3. Reviews on Linear algebra

- matrices and vectors
- linear equations
- range and nullspace of matrices
- norm and inner product spaces
- matrix factorizations
- function of vectors, gradient and Hessian
- function of matrices

Vector notation

n-vector x:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- also written as $x = (x_1, x_2, \ldots, x_n)$
- set of n-vectors is denoted $\mathbf{R}^n$ (Euclidean space)
- x_i: ith **element** or **component** or **entry** of x
- x is also called a column vector
- $y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$ is called a row vector

unless stated otherwise, a vector typically means a column vector

Special vectors

zero vectors: $x = (0, 0, \ldots, 0)$

all-ones vectors: $x = (1, 1, \cdots, 1)$ (we will denote it by **1**)

standard unit vectors: e_k has only 1 at the kth entry and zero otherwise

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

(standard unit vectors in $\mathbf{R}^3$)

unit vectors: any vector u whose norm (magnitude) is 1, *i.e.*,

$$\|u\| \triangleq \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = 1$$

example: $u = (1/\sqrt{2}, 2/\sqrt{6}, -1/\sqrt{2})$

Inner products

definition: the inner product of two n-vectors x, y is

$$x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

also known as the **dot product** of vectors x, y

notation: $x^T y$

properties ✎

- $(\alpha x)^T y = \alpha(x^T y)$ for scalar α
- $(x + y)^T z = x^T z + y^T z$
- $x^T y = y^T x$

Euclidean norm

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$$

properties

- also written $\|x\|_2$ to distinguish from other norms
- $\|\alpha x\| = |\alpha|\|x\|$ for scalar α
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
- $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$

interpretation

- $\|x\|$ measures the *magnitude* or length of x
- $\|x - y\|$ measures the *distance* between x and y

Matrix notation

an $m \times n$ matrix A is defined as

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad \text{or} \quad A = [a_{ij}]_{m \times n}$$

- a_{ij} are the **elements**, or **coefficients**, or **entries** of A
- set of $m \times n$-matrices is denoted $\mathbf{R}^{m \times n}$
- A has m rows and n columns (m, n are the **dimensions**)
- the (i, j) entry of A is also commonly denoted by A_{ij}
- A is called a **square** matrix if $m = n$

Special matrices

zero matrix: $A = 0$

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$a_{ij} = 0$, for $i = 1, \ldots, m, j = 1, \ldots, n$

identity matrix: $A = I$

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

a square matrix with $a_{ii} = 1, a_{ij} = 0$ for $i \neq j$

diagonal matrix: a square matrix with $a_{ij} = 0$ for $i \neq j$

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_n \end{bmatrix}$$

triangular matrix:

a square matrix with zero entries in a triangular part

upper triangular

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

$a_{ij} = 0$ for $i \geq j$

lower triangular

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$a_{ij} = 0$ for $i \leq j$

Block matrix notation

example: 2×2-block matrix A

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

for example, if B, C, D, E are defined as

$$B = \begin{bmatrix} 2 & 1 \\ 3 & 8 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 7 \\ 1 & 9 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} -4 & 1 & -1 \end{bmatrix}$$

then A is the matrix

$$A = \begin{bmatrix} 2 & 1 & 0 & 1 & 7 \\ 3 & 8 & 1 & 9 & 1 \\ 0 & 1 & -4 & 1 & -1 \end{bmatrix}$$

note: dimensions of the blocks must be compatible

Column and Row partitions

write an $m \times n$-matrix A in terms of its columns or its rows

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_m^T \end{bmatrix}$$

- a_j for $j = 1, 2, \ldots, n$ are the columns of A
- b_i^T for $i = 1, 2, \ldots, m$ are the rows of A

example: $A = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 9 & 0 \end{bmatrix}$

$$a_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ 9 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad b_1^T = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}, \quad b_2^T = \begin{bmatrix} 4 & 9 & 0 \end{bmatrix}$$

Matrix-vector product

product of $m \times n$-matrix A with n-vector x

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix}$$

- dimensions must be compatible: $\#$ columns in $A = \#$ elements in x

if A is partitioned as $A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$, then

$$Ax = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

- Ax is a linear combination of the column vectors of A
- the coefficients are the entries of x

Product with standard unit vectors

post-multiply with a column vector

$$Ae_k = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} = \text{the } k\text{th column of } A$$

pre-multiply with a row vector

$$\begin{aligned} e_k^T A &= \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} = \text{the } k\text{th row of } A \end{aligned}$$

Trace

Definition: trace of a square matrix A is the sum of the diagonal entries in A

$$\mathbf{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

example:

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

trace of A is $2 - 1 + 6 = 7$

properties ✎

- $\mathbf{tr}(A^T) = \mathbf{tr}(A)$
- $\mathbf{tr}(\alpha A + B) = \alpha\,\mathbf{tr}(A) + \mathbf{tr}(B)$
- $\mathbf{tr}(AB) = \mathbf{tr}(BA)$

Eigenvalues

$\lambda \in \mathbf{C}$ is called an **eigenvalue** of $A \in \mathbf{C}^{n\times n}$ if

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

equivalent to:

- there exists nonzero $x \in \mathbf{C}^n$ s.t. $(\lambda I - A)x = 0$, *i.e.*,

$$Ax = \lambda x$$

  any such x is called an **eigenvector** of A (associated with eigenvalue λ)

- there exists nonzero $w \in \mathbf{C}^n$ such that

$$w^T A = \lambda w^T$$

  any such w is called a **left eigenvector** of A

Computing eigenvalues

- $\mathcal{X}(\lambda) = \det(\lambda I - A)$ is called the **characteristic polynomial** of A
- $\mathcal{X}(\lambda) = 0$ is called the **characteristic equation** of A
- eigenvalues of A are the root of characteristic polynomial

Properties

- if A is $n \times n$ then $\mathcal{X}(\lambda)$ is a polynomial of order n
- if A is $n \times n$ then there are n eigenvalues of A
- even when A is real, eigenvalues and eigenvectors can be complex, *e.g.*,

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}, \quad A = \begin{bmatrix} -2 & 0 & 1 \\ -6 & -2 & 0 \\ 19 & 5 & -4 \end{bmatrix}$$

- if A and λ are real, we can choose the associated eigenvector to be real
- if A is real then eigenvalues must occur in complex conjugate pairs
- if x is an eigenvector of A, so is αx for any $\alpha \in \mathbf{C}$, $\alpha \neq 0$
- an eigenvector of A associated with λ lies in $\mathcal{N}(\lambda I - A)$

Important facts

denote $\lambda(A)$ an eigenvalue of A

- $\lambda(\alpha A) = \alpha\lambda(A)$ for any $\alpha \in \mathbf{C}$
- $\mathbf{tr}(A)$ is the sum of eigenvalues of A
- $\det(A)$ is the product of eigenvalues of A
- A and A^T share the same eigenvalues ✎
- $\lambda(\overline{A^T}) = \overline{\lambda(A)}$ ✎
- $\lambda(A^T A) \geq 0$
- $\lambda(A^m) = (\lambda(A))^m$ for any integer m
- A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A ✎

Eigenvalue decomposition

if A is diagonalizable then A admits the decomposition

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

- D is diagonal containing the eigenvalues of A
- columns of T are the corresponding eigenvectors of A
- note that such decomposition is not unique (up to scaling in T)

recall: A is diagonalizable iff all eigenvectors of A are independent

Inverse of matrices

Definition:

a *square* matrix A is called **invertible** or **nonsingular** if there exists B s.t.

$$AB = BA = I$$

- B is called an **inverse** of A
- it is also true that B is invertible and A is an inverse of B
- if no such B can be found A is said to be **singular**

assume A is invertible

- an inverse of A is unique
- the inverse of A is denoted by A^{-1}

assume A, B are invertible

Facts ✎

- $(\alpha A)^{-1} = \alpha^{-1} A^{-1}$ for nonzero α
- A^T is also invertible and $(A^T)^{-1} = (A^{-1})^T$
- AB is invertible and $(AB)^{-1} = B^{-1} A^{-1}$
- $(A + B)^{-1} \neq A^{-1} + B^{-1}$

Inverse of 2×2 matrices

the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is invertible if and only if

$$ad - bc \neq 0$$

and its inverse is given by

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

example:

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}, \quad A^{-1} = \frac{1}{7} \begin{bmatrix} 3 & -1 \\ 1 & 2 \end{bmatrix}$$

Invertible matrices

✌ **Theorem:** for a square matrix A, the following statements are equivalent

1. A is invertible
2. $Ax = 0$ has only the trivial solution ($x = 0$)
3. the reduced echelon form of A is I
4. A is invertible if and only if $\det(A) \neq 0$

Inverse of special matrices

diagonal matrix

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_n \end{bmatrix}$$

a diagonal matrix is invertible iff the diagonal entries are all nonzero

$$a_{ii} \neq 0, \quad i = 1, 2, \ldots, n$$

the inverse of A is given by

$$A^{-1} = \begin{bmatrix} 1/a_1 & 0 & \cdots & 0 \\ 0 & 1/a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1/a_n \end{bmatrix}$$

the diagonal entries in A^{-1} are the inverse of the diagonal entries in A

triangular matrix:

upper triangular **lower triangular**

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$a_{ij} = 0$ for $i \geq j$ $\qquad$ $a_{ij} = 0$ for $i \leq j$

a triangular matrix is invertible iff the diagonal entries are all nonzero

$$a_{ii} \neq 0, \quad \forall i = 1, 2, \ldots, n$$

- product of lower (upper) triangular matrices is lower (upper) triangular
- the inverse of a lower (upper) triangular matrix is lower (upper) triangular

symmetric matrix: $A = A^T$

✎

- for any square matrix A, AA^T and A^TA are always symmetric
- if A is symmetric and invertible, then A^{-1} is symmetric
- if A is invertible, then AA^T and A^TA are also invertible

Symmetric matrix

$A \in \mathbf{R}^{n\times n}$ is called *symmetric* if $A = A^T$

Facts: if A is symmetric

- all eigenvalues of A are real
- all eigenvectors of A are orthogonal
- A admits a decomposition
$$A = UDU^T$$
where $U^TU = UU^T = I$ (U is unitary) and D is diagonal

(of course, the diagonals of D are eigenvalues of A)

Unitary matrix

a matrix $U \in \mathbf{R}^{n\times n}$ is called **unitary** if

$$U^TU = UU^T = I$$

example: $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Facts:

- a real unitary matrix is also called **orthogonal**
- a unitary matrix is always invertible and $U^{-1} = U^T$
- columns vectors of U are mutually orthogonal
- norm is preserved under a unitary transformation:

$$y = Ux \quad \Longrightarrow \quad \|y\| = \|x\|$$

Idempotent Matrix

$A \in \mathbf{R}^{n \times n}$ is an **idempotent** (or projection) matrix if

$$A^2 = A$$

examples: identity matrix

Facts: Let A be an idempotent matrix

- eigenvalues of A are all equal to 0 or 1
- $I - A$ is idempotent
- if $A \neq I$, then A is singular

Projection matrix

a square matrix P is a **projection** matrix if and only if $P^2 = P$

- P is a linear transformation from $\mathbf{R}^n$ to a subspace of $\mathbf{R}^n$, denoted as S
- columns of P are the projections of standard basis vectors
- S is the range of P
- from $P^2 = P$, it means if P is applied twice on a vector in S, it gives the same vector
- examples:

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

Orthogonal projection matrix

a projection matrix is called **orthogonal** if and only if $P = P^T$

- P is bounded, *i.e.*, $\|Px\| \leq \|x\|$

$$\|Px\|_2^2 = x^T P^T P x = x^T P^2 x = x^T P x \leq \|Px\| \|x\|$$

  (by Cauchy-Schwarz inequality – more on this later)

- if P is an orthogonal projection onto a line spanned by a unit vector u,

$$P = uu^T$$

  (we see that $\mathbf{rank}(P) = 1$ as the dimension of a line is 1)

- another example: $P = A(A^T A)^{-1} A^T$ for any matrix A

Nilpotent matrix

$A \in \mathbf{R}^{n\times n}$ is *nilpotent* if

$$A^k = 0, \quad \text{for some positive integer } k$$

Example: any triangular matrices with 0's along the main diagonal

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \text{(shift matrix)}$$

also related to deadbeat control for linear discrete-time systems

Facts:

- the characteristic equation for A is $\lambda^n = 0$
- all eigenvalues are 0

Positive definite matrix

a symmetric matrix A is **positive semidefinite**, written as $A \succeq 0$ if

$$x^T A x \geq 0, \quad \forall x \in \mathbf{R}^n$$

and **positive definite**, written as $A \succ 0$ if

$$x^T A x > 0, \quad \text{for all } \textit{nonzero } x \in \mathbf{R}^n$$

Facts: $A \succeq 0$ if and only if

- all eigenvalues of A are non-negative
- all principle minors of A are non-negative

example: $A = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \succeq 0$ because

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

or we can check from

- eigenvalues of A are 0.38 and 2.61 (real and positive)
- the principle minors are 1 and $\begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} = 1$ (all positive)

note: $A \succeq 0$ does not mean all entries of A are positive!

Properties: if $A \succeq 0$ then

- all the diagonal terms of A are nonnegative
- all the leading blocks of A are positive semidefinite
- $BAB^T \succeq 0$ for any B
- if $A \succeq 0$ and $B \succeq 0$, then so is $A + B$
- A has a square root, denoted as a symmetric $A^{1/2}$ such that

$$A^{1/2}A^{1/2} = A$$

Schur complement

a consider a symmetric matrix X partitioned as

$$X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

- Schur complement of D in X is defined as

$$S = A - BD^{-1}C, \quad \text{if } \det D \neq 0$$

  we can show that $\det X = \det D \det S$

- Schur complement of A in X is defined as

$$S = D - CA^{-1}B, \quad \text{if } \det A \neq 0$$

  we can show that $\det X = \det A \det S$

Matrix inversion lemma

the inverse of X can be expressed with the terms involving Schur complement

an LDU decomposition is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$

this proves that the inverse of the whole box is $\det(A - BD^{-1}C) \det D$

if D and $S = A - BD^{-1}C$ are invertible, the inverse of the whole block is

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} S^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} S^{-1} & -S^{-1}BD^{-1} \\ -D^{-1}CS^{-1} & D^{-1} + D^{-1}CS^{-1}BD^{-1} \end{bmatrix} \end{aligned}$$

Schur complement of positive semidefinite matrix

$$X = \begin{bmatrix} A & B \\ B^T & D \end{bmatrix}, \quad S_1 = A - BD^{-1}B^T, \quad S_2 = D - B^T A^{-1} B,$$

Facts:

- $X \succ 0$ if and only if $D \succ 0$ and $S_2 \succ 0$
- if $D \succ 0$ then $X \succeq 0$ if and only if $S_2 \succeq 0$
- $\det X = \det D \det S_1 = \det A \det S_2$

analogous results for S_2

- $X \succ 0$ if and only if $A \succ 0$ and $S_1 \succ 0$
- if $A \succ 0$ then $X \succeq 0$ if and only if $S_1 \succeq 0$

Linear equations

a general linear system of m equations with n variables is described by

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

where a_{ij}, b_j are constants and $x_1, x_2, \ldots, x_n$ are unknowns

- equations are linear in $x_1, x_2, \ldots, x_n$
- existence and uniqueness of a solution depend on a_{ij} and b_j

Linear equation in matrix form

the linear system of m equations in n variables

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

in matrix form: $Ax = b$ where

$$
A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}
$$

Three types of linear equations

- **square** if $m = n$ (A is square)

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

- **underdetermined** if $m < n$ (A is fat)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

- **overdetermined** if $m > n$ (A is skinny)

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Existence and uniqueness of solutions

existence:

- no solution
- a solution exists

  uniqueness:

  - the solution is unique
  - there are infinitely many solutions

every system of linear equations has zero, one, or infinitely many solutions

there are no other possibilities

Nullspace

the **nullspace** of an $m \times n$ matrix is defined as

$$\mathcal{N}(A) = \{x \in \mathbf{R}^n \mid Ax = 0\}$$

- the set of all vectors that are mapped to zero by $f(x) = Ax$
- the set of all vectors that are orthogonal to the rows of A
- if $Ax = b$ then $A(x + z) = b$ for all $z \in \mathcal{N}(A)$
- also known as **kernel** of A
- $\mathcal{N}(A)$ is a subspace of $\mathbf{R}^n$ ✎

Zero nullspace matrix

- A has a zero nullspace if $\mathcal{N}(A) = \{0\}$
- if A has a zero nullspace and $Ax = b$ is solvable, the solution is unique
- columns of A are independent

✌ **equivalent conditions:** $A \in \mathbf{R}^{n \times n}$

- A has a zero nullspace
- A is invertible or nonsingular
- columns of A are a basis for $\mathbf{R}^n$

Range space

the **range** of an $m \times n$ matrix A is defined as

$$\mathcal{R}(A) = \{y \in \mathbf{R}^m \mid y = Ax \text{ for some } x \in \mathbf{R}^n \}$$

- the set of all m-vectors that can be expressed as Ax
- the set of all linear combinations of the columns of $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$

$$\mathcal{R}(A) = \{y \mid y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n, \quad x \in \mathbf{R}^n\}$$

- the set of all vectors b for which $Ax = b$ is solvable
- also known as the **column space** of A
- $\mathcal{R}(A)$ is a subspace of $\mathbf{R}^m$ ✎

Full range matrices

A has a full range if $\mathcal{R}(A) = \mathbf{R}^m$

✌ **equivalent conditions:**

- A has a full range
- columns of A span $\mathbf{R}^m$
- $Ax = b$ is solvable for *every* b
- $\mathcal{N}(A^T) = \{0\}$

Rank and Nullity

rank of a matrix $A \in \mathbf{R}^{m \times n}$ is defined as

$$\mathbf{rank}(A) = \dim \mathcal{R}(A)$$

nullity of a matrix $A \in \mathbf{R}^{m \times n}$ is

$$\mathbf{nullity}(A) = \dim \mathcal{N}(A)$$

Facts ✌

- $\mathbf{rank}(A)$ is maximum number of independent columns (or rows) of A

$$\mathbf{rank}(A) \leq \min(m, n)$$

- $\mathbf{rank}(A) = \mathbf{rank}(A^T)$

Full rank matrices

for $A \in \mathbf{R}^{m \times n}$ we always have $\mathbf{rank}(A) \leq \min(m, n)$

we say A is **full rank** if $\mathbf{rank}(A) = \min(m, n)$

- for **square** matrices, full rank means nonsingular (invertible)
- for **skinny** matrices ($m \geq n$), full rank means columns are independent
- for **fat** matrices ($m \leq n$), full rank means rows are independent

Theorems

- Rank-Nullity Theorem: for any $A \in \mathbf{R}^{m \times n}$,

$$\mathbf{rank}(A) + \dim \mathcal{N}(A) = n$$

- the system $Ax = b$ has a solution if and only if $b \in \mathcal{R}(A)$
- the system $Ax = b$ has a unique solution if and only if

$$b \in \mathcal{R}(A), \quad \text{and} \quad \mathcal{N}(A) = \{0\}$$

Vector space

a vector space or linear space (over $\mathbf{R}$) consists of

- a set $\mathcal{V}$
- a vector sum $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$
- a scalar multiplication : $\mathbf{R} \times \mathcal{V} \to \mathcal{V}$
- a distinguished element $0 \in \mathcal{V}$

which satisfy a list of properties

$\mathcal{V}$ is called a vector space over **R**, denoted by $(\mathcal{V}, \mathbf{R})$

if elements, called *vectors* of $\mathcal{V}$ satisfy the following main operations:

1. **vector addition:**

$$x, y \in \mathcal{V} \quad \Rightarrow \quad x + y \in \mathcal{V}$$

2. **scalar multiplication:**

$$\text{for any } \alpha \in \mathbf{R}, x \in \mathcal{V} \quad \Rightarrow \quad \alpha x \in \mathcal{V}$$

- the definition 2 implies that a vector space contains the **zero vector**

$$0 \in \mathcal{V}$$

- the two conditions can be combined into one operation:

$$x, y \in \mathcal{V}, \;\; \alpha \in \mathbf{R} \quad \Rightarrow \quad \alpha x + \alpha y \in \mathcal{V}$$

Inner product space

a vector space with an additional structure called *inner product*

an inner product space is a vector space $\mathcal{V}$ over **R** with a map

$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbf{R}$$

for all $x, y, z \in V$ and all scalars $a \in \mathbf{R}$, it satisfies

- conjugate symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- linearity in the first argument:

$$\langle ax, y \rangle = a\langle x, y \rangle, \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$$

- positive definiteness

$$\langle x, x \rangle \geq 0, \quad \text{and} \quad \langle x, x \rangle = 0 \Longleftrightarrow x = 0$$

Examples of inner product spaces

- $\mathbf{R}^n$

$$\langle x, y\rangle = y^T x = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

- $\mathbf{R}^{m\times n}$

$$\langle X, Y\rangle = \mathbf{tr}(Y^T X)$$

- $\mathcal{L}_2(a,b)$: space of real functions defined on (a,b) for which its second-power of the absolute value is Lebesgue integrable, $i.e.$,

$$f \in \mathcal{L}_2(a,b) \quad \Longrightarrow \quad \sqrt{\int_a^b |f(t)|^2 dt} < \infty$$

  the inner product of this space is

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt$$

Orthogonality

let $(\mathcal{V}, \mathbf{R})$ be an inner product space

- x and y are **orthogonal**:

$$x \perp y \quad \Longleftrightarrow \quad \langle x, y \rangle = 0$$

- **orthogonal complement** in $\mathcal{V}$ of $S \subset \mathcal{V}$, denoted by $S^{\perp}$, is defined by

$$S^{\perp} = \{x \in \mathcal{V} \mid \langle x, s \rangle = 0, \ \ \forall s \in S\}$$

- $\mathcal{V}$ admits the **orthogonal decomposition**:

$$\mathcal{V} = \mathcal{M} \oplus \mathcal{M}^{\perp}$$

  where $\mathcal{M}$ is a subspace of $\mathcal{V}$

Orthonormal basis

$\{\phi_n, n \geq 0\} \subset \mathcal{V}$ is an **orthonormal (ON)** set if

$$\langle \phi_i, \phi_j \rangle = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

and is called an **orthonormal basis** for $\mathcal{V}$ if

1. $\{\phi_n, n \geq 0\}$ is an ON set
2. span$\{\phi_n, n \geq 0\} = \mathcal{V}$

we can construct an orthonormal basis from the *Gram-Schmidt* orthogonalization

Orthogonal expansion

let $\{\phi_i\}_{i=1}^n$ be an orthonormal basis for a vector $\mathcal{V}$ of dimension n

for any $x \in \mathcal{V}$, we have the orthogonal expansion:

$$x = \sum_{i=1}^{n} \langle x, \phi_i \rangle \phi_i$$

meaning: we can project x into orthogonal subspaces spanned by each ϕ_i

the norm of x is given by

$$\|x\|^2 = \sum_{i=1}^{n} |\langle x, \phi_i \rangle|^2$$

can be easily calculated by the sum square of projection coefficients

Adjoint of a Linear Transformation

let $A : \mathcal{V} \to \mathcal{W}$ be a linear transformation

the **adjoint** of A, denoted by A^* is defined by

$$\langle Ax, y\rangle_{\mathcal{W}} = \langle x, A^* y\rangle_{\mathcal{V}}, \quad \forall x \in \mathcal{V}, y \in \mathcal{W}$$

A^* is a linear transformation from $\mathcal{W}$ to $\mathcal{V}$

one can show that

$$\begin{aligned} \mathcal{W} &= \mathcal{R}(A) \oplus \mathcal{N}(A^*) \\ \mathcal{V} &= \mathcal{R}(A^*) \oplus \mathcal{N}(A) \end{aligned}$$

Example

$A : \mathbf{C}^n \to \mathbf{C}^m$ and denote $A = \{a_{ij}\}$

for $x \in \mathbf{C}^n$ and $y \in \mathbf{C}^m$, and with the usual inner product in $\mathbf{C}^m$, we have

$$
\begin{aligned}
\langle Ax, y\rangle_{\mathbf{C}^m} &= \sum_{i=1}^{m} (Ax)_i \, \overline{y}_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \overline{y}_i \\
&= \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij} \overline{y}_i \right) = \sum_{j=1}^{n} x_j \overline{\left(\sum_{i=1}^{m} \overline{a_{ij}} y_i \right)} \\
&= \sum_{j=1}^{n} x_j \overline{\left(\overline{A}^T y \right)_j} \triangleq \langle x, \overline{A}^T y \rangle_{\mathbf{C}^n}
\end{aligned}
$$

hence, $A^* = \overline{A}^T$

Basic properties of A^*

Let $A^* : \mathcal{W} \to \mathcal{V}$ be the adjoint of A

facts:

- $\langle A^* y, x \rangle = \langle y, Ax \rangle \Leftrightarrow (A^*)^* = A$
- A^* is a linear transformation
- $(\alpha A)^* = \overline{\alpha} A^*$ for $\alpha \in \mathbf{C}$
- let A and B be linear transformations, then

$$(A + B)^* = A^* + B^* \quad \text{and} \quad (AB)^* = B^* A^*$$

Normed vector space

a **normed vector space** is a vector space $\mathcal{V}$ over a **R** with a map

$$\|\cdot\|:\mathcal{V}\rightarrow\mathbf{R}$$

called **norm** that satisfies

- homogenity

$$\|\alpha x\|=|\alpha|\|x\|,\qquad\forall x\in\mathcal{V},\forall\alpha\in\mathbf{R}$$

- triangular inequality

$$\|x+y\|\leq\|x\|+\|y\|,\qquad\forall x,y\in\mathcal{V}$$

- positive definiteness

$$\|x\|\geq 0,\quad\|x\|=0\iff x=0,\qquad\forall x\in\mathcal{V}$$

Cauchy-Schwarz inequality

for any x, y in an inner product space $(\mathcal{V}, \mathbf{R})$

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

moreover, for $y \neq 0$,

$$\langle x, y\rangle = \|x\|\|y\| \quad \Longleftrightarrow \quad x = \alpha y, \quad \exists \alpha \in \mathbf{R}$$

proof. for any scalar α

$$0 \leq \|x + \alpha y\|^2 = \|x\|^2 + \alpha^2\|y\|^2 + \overline{\alpha}\langle x, y\rangle + \alpha\langle y, x\rangle$$

if $y = 0$ then the inequality is trivial

if $y \neq 0$, then we can choose $\alpha = -\dfrac{\langle x, y\rangle}{\|y\|^2}$

and the C-S inequality follows

Example of vector and matrix norms

$x \in \mathbf{R}^n$ and $A \in \mathbf{R}^{m \times n}$

- 2-norm

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

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

- 1-norm

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|, \quad \|A\|_1 = \textstyle\sum_{ij} |a_{ij}|$$

- ∞-norm

$$\|x\|_\infty = \max_k \{|x_1|, |x_2|, \ldots, |x_n|\}, \quad \|A\|_\infty = \max_{ij} |a_{ij}|$$

Operator norm

matrix operator norm of $A \in \mathbf{R}^{m \times n}$ is defined as

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

also often called **induced norm**

properties:

1. for any x, $\|Ax\| \leq \|A\|\|x\|$
2. $\|aA\| = |a|\|A\|$ (scaling)
3. $\|A + B\| \leq \|A\| + \|B\|$ (triangle inequality)
4. $\|A\| = 0$ if and only if $A = 0$ (positiveness)
5. $\|AB\| \leq \|A\|\|B\|$ (submultiplicative)

examples of operator norms

- **2-norm** or **spectral norm**

$$\|A\|_2 \triangleq \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{\max}(A^T A)}$$

- **1-norm**

$$\|A\|_1 \triangleq \max_{\|x\|_1=1} \|Ax\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}|$$

- **∞-norm**

$$\|A\|_\infty \triangleq \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|$$

note that the notation of norms may be duplicative

Matrix factorizations

- LU factorization
- QR factorization
- singular value decomposition
- Cholesky factorization

LU factorization

for any $n \times n$ matrix A, it admits a decomposition

$$A = PLU$$

with row pivoting

- P permutation matrix, L unit lower triangular, U upper triangular
- the decomposition exists if and only if A is nonsingular
- it is obtained from the Gaussian elimination process

QR factorization

a tall matrix $A \in \mathbf{R}^{m \times n}$ with $m \geq n$ is decomposed as

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$$

- $Q \in \mathbf{R}^{m \times n}$ is an orthogonal matrix ($Q^T Q = I$)
- $R \in \mathbf{R}^{n \times n}$ is an upper triangular
- if $\mathbf{rank}(A) = n$, then n columns in $Q_1 \in \mathbf{R}^{m \times n}$ forms an orthonormal basis for $\mathcal{R}(A)$ and that R_1 is invertible
- if $\mathbf{rank}(A) < n$ then R_1 contains a zero in the diagonal
- QR is obtained by many methods, *e.g.*, Gram Schmidt, Householder transform

Singular value decomposition

ler $A \in \mathbf{R}^{m\times n}$ with $\mathbf{rank}(A) = r \le \min(m,n)$ then

$$A = U \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} V^T, \quad \Sigma_+ = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

$$U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}, \quad U_1 \in \mathbf{R}^{m\times r}, U_2 \in \mathbf{R}^{m\times(m-r)}, \quad U^TU = I_m$$

$$V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}, \quad V_1 \in \mathbf{R}^{n\times r}, V_2 \in \mathbf{R}^{n\times(n-r)}, \quad V^TV = I_n$$

- the singular values of A:

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0, \quad p = \min(m,n)$$

  are the square root of the eigenvalues of A^TA

- columns of U are the eigenvectors of A^TA
- columns of V are the eigenvectors of AA^T
- the reduced form of SVD is $A = U_1\Sigma_+V_1^T$
- the Frobenious norm of A is $\|A\|_F = \mathbf{tr}(\Sigma_+)$
- $\|A\|_2$ is the maximum singular value of A
- $\mathbf{rank}(A)$ is the number of *nonzero* singular value of A

Cholesky factorization

every **positive definite** matrix A can be factored as

$$A = LL^T$$

where L is lower triangular with positive diagonal elements

- L is called the *Cholesky factor* of A
- can be interpreted as 'square root' of a positive define matrix

Derivative and Gradient

Suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ and $x \in \mathbf{int\,dom}\, f$

the **derivative** (or **Jacobian**) of f at x is the matrix $Df(x) \in \mathbf{R}^{m\times n}$:

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

- when f is scalar-valued ($i.e.$, $f : \mathbf{R}^n \to \mathbf{R}$), the derivative $Df(x)$ is a row vector
- its transpose is called the **gradient** of the function:

$$\nabla f(x) = Df(x)^T, \qquad \nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, \dots, n$$

which is a column vector in $\mathbf{R}^n$

Second Derivative

suppose f is a scalar-valued function ($i.e.$, $f : \mathbf{R}^n \to \mathbf{R}$)

the second derivative or **Hessian matrix** of f at x, denoted $\nabla^2 f(x)$ is

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n$$

example: the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$

$$f(x) = (1/2)x^T P x + q^T x + r,$$

where $P \in \mathbf{S}^n, q \in \mathbf{R}^n$, and $r \in \mathbf{R}$

- $\nabla f(x) = Px + q$
- $\nabla^2 f(x) = P$

Chain rule

assumptions:

- $f : \mathbf{R}^n \to \mathbf{R}^m$ is differentiable at $x \in \mathbf{int\,dom}\, f$
- $g : \mathbf{R}^m \to \mathbf{R}^p$ is differentiable at $f(x) \in \mathbf{int\,dom}\, g$
- define the composition $h : \mathbf{R}^n \to \mathbf{R}^p$ by

$$h(z) = g(f(z))$$

then h is differentiable at x, with derivative

$$Dh(x) = Dg(f(x))Df(x)$$

special case: $f : \mathbf{R}^n \to \mathbf{R}$, $g : \mathbf{R} \to \mathbf{R}$, and $h(x) = g(f(x))$

$$\nabla h(x) = g'(f(x))\nabla f(x)$$

example: $h(x) = f(Ax + b)$

$$Dh(x) = Df(Ax + b)A \quad \Rightarrow \quad \nabla h(x) = A^T \nabla f(Ax + b)$$

example: $h(x) = (1/2)(Ax - b)^T P(Ax - b)$

$$\nabla h(x) = A^T P(Ax - b)$$

Function of matrices

we typically encounter some scalar-valued functions of matrix $X \in \mathbf{R}^{m \times n}$

- $f(X) = \mathbf{tr}(A^T X)$ (linear in X)
- $f(X) = \mathbf{tr}(X^T A X)$ (quadratic in X)

definition: the derivative of f (scalar-valued function) with respect to X is

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

note that the differential of f can be generalized to

$$f(X + dX) - f(X) = \langle \frac{\partial f}{\partial X}, dX \rangle + \text{higher order term}$$

Derivative of a trace function

let $f(X) = \mathbf{tr}(A^T X)$

$$\begin{aligned} f(X) &= \sum_i (A^T X)_{ii} = \sum_i \sum_k (A^T)_{ki} X_{ki} \\ &= \sum_i \sum_k A_{ki} X_{ki} \end{aligned}$$

then we can read that $\frac{\partial f}{\partial X} = A$ (by the definition of derivative)

we can also note that

$$f(X + dX) - f(X) = \mathbf{tr}(A^T(X + dX)) - \mathbf{tr}(A^T X) = \mathbf{tr}(A^T dX) = \langle dX, A \rangle$$

then we can read that $\frac{\partial f}{\partial X} = A$

- $f(X) = \mathbf{tr}(X^T A X)$

$$\begin{aligned} f(X+dX) - f(X) &= \mathbf{tr}((X+dX)^T A (X+dX)) - \mathbf{tr}(X^T A X) \\ &\approx \mathbf{tr}(X^T A dX) + \mathbf{tr}(dX^T A X) \\ &= \langle dX, A^T X\rangle + \langle AX, dX\rangle \end{aligned}$$

then we can read that $\frac{\partial f}{\partial X} = A^T X + AX$

- $f(X) = \|Y - XH\|_F^2$ where Y and H are given

$$\begin{aligned} f(X+dX) &= \mathbf{tr}((Y - XH - dXH)^T (Y - XH - dXH)) \\ f(X+dX) - f(X) &\approx -\mathbf{tr}(H^T dX^T (Y - XH)) - \mathbf{tr}((Y-XH)^T dXH) \\ &= -\mathbf{tr}((Y-XH)H^T dX^T) - \mathbf{tr}(H(Y-XH)^T dX) \\ &= -2\langle (Y-XH)H^T, dX\rangle \end{aligned}$$

then we identifiy that $\frac{\partial f}{\partial X} = -2(Y - XH)H^T$

Derivative of a $\log\det$ function

let $f : \mathbf{S}^n \to \mathbf{R}$ be defined by $f(X) = \log\det(X)$

$$\begin{aligned}\log\det(X + dX) &= \log\det(X^{1/2}(I + X^{-1/2}dX X^{-1/2})X^{1/2}) \\ &= \log\det X + \log\det(I + X^{-1/2}dX X^{-1/2}) \\ &= \log\det X + \sum_{i=1}^{n} \log(1 + \lambda_i)\end{aligned}$$

where λ_i is an eigenvalue of $X^{-1/2}dX X^{-1/2}$

$$\begin{aligned}f(X + dX) - f(X) &\approx \sum_{i=1}^{n} \lambda_i \quad (\log(1+x) \approx x, \;\; x \to 0) \\ &= \mathbf{tr}(X^{-1/2}dX X^{-1/2}) \\ &= \mathbf{tr}(X^{-1}dX)\end{aligned}$$

we identify that $\frac{\partial f}{\partial X} = X^{-1}$

References

H. Anton, *Elementary Linear Algebra*, 10th edition, Wiley, 2010

K.B. Petersen and M.S. Pedersen, et.al., *The Matrix Cookbook*, Technical University of Denmark, 2008

S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, 2004

R.A. Horn and C.R. John Son, *Matrix Analysis*, 2nd edition, Cambridge Press, 2013

Chapter 2 in

T. Katayama, *Subspace methods for system identification*, Springer, 2006