathbf{u}_i$ and $\mathbf{v}_i$, but 6.10 must hold.

– A procedure to compute SVD

* Compact SVD

1. Compute the eigenvalues and eigenvectors of $\mathbf{A}^T\mathbf{A}$.
2. The nonzero eigenvalues and corresponding eigenvectors of $\mathbf{A}^T\mathbf{A}$ forms columns of $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \cdots \mathbf{v}_r\end{bmatrix}$.
3. Compute the r columns of $\mathbf{U}$ by

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i = \frac{1}{\sqrt{\lambda_i}}\mathbf{A}\mathbf{v}_i. \tag{6.11}$$

* Full SVD

1. Compute the eigenvalues and eigenvectors of $\mathbf{A}^T\mathbf{A}$
2. The nonzero eigenvalues and corresponding eigenvectors of $\mathbf{A}^T\mathbf{A}$ forms columns of $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \cdots \mathbf{v}_r\end{bmatrix}$.
3. Compute the r columns of $\mathbf{U}$ by 6.11.

4. Put the corresponding eigenvectors of zero eigenvalues of $\mathbf{A}^T\mathbf{A}$ in $\mathbf{V}$.
5. Put the corresponding eigenvectors of zero eigenvalues of $\mathbf{A}\mathbf{A}^T$ in $\mathbf{U}$.

- ***Theorem 6.6 (Basis for 4 Fundamental Subspaces)***

  Let $\mathbf{A}$ be an $m \times l$ matrix with rank r. Let $\mathbf{A} = \mathbf{U\Sigma V}^T$ be a singular value decomposition. Then $\mathbf{U}$ and $\mathbf{V}$ give orthonormal basis for all four fundamental subspaces:

  – First r columns of $\mathbf{U}$ forms a basis for $C(\mathbf{A})$

  – Last $m - r$ columns of $\mathbf{U}$ forms a basis for $N(\mathbf{A}^T)$

  – First r columns of $\mathbf{V}$ forms a basis for $C(\mathbf{A}^T)$

  – Last $l - r$ columns of $\mathbf{V}$ forms a basis for $\mathbf{N}(\mathbf{A})$

- ***Theorem 6.7 (Polar Decomposition)***

  Every real square matrix can be factorized into $\mathbf{A} = \mathbf{QS}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{S}$ is symmetric positive semidefinite. If $\mathbf{A}$ is intertible, then $\mathbf{S}$ is symmetric positive definite.

  Proof. Use SVD on $\mathbf{A}$, we have $\mathbf{A} = \mathbf{U\Sigma V}^T = \mathbf{UV}^T\mathbf{V\Sigma V}^T$. Therefore, we can take $\mathbf{Q} = \mathbf{UV}^T$ and $\mathbf{S} = \mathbf{V\Sigma V}^T$. On page 17, we proved the nullspace of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}$ are the same. Therefore, if $\mathbf{A}$ is invertible, $\mathbf{A}^T\mathbf{A}$ is also invertible. Therefore, there would not be any zero on the diagonal of $\mathbf{\Sigma}$. That is, $\mathbf{S}$ is positive definite.

  Remark. If $\mathbf{A}$ is invertible, so is $\mathbf{S} = \mathbf{V\Sigma V}^T$. Then, $\mathbf{Q} = \mathbf{AS}^{-1}$.

- Let $\mathbf{A}$ be an $m \times l$ matrix with SVD $\mathbf{A} = \mathbf{U\Sigma V}^T$. Then its pseudoinverse $\mathbf{A}^+$ is given by $\mathbf{A}^+ = \mathbf{V\Sigma}^+\mathbf{U}^T$, where

$$\mathbf{\Sigma}^+ = \begin{bmatrix} \sigma_1^{-1} & & & \\ & \ddots & & \\ & & \sigma_r^{-1} & \\ & & & \end{bmatrix}. \tag{6.12}$$

  The size of $\mathbf{\Sigma}^+$ is equal to that of $\mathbf{\Sigma}^T$.

- ***Theorem 6.8 (Characterization of Pseudoinverse)***

  Let $\mathbf{A}$ be an $m\times l$ matrix. Its pseudoinverse $\mathbf{A}^+$ is an $l\times m$ matrix characterized by the following properties:

– $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$

– $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$

– Both $\mathbf{A}^+\mathbf{A}$ and $\mathbf{A}\mathbf{A}^+$ are symmetric.

Even though $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ is not unique, $\mathbf{A}^+$ is unique.

– Properties

 * If $\mathbf{A}$ is invertible, then $\mathbf{A}^+ = \mathbf{A}^{-1}$.

 Proof. Because $\mathbf{A}$ is invertible, we know $\mathbf{A}\mathbf{A}^+ = \mathbf{I}$. Therefore, $\mathbf{A}^+ = \mathbf{A}^{-1}$ since the inverse of $\mathbf{A}$ is unique.

 * $(\mathbf{A}^+)^+ = \mathbf{A}$

 * $(\mathbf{A}^T)^+ = (\mathbf{A}^+)^T$

 * $(c\mathbf{A})^+ = c^{-1}\mathbf{A}^+$ if $c \neq 0$

- The optimal solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is the minimum length solution of $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$. It is called shortest least-squares solution, which is denoted by $\mathbf{x}^+$.

Remark.

– The shortest solution $\mathbf{x}^+$ is unique and in the row space of $\mathbf{A}$.

– The shortest solution $\mathbf{x}^+$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\mathbf{x}^+ = \mathbf{A}^+\mathbf{b}$.

7 Computation with Matrices

- Condition number and Relative Error

In $\mathbf{A}\mathbf{x} = \mathbf{b}$, suppose we add some small perturbation $\delta\mathbf{A}$ to $\mathbf{x}$. We want to analyze how much the product changes with respect to $\delta\mathbf{A}$. That is, how big $\delta\mathbf{b}$ is in

$$\mathbf{A}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}. \tag{7.1}$$

We model the change by comparing the relative errors $\frac{||\delta\mathbf{x}||}{||\mathbf{x}||}$ and $\frac{||\delta\mathbf{b}||}{||\mathbf{b}||}$. A *good* linear system should guarantee that $\frac{||\delta\mathbf{x}||}{||\delta\mathbf{x}||}$ is not large compared to $\frac{||\delta\mathbf{b}||}{||\mathbf{b}||}$.

– **Theorem 7.1 (Condition Number for Positive Definite Matrices)**

Let $\mathbf{A}$ be an $n \times n$ symmetric positive definite matrix. The solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ and the error $\delta\mathbf{x} = \mathbf{A}^{-1}(\delta\mathbf{b})$ always satisfy

$$||\mathbf{x}|| \leq \frac{||\mathbf{b}||}{\lambda_{\max}}, \tag{7.2}$$

$$||\delta\mathbf{x}|| \leq \frac{||\mathbf{b}||}{\lambda_{\min}}, \tag{7.3}$$

$$\frac{||\delta\mathbf{x}||}{||\mathbf{x}||} \leq \frac{\lambda_{\max}}{\lambda_{\min}} \frac{||\delta\mathbf{b}||}{||\mathbf{b}||}. \tag{7.4}$$

The ratio $c = \frac{\lambda_{\max}}{\lambda_{\min}}$ is the condition number of a positive definite matrix $\mathbf{A}$.

Proof. Because $\mathbf{A}$ is positive definite, it is invertible. Suppose its eigenvectors are given by

$$\lambda_1^{-1} \geq \lambda_2^{-1} \geq \cdots \geq \lambda_n^{-1} > 0, \tag{7.5}$$

and the corresponding orthonormal eigenvectors are given by $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Let $\delta\mathbf{b} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, then

$$\delta\mathbf{x} = \mathbf{A}^{-1}(\delta\mathbf{b}) = c_1\mathbf{A}^{-1}\mathbf{v}_1 + c_2\mathbf{A}^{-1}\mathbf{v}_2 + \cdots + c_n\mathbf{A}^{-1}\mathbf{v}_n \tag{7.6}$$

$$= c_1\lambda_1^{-1}\mathbf{v}_1 + c_2\lambda_2^{-1}\mathbf{v}_2 + \cdots + c_n\lambda_n^{-1}\mathbf{v}_n. \tag{7.7}$$

It can be seen that

$$||\delta\mathbf{b}||^2 = c_1^2 + c_2^2 + \cdots + c_n^2, \tag{7.8}$$

$$||\delta\mathbf{x}||^2 = c_1^2\lambda_1^{-2} + c_2^2\lambda_2^{-2} + \cdots + c_n^2\lambda_n^{-2}. \tag{7.9}$$

Therefore, we can write

$$(c_1^2 + c_2^2 + \cdots + c_n^2)\lambda_n^{-2} \leq ||\delta\mathbf{x}||^2 \leq (c_1^2 + c_2^2 + \cdots + c_n^2)\lambda_1^{-2} \tag{7.10}$$

$$\Rightarrow ||\delta\mathbf{b}||^2\lambda_n^{-2} \leq ||\delta\mathbf{x}||^2 \leq ||\delta\mathbf{b}||^2\lambda_1^{-2} \tag{7.11}$$

$$\Rightarrow ||\delta\mathbf{b}||\lambda_n^{-1} \leq ||\delta\mathbf{x}|| \leq ||\delta\mathbf{b}||\lambda_1^{-1}. \tag{7.12}$$

– The norm of matrix $\mathbf{A}$ is the number $||\mathbf{A}||$ defined by

$$||\mathbf{A}|| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{||\mathbf{A}\mathbf{x}||}{||\mathbf{x}||}. \tag{7.13}$$

Remark.

* The matrix norm bounds the amplifying power of a matrix. That is, $||\mathbf{Ax}|| \leq ||\mathbf{A}|| \cdot ||\mathbf{x}||$.
* For positive definite matrix $\mathbf{A}$, $||\mathbf{A}|| = \lambda_{\max}$.

– The condition number of $\mathbf{A}$ is

$$c = \begin{cases} ||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}||, & \text{if } \mathbf{A} \text{ is invertible,} \\ \infty, & \text{if } \mathbf{A} \text{ is singular.} \end{cases} \tag{7.14}$$

– ***Theorem 7.2 (Condition Number and Relative Error I)***

The relative error $\delta\mathbf{x}$ from $\delta\mathbf{b} : \mathbf{A}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$ satisfies

$$\frac{||\delta\mathbf{x}||}{||\mathbf{x}||} \leq c\frac{||\delta\mathbf{b}||}{||\mathbf{b}||}, \tag{7.15}$$

where c is the condition number of $\mathbf{A}$.

– ***Theorem 7.3 (Condition Number and Relative Error II)***

The relative error $\delta\mathbf{x}$ from $\delta\mathbf{A} : (\mathbf{A} + \delta\mathbf{A})(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b}$ satisfies

$$\frac{||\delta\mathbf{x}||}{||\mathbf{x} + \delta\mathbf{x}||} \leq c\frac{||\delta\mathbf{A}||}{||\mathbf{A}||}, \tag{7.16}$$

where c is the condition number of $\mathbf{A}$.

– ***Theorem 7.4 (Computation of Matrix Norm)***

The norm of matrix $\mathbf{A}$ is given by

$$||\mathbf{A}|| = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}. \tag{7.17}$$

The norm of matrix $\mathbf{A}^{-1}$ is given by

$$||\mathbf{A}^{-1}|| = \sqrt{\lambda_{\min}(\mathbf{A}^T\mathbf{A})^{-1}}. \tag{7.18}$$

A Concepts

B Theorems