2. Linear algebra background

- subspace, dimension, and rank
- matrix inverses
- solution of linear equations
- eigenvalues and eigenvectors
- positive definite matrices
- norms

Subspace

a nonempty set $\mathcal{V}$ of $\mathbb{R}^n$ is a *subspace* of $\mathbb{R}^n$ if

$$\alpha x + \beta y \in \mathcal{V}, \quad \text{for all } x, y, \in \mathcal{V} \text{ and } \alpha, \beta \in \mathbb{R}$$

(closed under vector addition and scalar multiplication)

- all linear combination of elements of $\mathcal{V}$ are in $\mathcal{V}$
- every subspace includes the zero vector 0
- geometrically, a subspace is a flat (plane) that passes through the origin

Examples

- $\{0\}$ and $\mathbb{R}^n$ are subspaces
- for $m \in \mathbb{R}$, the line $\{(x, mx) \mid x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$
- $\mathbb{R}^2_+$ is not a subspace; for instance, $(1, 1) \in \mathbb{R}^2_+$ but $-1(1, 1) \notin \mathbb{R}^2_+$

Span

the *span* of a collection of vectors $\{a_1, a_2, \ldots, a_k\}$, with $a_i \in \mathbb{R}^n$ is

$$\mathrm{span}(a_1, \ldots, a_k) = \{\alpha_1 a_1 + \cdots + \alpha_k a_k \mid \alpha_i \in \mathbb{R}\}$$

- the set of all linear combinations of $\{a_1, a_2, \ldots, a_k\}$
- a subspace called *subspace generated or spanned by* $\{a_1, a_2, \ldots, a_k\}$
- if $x = \alpha_1 a_1 + \cdots + \alpha_k a_k$, then,

$$\mathrm{span}(a_1, \ldots, a_k, x) = \mathrm{span}(a_1, \ldots, a_k)$$

Operations on subspaces

let $\mathcal{V} \subseteq \mathbb{R}^n$ and $\mathcal{W} \subseteq \mathbb{R}^n$ be subspaces

- *intersection*

$$\mathcal{V} \cap \mathcal{W} = \{x \mid x \in \mathcal{V},\ x \in \mathcal{W}\}$$

- *sum*

$$\mathcal{V} + \mathcal{W} = \{x + y \mid x \in \mathcal{V},\ y \in \mathcal{W}\}$$

  when $\mathcal{V} \cap \mathcal{W} = \{0\}$, then their sum is called *direct sum*, written as $\mathbb{X} \oplus \mathbb{Y}$

- *orthogonal complement*

$$\mathcal{V}^{\perp} = \{x \in \mathbb{R}^n \mid y^T x = 0 \text{ for all } y \in \mathcal{V}\}$$

  (the set of all vectors $x \in \mathbb{R}^n$, each of which is orthogonal to every vector in $\mathcal{X}$)

results of these operations is a subspace ($\mathcal{X}^{\perp}$ is a subspace even if $\mathcal{X}$ is not)

Range and nullspace

suppose that A is an $m \times n$ matrix with columns $a_1, \ldots, a_n$

Range space: the span of the columns vectors (a subspace of $\mathbb{R}^m$):

$$\begin{aligned}\mathrm{range}(A) = \mathrm{span}(a_1, \ldots, a_n) &= \{x_1 a_1 + \cdots + x_n a_n \mid x \in \mathbb{R}^n\} \\ &= \{Ax \mid x \in \mathbb{R}^n\}\end{aligned}$$

- 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 $\mathbb{R}^n$

Nullspace: a subspace of $\mathbb{R}^n$ defined as

$$\mathrm{null}(A) = \{x \mid Ax = 0\}$$

- the nullspace is also called *kernal* of A
- the set of vectors orthogonal to the rows of the matrix
- gives ambiguity in x given $y = Ax$ since $y = Ax = A(x + \tilde{x})$ for any $\tilde{x} \in \mathrm{null}(A)$
- $\mathrm{null}(A^T)$ is called the *left nullspace* of A

Orthogonal decomposition

the nullspace of a matrix A is the orthogonal complement of the row space:

$$\text{null}(A) = \text{range}(A^T)^\perp \quad \text{and} \quad \text{null}(A)^\perp = \text{range}(A^T)$$

Orthogonal decomposition induced by a matrix

- every vector $y \in \mathbb{R}^m$ can be represented uniquely as

$$y = y_1 + y_2$$

  where $y_1 \in \text{range}(A)$ and $y_2 \in \text{range}(A)^\perp = \text{null}(A^T)$

- every vector $x \in \mathbb{R}^n$ can be represented uniquely as

$$x = x_1 + x_2$$

  where $x_1 \in \text{null}(A)$ and $x_2 \in \text{null}(A)^\perp = \text{range}(A^T)$

Linear independence

a set of vectors $\{a_1, \ldots, a_k\}$ is *linearly independent* if the equality

$$\alpha_1 a_1 + \alpha_2 a_2 + \cdots + \alpha_k a_k = 0$$

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
  - set of vectors is lin. dep. iff (at least) one of them is a linear combination of the others
- any set of vectors that contains the zero vector is linearly dependent
- adding vectors to a linearly depen. set of vectors preserves its linear dependence
- removing vectors from a linearly indep. set of vectors preserves its linear indep.
- saying $a_1, \ldots, a_k$ are linearly (in)depen. means the *set* $\{a_1, \ldots, a_k\}$ being so

Examples

- vectors $a_1 = (1, 2)$ and $a_2 = (2, 1)$ are linearly independent:

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

- the unit vectors $e_1, e_2, \ldots, e_n$ are linearly independent:

$$0 = \alpha_1 e_1 + \cdots + \alpha_n e_n = (\alpha_1, \alpha_2, \ldots, \alpha_n) \Rightarrow \alpha_1 = \cdots = \alpha_n = 0$$

- $a_1 = (1, 1, 0)$, $a_2 = (2, 2, 0)$, $a_3 = (0, 0, 1)$ are dependent:

$$-2a_1 + a_2 + 0a_3 = 0$$

- $a_1 = (0.2, -7, 8.6)$, $a_2 = (-0.1, 2, -1)$, $a_3 = (0, -1, 2.2)$ are dependent:

$$a_1 + 2a_2 - 3a_3 = 0$$

Linear independence in matrix notation

for an $m \times n$ matrix A with columns $a_1, \ldots, a_n$ and an n-vector x, we have

$$Ax = \begin{bmatrix} a_1 \; a_2 \; \cdots \; a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 a_1 + \cdots + x_n a_n$$

- the columns of a matrix A are linearly independent if

$$Ax = 0 \text{ holds only if } x = 0$$

- they are linearly dependent if $Ax = 0$ for some $x \neq 0$

Linear combination of independent set of vectors

let x be a vector that can be expressed as a linear combination of $a_1, \ldots, a_k$:

$$x = \alpha_1 a_1 + \cdots + \alpha_k a_k$$

- if $a_1, \ldots, a_k$ are linearly independent, then $\alpha_1, \ldots, \alpha_k$ are unique
- too see this, assume there exist coefficients $\beta_1, \ldots, \beta_k$ such that

  $$x = \beta_1 a_1 + \cdots + \beta_k a_k$$

  subtracting the two equations, we get:

  $$0 = (\alpha_1 - \beta_1)a_1 + \cdots + (\alpha_k - \beta_k)a_k$$

  given that $a_1, \ldots, a_k$ are linearly independent, we must have $\alpha_i = \beta_i$ for each i.

Basis

given a subspace $\mathcal{V}$, the set of vectors $\{v_1, v_2, \ldots, v_k\} \in \mathcal{V}$ is a *basis for* $\mathcal{V}$ if

1. the set $\{v_1, v_2, \ldots, v_k\}$ is linearly independent
2. $\mathcal{V} = \mathrm{span}(v_1, \ldots, v_k)$

- every $x \in \mathcal{V}$ can be expressed uniquely as

$$x = \alpha_1 v_1 + \cdots + \alpha_k v_k$$

  for some coefficients $\alpha_1, \ldots \alpha_k$ called *coordinates* or *components*

- any set of n-linearly independent vectors $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$ is a *basis of* $\mathbb{R}^n$

Examples

- $e_1, \ldots, e_n$ are basis (called *natural basis*) for $\mathbb{R}^n$; any $x \in \mathbb{R}^n$ can be written as

$$x = x_1 e_1 + \cdots + x_n e_n$$

  and this expansion is unique

- vectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

  are bases for $\mathbb{R}^3$ since they are linearly independent; another basis is

$$\begin{bmatrix} 0.5 \\ 0.8 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 1.8 \\ 0.3 \\ 0.3 \end{bmatrix}, \begin{bmatrix} -2.2 \\ -1.3 \\ 3.5 \end{bmatrix}$$

Orthonormal basis

recall that a set of vectors $a_1, a_2, \ldots, a_n$ is *orthonormal* if:

$$a_i^T a_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

- orthonormal set of vectors are linearly independent and form an *orthonormal basis*
- we have for any n-vector x

$$x = (a_1^T x) a_1 + \cdots + (a_n^T x) a_n$$

this is called *orthonormal expansion* of x (in the orthonormal basis)

Dimension

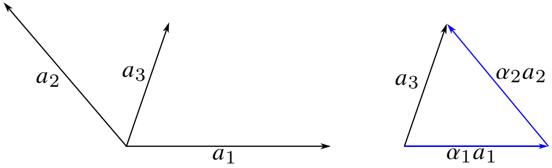
Dimension

- the number of vectors in any basis of subspace $\mathcal{V}$ is the same
- this number is called the *dimension* of $\mathcal{V}$, denoted as $\dim \mathcal{V}$

Dimension inequality: let $a_1, \ldots, a_k$ be linearly independent vectors in $\mathbb{R}^n$, then

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
- if A is an $m \times n$ wide matrix, then its columns are linearly dependent



Matrix rank

the *rank* of a matrix A is

$$\mathrm{rank}(A) = \dim(\mathrm{range}(A)) = \text{max no. of linearly independent columns}$$

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

Rank of matrix transpose

$$\mathrm{rank}(A) = \mathrm{rank}(A^T)$$

i.e., max no. of linearly indep. columns is equal to max no. of linearly indep. rows

Example

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

- 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 and $\operatorname{rank} A = 2$

(we can find rank systemically using determinant test or via Gaussian elimination)

Rank-nullity theorem

the dimension of the nullspace is called *nullity* of A; we have

$$n = \dim(\mathrm{null}(A)) + \mathrm{rank}(A)$$

and

$$m = \dim(\mathrm{null}(A^T)) + \mathrm{rank}(A)$$

- $\mathrm{null}(A) = \{0\}$ if and only if $\mathrm{rank}(A) = n$
- $\mathrm{range}(A) = \mathbb{R}^m$ if and only if $\mathrm{rank}(A) = m$

Outline

- subspace, dimension, and rank
- **matrix inverses**
- solution of linear equations
- eigenvalues and eigenvectors
- positive definite matrices
- norms

Left and right inverse

suppose A is an $m \times n$ matrix

Left inverse: X is a *left inverse* of A if

$$XA = I$$

a left inverse of an $m \times n$ matrix must have size $n \times m$

Right inverse: X is a *right inverse* of A if

$$AX = I$$

A is *right-invertible* if it has at least one right inverse

Immediate properties

- a left or right inverse of an $m \times n$ matrix must have size $n \times m$
- X is a left (right) inverse of A if and only if X^T is a right (left) inverse of A^T

Examples

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

- A is left-invertible; the following matrices are left inverses:

$$\frac{1}{9}\begin{bmatrix} -11 & -10 & 16 \\ 7 & 8 & -11 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1/2 & 3 \\ 0 & 1/2 & -2 \end{bmatrix}$$

- B is right-invertible; the following matrices are right inverses:

$$\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

- for n-vector a ($n \times 1$ matrix), $x = (1/a_i)\, e_i^T$ is left-inverse for any i with $a_i \neq 0$

Column and row independence

Left inverse: a matrix is left-invertible iff its **columns** are linearly independent

- to see this: if $Ax = 0$ and $CA = I$ then

$$0 = C0 = C(Ax) = (CA)x = Ix = x$$

- the converse is also true
- left-invertible matrices are tall or square (by dimension inequality)

Right inverse: A is right-invertible iff its **rows** are linearly independent

- A is right-invertible if and only if A^T is left-invertible:

$$AX = I \iff (AX)^T = I \iff X^T A^T = I$$

- hence, A is right-invertible if and only if its rows are linearly independent
- right-invertible matrices are wide or square

Matrix with orthonormal columns

$A \in \mathbb{R}^{m\times n}$ has orthonormal columns if its Gram matrix is the identity matrix:

$$A^TA = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \cdots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \cdots & a_2^Ta_n \\ \vdots & \vdots & & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \cdots & a_n^Ta_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

there is no standard short name for "matrix with orthonormal columns"

- A is left-invertible with left inverse A^T
- A has linearly independent columns: $Ax = 0 \Longrightarrow A^TAx = x = 0$
- A is tall or square: $m \geq n$
- if A is tall $m > n$, then A has no right inverse; in particular

$$AA^T \neq I$$

Matrix-vector product

if $A \in \mathbb{R}^{m \times n}$ has orthonormal columns, then the linear function $f(x) = Ax$

- preserves inner products:

$$(Ax)^T(Ay) = x^T A^T A y = x^T y$$

- preserves norms:

$$\|Ax\| = \left((Ax)^T(Ax)\right)^{1/2} = (x^T x)^{1/2} = \|x\|$$

- preserves distances: $\|Ax - Ay\| = \|x - y\|$

- preserves angles:

$$\angle(Ax, Ay) = \arccos\left(\frac{(Ax)^T(Ay)}{\|Ax\|\|Ay\|}\right) = \arccos\left(\frac{x^T y}{\|x\|\|y\|}\right) = \angle(x, y)$$

Inverse

if A has a left and a right inverse, then they are equal and unique:

$$XA = I, \quad AY = I \quad \Longrightarrow \quad X = X(AY) = (XA)Y = Y$$

- in this case, we call $X = Y$ the inverse of A, denoted A^{-1}
- A is *invertible* or *nonsingular* if its inverse exists
- invertible matrices must be square

Properties

- $(AB)^{-1} = B^{-1}A^{-1}$ (provided inverses exist)
- $(A^T)^{-1} = (A^{-1})^T$ (sometimes denoted A^{-T})
- negative matrix powers: $(A^{-1})^k$ is denoted A^{-k}
- with $A^0 = I$, identity $A^kA^l = A^{k+l}$ holds for any integers k, l
- for invertible matrix A, we have $\det A^{-1} = 1/\det A$

Examples

- inverse of identity is simply the identity $I^{-1} = I$

- $A = \mathrm{diag}(a_1, \ldots, a_n)$ has inverse $A = \mathrm{diag}(\frac{1}{a_1}, \ldots, \frac{1}{a_n})$ if and only if $a_i \neq 0$

- 2×2 matrix A is invertible if and only $a_{11}a_{22} \neq a_{12}a_{21}$

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

- a non-obvious example:

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & 2 \\ -3 & -4 & -4 \end{bmatrix}, \qquad A^{-1} = \frac{1}{30} \begin{bmatrix} 0 & -20 & -10 \\ -6 & 5 & -2 \\ 6 & 10 & 2 \end{bmatrix}$$

verified by checking $AA^{-1} = I$ (or $A^{-1}A = I$)

Equivalent conditions for invertibility

for a square invertible (nonsingular) matrix $A \in \mathbb{R}^{n \times n}$, the following are equivalent

- A is invertible
- A is left/right invertible
- columns/rows of A are linearly independent ($\mathrm{rank}(A) = n$)
- $\mathrm{null}(A) = \{0\}$
- $\mathrm{range}(A) = \mathbb{R}^n$
- $\det(A) \neq 0$

Example: Vandermonde matrix

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix} \quad \text{with } t_i \neq t_j \text{ for } i \neq j$$

the Vandermonde matrix is nonsingular

Proof

- $Ax = 0$ implies $p(t_1) = p(t_2) = \cdots = p(t_n) = 0$ where

$$p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

  $p(t)$ is a polynomial of degree $n - 1$ or less

- for $x \neq 0$, $p(t)$ can not have more than $n - 1$ distinct real roots
- therefore $p(t_1) = \cdots = p(t_n) = 0$ is only possible if $x = 0$

Orthogonal matrix

a **square** real matrix with orthonormal columns is called *orthogonal*

Nonsingularity: if A is orthogonal, then

- A is invertible, with inverse A^T:

$$\left.\begin{array}{l} A^T A = I \\ A \text{ is square} \end{array}\right\} \quad \Longrightarrow \quad A A^T = I$$

- A^T is also an orthogonal matrix
- rows of A are orthonormal (have norm one and are mutually orthogonal)

Example: permutation matrices

- permutation matrix is square with exactly one entry of each row/column is one
- let $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ be a *permutation* (reordering) of $(1, 2, \dots, n)$
- permutation matrix A,

$$A_{i\pi_i} = 1, \quad A_{ij} = 0 \text{ if } j \neq \pi_i$$

  is orthogonal
- Ax is a permutation of the elements of x: $Ax = (x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_n})$

Proof

- $A^T A = I$ because A has one element equal to one in each row and column

$$(A^T A)_{ij} = \sum_{k=1}^{n} A_{ki} A_{kj} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

- $A^T = A^{-1}$ is the inverse permutation matrix

Example: permutation on $\{1, 2, 3, 4\}$

$$(\pi_1, \pi_2, \pi_3, \pi_4) = (2, 4, 1, 3)$$

- corresponding permutation matrix and its inverse

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad A^{-1} = A^T = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

- A^T is permutation matrix associated with the permutation

$$(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3, \tilde{\pi}_4) = (3, 1, 4, 2)$$

Nonsingular Gram matrix

A^TA is nonsingular iff $A \in \mathbb{R}^{m\times n}$ has linearly independent columns ($\mathrm{rank}(A) = n$)

- suppose A has linearly independent columns:

$$\begin{aligned} A^TAx = 0 &\Longrightarrow x^TA^TAx = (Ax)^T(Ax) = \|Ax\|^2 = 0 \\ &\Longrightarrow Ax = 0 \Longrightarrow x = 0 \end{aligned}$$

  thus A^TA is nonsingular

- assume A^TA is nonsingular but the columns of A are linearly dependent, then

$$\text{there exists } x \neq 0,\ Ax = 0 \quad \Longrightarrow \quad A^TAx = 0$$

  therefore A^TA is singular, a contradiction

Pseudo-inverse of matrix with independent columns

- suppose $A \in \mathbb{R}^{m \times n}$ has linearly independent columns
- this implies that A is tall or square $(m \geq n)$

Pseudo-inverse

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

- this matrix exists, because the Gram matrix $A^T A$ is nonsingular
- $A^{\dagger}$ is a left inverse of A:

$$A^{\dagger} A = (A^T A)^{-1} (A^T A) = I$$

- reduces to the inverse when A is square

Pseudo-inverse of matrix with independent rows

- suppose $A \in \mathbb{R}^{m \times n}$ has linearly independent rows
- this implies that A is wide or square ($m \leq n$)

Pseudo-inverse

$$A^{\dagger} = A^T (AA^T)^{-1}$$

- A^T has linearly independent columns
- hence its Gram matrix AA^T is nonsingular, so $A^{\dagger}$ exists
- $A^{\dagger}$ is a right inverse of A:

$$AA^{\dagger} = (AA^T)(AA^T)^{-1} = I$$

- reduces to the inverse when A is square

Outline

- subspace, dimension, and rank
- matrix inverses
- **solution of linear equations**
- eigenvalues and eigenvectors
- positive definite matrices
- norms

Existence of solution

for $A \in \mathbb{R}^{m \times n}$, the linear equation $Ax = b$ has a solution if and only if

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

this implies that $b \in \operatorname{range}(A)$ (no solution exists if $b \notin \operatorname{range}(A)$)

Cases for solution existence

- a solution exists for *any* b iff $\operatorname{rank} A = m$ (implies $\operatorname{range}(A) = \mathbb{R}^m$)
- *unique* solution if and only if $\operatorname{rank} A = \operatorname{rank}[A \; b] = n$
  - this implies $b \in \operatorname{range}(A)$ and the columns of A are lin. indep ($\operatorname{null}(A) = \{0\}$)
  - a unique solution exists for any b if and only if A is nonsingular
- infinitely many solutions if and only if $\operatorname{rank} A = \operatorname{rank}[A \; b] < n$
  - this implies $b \in \operatorname{range}(A)$ and the $\operatorname{null}(A)$ is nonempty
  - infinitely many solutions for any b if and only if $\operatorname{rank} A = m < n$

Example

$$A = \begin{bmatrix} -3 & -4 \\ 4 & 6 \\ 1 & 1 \end{bmatrix} \quad \text{with } \operatorname{rank} A = n = 2$$

- for $b = (1, -2, 0)$,

$$\operatorname{rank} A = \operatorname{rank}[A\ b] = 2 \Rightarrow \text{unique solution } x = (1, -1)$$

- for $b = (1, -1, 0)$,

$$\operatorname{rank} A = 2 \neq \operatorname{rank}[A\ b] = 3 \Rightarrow \text{no solution}$$

- for the system $A^T x = (1, 2)$,

$$\operatorname{rank} A^T = 2 < 3 \Rightarrow \text{infinitely many solutions}$$

two solutions are $x_1 = (\frac{1}{3}, \frac{2}{3}, \frac{38}{9})$, $x_2 = (0, \frac{1}{2}, -1)$

Example

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 b, the equation $Ax = b$ may or may not have solutions

- if b does not belong to the range of A, then no solution exists
- if b is a multiple of the vector $(1, 3)$ there are infinitely many solutions

Linear equations and matrix inverses

Left inverse: if X is a left inverse of A and $Ax = b$, then

$$x = XAx = Xb$$

- if there is a solution ($b \in \mathrm{range}(A)$), it must be equal to Xb
- if $A(Xb) \neq b$, then there is no solution

Right inverse: if X is a right inverse of A, then

$$x = Xb \quad \Longrightarrow \quad Ax = AXb = b$$

- there is at least one solution, $x = Xb$, for any b
- there can be other solutions

Inverse: if A is invertible, then $x = A^{-1}b$ is the *unique* solution to $Ax = b$

Example

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

$$(0, 1), (0, 4), (0.1, -0.9), (0.8, 10).$$

our objective is to fit these data points using the function

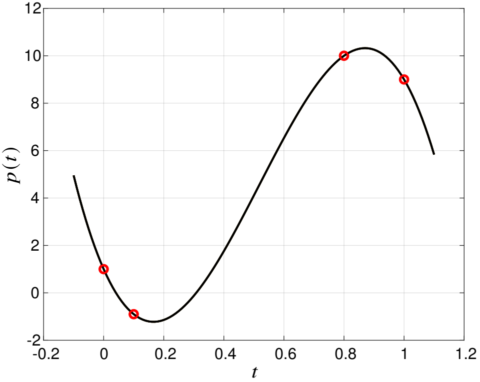
$$p(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$$

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

this can be represented as the linear system $Ax = b$, where

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

```
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
```



Solving underdetermined linear equations

consider an an underdtermined linear equations

$$Ax = b$$

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

the matrix A has linearly independent rows, i.e., $\mathrm{rank}\, A = m$

- there is at least one solution and there can be many solutions
- the matrix A also has m linearly independent *columns*
- assume columns of A are reordered such that the first m columns are lin. indep.

Partitioned system

let us partition A and x as

$$A = [B\ D] \qquad x = \begin{bmatrix} x_B \\ x_D \end{bmatrix}$$

- B is an $m \times m$ invertible matrix (since first m columns are linearly independent)
- D is an $m \times (n - m)$ matrix
- x_B is an m vector; x_D is an $n - m$ vector

we can then write

$$Ax = [B\ D] \begin{bmatrix} x_B \\ x_D \end{bmatrix} = Bx_B + Dx_D = b$$

General and basic solutions

solving for x_B, we have $x_B = B^{-1}b - B^{-1}Dx_D$; thus

$$x = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix} + \begin{bmatrix} -B^{-1}Dx_D \\ x_D \end{bmatrix}$$

is a solution to $Ax = b$ for any arbitrary $x_D \in \mathbb{R}^{(n-m)}$

the general solution (set of all solutions) can be written as

$$x = \hat{x} + Fx_D$$

where

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

- the columns of the matrix F form a basis for the nullspace of A
- for $x_D = 0$, we get $x = (B^{-1}b, 0)$, which is called a *basic solution* w.r.t. basis B

Example

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

- selecting the 1st and 2nd columns, we have $x_B = (x_1, x_2)$, $x_D = (x_3, x_4)$ and

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

hence,

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

thus, a basic solution is $x = (\frac{4}{5}, -\frac{1}{5}, 0, 0)$ and the general solution is

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

- if we select the 1st and 3rd columns, then $x_B = (x_1, x_3)$, $x_D = (x_2, x_4)$ 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

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

thus, a basic solution is $x = (\frac{2}{3}, 0, \frac{1}{3}, 0)$ and the general solution is

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

Outline

- subspace, dimension, and rank
- matrix inverses
- solution of linear equations
- **eigenvalues and eigenvectors**
- positive definite matrices
- norms

Eigenvalues and eigenvectors

scalar λ is an *eigenvalue* of a square $n \times n$ matrix A if

$$Av = \lambda v \quad \text{for} \quad v \neq 0$$

- v is an *eigenvector* associated with eigenvalue λ
- together, (λ, v) is an *eigenpair*; set of all eigenvalues is called *spectrum* of A
- matrix expands/shrinks any vector lying in eigenvector direction by a scalar
- eigenvalues are useful in analyzing numerical methods
  - analysis of iterative methods for solving systems of equations and optimization problems
  - analysis of numerical methods for solving differential equations

Left eigenvector

- w is a left *eigenvector*, associated with eigenvalue λ, if $w^T A = \lambda w^T$
- a left eigenvector of A is a (right) eigenvector of A^T

Characteristic equation

- we can write the eigenvalue problem $Ax = \lambda x$ as a homogeneous linear system

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

  since we want a nontrivial x, this means that $\lambda I - A$ must be singular

- we can find λ by finding the roots of the *characteristic equation*:

$$p(\lambda) = \det(\lambda I - A) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) = 0$$

  $p(\lambda)$ is the *characteristic polynomial*

  – a polynomial of degree n with n roots counting multiplicities

- eigenvalues (and eigenvectors) can be complex even if A is real

  – complex eigenvalues of real A appear as conjugate pairs

- eigenvalues are typically computed using an iterative process

  – no closed-form formula exists for a polynomial of degree greater than or equal to 4

Eigenvalues of a 2×2 matrix

for a 2×2 matrix, the characteristic equation is

$$\begin{aligned}\det(\lambda I - A) &= \det \begin{bmatrix} \lambda - A_{11} & -A_{12} \\ -A_{21} & \lambda - A_{22} \end{bmatrix} \\ &= \lambda^2 - (A_{11} + A_{22})\,\lambda + (A_{11}A_{22} - A_{12}A_{21})\end{aligned}$$

we therefore have to solve a quadratic equation of the form

$$\lambda^2 - b\lambda + c = 0$$

solving gives

$$\lambda_{1,2} = \frac{1}{2}(b \pm \sqrt{\Delta}) = \frac{1}{2}\left(A_{11} + A_{22} \pm \sqrt{\Delta}\right)$$

where $\Delta = b^2 - 4c = (A_{11} - A_{22})^2 + 4A_{12}A_{21}$

- if $\Delta > 0$, then there are two real eigenvalues
- if $\Delta = 0$, then there is the double real eigenvalue
- if $\Delta < 0$, then there are two complex eigenvalues

Example

- for the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

  - $\Delta = 3^2 + 4^2 = 25$
  - $\lambda_1 = (5+5)/2 = 5$
  - $\lambda_2 = (5-5)/2 = 0$

- for the matrix

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

  - $\Delta = -4$
  - $\lambda_1 = \mathrm{j}$
  - $\lambda_2 = -\mathrm{j}$

Some properties

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

- if v is an eigenvector, then γv is an also an eigenvector for any scalar $\gamma \neq 0$
- eigenvalues of $A + \alpha I$ are $\lambda_1 + \alpha, \ldots, \lambda_n + \alpha$
- eigenvalues of A^k are $\lambda_1^k, \ldots, \lambda_n^k$
- eigenvalues of A^T are equal to the eigenvalues of A
- eigenvalues of A^{-1} are $1/\lambda_1, \ldots, 1/\lambda_n$
- if A is a triangular matrix, then its eigenvalues are equal to its diagonal elements
- $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$
- $\det(A) = \prod_{i=1}^n \lambda_i$
- if $v_1, \ldots, v_k$ are eigenvectors for k different eigenvalues:

$$Av_1 = \lambda_1 v_1, \quad \ldots, \quad Av_k = \lambda_k v_k$$

  then $v_1, \ldots, v_k$ are linearly independent (converse is not true)

Similar matrices

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

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

- we call the transformation $A \to T^{-1}AT$ a *similarity transformation* of A
- similar matrices have the same eigenvalues:

$$\det(\lambda I - B) = \det(\lambda I - T^{-1}AT) = \det(T^{-1}(\lambda I - A)T) = \det(\lambda I - A)$$

- if v is an eigenvector of A then $y = T^{-1}v$ is an eigenvector of B:

$$By = (T^{-1}AT)(T^{-1}v) = T^{-1}Av = T^{-1}(\lambda v) = \lambda y$$

- a matrix $A \in \mathbb{R}^{n\times n}$ is *diagonalizable* if it is similar to a diagonal matrix:

$$D = P^{-1}AP$$

  for some invertible $P \in \mathbb{R}^{n\times n}$ where D is a diagonal matrix

Diagonalizbale matrices

if (λ_j, v_j) is an eigenpair, then

$$\begin{aligned} AV &= A\,[v_1 \;\; v_2 \;\; \cdots \;\; v_n] \\ &= [\lambda_1 v_1 \;\; \lambda_2 v_2 \;\; \cdots \;\; \lambda_n v_n] \\ &= V\Lambda \end{aligned}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$

Eigendecomposition: if the eigenvectors are linearly independent, then

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

- rows of V^{-1} are linearly independent left eigenvectors
- this decomposition is the *eigendecomposition* of A
- not all matrices are diagonalizable

Symmetric eigendecomposition

let A be a real symmetric matrix ($A = A^T \in \mathbb{R}^{n\times n}$), then

- all eigenvalues of A are real
- A has n linearly independent eigenvectors
- there is a set of n orthonormal eigenvectors of A

Symmetric eigendecomposition: let $A \in \mathbb{R}^{n\times n}$ be symmetric, then

$$A = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

- $Q \in \mathbb{R}^{n\times n}$ is *orthogonal* ($Q^T Q = I$)
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$
- columns of Q forms an orthonormal set of eigenvectors of A
- this is also known as the *spectral decomposition*

Singular value decomposition (SVD)

every $m \times n$ matrix A can be factored as

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

- U is $m \times m$ and orthogonal, V is $n \times n$ and orthogonal
- Σ is $m \times n$ and "diagonal":

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \qquad \text{if } m = n$$

$$\Sigma = \begin{bmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \\ 0_{(m-n)\times n} \end{bmatrix} \qquad \text{if } m > n$$

$$\Sigma = \begin{bmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_m) & 0_{m\times(n-m)} \end{bmatrix} \qquad \text{if } m < n$$

- diagonal entries of Σ are nonnegative and ordered:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$$

- in MATLAB, the command is `[U,Sigma,V] = svd(A)`

Singular values and singular vectors

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

- columns of U are called *left singular vectors of* A
- columns of V are *right singular vectors of* A
- numbers σ_i are the *singular values of* A

if we write the factorization $A = U\Sigma V^T$ as

$$AV = U\Sigma, \quad A^T U = V\Sigma^T$$

and compare the ith columns on the left- and right-hand sides, we see that

$$Av_i = \sigma_i u_i \quad \text{and} \quad A^T u_i = \sigma_i v_i \quad \text{for } i = 1, \ldots, \min\{m, n\}$$

- if $m > n$ the additional $m - n$ vectors u_i satisfy $A^T u_i = 0$ for $i = n + 1, \ldots, m$
- if $n > m$ the additional $n - m$ vectors v_i satisfy $Av_i = 0$ for $i = m + 1, \ldots, n$

Rank and compact SVD

- rank of a matrix is the maximum number of linearly independent columns
- number of positive singular values is the rank of a matrix

Compact-form of SVD: suppose there are r positive singular values:

$$\sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}}$$

partition the matrices in a full SVD of A as

$$\begin{aligned} A &= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \left[\begin{array}{c|c} \Sigma_1 & 0_{r\times(n-r)} \\ \hline 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T \\ &= U_1 \Sigma_1 V_1^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \end{aligned}$$

- Σ_1 is $r \times r$ with the positive singular values $\sigma_1, \ldots, \sigma_r$ on the diagonal
- U_1 is $m \times r$ and V_1 is $n \times r$ have orthonormal columns

SVD and Gram matrix

the SVD gives the eigendecomposition of A^TA:

$$A^TA = V\Sigma^T\Sigma V^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^2 & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

- the nonzero eigenvalues of A^TA are the squared singular values of A
- the associated eigenvectors of A^TA are the right singular vectors of A

the SVD also gives the eigendecomposition of AA^T:

$$AA^T = U\Sigma\Sigma^T U^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^2 & 0_{r\times(m-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(m-r)} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}$$

- the nonzero eigenvalues of AA^T are the squared singular values of A
- the associated eigenvectors of are the left singular vectors of A

Four subspaces

the SVD of A

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T$$

provides orthonormal bases for the four subspaces associated with A

- columns of the $m \times r$ matrix U_1 are a basis of $\mathrm{range}(A)$
- columns of the $m \times (m-r)$ matrix U_2 are a basis of $\mathrm{range}(A)^\perp = \mathrm{null}(A^T)$
- columns of the $n \times r$ matrix V_1 are a basis of $\mathrm{range}(A^T)$
- columns of the $n \times (n-r)$ matrix V_2 are a basis of $\mathrm{null}(A)$

Outline

- subspace, dimension, and rank
- matrix inverses
- solution of linear equations
- eigenvalues and eigenvectors
- **positive definite matrices**
- norms

Positive (semi)definite matrix

- a *symmetric* matrix $A \in \mathbb{R}^{n \times n}$ is *positive semidefinite* if

$$x^T A x \geq 0 \quad \text{for all } x$$

- a *symmetric matrix* $A \in \mathbb{R}^{n \times n}$ is *positive definite* if

$$x^T A x > 0 \quad \text{for all } x \neq 0$$

  this is a subset of the positive semidefinite matrices

note: if A is symmetric and $n \times n$, then the function

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{i>j} a_{ij} x_i x_j$$

is called a *quadratic form*

Other notions of definiteness

- a symmetric matrix A is *negative semidefinite* if $-A$ is positive semidefinite

$$x^T A x \leq 0 \quad \text{for all } x$$

- a symmetric matrix A is *negative definite* if $-A$ is positive definite

$$x^T A x < 0 \quad \text{for all } x \neq 0$$

- a symmetric matrix A is is *indefinite* if $x^T A x$ has both positive and negative values

Notation

- we use the notation $A \succeq 0$ to indicate that A is positive semidefinite
- we use $A \succ 0$ to indicate that A is positive definite
- we use $A \preceq 0$ and $A \prec 0$ for negative semidefinite and negative definite matrices
- $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ denote the set of positive semidefinite and positive definite matrices

Examples

- the identity matrix I is positive definite $x^T I x = \|x\|^2 > 0$ for all $x \neq 0$
- a diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$ is
  - positive definite if $d_i > 0$
  - positive semidefinite if $d_i \geq 0$
  - indefinite some d_i is positive and some other d_j is negative
- the matrix

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

  is positive definite since

$$x^T \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} x = x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 > 0$$

  for all $x \neq 0$

Gram matrix

recall the definition of Gram matrix of a matrix B

$$A = B^T B$$

- every Gram matrix is positive semidefinite

$$x^T A x = x^T B^T B x = \|Bx\|^2 \geq 0 \quad \forall\, x$$

- a Gram matrix is positive definite if

$$x^T A x = x^T B^T B x = \|Bx\|^2 > 0 \quad \forall\, x \neq 0,$$

  i.e., B has linearly independent columns

Properties

- every positive definite matrix A is nonsingular

$$Ax = 0 \implies x^T Ax = 0 \implies x = 0$$

  (last step follows from positive definiteness)

- every positive definite matrix A has positive diagonal elements

$$A_{ii} = e_i^T Ae_i > 0$$

- every positive semidefinite matrix A has nonnegative diagonal elements

$$A_{ii} = e_i^T Ae_i \geq 0$$

- a symmetric matrix is positive definite iff its eigenvalues are positive
- a symmetric matrix is positive semidefinite iff its eigenvalues are nonnegative
- a symmetric matrix is indefinite iff it has some positive and negative eigenvalues

Singular positive semidefinite matrices

if A is positive semidefinite, but not positive definite, then it is singular

to see this, suppose A is positive semidefinite but not positive definite

- there exists a nonzero x with $x^TAx = 0$
- since A is positive semidefinite the following function is nonnegative:

$$\begin{aligned} f(t) &= (x - tAx)^T A(x - tAx) \\ &= x^TAx - 2tx^TA^2x + t^2x^TA^3x \\ &= -2t\|Ax\|^2 + t^2x^TA^3x \end{aligned}$$

- $f(t) \geq 0$ for all t is only possible if $\|Ax\| = 0$; therefore $Ax = 0$
- hence there exists a nonzero x with $Ax = 0$, so A is singular

Principle submatrices

a **principle submatrix** of an $n \times n$ matrix A is the $(n-k) \times (n-k)$ matrix 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

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

and the leading principle submatrices are A_1, A_2, A_3

Determinant positive (semi)definite test

Sylvester's criterion

- a symmetric matrix is positive semidefinite if and only if the determinants of all principal sub-matrices are nonnegative
- a symmetric matrix is positive definite if and only if the determinants of all leading principal sub-matrices are positive, *i.e.*, $\det A_k > 0$

Remarks

- determinant of principal sub-matrices of negative definite matrix may be both positive and negative (*e.g.*, $-I_2$)
- to check if A is (semi)negative definite we check that $-A$ is positive (semi)definite

Examples

- the matrix

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

has $\det A_1 = 3 > 0$; thus not positive semidefinite

it is also not negative semidefinite since $-A$ is not positive semidefinite (negative diagonals) thus, it is indefinite

- for the matrix

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

the determinant of all leading principle submatrices are positive

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

thus it is positive definite

Outline

- subspace, dimension, and rank
- matrix inverses
- solution of linear equations
- eigenvalues and eigenvectors
- positive definite matrices
- **norms**

Vector norms

a *norm* on $\mathbb{R}^n$ is a function $\|\cdot\| : \mathbb{R}^n \longrightarrow \mathbb{R}$ satisfying the following properties

- *positive definiteness:* $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$
- *homogeneity:* $\|\alpha x\| = |\alpha|\|x\|$, $\alpha \in \mathbb{R}$
- *triangle inequality:* $\|x + y\| \leq \|x\| + \|y\|$

p**-norm** is defined as:

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

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

$$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$$

- this is the default norm, written as $\|\cdot\|$, without the subscript 2

Common vector norms

ℓ_1**-norm** (Manhattan norm)

$$\|x\|_1 = |x_1| + \cdots + |x_n|$$

ℓ_∞**-norm** (Chebyshev norm)

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

Quadratic norms

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

- $P > 0$ is any positive-definite matrix
- $P^{1/2}$ 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 $\|A\| = 0 \iff A = 0$

an example is the **Frobenius norm**

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} = \left(\mathrm{tr}(A^T A)\right)^{\frac{1}{2}}$$

- $\|A\|_F = \|A^T\|_F = \sqrt{\|a_1\|^2 + \cdots + \|a_n\|^2}$, where a_j is jth column of A
- Frobenius norm is a submultiplicative norm: $\|AB\|_F \leq \|A\|_F \|B\|_F$

Induced norms

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

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

gives the maximum amplification factor or gain of A in the direction x

- **spectral norm** or ℓ_2 *norm* of A

  $$\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^T A)}$$

  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

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

Rank-r approximation

let A be an $m \times n$ matrix with $\mathrm{rank}(A) > r$ and full SVD

$$A = U\Sigma V^T = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T, \quad \sigma_1 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0, \quad \sigma_{r+1} > 0$$

the best rank-r approximation of A is the sum of the first r terms in the SVD:

$$B = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

- B is the best approximation for the Frobenius norm: for every C with rank r,

$$\|A - C\|_F \geq \|A - B\|_F = \left(\sum_{i=r+1}^{\min\{m,n\}} \sigma_i^2 \right)^{1/2}$$

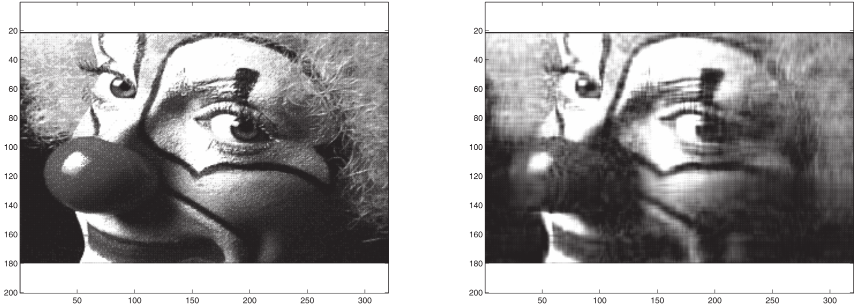
- B is also the best approximation for the 2-norm: for every C with rank r,

$$\|A - C\|_2 \geq \|A - B\|_2 = \sigma_{r+1}$$

Image compression

rank-r approximation makes it possible to devise a compression scheme:

- by storing the first r columns of U and V, as well as the first r singular values
- we obtain an approximation of A using only $r(m+n+1)$ locations instead of mn



- image of size $200 \times 320 = 64,000$
- rank-20 SVD approximation of size $20 \times (200 + 320 + 1) \approx 10,000$

References and further readings

- E. K.P. Chong, Wu-S. Lu, and S. H. Zak, *An Introduction to Optimization: With Applications to Machine Learning*. John Wiley & Sons, 2023. (Ch. 2,3)
- S. Boyd and L. Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares.* Cambridge University Press, 2018. (Ch. 5,8,11)
- S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004. (Appendix A.1, A.5, C.5)