Lecture 1

Transform of a system of equation

Three operations are used to solve systems (applied to matrices):

1. swap any two equations (switching rows)
2. multiply any equation by any $\lambda \neq 0$
3. add any multiple of any row to any other row

Definition 1.1: (row) echelon form

A matrix A is in (row) echelon form if both :

- any all-zero row lies at the very bottom of the matrix
- the pivots of the non-zero rows go strictly to the right as we go down

Pivot : left-most non-zero entry of some non-zero row of A

Definition 1.2: ***reduced*** (row) echelon form (REF)

A matrix is called reduced (row) echelon form if :

- it is in echelon form
- all of its pivots (leading entries) are 1
- any coefficient above a pivot is 0

Lecture 2

Theorem 2.1: uniqueness of reduced echelon form

Any matrix can be ***uniquely*** put in reduced echelon form using operations.

Information given by the echelon form of $\left(\begin{array}{c|c} A & b \end{array} \right)$

The row echelon form of $\left(\begin{array}{c|c} A & b \end{array} \right)$ gives us information about the number of solutions of the associated system :

- when the REF has a row $\left(\begin{array}{ccc|c} 0 & \cdots & 0 & b \end{array} \right)$ with $b \neq 0$ then no solution exists.
- if any row of the REF looks like $\left(\begin{array}{ccc|c} 0 & \cdots & 0 & 0 \end{array} \right)$, then $\exists \geq 1$ solution
  - if there are no free variables, $\exists 1$ solution

- if there are free variables, $\exists\infty$ solution.
  (if $\exists k$ free variables, this infinity is k-dimensional)

Fact : (# of free variables) + (# of pivot variables) = (# of column of A) = (# of variables)

Definition 2.2: linear combination

A ***linear combination*** of vectors $v_1, ..., v_n \in \mathbb{R}^n$ is any vector of the form :

$$C_1v_1 + C_2v_2 + \cdots + C_mv_m$$

where $C_1, C_2, ..., C_m$ are arbitrary real numbers.

Definition 2.3: span

The ***span*** of vectors $v_1, ..., v_m \in \mathbb{R}^n$ is the set of all possible combinations of $v_1, ..., v_m$ denoted *span* $\{v_1, ..., v_m\}$ this can be a point, a line, a plane, ...

Lecture 3

Parametric Form of the Solution

Let $A \in \mathbb{R}^{m\times n}$ and consider the linear system

$$Ax = b.$$

- If the system is *consistent*, then the solution set can be written in *parametric vector form*, for example:

$$x = \begin{bmatrix}1\\0\\2\end{bmatrix} + t_1\begin{bmatrix}1\\1\\0\end{bmatrix} + t_2\begin{bmatrix}-1\\0\\1\end{bmatrix}, \quad t_1, t_2 \in \mathbb{R}.$$

- If the system has *no solution*, we call it *inconsistent*, and the solution set is denoted by the empty set:

$$\{x \in \mathbb{R}^n \mid Ax = b\} = \emptyset.$$

Definition 3.1: Ax

Let

$$A = [A_1\, A_2\, \ldots\, A_n] \in \mathbb{R}^{m\times n}, \quad x = \begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} \in \mathbb{R}^n.$$

A *linear combination* of the columns of A is any vector of the form

$$Ax = x_1 A_1 + x_2 A_2 + \cdots + x_n A_n, \quad \text{with } A_i \in \mathbb{R}^m.$$

Explicitly, writing the columns as vectors with entries, we have

$$A_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix}, \quad Ax = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Theorem 3.2: Consistency of $Ax = b$

Let $A \in \mathbb{R}^{m \times n}$ be a fixed matrix with columns $A_1, \ldots, A_n$, and let $b \in \mathbb{R}^m$. Then the following statements are equivalent:

1. The equation $Ax = b$ is consistent for all $b \in \mathbb{R}^m$, i.e.,

$$b \in \operatorname{span}\{A_1, \ldots, A_n\}.$$

2. The row echelon form (REF) of A has no all-zero rows.
3. Some echelon form of A has no all-zero rows.

Lecture 4

Definition: Types of Matrix Equations

A matrix equation has the form

$$Ax = b, \quad A \in \mathbb{R}^{m \times n},\ x \in \mathbb{R}^n,\ b \in \mathbb{R}^m.$$

- The equation is called *homogeneous* if $b = 0$, i.e.,

$$Ax = 0.$$

  Such a system always has at least the trivial solution $x = 0$.

- The equation is called *inhomogeneous* if $b \neq 0$, i.e.,

$$Ax = b.$$

  The system may be consistent (having one or infinitely many solutions) or inconsistent (no solution).

Theorem 4.1: General Solution of a Linear System

Any solution $\triangle \in S$ of the system $Ax = b$ can be written as

$$\triangle = \triangle_{\text{particular}} + \triangle_{\text{homogeneous}},$$

where

- $\triangle_{\text{particular}}$ is one fixed solution of the equation $Ax = b$,
- $\triangle_{\text{homogeneous}}$ is an arbitrary solution of the homogeneous system $Ax = 0$.

Principles of Solving $Ax = b$ in Parametric Form

To solve a linear system $Ax = b$ in parametric form, follow these steps:

1. Form the augmented matrix $(A \mid b)$ and apply Gaussian elimination to obtain the row echelon form $(A^{\text{ref}} \mid b^{\text{ref}})$.
2. Identify the free variables (columns without pivots) and assign parameters to them, e.g.,
$$x_j = t, \ x_k = s, \ldots, \quad t, s, \dots \in \mathbb{R}.$$
3. Use the REF to solve for the basic (pivot) variables in terms of the free variables.
4. Separate the contributions of each parameter to form the solution as a sum of vectors:
$$x = x_{x_{\text{particular}}} + tv_1 + sv_2 + \dots$$
where $x_{\text{particular}}$ is one fixed solution of $Ax = b$, and $tv_1 + sv_2 + \dots$ is a general solution of the homogeneous system $Ax = 0$.

Definition 4.2: Linear (In)dependence of Vectors

A set of vectors $v_1, \dots, v_m \in \mathbb{R}^n$ is called *linearly dependent* if there exist scalars $c_1, \dots, c_m \in \mathbb{R}$, not all zero, such that

$$c_1v_1 + c_2v_2 + \dots + c_mv_m = 0.$$

Otherwise, the set is called *linearly independent*.

How to check linear (in)dependence: Form the matrix

$$A = (v_1 \mid v_2 \mid \ \dots \ \mid v_m) \in \mathbb{R}^{n \times m}, \quad x = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}.$$

Then consider the homogeneous system

$$Ax = 0.$$

- The vectors are *linearly independent* if this system has only the trivial solution $x = 0$.
- The vectors are *linearly dependent* if this system has infinitely many non-trivial solutions.

Theorem 5.1

Let $v_1, \dots, v_m \in \mathbb{R}^n$. Then the set $\{v_1, \dots, v_m\}$ is *linearly dependent* if and only if one of the vectors v_j can be written as a linear combination of the others.

Lecture 5

Theorem 5.2:

Assume $A \in \mathbb{R}^{m \times n}$. Then the system $Ax = b$ has

1. at least one solution $x \in \mathbb{R}^n$, $\forall b \in \mathbb{R}^m$ if and only if
$$\text{span}\{\text{columns of } A\} = \mathbb{R}^m$$
2. at most one solution $x \in \mathbb{R}^n$, $\forall b \in \mathbb{R}^m$ if and only if the columns of A are linearly independent.

Definition 5.3

Given sets S and S', a *function*

$$f : S \to S'$$

is an assignment of some $f(s) \in S'$ for all $s \in S$. The image of f is defined as

$$\text{Im}(f) = \{s' \in S' \mid \exists s \in S \text{ such that } f(s) = s'\} \subseteq S'.$$

Definition 5.4

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called *linear* if it preserves vector addition and vector scaling, i.e.

1. $f(v_1 + v_2) = f(v_1) + f(v_2)$
2. $f(cv) = cf(v)$

for all $v_1, v_2, v \in \mathbb{R}^n$ and for all $c \in \mathbb{R}$. Equivalently,

$$f(c_1 v_1 + c_2 v_2) = c_1 f(v_1) + c_2 f(v_2)$$

for all $v_1, v_2 \in \mathbb{R}^n$ and for all $c_1, c_2 \in \mathbb{R}$.

Theorem 5.5

Any linear function
$$f : \mathbb{R}^n \to \mathbb{R}^m$$
can be written uniquely as
$$f(x) = Ax, \quad \forall x \in \mathbb{R}^n,$$
for some $m \times n$ matrix A depending on f.

Lecture 6

Definition 6.1 (Injective function)

A function $f : X \to Y$ is called *injective* if it sends different elements of X to different elements of Y, i.e.
$$\forall x \neq x' \in X, \quad f(x) \neq f(x').$$
Equivalently,
$$f(x) = f(x') \;\; \Rightarrow \;\; x = x'.$$
To have a chance at f being injective, we must have
$$|X| \leq |Y|.$$

Definition 6.2 (Surjective function)

A function $f : X \to Y$ is called *surjective* if every element of Y is mapped onto by at least one element of X, i.e.
$$\forall y \in Y, \; \exists x \in X \;\text{ such that }\; f(x) = y.$$
To have a chance at f being surjective, we must have
$$|X| \geq |Y|.$$

Definition 6.3 (Bijective function)

A function $f : X \to Y$ is called *bijective* if it is both injective and surjective.
$$f \text{ is bijective} \quad \Longleftrightarrow \quad f \text{ is invertible.}$$
To have a chance at f being bijective, we must have
$$|X| = |Y|.$$

Theorem 6.4

Assume $f(x) = Ax$ for some $A \in \mathbb{R}^{m\times n}$. Then:

1. f is injective $\iff$ the columns of A are linearly independent $\iff$ REF(A) has n pivots (this can only happen if $m \geq n$).
2. f is surjective $\iff$ the columns of A span $\mathbb{R}^m$ $\iff$ REF(A) has m pivots (this can only happen if $m \leq n$).
3. f is bijective $\iff$ the columns of A are linearly independent and span $\mathbb{R}^m$ $\iff$ $m = n$ and REF(A) has $m = n$ pivots.

For any matrix, the number of pivots is $\leq m$ and $\leq n$. The situations above pertain to when we might have equalities in these inequalities.

Lecture 7

Remark : Addition of linear functions corresponds to the addition of matrices.

Properties of addition of matrices ($\forall A, B, C \in \mathbb{R}^{m\times n}$)

- $A + B = B + A$
- $A + (B + C) = (A + B) + C$
- $A + 0 = A$
- $\lambda(A + B) = \lambda A + \lambda B$
- $\lambda(\mu A) = (\lambda\mu)A$
- $(\lambda + \mu)A = \lambda A + \mu A$
- $0A = 0$

Definition 7.1: Compositions of Functions

Given two functions $f : Y \to Z$ and $g : X \to Z$, a composition is only defined when **domain**(f) **= codomain**(g).

$(f \circ g) : X \to Z$ is given by

$$(f \circ g)(x) = f(g(x)) \in Z, \quad \forall x \in X$$

Remark : For every set X, the identity function is

$$\begin{aligned} Id_X : X &\to X \\ x &\mapsto x \end{aligned}$$

Theorem 7.2: Compositions of Linear Functions

If $f : \mathbb{R}^m \to \mathbb{R}^d$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are linear, then $f \circ g : \mathbb{R}^n \to \mathbb{R}^d$ is *also linear.*

The composition of linear functions corresponds to multiplication of matrices, i.e., if $f(x) = Ax$ and $g(x) = Bx$, then

$$(f \circ g)(x) = (AB)x$$

Matrix multiplication

Rule: $A \in \mathbb{R}^{d\times m}$ and $B \in \mathbb{R}^{m\times n}$ two matrices.
Then $M = AB \in \mathbb{R}^{d\times n}$ is the matrix whose columns are $AB_1, ..., AB_n$, where $B_1, ..., B_n$ are the columns of B.
The closed formula for matrix multiplication is :

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{im}b_{mj}$$

Properties of matrix multiplication

- $A(BC) = (AB)C$ (associativity)
- $A(B + C) = AB + AC$ (distributivity)
- $(B + C)A = BA + CA$ (distributivity)
- $\lambda(AB) = (\lambda A)B = A(\lambda B)$
- $I_nA = A = AI_n$ (identity property)

  where

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

is the $n \times n$ identity matrix (square).

Important Remarks :

- $AB \neq BA$ (not commutative for general A,B)
- AB can be equal to 0 even if $A \neq 0 \neq B$
- $AB = AC$ and $BA = CA \quad \not\Rightarrow \quad B = C$

Lecture 8

Definition 8.1: Square matrices

A $m \times n$ matrix is *square* if $m = n$.

Definition 8.2: Diagonal matrices

A square matrix is *diagonal* if its off-diagonal coefficients (a_{ij} for $i \neq j$) are all 0.

Definition 8.3: Lower/Upper Triangular matrices

A square matrix is *upper (lower)* triangular if all entries below (above) diagonal, i.e. , a_{ij} for $i > j$ (a_{ij} for $i < j$), are 0. ("Strictly" triangular: triangular and diagonal is all 0).

Definition 8.4: Transposed matrices

Given a matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$, where $1 \leq i \leq m$ and $1 \leq j \leq n$, its transpose is $A^T = (a_{ji}) \in \mathbb{R}^{n \times m}$. Important properties

- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$

Lecture 9

Definition 9.1 and 9.3: Invertible Matrix

An $n \times n$ matrix A is *invertible* if there exists an $n \times n$ matrix A^{-1} such that

$$AA^{-1} = I_n \quad \text{and} \quad A^{-1}A = I_n.$$

Theorem 9.2: Linear Maps, Bijectivity and Inverses

A function is bijective if and only if it has an inverse function. In particular:

- If a function is not injective, it cannot have an inverse (one-to-one failure gives ambiguity).
- If a function is not surjective, it cannot have an inverse (some targets are not hit).

In particular, a linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ has a chance of being bijective/invertible only if $m = n$; however, this may not be enough.

Properties of Matrix Inverses

Suppose A and B are invertible square matrices of the same size.

1. **Uniqueness:** If A has an inverse, it is unique.
2. **Inverse of a product:** $(AB)^{-1} = B^{-1}A^{-1}$.

3. **Inverse of an inverse:** $(A^{-1})^{-1} = A$.
4. **Inverse of a transpose:** $(A^T)^{-1} = (A^{-1})^T$.
5. **Inverse of a chain product:** $(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1} A_1^{-1}$.
6. **Inverse of a matrix multiplied by a constant**: $(\lambda A)^{-1} = \lambda^{-1} A^{-1}$ for any $\lambda \in \mathbb{R} \setminus \{0\}$
7. **Diagonal matrix inverse:** A diagonal matrix $\operatorname{diag}(a_1, \ldots, a_n)$ is invertible iff all $a_i \neq 0$. In that case
$$\operatorname{diag}(a_1, \ldots, a_n)^{-1} = \operatorname{diag}\left(\tfrac{1}{a_1}, \ldots, \tfrac{1}{a_n}\right).$$

Caveat: $(A+B)^{-1}$ might not exist even if A^{-1}, B^{-1} exist and it does not have any reasonable formula in terms of A^{-1}, B^{-1}

Inverses and Solving Linear Systems

If A is an invertible $n \times n$ matrix, then the linear system $A\mathbf{X} = \mathbf{B}$ has the unique solution
$$\mathbf{X} = A^{-1}\mathbf{B}.$$

Computing the Inverse via Elementary Matrices

The process of Gaussian elimination, which transforms A into I_n, can be represented as a sequence of matrix multiplications by **elementary matrices**:

$$D_i^{(\lambda)} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 1 \end{pmatrix}_{n\times n} \qquad \text{(diagonal entry at position } (i,i) \text{ replaced by } \lambda)$$

$$S_{ij} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 1 \end{pmatrix}_{n\times n} \qquad \text{(rows } i \text{ and } j \text{ swapped)}$$

$$T_{ji}^{(\lambda)} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & \cdots & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 1 \end{pmatrix}_{n\times n} \qquad \text{entry of } \lambda \text{ at position } (j,i)$$

(for the latter matrix, the lone λ is on row j and column i, as indicated by the subscripts of T). The inverses are of these matrices are

$$\left(D_i^{(\lambda)}\right)^{-1} = D_i^{(1/\lambda)}, \quad (S_{ij})^{-1} = S_{ij} \quad \left(T_{ji}^{(\lambda)}\right)^{-1} = T_{ji}^{(-\lambda)}$$

Solution via Elementary Matrices

If A is expressed as a product of elementary matrices, $A = M_1 M_2 \cdots M_k$, the solution to $AX = B$ is:

$$X = (M_1 M_2 \cdots M_k)^{-1} B = M_k^{-1} \cdots M_2^{-1} M_1^{-1} B.$$

The inverses of the elementary matrices (M_i^{-1}) are also elementary matrices.

Lecture 10

Equivalent Characterizations of an Invertible Square Matrix

For a square matrix $A \in \mathbb{R}^{n\times n}$, the following are equivalent to A being invertible:

1. A is a product of elementary matrices.
2. REF(A) has n pivots.
3. The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has $\mathbf{x} = \mathbf{0}$ as its only solution.
4. The columns of A are linearly independent.
5. The linear function $f : \mathbb{R}^n \to \mathbb{R}^n$, $f(\mathbf{x}) = A\mathbf{x}$, is injective.
6. The equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution $\mathbf{x} \in \mathbb{R}^n$ for all $\mathbf{b} \in \mathbb{R}^n$.
7. The columns of A span $\mathbb{R}^n$.

8. The linear function $f : \mathbb{R}^n \to \mathbb{R}^n$, $f(\mathbf{x}) = A\mathbf{x}$, is surjective.
9. There exists an $n \times n$ matrix C such that $CA = I_n$.
10. There exists an $n \times n$ matrix C such that $AC = I_n$.
11. The transpose A^T is invertible.
12. The rows of A are linearly independent.
13. The rows of A span $\mathbb{R}^n$.

Column Operations

Column operations on a matrix $A \in \mathbb{R}^{m\times n}$ are achieved by multiplying A on the right by elementary matrices:

- Multiply j-th column by λ: $A \rightsquigarrow A \cdot D_j^{(\lambda)}$.
- Swap columns i and j: $A \rightsquigarrow A \cdot S_{ij}$.
- Add $\lambda \cdot$ (column i) to column j: $A \rightsquigarrow A \cdot T_{ij}^{(\lambda)}$.

Determinants: definition and motivation

The **determinant** $\det(A) \in \mathbb{R}$ of a square matrix A indicates whether A is invertible.

- **Theorem 10.1:** $\det(A) \neq 0 \iff A$ is invertible.

The determinant represents the factor by which the linear function $f(\mathbf{x}) = A\mathbf{x}$ (where $A \in \mathbb{R}^{n\times n}$) changes the volume of a region C in $\mathbb{R}^n$.

- **Definition 10.2:** $\det(A) = \pm\dfrac{\text{vol}(C')}{\text{vol}(C)}$

where $C' = f(C)$ and vol means length/area/volume/hypervolume. The sign ($\pm$) accounts for changes in orientation (e.g., reflection changes orientation, $\det(A) = -1$).

Properties of determinants

- $\det(a) = a$ (for 1×1 matrix).
- $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ (for 2×2 matrix).
- $\det(I_n) = 1$ (the identity matrix doesn't rescale).
- **Multiplicative Property, Theorem 10.3:** $\det(AB) = \det(A)\det(B)$.

- $\det(A^{-1}) = \frac{1}{\det(A)}$.
- The determinant of a triangular (upper or lower) or diagonal matrix is the product of its diagonal entries:

$$\det \begin{pmatrix} a_1 & * & \\ & \ddots & \\ 0 & & a_n \end{pmatrix} = \det \begin{pmatrix} a_1 & 0 & \\ & \ddots & \\ 0 & & a_n \end{pmatrix} = a_1 a_2 \dots a_n.$$

Determinants of Elementary Matrices

Using $\det(AB) = \det(A)\det(B)$ and the factorization $A = M_1 \dots M_k \cdot \text{REF}(A)$, the determinant can be computed via Gaussian elimination:

- $\det(D_i^{(\lambda)}) = \lambda$ (rescaling).
- $\det(S_{ij}) = -1$ (row swap/reflection).
- $\det(T_{ij}^{(\lambda)}) = 1$ (shearing).

In general, $\det(A) = (-1)^{(\#\text{ row swaps})} \cdot$ (product of all row rescaling factors) during Gaussian elimination to REF(A).

Lecture 11

Theorem 11.1

If matrices A, B, and C have the same rows $1, \dots, i-1, i+1, \dots, n$, but the i-th row of C is the sum of the i-th rows of A and B:

$$A = \begin{pmatrix} & \text{same rows as } B, C & \\ a_{i1} & \dots & a_{in} \\ & \text{same rows as } B, C & \end{pmatrix}, \quad B = \begin{pmatrix} & \text{same rows as } A, C & \\ b_{i1} & \dots & b_{in} \\ & \text{same rows as } A, C & \end{pmatrix},$$

$$C = \begin{pmatrix} & \text{same rows as } A, B & \\ a_{i1} + b_{i1} & \dots & a_{in} + b_{in} \\ & \text{same rows as } A, B & \end{pmatrix}.$$

Then, under this setup,

$$\det(C) = \det(A) + \det(B).$$

Theorem 11.2 (Laplace Cofactor Expansion)

Let $A \in \mathbb{R}^{n \times n}$. The determinant of A can be expanded along any row or column.

Expansion along the i-th row:

$$\det(A) = (-1)^{i+1} a_{i1} \det(A_{i1}) + (-1)^{i+2} a_{i2} \det(A_{i2}) + \dots + (-1)^{i+n} a_{in} \det(A_{in})$$

where A_{ij} denotes the matrix obtained from A by removing the i-th row and the j-th column.

Expansion along the j-th column:

$$\det(A) = (-1)^{1+j}a_{1j}\det(A_{1j})+(-1)^{2+j}a_{2j}\det(A_{2j})+\cdots+(-1)^{n+j}a_{nj}\det(A_{nj})$$

Lecture 12

Definition 12.1 (Vector Space)

A **vector space** is a set V together with two operations:

- **Addition:** for all $v, w \in V$, there exists $v + w \in V$,
- **Scalar multiplication:** for all $v \in V$ and all $\lambda \in \mathbb{R}$, there exists $\lambda v \in V$,

satisfying the following axioms for all $u, v, w \in V$ and all scalars $\lambda, \mu \in \mathbb{R}$:

1. $v + w = w + v$ (commutativity)
2. $(v + w) + u = v + (w + u)$ (associativity)
3. There exists a special element $0_V \in V$ such that $v + 0_V = 0_V + v = v$ (neutral element)
4. For every $v \in V$, there exists an element $-v \in V$ such that $v + (-v) = (-v) + v = 0_V$ (opposite vector)
5. $(\lambda + \mu)v = \lambda v + \mu v$ (distributivity 1)
6. $\lambda(v + w) = \lambda v + \lambda w$ (distributivity 2)
7. $(\lambda\mu)v = \lambda(\mu v)$ (associativity of scalar multiplication)
8. $1v = v$ (unit property)

Elements of V are called **vectors**.

Definition 12.2 (Subspace)

Let V be a vector space (over $\mathbb{R}$). A nonempty subset $W \subseteq V$ is called a **subspace** of V if it is closed under the two operations:

$$\begin{cases} \textbf{(addition)} & \forall\, v, w \in W: \; v + w \in W, \\ \textbf{(scalar multiplication)} & \forall\, v \in W,\; \forall\, \lambda \in \mathbb{R}: \; \lambda v \in W. \end{cases}$$

Note. Such a W must contain the zero vector 0_V: pick any $v \in W$; since $0 \in \mathbb{R}$ and W is closed under scalar multiplication, $0 \cdot v = 0_V \in W$.

Theorem 12.3

If W is a subspace of a vector space V, then W is itself a vector space in its own right, with respect to addition and scalar multiplication inherited from V.

That is,

$$\begin{aligned} w, w' \in W \subseteq V \quad &\Longrightarrow \quad w + w' \in W \subseteq V, \\ \lambda \in \mathbb{R},\ w \in W \subseteq V \quad &\Longrightarrow \quad \lambda w \in W \subseteq V. \end{aligned}$$

Definition 12.4 (Span)

In any vector space V, take vectors $v_1, \ldots, v_k \in V$ for $k \geq 1$. The **span** of these vectors is defined as

$$\operatorname{span}\{v_1, \ldots, v_k\} = \{\, c_1 v_1 + \cdots + c_k v_k \mid c_1, \ldots, c_k \in \mathbb{R} \,\} \subseteq V.$$

For any vectors $v_1, \ldots, v_k \in V$, the set $\operatorname{span}\{v_1, \ldots, v_k\}$ is a subspace of V.

Lecture 13

Definition 13.1

We call vectors $v_1, \ldots, v_k \in V$ a **basis** of V if:

- $\operatorname{span}\{v_1, \ldots, v_k\} = V$
- the set $\{v_1, \ldots, v_k\}$ is minimal with this property

This is equivalent to saying that $v_1, \ldots, v_k$ are **linearly independent**, i.e.

$$\nexists\, \lambda_1, \ldots, \lambda_k \in \mathbb{R},\ \text{not all zero, such that } \lambda_1 v_1 + \cdots + \lambda_k v_k = 0_V.$$

Conversely, $v_1, \ldots, v_k$ are called **linearly dependent** if such λ_i exist.

Theorem 13.2

The change of basis matrices A and B from one basis to another and back are **inverses** of each other.

Corollary 13.3

Any two bases of a vector space have the same number of vectors.

Definition 13.4

The **dimension** of a vector space V is the number of vectors in any basis of V.

Coordinates

Given a basis $\underline{v} = \{v_1, \ldots, v_n\}$ of a vector space V, any vector $z \in V$ is completely determined by its coordinate vector

$$[z]_{\underline{v}} \in \mathbb{R}^n$$

with respect to the given basis.

Theorem 13.5 (Coordinate Change Formula)

Consider bases $\underline{v} = \{v_1, \ldots, v_n\}$ and $\underline{w} = \{w_1, \ldots, w_n\}$ of one and the same vector space V. Consider the matrix

$$A \in \mathbb{R}^{n \times n}$$

whose columns are the coordinate vectors of $w_1, \ldots, w_n$ in the basis $\underline{v}$. The above matrix is called a "change of coordinate matrix from the basis $\underline{w}$ to the basis $\underline{v}$ because

$$[z]_{\underline{v}} = A[z]_{\underline{w}}$$

for any vector $z \in V$.

Theorem 13.6

Let A the change of coordinate matrix from $\underline{w}$ to $\underline{v}$, and B the change of coordinate matrix from $\underline{z}$ to $\underline{w}$.
Then the matrix $C = AB$ changes from the basis $\underline{z}$ directly to the basis $\underline{v}$.

Lecture 14

Definition 14.1

Given vector spaces V and W, a function $f : W \to V$ is called **linear** if:

- $f(w + w') = f(w) + f(w'), \quad \forall w, w' \in W$
- $f(\lambda w) = \lambda f(w), \quad \forall \lambda \in \mathbb{R}, \forall w \in W$

Theorem 14.2

If a matrix A corresponds to a linear function $f : W \to V$ with respect to bases $\underline{v}$ and $\underline{w}$, then

$$[f(z)]_{\underline{v}} = A[z]_{\underline{w}}, \quad \forall z \in W$$

Theorem 14.3

If

$$Z \xrightarrow{g} W \xrightarrow{f} V$$

and we fix bases of Z, W, V (respectively), then:

- $f \circ g$ corresponds to the matrix AB
- f^{-1} corresponds to the matrix A^{-1}

Definition 14.4

$$\mathrm{Ker}(f) = \{\, w \in W \mid f(w) = 0_V \,\} \subseteq W$$

$$\mathrm{Im}(f) = \{\, v \in V \mid \exists w \in W \text{ such that } f(w) = v \,\} \subseteq V$$

Both $\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$ are subspaces of W and V respectively. Proof for $\mathrm{Ker}(f)$:

- $0_W \in \ker(f)$ because $f(0_W) = 0_V$
- If $w, w' \in \ker(f)$, then $f(w + w') = f(w) + f(w') = 0_V \Rightarrow w + w' \in \ker(f)$
- If $w \in \ker(f)$ and $\lambda \in \mathbb{R}$, then $f(\lambda w) = \lambda f(w) = 0_V \Rightarrow \lambda w \in \ker(f)$

Similarly, $\mathrm{Im}(f)$ is a subspace.

Theorem 14.5

- f is **injective** $\Leftrightarrow \mathrm{Ker}(f) = \{0_W\}$
- f is **surjective** $\Leftrightarrow \mathrm{Im}(f) = V$

Definition 14.6

$\mathbb{P}_d = \{$vector space of polynomials of degree $\leq d\}$. A basis of $\mathbb{P}_d$ is given by $\{x^0, x^1, x^2, \ldots, x^d\}$, so

$$\dim(\mathbb{P}_d) = d + 1$$

Let $f : \mathbb{P}_d \to \mathbb{P}_d$ be the derivative operator, i.e.

$$f(x^k) = (x^k)' = kx^{k-1}$$

Claim: f is a linear function on $\mathbb{P}_d$ since:

$$(P + Q)' = P' + Q', \quad (\lambda P)' = \lambda P'$$

Lecture 15

Definition 15.1 (Isomorphism)

An **isomorphism** is a bijective/invertible linear function $f : V \to W$.
We write $V \cong W$ (and say "V is isomorphic to W").

- $V \cong V$
- if $V \cong W$, then $W \cong V$ (if $\exists f$ going one way, then f^{-1} goes the other way)
- if $V \cong W$ and $W \cong Z$, then $V \cong Z$

Theorem 15.2

If $f : V \xrightarrow{\cong} W$, then:

- $v_1, \ldots, v_n$ are linearly independent in $V \iff f(v_1), \ldots, f(v_n)$ are linearly independent in W
- $v_1, \ldots, v_n$ span/generate $V \iff f(v_1), \ldots, f(v_n)$ span/generate W
- $v_1, \ldots, v_n$ are a basis of $V \iff f(v_1), \ldots, f(v_n)$ are a basis of W
$\implies dim(V) = dim(W)$

Based on the following fact:
$v = c_1 v_1 + \ldots + c_n v_n \iff f(v) = c_1 f(v_1) + \ldots c_n f(v_n)$

Theorem 15.3

Any vector space V of dimension n is **isomorphic** to $\mathbb{R}^n$.

$$\begin{aligned}
\implies \text{ up to isomorphism, } & \text{any v.s of dim 0 is } \mathbb{R}^0 = \{0\} \\
& \text{any v.s of dim 1 is } \mathbb{R}^1 \\
& \text{any v.s of dim 2 is } \mathbb{R}^2 \\
& \vdots \\
& \text{any v.s of dim } \infty \text{ is } \mathbb{R}^\infty
\end{aligned}$$

Big picture

W of dim $n \rightsquigarrow$ basis $\underline{w} = \{w_1, \ldots, w_n\}$
V of dim $m \rightsquigarrow$ basis $\underline{v} = \{v_1, \ldots, v_m\}$

Suppose you have a linear function $f : W \xrightarrow{f} V$ and you want to find the $m \times n$ matrix A which represents it.

$$\Phi : \mathbb{R}^n \xleftarrow[h]{\cong} W \xrightarrow{f} V \xrightarrow[g]{\cong} \mathbb{R}^m$$

$$\text{where} \quad \begin{array}{ll} g : V \xrightarrow{\cong} \mathbb{R}^m, & g(z) = [z]_{\underline{v}}, \\ h : W \xrightarrow{\cong} \mathbb{R}^n, & h(z) = [z]_{\underline{w}} \end{array}$$

$$\Phi = g \circ f \circ h^{-1}, \quad \mathbb{R}^n \to \mathbb{R}^m \text{ linear}$$

$\Longrightarrow$ corresponds to $A \in \mathbb{R}^{m \times n}$, this is precisely the matrix which represents f

Lecture 16

Change of basis formula for matrices

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be given by $f(x) = Bx$. Consider bases $\underline{v} = \{v_1, \ldots, v_m\}$ and $\underline{w} = \{w_1, \ldots, w_n\}$ of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, and let

$$\Phi : \mathbb{R}^n \xrightarrow{h^{-1}} \mathbb{R}^n \xrightarrow{f} \mathbb{R}^m \xrightarrow{g} \mathbb{R}^m$$

where $g(v_i) = e_i$ and $h(w_i) = e_i$ for all i (which implies $g(x) = P_{\underline{v}}^{-1}x$ and $h(x) = P_{\underline{w}}^{-1}x$). The change of basis formula for matrices says that $\Phi(x) = Ax$, where

$$A = P_{\underline{v}}^{-1} B P_{\underline{w}}$$

In other words,

$$A = P_{\underline{v} \leftarrow \underline{e}} \, B \, P_{\underline{e} \leftarrow \underline{w}},$$

so this amounts to "changing B from the $\underline{e} - \underline{e}$ basis to A in the $\underline{v} - \underline{w}$ basis", i.e.

$$A[z]_{\underline{w}} = P_{\underline{v} \leftarrow \underline{e}} \, B \, P_{\underline{e} \leftarrow \underline{w}}[z]_{\underline{w}} = P_{\underline{v} \leftarrow \underline{e}} \, Bz = [Bz]_{\underline{v}}.$$

Theorem 16.1

$$A[z]_{\underline{w}} = [Bz]_{\underline{v}} \qquad , \forall z \in \mathbb{R}^n$$

Corollary 16.2

Any two vector spaces V and W of the same dimension are isomorphic: $V \cong W$.
$\Longrightarrow$ The dimension is the most important property of a vector space

Definition 16.3

$\forall A \in \mathbb{R}^{m \times n}$

- $rank(A) := \dim \operatorname{Col}(A)$
- $nullity(A) := \dim \operatorname{Ker}(A)$

Theorem 16.4 (Rank-Nullity)

$\forall A \in \mathbb{R}^{m\times n}$ with reduced echelon form $\tilde{A}$,

$$n = \text{rank}(A) + \text{nullity}(A) \;=\; \#\text{pivots of } \tilde{A} \;+\; \#\text{free columns of } \tilde{A}$$

Theorem 16.5

$0 \leq rank(A) \leq min(m, n)$
$max(0, n - m) \leq nullity(A) \leq n$

$n = n - 0 \geq nullity(A) = n - rank(A) \geq n - min(m, n) = max(0, n - m)$

Theorem 16.6

$\forall r \in \{0, \ldots, \min(m, n)\}$, $\exists$ a matrix $A \in \mathbb{R}^{m\times n}$ of rank r.

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix}$$

(the only matrix of rank 0 is the all-zero matrix)

Theorem 16.7

For any square matrix $A \in \mathbb{R}^{n\times n}$, the following statements are equivalent:

- $rank(A) = n$
- A invertible
- $det(A) \neq 0$
- columns of A form a basis of $\mathbb{R}^n$
- $\text{Ker}(A) = \{0\}$

Theorem 16.8

For any $B \in \mathbb{R}^{m\times n}$ of rank r, there exists invertible matrices $P \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{n\times n}$ such that

$$B = P \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{pmatrix} Q$$

Proposition

$\forall$ vector space V of dimension m, suppose you have linearly independent vectors $\{v_1, \ldots, v_k\} \in V$; then we can find $v_{k+1}, \ldots, v_m \in V$ such that $\{v_1, \ldots, v_m\}$ is a basis of V.

Step 1) Pick $v_{k+1} \notin \text{span}\{v_1, \ldots, v_k\}$.
Is $V = \text{span}\{v_1, \ldots, v_k, v_{k+1}\}$? If yes, we are done; If no, go to Step 2.

Step 2) Pick $v_{k+2} \notin \text{span}\{v_1, \ldots, v_k, v_{k+1}\}$.
Is $V = \text{span}\{v_1, \ldots, v_k, v_{k+1}, v_{k+2}\}$? If yes, we are done; If no, go to Step 3.

$\vdots$

Step $m - k$) Pick $v_m \notin \text{span}\{v_1, \ldots, v_{m-1}\}$.

Then $v_1, \ldots, v_m$ are linearly independent, and so they form a basis of the m-dimensional space V.

Lecture 17

Definition 17.1 (Diagonalization)

A square matrix B is called *diagonalizable* if it is similar to a diagonal matrix, i.e.

$$B = P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} P^{-1},$$

where P is an invertible matrix and $\lambda_1, \ldots, \lambda_n$ are numbers.

Theorem 17.2

Almost any square matrix is diagonalizable.

Principle. A matrix $B \in \mathbb{R}^{n \times n}$ is diagonalizable if and only if there exists a basis $v_1, \ldots, v_n$ of $\mathbb{R}^n$ such that

$$Bv_i = \lambda_i v_i, \qquad \forall\, i \in \{1, \ldots, n\}.$$

That is, B dilates each basis vector v_i by a factor of λ_i.

Definition 17.3 (Complex Number)

A *complex number* is an expression of the form

$$z = a + bi = a + ib,$$

where i is a symbol such that $i^2 = -1$, and $a, b \in \mathbb{R}$.

Let $\mathbb{C}$ denote the set of all complex numbers.

Definition 17.4 (Conjugate and Absolute Value)

For every complex number $z = a + bi \in \mathbb{C}$:

- **Conjugate:** The conjugate of z is
$$\overline{z} = a - bi \in \mathbb{C}.$$
For example,
$$\overline{2 + 3i} = 2 - 3i.$$
- **Absolute value:** The absolute value of z is
$$|z| = \sqrt{a^2 + b^2} \in \mathbb{R}_{\geq 0}.$$

Proposition 17.5

For every complex number z, we have
$$z \cdot \overline{z} = |z|^2.$$

Theorem 17.6

Let $z, w \in \mathbb{C}$. Then:
$$\overline{z \pm w} = \overline{z} \pm \overline{w}, \qquad \overline{zw} = \overline{z}\,\overline{w}, \qquad \overline{\left(\frac{z}{w}\right)} = \frac{\overline{z}}{\overline{w}} \quad (w \neq 0).$$
$$|zw| = |z| \cdot |w|, \qquad \left|\frac{z}{w}\right| = \frac{|z|}{|w|} \quad (w \neq 0).$$
$$\overline{(\overline{z})} = z.$$

Geometric Interpretation of Complex Numbers

A complex number $z = a + bi$ can be represented as a point in the plane, with horizontal coordinate a (real part) and vertical coordinate b (imaginary part). This identifies $\mathbb{C}$ with the plane $\mathbb{R}^2$.

Polar coordinates. Every complex number $z \neq 0$ can also be described using its *modulus* r and its *argument* θ:
$$r = |z| \in \mathbb{R}_{\geq 0}, \qquad \theta = \arg(z) \in [0, 2\pi).$$

These satisfy
$$a = r\cos\theta, \qquad b = r\sin\theta.$$

Polar form of a complex number. Using these coordinates, we can write
$$z = r(\cos\theta + i\sin\theta).$$

Here:

- $r = |z|$ is the distance of z from the origin,
- $\theta = \arg(z)$ is the angle between the positive real axis and the line joining 0 to z, called the *argument* of z.

Proposition 17.7

For all $z, w \in \mathbb{C}$:

$$|zw| = |z|\,|w|,$$

and

$$\arg(zw) = \arg(z) + \arg(w) \mod 2\pi.$$

Lecture 18

Theorem 18.1

Any polynomial $P(t)$ of degree n has a complete set of n roots in $\mathbb{C}$. That is, there exist complex numbers $\lambda_1, \dots, \lambda_n$ such that

$$P(t) = \text{const} \cdot (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n).$$

The roots λ_i may repeat, i.e. one may have $\lambda_i = \lambda_j$.

Definition 18.2 (Eigenvalues and Eigenvectors)

Let $B \in \mathbb{R}^{n \times n}$. We call a number $\lambda \in \mathbb{C}$ an *eigenvalue* of B if there exists a nonzero vector $v \in \mathbb{C}^n$ such that

$$Bv = \lambda v.$$

Any such nonzero vector v is called an *eigenvector* of B corresponding to λ.

A matrix $B \in \mathbb{R}^{n \times n}$ can have $1, 2, \dots$ or n distinct eigenvalues.

A matrix $B \in \mathbb{R}^{n \times n}$ can have $1, 2, \dots$ or n linearly independent eigenvectors.

Theorem 18.3

An $n \times n$ matrix is diagonalizable if and only if it has n linearly independent eigenvectors.

Definition 18.4 (Characteristic Polynomial)

Given an $n \times n$ matrix B, its *characteristic polynomial* is the degree-n polynomial

$$\chi_B(t) = \det(tI_n - B).$$

Corollary 18.5

The eigenvalues of B are the roots of its characteristic polynomial $\chi_B(t)$.
In particular,

$$\chi_B(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n),$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of B (they may repeat, i.e. $\lambda_i = \lambda_j$).

How to find eigenvalues and eigenvectors

Step 1: Compute the characteristic polynomial of B:

$$\chi_B(t) = \det(tI_n - B).$$

Step 2: The eigenvalues of B are the roots of $\chi_B(t)$.

Step 3: For each eigenvalue λ_i, find the corresponding eigenvectors v_i by solving

$$Bv_i = \lambda_i v_i \quad \iff \quad (B - \lambda_i I_n)v_i = 0.$$

This is the same thing as calculating $\mathrm{Ker}(B - \lambda_i I_n)$.

Lecture 19

Defintion 19.1: Spectrum of a Matrix

Let B be a matrix. The *spectrum* of B, denoted by $\mathrm{Spec}(B)$, is the **multiset of its eigenvalues**. That is, the eigenvalues are listed with their algebraic multiplicities.

$$\text{Example: } \{2, 5, 2\} = \{5, 2, 2\} = \{2, 2, 5\}.$$

Definition 19.2: Similarity of Matrices

Let A and B be two square matrices of the same size. They are called *similar* (or *conjugate*) if

$$B = PAP^{-1}$$

for some invertible matrix P. In this case, we write

$$A \sim B.$$

Theorem 19.3

If A and B are similar, then

$$\mathrm{Spec}(A) = \mathrm{Spec}(B).$$

Caution. The converse is not true. For example, let

$$A = \lambda I_n, \qquad B = J_n^{(\lambda)} = \begin{pmatrix} \lambda & 1 & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix},$$

Both have spectrum $\{\lambda, \dots, \lambda\}$, but they are *not* similar.

Theorem 19.4

The complex roots of a real polynomial $f(t)$ always come in complex conjugate pairs. That is,

$$f(\lambda) = 0 \iff f(\overline{\lambda}) = 0.$$

Consequence. If a real matrix has $a + bi$ as an eigenvalue, then it also has $a - bi$ as an eigenvalue.

Theorem 19.4 (continued)

If v is an eigenvector of a real matrix B corresponding to

$$\lambda = a + bi,$$

then $\overline{v}$ is an eigenvector of B corresponding to

$$\overline{\lambda} = a - bi.$$

Definition 19.5

Assume that the distinct eigenvalues of a matrix $B \in \mathbb{R}^{n \times n}$ are $\lambda_1, \dots, \lambda_r$ (with $r \leq n$).
If the characteristic polynomial of B factors as

$$\chi_B(t) = (t - \lambda_1)^{d_1}(t - \lambda_2)^{d_2} \cdots (t - \lambda_r)^{d_r},$$

then the integers

$$d_1, \dots, d_r$$

are called the *algebraic multiplicities* of

$$\lambda_1, \dots, \lambda_r.$$

Theorem 19.6

The sum of the algebraic multiplicities of all eigenvalues of an $n \times n$ matrix is equal to n.

Moreover, if an eigenvalue λ has algebraic multiplicity d, then its complex conjugate $\overline{\lambda}$ also has algebraic multiplicity d.

Definition 19.7

Let λ be an eigenvalue of a matrix B. The *geometric multiplicity* of λ is defined as

$$\dim \ker(B - \lambda I_n).$$

Equivalently,

$$\dim\{\, v \in \mathbb{R}^n \mid Bv = \lambda v \,\} \;=\; \dim\bigl(\{\text{set of eigenvectors for } \lambda\} \cup \{0\}\bigr).$$

Theorem 19.8

For any eigenvalue λ of a matrix B,

$$1 \;\leq\; \text{geometric multiplicity of } \lambda \;\leq\; \text{algebraic multiplicity of } \lambda.$$

In words: the geometric multiplicity is the maximum number of linearly independent eigenvectors associated with λ, and it can never exceed its algebraic multiplicity.

A matrix is *diagonalizable* if and only if all its eigenvalues satisfy

$$\text{geometric multiplicity} = \text{algebraic multiplicity}.$$

Lecture 20

Definition 20.1

The **eigenspace** of a matrix B with respect to the eigenvalue λ is

$$V_\lambda = \ker(B - \lambda I_n) = \{0\} \cup \{\text{eigenvectors for } \lambda\}.$$

A vector $v \in \mathbb{R}^n$ belongs to V_λ exactly when

$$(B - \lambda I_n)v = 0 \quad \Longleftrightarrow \quad Bv = \lambda v \quad \Longleftrightarrow \quad v \text{ is an eigenvector for } \lambda \text{ or } v = 0.$$

The **geometric multiplicity** of λ is

$$\dim V_\lambda,$$

the maximal number of linearly independent eigenvectors corresponding to λ.

Note. V_λ is a subspace of $\mathbb{R}^n$ for every eigenvalue λ.

Theorem 20.2

The different eigenspaces of any matrix B are linearly independent. That is, if you choose one nonzero eigenvector from each eigenspace, then all these eigenvectors will be linearly independent.

Suppose $\lambda_1, \dots, \lambda_r$ are the distinct eigenvalues of B, and let

$$V_{\lambda_1}, \ \dots, \ V_{\lambda_r}$$

be the corresponding eigenspaces, which are subspaces of $\mathbb{R}^n$.
Then for any choice of nonzero vectors

$$0 \neq v_1 \in V_{\lambda_1}, \quad \dots, \quad 0 \neq v_r \in V_{\lambda_r},$$

the set

$$\{v_1, \dots, v_r\}$$

is linearly independent.

Corollary 20.3

Any two eigenspaces intersect only in $\{0\}$. Equivalently, no nonzero vector v can be an eigenvector for two distinct eigenvalues $\lambda \neq \mu$.

Definition 20.4

Let

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

The **trace** of B is

$$\operatorname{tr}(B) = b_{11} + b_{22} + \cdots + b_{nn}.$$

Theorem 20.5

For every matrix $B \in \mathbb{R}^{n \times n}$ (not necessarily diagonalizable), the sum of its eigenvalues (counted with algebraic multiplicities) satisfies

$$\lambda_1 + \cdots + \lambda_n = \operatorname{tr}(B).$$

Proposition 20.6

Let $X \in \mathbb{R}^{m \times n}$ and $Y \in \mathbb{R}^{n \times m}$. Then

$$\operatorname{tr}(XY) = \operatorname{tr}(YX),$$

where XY is an $m \times m$ matrix and YX is an $n \times n$ matrix.

Lecture 21

Theorem 21.1

Recall that similar matrices A and B ($A \sim B$ if $\exists P$ s.t. $B = PAP^{-1}$) have the same spectra :

$$\{\text{eigenvalues of } A\} = \{\text{eigenvalues of } B\}$$

Moreover, the eigenspace $V_\lambda^{(A)}$ of A for the eigenvalue λ is isomorphic to the eigenspace $V_\lambda^{(B)}$ of B for the eigenvalue λ.

Definition 21.2

The **inner/dot product** of $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ and $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ is :

$$u \cdot v = u_1 v_1 + \cdots + u_n v_n$$

- $u \cdot v = u^\top v = \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 + \cdots + u_n v_n \end{pmatrix}$
- $u \cdot v = v \cdot u$
- $u \cdot (v + v') = u \cdot v + u \cdot v'$
- $(u + u') \cdot v = u \cdot v + u' \cdot v$
- $u \cdot (\lambda v) = (\lambda u) \cdot v = \lambda (u \cdot v) \quad \forall \lambda \in \mathbb{R}$
- $u \cdot \underset{\in \mathbb{R}^n}{0} = \underset{\in \mathbb{R}^n}{0} \cdot u = \underset{\in \mathbb{R}}{0}$

Definition & Proposition 21.3

$\forall$ vector $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ we have :

$v \cdot v = v_1^2 + v_2^2 + \cdots + v_n^2 \geq 0$

and $\|v\| = \sqrt{v \cdot v} \in \mathbb{R}_{\geq 0}$ is called the **length/norm** of v

- the above inequality is an equality $\iff v = 0$

- given $u, v \in \mathbb{R}^n$, the **distance** between them is $\|u - v\| \in \mathbb{R}_{\geq 0}$

Note : distance $= 0 \iff u = v$

Theorem 21.4

The angle θ between vectors $u, v \in \mathbb{R}^n$ satisfies $\cos\theta = \frac{u \cdot v}{\|u\|\|v\|}$

- $\cos\theta = \frac{u \cdot v}{\|u\|\|v\|}$ is unchanged by scaling u and v (up to a sign).

  $\alpha, \beta \in \mathbb{R}$, $\frac{(\alpha u) \cdot (\beta v)}{\|\alpha u\|\|\beta v\|} = \frac{\alpha\beta\ u \cdot v}{|\alpha\beta|\ \|u\|\|v\|} = sign(\alpha\beta)\frac{u \cdot v}{\|u\|\|v\|}$
- $\forall v \neq 0$ in $\mathbb{R}^n$, **normalization/unit vector** of v is $\frac{v}{\|v\|} \in \mathbb{R}^n$
- Cauchy-Schwartz inequality : $-1 \leq \cos\theta \leq 1 \Rightarrow -\|u\|\|v\| \leq u \cdot v \leq \|u\|\|v\|$
- Equality in C-S, i.e. $u \cdot v = \begin{cases} \|u\|\|v\| & \text{if } \cos\theta = 1 \iff \theta = 0 \\ -\|u\|\|v\| & \text{if } \cos\theta = -1 \iff \theta = \pi \end{cases}$
- $u \perp v$ $\underset{\text{(orthogonality)}}{\text{perpendicular}} \iff u \cdot v = 0$

Definition 21.5

Suppose you have subspaces $V, W \subseteq \mathbb{R}^n$. We say that

$$V \perp W$$

if $v \perp w,\ \forall v \in V,\ \forall w \in W$

Theorem 21.6

Suppose we have subspaces $V \subset \mathbb{R}^n$ with basis $v_1, \ldots, v_k$, $W \subset \mathbb{R}^n$ with basis $w_1, \ldots, w_l$

$$V \perp W \iff v_i \perp w_j, \quad \forall 1 \leq i \leq k,\ 1 \leq j \leq l$$

Note : $v_i \perp w_j \iff v_i \cdot w_j = 0$

Corollary 21.7

If V and W are orthogonal subspaces of $\mathbb{R}^n$ (i.e. $V \perp W$), then $V \cap W = \{0\}$

Definition 21.8

Given a subspace $V \subset \mathbb{R}^n$, its **orthogonal complement**

$$V^\perp = \{u \in \mathbb{R}^n \text{ s.t. } u \perp V, \text{ i.e. } u \perp v,\ \forall v \in V\}$$

a.k.a. the biggest subspace of $\mathbb{R}^n$ which is $\perp V$

Proposition 21.9

$$\left(V^{\perp}\right)^{\perp} = V$$

Lecture 22

Theorem 22.1

For any matrix A, we have

$$\text{Col}(A)^{\perp} = \text{Ker}(A^{\top})$$

or the equivalent formula

$$\text{Row}(A)^{\perp} = \text{Ker}(A)$$

Theorem 22.2

$\forall$ subspaces $V \in \mathbb{R}^n$,

$$\dim(V) + \dim(V^{\perp}) = n$$

Definition 22.3

A set of vector $\{v_1, \ldots, v_k\}$ of $\mathbb{R}^n$ is called an **orthogonal set** if $v_i \perp v_j \quad \forall i \neq j$

Proposition 22.4

Any orthogonal set (that does not contain 0) is linearly independent.

Definition 22.5

Let $V \subset \mathbb{R}^n$; an **orthogonal basis** $\{v_1, \ldots, v_k\}$ of V is a basis of V consisting of mutually orthogonal vectors.

Proposition 22.6

Any subspace $V \subset \mathbb{R}^n$ has (infinitely many) orthogonal bases.

Theorem 22.7

Let subspace $V \subseteq \mathbb{R}^n$ with **orthogonal** basis $v_1, \ldots, v_k$ of V. For an arbitrary $w \in V$, the constants $c_1, \ldots c_k$ such that $w = c_1 v_1 + \cdots + c_k v_k$ are given by

$$c_1 = \frac{w \cdot v_1}{\|v_1\|^2}, \quad \cdots, \quad c_k = \frac{w \cdot v_k}{\|v_k\|^2}$$

Theorem 22.8

Let $V \subset \mathbb{R}^n$ be any subspace. $\forall$ vector $w \in \mathbb{R}^n$, we can decompose it uniquely as

$$w = \underbrace{w_V}_{proj_V(w)} + \underbrace{w_{V^\perp}}_{proj_{V^\perp}(w)}$$

where $w_V \in V$ and $w_{V^\perp} \in V^\perp$

If you have an orthogonal basis $\{v_1, \dots, v_k\}$ of V:

$$proj_V(w) = \frac{w \cdot v_1}{\|v_1\|^2}\, v_1 + \cdots + \frac{w \cdot v_k}{\|v_k\|^2}\, v_k$$

Note : $\frac{w \cdot v_i}{\|v_i\|^2}$ are coefficients in $\mathbb{R}$. Also, the summands

$$\frac{w \cdot v_i}{\|v_i\|^2}\, v_i$$

are unchanged by rescaling $v_1, \dots, v_k$ independently

Definition 22.9

An **orthonormal basis** is an orthogonal basis where all vectors have length 1. (e.g. $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$)

$$v_i \cdot v_i = 1 \iff \|v_1\| = \cdots = \|v_k\| = 1$$
$$v_i \cdot v_j = 0 \iff v_i \perp v_j, \quad \forall i \neq j$$

Note : orthonormal bases do not need denominators in the formula for $proj_V(w)$.

If $\{v_1, \dots v_k\}$ is an orthonormal basis, then for $A = (v_1| \cdots |v_k)$ we have

$$A^\top A = \begin{pmatrix} v_1 \cdot v_1 & \cdots & v_1 \cdot v_k \\ \vdots & \ddots & \vdots \\ v_k \cdot v_1 & \cdots & v_k \cdot v_k \end{pmatrix} = I_k$$

Definition 22.10

Square matrices $A \in \mathbb{R}^{n \times n}$ s.t. $A^\top A = I_n$ are called **orthogonal**.

Lecture 23

Theorem 23.1

Let V be a subspace of $\mathbb{R}^n$ and let $v_1, \dots, v_k$ be an orthonormal basis of V. Define

$$A = (v_1 \ \dots \ v_k), \qquad A^T A = I_k.$$

Then the projection of a vector $w \in \mathbb{R}^n$ onto V is given by

$$\text{proj}_V(w) = AA^T w.$$

Moreover, the image of the projection map is

$$\text{Im}(\text{proj}_V) = V = \text{Col}(A).$$

Gram–Schmidt Orthogonalization / Orthonormalization

The Gram–Schmidt process converts any basis of a subspace $V \subseteq \mathbb{R}^n$ into an orthogonal (or orthonormal) basis.

Step 0: Pick any basis $v_1, \ldots, v_k$ of V.

Step 1: Define

$$b_1 = v_1.$$

Step 2: Define

$$b_2 = v_2 - \text{proj}_{b_1}(v_2) = v_2 - \frac{v_2 \cdot b_1}{b_1 \cdot b_1}\, b_1.$$

Step 3: Define

$$b_3 = v_3 - \text{proj}_{\text{span}\{b_1, b_2\}}(v_3) = v_3 - \frac{v_3 \cdot b_1}{b_1 \cdot b_1}\, b_1 - \frac{v_3 \cdot b_2}{b_2 \cdot b_2}\, b_2.$$

In general: Step k

$$b_k = v_k - \text{proj}_{\text{span}\{b_1, \ldots, b_{k-1}\}}(v_k) = v_k - \frac{v_k \cdot b_1}{b_1 \cdot b_1}\, b_1 - \cdots - \frac{v_k \cdot b_{k-1}}{b_{k-1} \cdot b_{k-1}}\, b_{k-1}.$$

Orthonormal basis: Normalize each b_i to obtain

$$q_i = \frac{b_i}{\|b_i\|}, \qquad i = 1, \ldots, k,$$

giving an orthonormal basis $\{q_1, \ldots, q_k\}$ of V.

Theorem 23.2

Let $\{v_1, \ldots, v_k\}$, $\{b_1, \ldots, b_k\}$, $\{q_1, \ldots, q_k\}$ be the input, middle and output of the Gram-Schmidt algorithm, as explained above.

- $\{v_1, \ldots, v_k\}$, $\{b_1, \ldots, b_k\}$, $\{q_1, \ldots, q_k\}$ are bases of V.
- The vectors $b_1, \ldots, b_k$ form an orthogonal basis of V.
- The vectors $q_1, \ldots, q_k$ form an orthonormal basis of V.

Theorem 23.3 (QR Decomposition)

Assume $A \in \mathbb{R}^{n\times k}$ has linearly independent columns. Then there is a unique way to write

$$A = QR,$$

where

- $Q \in \mathbb{R}^{n\times k}$ has orthonormal columns,
- $R \in \mathbb{R}^{k\times k}$ is upper triangular with positive diagonal entries.

The matrices Q and R come from applying the Gram–Schmidt process to the columns of A.

Lecture 24

Theorem 24.1

Suppose

$$A = (v_1, \ldots, v_k), \qquad Q = (q_1, \ldots, q_k)$$

with Q having orthonormal columns.
And that the QR decomposition is obtained with

$$A = QR$$

Then the entries of R are given by

$$r_{ij} = q_i \cdot v_j,$$

i.e. the coefficients of R are the inner products of the orthonormal vectors q_i with the original basis vectors v_j.

Definition 24.2 (Least Squares Solution)

A *least squares solution* to the equation

$$Ax = b$$

is a vector x^* such that

$$Ax^* = \mathrm{proj}_{\mathrm{Col}(A)}(b).$$

This choice of x^* minimizes the error:

$$\|Ax^* - b\| \;\le\; \|Ax - b\| \quad \text{for all } x.$$

A least squares solution always exists because

$$\mathrm{proj}_{\mathrm{Col}(A)}(b) \in \mathrm{Col}(A),$$

so the equation

$$Ax^* = \mathrm{proj}_{\mathrm{Col}(A)}(b)$$

always has at least one solution.

Theorem 24.3

A vector x^* is a least squares solution to

$$Ax = b$$

if and only if

$$A^T A\, x^* = A^T b.$$

Here $Ax = b$ may be rectangular and inconsistent, while the normal equation $A^T Ax^* = A^T b$ is square and always consistent.

Theorem 24.4

For a QR decomposition $A = QR$ as above, a least squares solution to $Ax = b$ is

$$x^* = R^{-1} Q^T b.$$

Theorem 24.5

A least squares solution x^* always exists, but it is unique if and only if

$$A^T A \text{ is invertible}$$

which is equivalent to saying that the columns of A are linearly independent.

Why is it called "least squares"? Because the solution comes from an approximation method that minimizes the sum of squared errors in the problem of fitting a line through a bunch of data points.

Lecture 25

Definition 25.1 (Orthogonally Diagonalizable Matrix)

A square matrix B is called *orthogonally diagonalizable* if there exist a diagonal matrix D and an orthogonal matrix U such that

$$B = UDU^{-1},$$

where $U^{-1} = U^T$.

Theorem 25.2

If a square matrix B is orthogonally diagonalizable, then B is symmetric, i.e.

$$B = B^T$$

Theorem 25.3

A real $n \times n$ matrix B is orthogonally diagonalizable if all of the following hold:

1. All eigenvalues $\lambda_1, \dots, \lambda_r$ are real.
2. For each eigenvalue λ_i, the geometric multiplicity equals the algebraic multiplicity:
$$\text{geom. mult}(\lambda_i) = \text{alg. mult}(\lambda_i).$$
3. The eigenspaces corresponding to distinct eigenvalues are orthogonal:
$$V_{\lambda_i} \perp V_{\lambda_j} \qquad \text{for all } \lambda_i \neq \lambda_j.$$

Theorem 25.4

If a matrix B is symmetric, then conditions (1)-(3) of Theorem 25.3 are satisfied, and therefore B is orthogonally diagonalizable.

Lemma

For any matrix A and any vectors v, w,

$$(Av) \cdot w \;=\; v \cdot (A^T w).$$

Lecture 26

Theorem 26.1 (Singular Value Decomposition)

Any matrix $B \in \mathbb{R}^{m \times n}$ admits a singular value decomposition (SVD):

$$B = U \Sigma V^T.$$

Here

- $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal matrices, i.e. $U^{-1} = U^T$ and $V^{-1} = V^T$;
- $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix whose diagonal entries $\sigma_1, \dots, \sigma_r > 0$ are the *singular values* of B, and all other entries of Σ are 0.

The columns of $U = (u_1, \dots, u_m)$ and $V = (v_1, \dots, v_n)$ are called *singular vectors*. The singular values are unique if we order them so that $\sigma_1 \geq \dots \geq \sigma_r$, while the singular vectors are not.

Proposition 26.2

All eigenvalues of $B^T B$ are nonnegative.

How to Compute the Singular Value Decomposition (SVD)

Let $B \in \mathbb{R}^{m\times n}$. We explain how to compute a singular value decomposition

$$B = U\Sigma V^T$$

step by step.

Step 1: Compute $B^T B$. Compute the symmetric matrix

$$B^T B \in \mathbb{R}^{n\times n}.$$

Step 2: Find the eigenvalues of $B^T B$. Compute the characteristic polynomial

$$\det(B^T B - tI)$$

and solve for the eigenvalues

$$\lambda_1, \dots, \lambda_n \geq 0.$$

Step 3: Find orthonormal eigenvectors. For each eigenvalue λ_i, compute

$$v_i \in \ker(B^T B - \lambda_i I).$$

Normalize the eigenvectors and choose them orthonormally. Define

$$V = (v_1, \dots, v_n),$$

which is an orthogonal matrix.

Step 4: Compute the singular values. The singular values are

$$\sigma_i = \sqrt{\lambda_i}.$$

Form the diagonal matrix

$$\Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{m\times n},$$

Step 5: Compute the left singular vectors. For each $\sigma_i > 0$, define

$$u_i = \frac{Bv_i}{\|Bv_i\|}.$$

The vectors $u_1, \dots, u_r$ are orthonormal.

Step 6: Complete to an orthonormal basis. If $r < m$, complete $\{u_1, \dots, u_r\}$ to an orthonormal basis of $\mathbb{R}^m$, for example using Gram–Schmidt. Define

$$U = (u_1, \dots, u_m).$$

Conclusion. The matrices U, Σ, and V satisfy

$$B = U\Sigma V^T,$$

which is the singular value decomposition of B.

Lecture 27 and 28

Application of Diagonalization: Solving Systems of Linear ODEs

We consider systems of linear ordinary differential equations of the form

$$\begin{cases} x_1'(t) = a_{11}x_1(t) + \cdots + a_{1n}x_n(t), \\ \vdots \\ x_n'(t) = a_{n1}x_1(t) + \cdots + a_{nn}x_n(t), \end{cases}$$

where the coefficients a_{ij} are known constants and the derivative is taken with respect to time t.

Equivalently, this system can be written in matrix form as

$$x'(t) = Ax(t),$$

where

$$x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}, \qquad A = (a_{ij}) \in \mathbb{R}^{n\times n}.$$

The goal is to solve for the functions

$$x_1(t), \dots, x_n(t) : \mathbb{R} \to \mathbb{R},$$

subject to the initial condition

$$x(0) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \qquad \text{i.e. } x_i(0) = c_i \text{ for } i = 1, \dots, n.$$

Matrix Exponential

Definition. For any $n \times n$ matrix $A \in \mathbb{R}^{n\times n}$, the matrix exponential is defined by

$$e^A := \exp(A) = I_n + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \frac{A^4}{4!} + \cdots \in \mathbb{R}^{n\times n}.$$

This is a convergent power series of $n \times n$ matrices.

Properties.

- If $AB = BA$, then
$$e^A e^B = e^{A+B} = e^B e^A.$$
- $e^0 = I_n$.
- e^A is invertible and
$$(e^A)^{-1} = e^{-A}, \qquad e^A e^{-A} = e^{A-A} = e^0 = I_n.$$
- If
$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix},$$
then
$$e^D = \begin{pmatrix} e^{d_1} & & 0 \\ & \ddots & \\ 0 & & e^{d_n} \end{pmatrix}.$$
- If A and B are similar, i.e.
$$A = PBP^{-1} \quad \text{for some invertible } P,$$
then
$$e^A = Pe^B P^{-1}.$$

Therefore, to compute e^A, you diagonalize

$$A = PDP^{-1}$$

and then $e^A = Pe^D P^{-1}$.

Theorem 27.1

Consider the system of differential equations

$$\begin{cases} v'(t) = A\,v(t), \\ v(0) = c, \end{cases}$$

where $A \in \mathbb{R}^{n\times n}$ and $c \in \mathbb{R}^n$ are given.

Then the unique solution is

$$v(t) = e^{At}\,c.$$

Theorem 28.1 (Euler's Formula)

For all $t \in \mathbb{R}$,

$$e^{it} = \cos(t) + i\sin(t).$$

Remark. In general, the solution to systems of linear ordinary differential equations (of any order) can be written in terms of:

- real exponentials e^{at},
- sines and cosines $\cos(bt)$, $\sin(bt)$,
- various constants.

Equivalently, these solutions can be expressed using complex exponentials

$$e^{(a+ib)t}.$$

Reducing Higher-Order Systems to First-Order Form

We illustrate how to set up a system of higher-order linear ODEs using matrices.

Consider the system

$$\begin{cases} x''(t) = 5y'(t) + x'(t) - 3y(t) + 6x(t), \\ y''(t) = 2x'(t) - 4x(t) + 17y(t). \end{cases}$$

Step 1: Introduce new variables. Define

$$\begin{aligned} x_0(t) &= x(t), & y_0(t) &= y(t), \\ x_1(t) &= x'(t), & y_1(t) &= y'(t), \\ x_2(t) &= x''(t). \end{aligned}$$

Collect these into a single vector

$$v(t) = \begin{pmatrix} x_0(t) \\ x_1(t) \\ x_2(t) \\ y_0(t) \\ y_1(t) \end{pmatrix} \in \mathbb{R}^5.$$

Step 2: Write the system as a first-order ODE. By construction,

$$\begin{aligned} x_0'(t) &= x_1(t), \\ x_1'(t) &= x_2(t), \\ x_2'(t) &= 6x_0(t) + x_2(t) - 3y_0(t) + 5y_1(t), \\ y_0'(t) &= y_1(t), \\ y_1'(t) &= -4x_0(t) + 2x_2(t) + 17y_0(t). \end{aligned}$$

This can be written in matrix form as

$$v'(t) = Av(t),$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 6 & 0 & 1 & -3 & 5 \\ 0 & 0 & 0 & 0 & 1 \\ -4 & 0 & 2 & 17 & 0 \end{pmatrix}.$$

Step 3: Solve the system. Given an initial condition $v(0)$, the solution is

$$v(t) = e^{At}\, v(0).$$

The matrix exponential e^{At} may be computed by diagonalizing A (when possible).