2. Linear algebra background

- linear independence
- matrix inverse
- matrix rank
- linear equations
- eigenvalues and eigenvectors
- positive semidefinite matrices
- norms

Linear independence

Linear independence: a set of vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ is *linearly independent* if the equality

$$\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k = \mathbf{0}$$

is satisfied only when all coefficients α_i are zero:

$$\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$$

- a set of vectors is *linearly dependent* if it's not linearly independent
- saying the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are linearly independent (or dependent) refers to the *set* $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ being so

Example 2.1

- vectors $\boldsymbol{a}_1 = (1, 2)$ and $\boldsymbol{a}_2 = (2, 1)$ are linearly independent:

$$\alpha_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \mathbf{0}$$

  holds only if $\alpha_1 = \alpha_2 = 0$

- the unit vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are linearly independent:

$$\mathbf{0} = \alpha_1 \boldsymbol{e}_1 + \cdots + \alpha_n \boldsymbol{e}_n = (\alpha_1, \alpha_2, \ldots, \alpha_n)$$

  only if $\alpha_1 = \cdots = \alpha_n = 0$

- $\boldsymbol{a}_1 = (1, 1, 0)$, $\boldsymbol{a}_2 = (2, 2, 0)$, and $\boldsymbol{a}_3 = (0, 0, 1)$ are linearly dependent:

$$-2\boldsymbol{a}_1 + \boldsymbol{a}_2 + 0\boldsymbol{a}_3 = \mathbf{0}$$

- $\boldsymbol{a}_1 = (0.2, -7, 8.6)$, $\boldsymbol{a}_2 = (-0.1, 2, -1)$, and $\boldsymbol{a}_3 = (0, -1, 2.2)$ are linearly dependent:

$$\boldsymbol{a}_1 + 2\boldsymbol{a}_2 - 3\boldsymbol{a}_3 = \mathbf{0}$$

Linear independence of matrix columns

for an $m \times n$ matrix A and an n-vector $\boldsymbol{x}$, we have

$$A\boldsymbol{x} = \begin{bmatrix} \boldsymbol{a}_1 \; \boldsymbol{a}_2 \; \cdots \; \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$$

$\boldsymbol{a}_j$ denote the jth column of A

- the columns of a matrix A are linearly independent if

$$A\boldsymbol{x} = \boldsymbol{0} \text{ holds only if } \boldsymbol{x} = \boldsymbol{0}$$

- they are linearly dependent if $A\boldsymbol{x} = \boldsymbol{0}$ for some $\boldsymbol{x} \neq \boldsymbol{0}$

Supersets and subsets

Superset

- a *superset* of a *linearly dependent* set remains linearly dependent
- if the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are linearly dependent, then for any a_{k+1}, the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{a}_{k+1}$ are linearly dependent as well

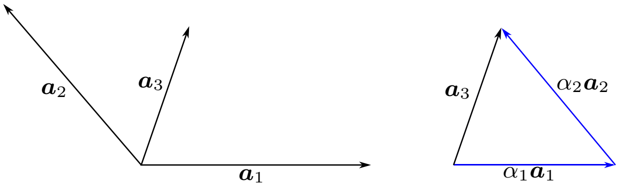
Subset

- a non-empty subset of a linearly independent set remains linearly independent
- removing vectors from a collection of vectors preserves its linear independence

Independence-dimension inequality

let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ be linearly independent vectors in $\mathbb{R}^n$

- the number of vectors is less than the vectors dimension $k \leq n$
- any collection of $n+1$ or more n-vectors is linearly dependent



Linear combination of independent vectors

suppose a vector $\boldsymbol{x}$ can be expressed as a linear combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$:

$$\boldsymbol{x} = \alpha_1 \boldsymbol{a}_1 + \cdots + \alpha_k \boldsymbol{a}_k$$

if the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are linearly independent, then the coefficients $\alpha_1, \ldots, \alpha_k$ are *unique*

proof:

- assume that we can find $\beta_1, \ldots, \beta_k$ such that

$$\boldsymbol{x} = \beta_1 \boldsymbol{a}_1 + \cdots + \beta_k \boldsymbol{a}_k$$

  subtracting the last two equations, we get:

$$0 = (\alpha_1 - \beta_1)\boldsymbol{a}_1 + \cdots + (\alpha_k - \beta_k)\boldsymbol{a}_k$$

- since $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are linearly independent, we must have $\alpha_i - \beta_i = 0$ and thus $\alpha_i = \beta_i$ for all $i = 1, \ldots, k$

Orthonormal vectors

a set of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ is *orthonormal* if:

$$\boldsymbol{a}_i^T \boldsymbol{a}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

- $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ are orthogonal and $\|\boldsymbol{a}_i\| = 1$ for $i = 1, \ldots, k$
- a vector of norm one is called *normalized*; dividing a vector by its norm is called *normalizing* it
- orthonormal set of vectors are linearly independent

Outline

- linear independence
- **matrix inverse**
- matrix rank
- linear equations
- eigenvalues and eigenvectors
- positive semidefinite matrices
- norms

Determinant of a matrix

the determinant of a square matrix for value of i ($i = 1, 2, \ldots, n$) is

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} (\det A_{ij})$$

- A_{ij} is the ij*th submatrix* of A obtained by removing row i and column j from A; for example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad A_{32} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$$

- $\det A_{ij}$ is called the ij *minor* of A

Determinant properties:

- $\det A = \det A^T$
- $\det \alpha A = \alpha^n \det A$ for any scalar α
- $\det AB = \det A \times \det B$ for square matrices A and B

Example 2.2

- for a scalar matrix $A = [a_{11}]$, we have $\det A = a_{11}$
- for a 2×2 matrix:

$$\det A = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

- for the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

we have for $i = 1$

$$A_{11} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad A_{13} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}$$

thus, the determinant is

$$\begin{aligned} \det A &= (-1)^2 a_{11}(\det A_{11}) + (-1)^3 a_{12}(\det A_{12}) + (-1)^4 a_{13}(\det A_{13}) \\ &= a_{11}(\det A_{11}) - a_{12}(\det A_{12}) + a_{13}(\det A_{13}) \\ &= 1(-3) - 2(-6) + 3(-3) = 0 \end{aligned}$$

Inverse of a matrix

an $n \times n$ matrix A^{-1} is the **inverse** of matrix A if:

$$AA^{-1} = A^{-1}A = I$$

- a matrix with an inverse is termed **invertible** or *nonsingular*
- only *square* matrices can be invertible
- invertibility implies $\det A \neq 0$
- an **orthogonal** matrix $A \in \mathbb{R}^{n \times n}$ satisfies: $A^T A = I$, meaning $A^{-1} = A^T$

Example 2.3

- the identity matrix I is invertible, with inverse $I^{-1} = I$ since $II = I$
- any 2×2 matrix A is invertible if and only if $a_{11}a_{22} \neq a_{12}a_{21}$, with inverse

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

- a diagonal matrix

$$D = \operatorname{diag}(d_1, \ldots, d_n) = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

is invertible if and only if $d_i \neq 0$ for $i = 1, \ldots, n$, and

$$D^{-1} = \operatorname{diag}(1/d_1, \ldots, 1/d_n)$$

Inverse properties

- *matrix transpose:* if A is invertible, then its transpose A^T is invertible:

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

- *matrix product:* for invertible square matrices A and B of the same size:

$$(AB)^{-1} = B^{-1}A^{-1}$$

- *negative matrix power:* for an invertible square matrix A and integer p:

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

  for any integer p

Linear independence and matrix inverse

for a square invertible matrix A, the following are equivalent

- A is invertible
- the columns of A are linearly independent
- the rows of A are linearly independent
- the determinant is nonzero

Outline

- linear independence
- matrix inverse
- **matrix rank**
- linear equations
- eigenvalues and eigenvectors
- positive semidefinite matrices
- norms

Subspace

a nonempty subset $\mathcal{V}$ of $\mathbb{R}^n$ is a *subspace* of $\mathbb{R}^n$ it's closed under vector addition and scalar multiplication, i.e.,

$$\alpha \boldsymbol{v} + \beta \boldsymbol{u} \in \mathcal{V} \quad \forall\, \boldsymbol{v}, \boldsymbol{u} \in \mathcal{V},\ \forall\, \alpha, \beta$$

- every subspace includes the zero vector $\mathbf{0}$
- examples:
  - $\{\mathbf{0}\}$ and $\mathbb{R}^n$ are subspaces
  - $\mathcal{V} = \left\{(v_1, v_2) \in \mathbb{R}^2 \mid 2v_1 = v_2\right\}$ is a subspace
  - $\mathcal{V} = \left\{(v_1, v_2) \in \mathbb{R}^2 \mid v_1 \geq 0 \text{ and } v_2 \geq 0\right\}$ is not a subspace; for instance, $(1,1) \in \mathcal{V}$ but $-1(1,1) \notin \mathcal{V}$

Span

given a collection of vectors $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ in a subspace with each $\boldsymbol{v}_i \in \mathcal{V}$, **span** of S is the set of all possible linear combinations of its elements:

$$\text{span}(S) = \text{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k) = \left\{ \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i \mid \alpha_i \in \mathbb{R} \right\}$$

- the span of any set of vectors is a subspace
- if $\boldsymbol{v}$ can be expressed as a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$, then the span remains unchanged upon its addition:

$$\text{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k, \boldsymbol{v}) = \text{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$$

Basis and dimension

Basis: for a subspace $\mathcal{V} \subseteq \mathbb{R}^n$, any set of k linearly independent vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \subset \mathcal{V}$ that spans $\mathcal{V}$ is termed a *basis of the subspace* $\mathcal{V}$

- every vector $\boldsymbol{x} \in \mathcal{V}$ has a unique representation:

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k$$

- coefficients $\alpha_1, \alpha_2, \ldots, \alpha_k$ are distinct and termed the *coordinates* of $\boldsymbol{x}$ relative to the basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$
- any set of n linearly independent vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$ defines a *basis of* $\mathbb{R}^n$

Dimension: The number of vectors in any basis of subspace $\mathcal{V}$ is constant; this number is called the *dimension* of $\mathcal{V}$, symbolized as $\dim \mathcal{V}$

Example 2.4

- the n unit vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are basis (called *natural basis*) for $\mathbb{R}^n$; any $\boldsymbol{x} \in \mathbb{R}^n$ can be written as

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n$$

  and this expansion is unique

- the vectors

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

  are a basis for $\mathbb{R}^2$ since they are 2 linearly independent vectors of size 2

Matrix rank

the *rank* of a matrix A is the maximal number of linearly independent columns of A, denoted by $\operatorname{rank} A$

- $\operatorname{rank} A \leq \min\{m, n\}$
- A has *full rank* if $\operatorname{rank} A = \min\{m, n\}$
- A has *full column rank* if $\operatorname{rank} A = n$ (linearly independent columns)
- A has *full row rank* if $\operatorname{rank} A = m$ (linearly independent rows)

Rank of matrix transpose

- the rank of a matrix A is equal to the rank of A^T
- in other words, the maximum number of linearly independent columns of a matrix is equal to the maximum number of linearly independent rows

Example 2.5

find the rank of the matrix

$$A = \begin{bmatrix} 3 & 0 & 2 & 2 \\ -6 & 42 & 24 & 54 \\ 21 & -21 & 0 & -15 \end{bmatrix}$$

- the first two columns are linearly independent
- it holds that

$$\begin{bmatrix} 2 \\ 24 \\ 0 \end{bmatrix} = 2/3 \begin{bmatrix} 3 \\ -6 \\ 21 \end{bmatrix} + 2/3 \begin{bmatrix} 0 \\ 42 \\ -21 \end{bmatrix}$$

and

$$\begin{bmatrix} 2 \\ 54 \\ -15 \end{bmatrix} = 2/3 \begin{bmatrix} 3 \\ -6 \\ 21 \end{bmatrix} + 29/21 \begin{bmatrix} 0 \\ 42 \\ -21 \end{bmatrix}$$

therefore, only two vectors are linearly independent, thus $\operatorname{rank} A = 2$

Outline

- linear independence
- matrix inverse
- matrix rank
- **linear equations**
- eigenvalues and eigenvectors
- positive semidefinite matrices
- norms

System of linear equations

consider a set or system of m linear equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

- $x_1, \ldots, x_n$ are called *variables*
- a_{ij} are called *coefficients*
- b_i are called *right-hand-sides*

Matrix vector representation

$$A\boldsymbol{x} = \boldsymbol{b}$$

- the $m \times n$ matrix A is called the *coefficient matrix*
- the m-vector $\boldsymbol{b}$ is called the *right-hand side*

System of linear equations

a system of linear equations is said to be

- *overdetermined* system if the number of equations is more than the number of unknowns: $m > n$
- *underdetermined* system if the number of equations is less than the number of unknowns: $m < n$
- *square* system if $m = n$

Solution

- any n vector $\hat{\boldsymbol{x}}$ satisfying $A\hat{\boldsymbol{x}} = \boldsymbol{b}$ called a *solution* of the linear equations $A\boldsymbol{x} = \boldsymbol{b}$
- a set of linear equations can have a unique solution, many solutions, or no solutions

Example 2.6

- the system of linear equations

$$x_1 + x_2 = 1, \quad , x_1 = -1, \quad x_1 - x_2 = 0$$

is an overdetermined system; it can be described as $A\boldsymbol{x} = \boldsymbol{b}$ with

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

this system has no solution

- the system of linear equations

$$x_1 + x_2 = 1, \quad x_2 + x_3 = 2$$

can be written as $A\boldsymbol{x} = \boldsymbol{b}$ with

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

which is an underdetermined system; this system has multiple solutions such as $\boldsymbol{x} = (1, 0, 2)$ and $\boldsymbol{x} = (0, 1, 1)$

Range space

(suppose that A is an $m \times n$ matrix with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$)

the *range space* of A is defined as the span of its column vectors (which is a subspace of $\mathbb{R}^m$):

$$\begin{aligned}\mathrm{range}(A) &= \text{span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n) \\ &= \{x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n \mid \boldsymbol{x} \in \mathbb{R}^n\} \\ &= \{A\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^n\}\end{aligned}$$

- range of A is also called the *column space* or *image* of A
- range of A^T is called the *row space* of A, which is a subspace of $\mathbb{R}^n$

Null space

the *null space* of A is a subspace of $\mathbb{R}^n$ defined as

$$\text{null}(A) = \{\boldsymbol{x} \mid A\boldsymbol{x} = \mathbf{0}\}$$

- the null space is also called *kernal* of A
- the null space of a matrix is the set of vectors orthogonal to the rows of the matrix
- the dimension of the null space of an $m \times n$ matrix A is

$$\text{dim}(\text{null}(A)) = n - \text{rank}\, A$$

Existence of solution

the fundamental theorem of linear systems state that the system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution if and only if

$$\operatorname{rank} A = \operatorname{rank}[A\ \boldsymbol{b}]$$

- this implies that $\boldsymbol{b} \in \operatorname{range}(A)$
- *unique* solution if and only if

$$\operatorname{rank} A = \operatorname{rank}[A\ \boldsymbol{b}] = n$$

  this implies that the solution is unique the columns are linearly independent ($\operatorname{null}(A) = 0$)
- *infinitely many solutions* for any $\boldsymbol{b}$ if and only if $\operatorname{rank} A = m < n$ ($\boldsymbol{b}$ is in the $\operatorname{range}(A)$ and the $\operatorname{null}(A)$ is nonempty)

Example 2.7

consider the matrix

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix},$$

with $\operatorname{rank} A = n = 2$

- for the system $A\boldsymbol{x} = (1, -2, 0)$, $\operatorname{rank} A = \operatorname{rank}[A\ \boldsymbol{b}] = 2$, and hence, there exists a unique solution: $\boldsymbol{x} = (1, -1)$

- for the system $A\boldsymbol{x} = (1, -1, 0)$ $\operatorname{rank} A = 2 \neq \operatorname{rank}[A\ \boldsymbol{b}] = 3$, hence it does not have a solution

- for the system $A^T\boldsymbol{x} = (1, 2)$, we have $\operatorname{rank} A^T = 2 < 3$, and there are multiple solutions, including

$$\boldsymbol{x}_1 = (\tfrac{1}{3}, \tfrac{2}{3}, \tfrac{38}{9}), \quad \boldsymbol{x}_2 = (0, \tfrac{1}{2}, -1)$$

Example 2.8

the matrix

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

is singular with null space

$$\text{null}(A) = \left\{ \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix} \mid \alpha \in \mathbb{R} \right\},$$

and range space

$$\text{range}(A) = \left\{ \beta \begin{bmatrix} 1 \\ 3 \end{bmatrix} \mid \beta \in \mathbb{R} \right\}.$$

for certain values of $\boldsymbol{b}$, the equation $A\boldsymbol{x} = \boldsymbol{b}$ may or may not have solutions

- if $\boldsymbol{b}$ does not belong to the range of A, then no solution exists
- if $\boldsymbol{b}$ is a multiple of the column vector

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

  there are infinitely many solutions

Example 2.9

consider four given measurements: $(t_1, b_1), (t_2, b_2), (t_3, b_3)$, and (t_4, b_4):

$$(0, 4), (0.1, -0.9), (0.8, 10).$$

our objective is to approximate these data points using the function

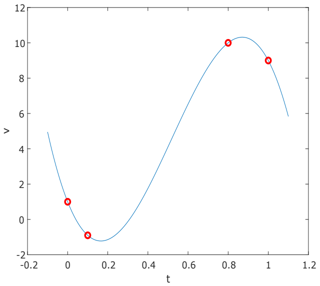
$$v(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$$

to satisfy $v(t_i) = b_i$ where c_i are parameters we want to find

this can be represented as the linear system $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ 1 & t_3 & t_3^2 & t_3^3 \\ 1 & t_4 & t_4^2 & t_4^3 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

```
t = [0,0.1,0.8,1]'; b = [1,-0.9,10,9]';
A = zeros(4,4); %
powers = 0:3;
for j=1:4
A(:,j) = t.^powers(j);
end
x = A \ b; % This solves the system Ax = b
tt = -0.1:.01:1.1;
pt = x(1) + x(2).*tt + x(3).*tt.^2 + x(4).*tt.^3;
plot(tt,pt); hold on
plot(t',b','ro','LineWidth',2); xlabel('t'); ylabel('v')
```



Particular solution

$$A\boldsymbol{x} = \boldsymbol{b}$$

where A is an $m \times n$ matrix with $m \leq n$; assume that

the matrix A has linearly independent rows, $\operatorname{rank} A = m$

- there is at least one solution and there can be many solutions
- the matrix A also has m linearly independent *columns*
- without loss of generality, we assume that the columns of the matrix A are reordered such that the first m columns are linearly independent

Finding a solution

let us partition A and $\boldsymbol{x}$ as

$$A = [B \ D] \qquad \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_D \end{bmatrix}$$

- B is an $m \times m$ invertible matrix (because the first m columns are linearly independent)
- D is an $m \times (n-m)$ matrix
- $\boldsymbol{x}_B$ is an m vector; $\boldsymbol{x}_D$ is an $n-m$ vector

we can then write

$$A\boldsymbol{x} = [B \ D] \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_D \end{bmatrix} = B\boldsymbol{x}_B + D\boldsymbol{x}_D = \boldsymbol{b}$$

Partitioned system

solving for $\boldsymbol{x}_B$, we have $\boldsymbol{x}_B = B^{-1}\boldsymbol{b} - B^{-1}D\boldsymbol{x}_D$; thus

$$\boldsymbol{x} = \begin{bmatrix} B^{-1}\boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} -B^{-1}D\boldsymbol{x}_D \\ \boldsymbol{x}_D \end{bmatrix}$$

is a solution to $A\boldsymbol{x} = \boldsymbol{b}$ for any arbitrary $\boldsymbol{x}_D \in \mathbb{R}^{(n-m)}$

the set of solutions can be written as

$$\boldsymbol{x} = \hat{\boldsymbol{x}} + F\boldsymbol{x}_D$$

where

$$\hat{\boldsymbol{x}} = \begin{bmatrix} B^{-1}\boldsymbol{b} \\ \boldsymbol{0} \end{bmatrix}, \quad F = \begin{bmatrix} -B^{-1}D \\ I \end{bmatrix}$$

- the columns of the matrix F form a basis for the nullspace of A
- if we set $\boldsymbol{x}_D = \boldsymbol{0}$, then we get the solution $\boldsymbol{x} = (B^{-1}\boldsymbol{b}, \boldsymbol{0})$, which is called a *basic solution* with respect to the basis B

Example 2.10

let us find a particular solution to the system of equations $Ax = \boldsymbol{b}$ given by

$$\begin{bmatrix} 2 & 3 & -1 & -1 \\ 4 & 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

we can select any two linearly independent columns of A as basis vectors to find a particular solution

- selecting the first and second columns, we have $\boldsymbol{x}_B = (x_1, x_2)$, $\boldsymbol{x}_D = (x_3, x_4)$ and

$$B = \begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} -1 & -1 \\ 1 & -2 \end{bmatrix}$$

hence,

$$\boldsymbol{x}_B = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = B^{-1}\boldsymbol{b} = \begin{bmatrix} \frac{4}{5} \\ -\frac{2}{5} \end{bmatrix}, \qquad B^{-1}D = \begin{bmatrix} \frac{2}{5} & -\frac{1}{2} \\ -\frac{3}{5} & 0 \end{bmatrix}$$

thus, a particular solution is $\boldsymbol{x} = (\frac{4}{5}, -\frac{1}{5}, 0, 0)$ and the set of all solutions can be written as

$$\boldsymbol{x} = \underbrace{\begin{bmatrix} \frac{4}{5} \\ -\frac{1}{5} \\ 0 \\ 0 \end{bmatrix}}_{\hat{\boldsymbol{x}}} + \underbrace{\begin{bmatrix} -\frac{2}{5} & \frac{1}{2} \\ \frac{3}{5} & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{F} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}$$

- if we select the first and third columns instead, then we have $\boldsymbol{x}_B = (x_1, x_3)$, $\boldsymbol{x}_D = (x_2, x_4)$ and

$$B = \begin{bmatrix} 2 & -1 \\ 4 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & -1 \\ 1 & -2 \end{bmatrix}$$

in this case, we have

$$\boldsymbol{x}_B = \begin{bmatrix} x_1 \\ x_3 \end{bmatrix} = B^{-1}\boldsymbol{b} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}, \qquad B^{-1}D = \begin{bmatrix} \frac{2}{3} & -\frac{1}{2} \\ -\frac{5}{3} & 0 \end{bmatrix}$$

therefore, a particular solution is $\boldsymbol{x} = (\frac{2}{3}, 0, \frac{1}{3}, 0)$ and the set of all solutions can be written as

$$\boldsymbol{x} = \underbrace{\begin{bmatrix} \frac{2}{3} \\ 0 \\ \frac{1}{3} \\ 0 \end{bmatrix}}_{\hat{\boldsymbol{x}}} + \underbrace{\begin{bmatrix} -\frac{2}{3} & \frac{1}{2} \\ 1 & 0 \\ \frac{5}{3} & 0 \\ 0 & 1 \end{bmatrix}}_{F} \begin{bmatrix} x_2 \\ x_4 \end{bmatrix}$$

Outline

- linear independence
- matrix inverse
- matrix rank
- linear equations
- **eigenvalues and eigenvectors**
- positive semidefinite matrices
- norms

Eigenvalues and eigenvectors

a scalar λ (possibly complex) is an *eigenvalue* of an $n \times n$ matrix A if

$$Av = \lambda v \quad \text{for some } \textit{non-zero} \text{ vector } v \in \mathbb{R}^n$$

- v is called an *eigenvector* of A (associated with eigenvalue λ)
- eigenvalues can be computed by solving the *characteristic equation*:

$$\det(\lambda I - A) = 0$$

- $\det(sI - A)$ is a polynomial of degree n known as the *characteristic polynomial*
- the characteristic equation must have n roots (possibly nondistinct), which are the eigenvalues of A

Properties and similar matrices

given $A \in \mathbb{R}^{n \times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$:

- for a triangular A, eigenvalues are its diagonal entries
- eigenvalues of $A + cI$ are $c + \lambda_i$
- eigenvalues of A^k are λ_i^k
- eigenvalues of A^{-1} are $1/\lambda_i$
- eigenvalues of A^T match those of A
- the trace of A is $\sum_{i=1}^n \lambda_i$ and the determinant is $\prod_{i=1}^n \lambda_i$
- if A is real and symmetric, then its eigenvalues are real

Similarity:

- matrices A and B are *similar* if a nonsingular matrix T exists such that

$$T^{-1}AT = B$$

- similar matrices share eigenvalues because $\det(\lambda I - B) = \det(\lambda I - A)$

Diagonalization

if A has n **distinct** eigenvalues, $\lambda_1, \ldots, \lambda_n$, then there exists n linearly independent eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, *i.e.*,

$$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \quad i = 1, \ldots, n$$

it follows that the matrix A is similar to a diagonal matrix

$$A = V \Lambda V^{-1}$$

where

$$\begin{aligned} \Lambda &= \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \\ V &= \begin{bmatrix} \boldsymbol{v}_1 \ \boldsymbol{v}_2 \ \cdots \ \boldsymbol{v}_n \end{bmatrix} \end{aligned}$$

Symmetric eigenvalue decomposition

suppose that $A \in \mathbb{R}^{n \times n}$ is symmetric, then A can be factored as

$$A = U \Lambda U^T$$

- $U \in \mathbb{R}^{n \times n}$ is *orthogonal* (*i.e.*, $U^T U = I$)
- $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ where λ_i denote the real eigenvalues of A
- the columns of U forms an orthonormal set of eigenvectors of U
- the above factorization is called *symmetric eigenvalue decomposition* or *spectral decomposition*

Singular value decomposition (SVD)

suppose that $A \in \mathbb{R}^{m \times n}$ with $\operatorname{rank} A = r$, then A can be factored as

$$A = \begin{bmatrix} U & \bar{U} \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^T \\ \bar{V}^T \end{bmatrix} = U \Sigma V^T$$

where $\begin{bmatrix} U & \bar{U} \end{bmatrix} \in \mathbb{R}^{m \times m}$ and $\begin{bmatrix} V & \bar{V} \end{bmatrix} \in \mathbb{R}^{n \times n}$ are orthogonal matrices and $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$ with

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

$$U \in \mathbb{R}^{m \times r},\ V \in \mathbb{R}^{n \times r},\ \bar{U} \in \mathbb{R}^{(m-r) \times m},\ \bar{V} \in \mathbb{R}^{n \times (n-r)}$$

- U satisfies $U^T U = I$, V satisfies $V^T V = I$
- the columns of U are called *left singular vectors of* A, the columns of V are *right singular vectors of* A
- the numbers σ_i are the *nonzero singular values of* A

the singular value decomposition can be written

$$A = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

where $\boldsymbol{u}_i \in \mathbb{R}^m$ are the left singular vectors, and $\boldsymbol{v}_i \in \mathbb{R}^n$ are the right singular vectors

Relation to A^TA and AA^T

$$A^TA = V\Sigma^2V^T = \begin{bmatrix} V & \bar{V} \end{bmatrix} \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^T \\ \bar{V}^T \end{bmatrix}$$

- this is the eigenvalue decomposition of A^TA
- the nonzero eigenvalues of A^TA are the squared singular values of A and the associated eigenvectors of A^TA are the right singular vectors of A
- similarly, the nonzero eigenvalues of AA^T are the squared singular values of A and the associated eigenvectors of are the left singular vectors of A

Outline

- linear independence
- matrix inverse
- matrix rank
- linear equations
- eigenvalues and eigenvectors
- **positive semidefinite matrices**
- norms

Definiteness of symmetric matrices

a square and symmetric matrix A is positive definite if

$$\boldsymbol{x}^T A \boldsymbol{x} > 0 \quad \forall \boldsymbol{x} \neq \mathbf{0}$$

(for any nonzero vector $\boldsymbol{x} = (x_1, \ldots, x_n)^T$, we require $\sum_{i,j=1}^{n} a_{i,j} x_i x_j > 0$)

other notions of definiteness are defined as follows:

- *Positive semidefinite* if $\boldsymbol{x}^T A \boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$; we use the notation $A \geq 0$ to indicate that A is positive semidefinite
- *Negative definite* if $-A$ is positive definite; we use the notation $A < 0$ to indicate that A is negative definite
- *Negative semidefinite* if $-A$ is positive semidefinite; we use the notation $A \leq 0$ to indicate that Q is negative semidefinite
- *Indefinite* if $\boldsymbol{x}^T A \boldsymbol{x}$ can take on both positive and negative values

Example 2.11

a) the identity matrix I is positive-definite as $\boldsymbol{x}^T I \boldsymbol{x} = \|\boldsymbol{x}\|^2 > 0$ for all $\boldsymbol{x} \neq 0$; conversely, $-I$ is negative definite

b) for a diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$:
- with $d_i > 0$: positive definite
- with $d_i \geq 0$: positive semidefinite
- with both positive and negative d_i: indefinite

c) any matrix $A = B^T B$ is positive semidefinite:

$$\boldsymbol{x}^T B^T B \boldsymbol{x} = (B\boldsymbol{x})^T (B\boldsymbol{x}) = \|B\boldsymbol{x}\|^2 \geq 0$$

If B has linearly independent columns, then A is positive definite since $B\boldsymbol{x} = \mathbf{0}$ only when $\boldsymbol{x} = \mathbf{0}$

Principle submatrices

a **principle submatrix** of an $n \times n$ matrix A is the $(n-k) \times (n-k)$ obtained by deleting k rows and the corresponding k columns of A

a **leading principle submatrix** of an $n \times n$ matrix A of order $n-k$, denoted by A_k, is the matrix obtained by deleting the last k rows and columns of A

example: the principle submatrices of

$$A = \begin{bmatrix} 3 & -4 & 4 \\ -4 & 6 & 5 \\ 4 & 5 & 7 \end{bmatrix}$$

are

$$3, 6, 7, \begin{bmatrix} 3 & -4 \\ -4 & 6 \end{bmatrix}, \begin{bmatrix} 6 & 5 \\ 5 & 7 \end{bmatrix}, \begin{bmatrix} 3 & 4 \\ 4 & 7 \end{bmatrix}, \begin{bmatrix} 3 & -4 & 4 \\ -4 & 6 & 5 \\ 4 & 5 & 7 \end{bmatrix}$$

and the leading principle submatrices are

$$A_1 = 3, \quad A_2 = \begin{bmatrix} 3 & -4 \\ -4 & 6 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 3 & -4 & 4 \\ -4 & 6 & 5 \\ 4 & 5 & 7 \end{bmatrix}$$

Properties of symmetric matrices

Definiteness and eigenvalues

- a symmetric matrix is positive (negative) definite if and only if its eigenvalues are positive (negative)
- a symmetric matrix is positive (negative) semidefinite if and only if its eigenvalues are nonnegative (nonpositive)
- a symmetric matrix is indefinite if and only if it has some positive and negative eigenvalues

Determinant positive definite test: we can test for positive definiteness using Sylvester's criterion

- a symmetric matrix is positive definite if and only if the determinants of all leading principal sub-matrices are positive ($\det A_k > 0$)
- a symmetric matrix is positive semidefinite if and only if the determinants of all principal sub-matrices are nonnegative

Example 2.12

a) the matrix

$$A = \begin{bmatrix} 3 & -4 & 4 \\ -4 & 6 & 5 \\ 4 & 5 & 7 \end{bmatrix}$$

is indefinite; this is because it is not positive semidefinite since $\det A_1 = 3 > 0$ and $\det A_3 = -317 < 0$; it is also not negative semidefinite since $-A$ is not positive semidefinite

b) the matrix

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

is positive definite since the determinant of all leading principle submatrices

$$\det A_1 = 2 > 1, \quad \det A_2 = 3 > 0, \quad \det A_3 = 4$$

are positive

Outline

- linear independence
- matrix inverse
- matrix rank
- linear equations
- eigenvalues and eigenvectors
- positive semidefinite matrices
- **norms**

Vector norms

the function $f : \mathbb{R}^n \to \mathbb{R}$ is a norm if it satisfies the following properties:

1. *positivity:* $f(\boldsymbol{x}) \geq 0$, $f(\boldsymbol{x}) = 0$ only if $\boldsymbol{x} = 0$
2. *homogeneity:* $f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$, $\alpha \in \mathbb{R}$
3. *triangle Inequality:* $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$

p**-norm** is defined as:

$$\|\boldsymbol{x}\|_p = \begin{cases} (|x_1|^p + \cdots + |x_n|^p)^{1/p} & \text{if } 1 \leq p < \infty \\ \max(|x_1|, \ldots, |x_n|) & \text{if } p = \infty \end{cases}$$

- an example is the *Euclidean norm or ℓ_2-norm*:

$$\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}, \quad (\textit{Euclidean norm or } \ell_2\textit{-norm})$$

- we use $\|.\| = \|.\|_2$ if there is no ambiguity

Common vector norms

ℓ_1-norm

$$\|\boldsymbol{x}\|_1 = |x_1| + \cdots + |x_n|$$

ℓ_∞-norm

$$\|\boldsymbol{x}\|_\infty = \max(|x_1|, \ldots, |x_n|)$$

quadratic norms:

$$\|\boldsymbol{x}\|_P = (\boldsymbol{x}^T P \boldsymbol{x})^{1/2} = \|P^{1/2}\boldsymbol{x}\|_2$$

where $P > 0$ is any positive-definite matrix and $P^{1/2}$ is the symmetric square root of P, *i.e.*, $P^{1/2}P^{1/2} = P$

Matrix norms

matrix norms $\|\cdot\|$ satisfies the properties of a norm:

1. $\|cA\| = |c|\|A\|$ for $c \in \mathbb{R}$
2. $\|A + B\| \leq \|A\| + \|B\|$
3. $\|A\| > 0$ and $\|A\| = 0 \iff A = 0$

Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

sum-absolute-value norm:

$$\|A\|_{\text{sav}} = \sum_{j=1}^{n} |a_{ij}|$$

maximum-absolute-value norm:

$$\|A\|_{\text{mav}} = \max\{|a_{ij}| \mid i = 1, \ldots, m, j = 1, \ldots, n\}$$

Induced norms

Induced p-norms: the *matrix p-norm* is

$$\|A\|_p = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|A\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p} = \max_{\|\boldsymbol{x}\|_p = 1} \|A\boldsymbol{x}\|_p$$

- *spectral norm* or ℓ_2 *norm* of A

$$\|A\|_2 = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^T A)}$$

  where $\sigma_{\max}(A)$ is the *maximum singular value of A*
- max-row-sum norm: $\|A\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|$
- max-column-sum norm: $\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}|$

the induced-norms satisfy the sub-multiplicative property

$$\|AB\|_p \leq \|A\|_p \|B\|_p$$

the Frobenius is not an induced norm but it satisfies the sub-multiplicative property

References and further readings

- Edwin KP Chong and Stanislaw H Zak. *An Introduction to Optimization,* John Wiley & Sons, 2013, chapters 2,3.
- Stephen Boyd and Lieven Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares,* Cambridge University Press, 2018, chapters 3,5,8.
- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization,* Cambridge University Press, 2004, appendices A.1, A.5, C.5.