Introduction to Convex Optimization

Lecture 1: Linear Algebra Review

Silin DU

Department of Management Science and Engineering
Tsinghua University
dsl21@mails.tsinghua.edu.cn

August 18, 2022

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix

- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

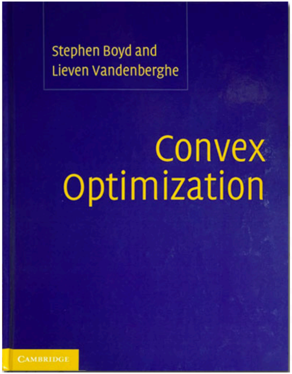
Outline of this Course

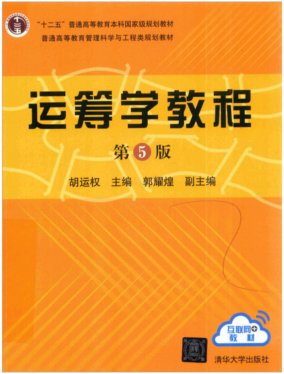
In this course, we focus on convex optimization, including three parts and the following modules:

1. Part I: Review and Preliminaries
   1.1 Linear Algebra Review
   1.2 Linear Programming Review
2. Part II: Theory
   2.1 Convex Sets
   2.2 Convex Functions
   2.3 Convex Optimization Problems
   2.4 Duality
3. Part III: Algorithm
   3.1 Unconstrained Minimization
   3.2 Interior-point Methods

References

1. This course follows the structure of the famous textbook *Convex Optimization* by Stephen Boyd and Lieven Vandenberghe.
2. We also borrow many contents from the course, *Advanced Operations Research*, presented by Prof. Yong Liang in Department of MS&E, Tsinghua University.
3. When reviewing the contents of linear programming, we follow the wonderful textbook in Chinese by Yunquan Hu.





Linear algebra is the basis of convex optimization. This lecture reviews several important concepts and theorems in linear algebra. In addition, basic knowledge in derivatives will be covered.

1. Matrix: Symmetric Matrix, Trace, Orthogonal Matrix
2. Norm: Norm Equivalence, Dual Norm
3. Rank: Inverse, Null Space, Range, Determinant
4. Eigenvalue and Eigenvector, Positive Semi-Definite Matrix
5. Matrix Decomposition
6. Vector Functions

We put some proofs in appendix.

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix

- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

Basic Notations

We first introduce some notations.

- $A \in \mathbb{R}^{m \times n}$: a matrix with m rows and n columns.
- a_{ij}: the element in the i-th row and the j-th column of matrix A.
- $A_{,j}$: the j-th column of matrix A.
- $A_{j,}$: the j-th row of matrix A.
- $x \in \mathbb{R}^n$: a vector with n elements.
- x_i: the i-th element of vector x.
- $\mathbb{R}$: the set of all real numbers.
- $\mathbb{N}$: the set of all natural numbers.
- $\mathbb{Z}$: the set of all integers.
- $\mathbb{R}_+, \mathbb{N}_+, \mathbb{Z}_+$: the set of all non-positive real numbers, natural numbers and integers.
- $\mathbb{C}$: the set of all complex numbers.

Symmetric Matrix I

Definition 1 (Symmetric Matrix)

For a square matrix $A \in \mathbb{R}^{n \times n}$, A is called a symmetric matrix iff $A = A^\top$.

Properties of Transpose:

1. $\left(A^\top\right)^\top = A$.
2. $(cA)^\top = cA^\top$ where c is a constant.
3. $(A \pm B)^\top = A^\top \pm B^\top$.
4. $(AB)^\top = B^\top A^\top$. This result extends to the general case of multiple matrices

$$(A_1 A_2 \ldots A_{k-1} A_k)^\top = A_k^\top A_{k-1}^\top \ldots A_2^\top A_1^\top$$

5. The determinant of a square matrix is the same as the determinant of its transpose. $|A| = |A^\top|$.
6. If A is invertible, then $\left(A^\top\right)^{-1} = \left(A^{-1}\right)^\top$.

Symmetric Matrix II

Definition 2 (Anti-symmetric Matrix)

For a square matrix $A \in \mathbb{R}^{n \times n}$, A is called an anti-symmetric matrix iff $A = -A^\top$.

Here are samples of symmetric matrices and anti-symmetric matrices.

$$\begin{bmatrix} 3 & 1 & 5 \\ 1 & 0 & 6 \\ 5 & 6 & 4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 5 \\ -1 & 0 & 6 \\ -5 & -6 & 0 \end{bmatrix}$$

Note that the diagonal elements of an anti-symmetric matrix must be 0.

Definition 3 (Hermitian Matrix)

For a square matrix $A \in \mathbb{C}^{n \times n}$, A is called a Hermitian matrix if the conjugate transpose of A is identical to itself, i.e. $A = A^*$.

The conjugate transpose of A, denoted by A^*, means:

1. taking the complex conjugate of each elements in A $(a + bi \to a - bi, \text{where } a, b \in \mathbb{R})$;
2. taking the transpose;

Note that the diagonal elements of a Hermitian matrix must be real numbers.

Properties of Symmetric Matrix:

1. For any square matrix $A \in \mathbb{R}^{n \times n}$, $A + A^\top$ is symmetric.
2. For any square matrix $A \in \mathbb{R}^{n \times n}$, $A - A^\top$ is symmetric.

Trace

Definition 4 (Trace)

The trace of a square matrix $A \in \mathbb{R}^{n \times n}$, denoted by $\operatorname{tr}(A)$, is the summation of all diagonal elements of A, i.e.

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

Properties of Trace:

1. $\operatorname{tr}(A) = \operatorname{tr}\left(A^\top\right)$.
2. $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
3. $\operatorname{tr}(cA) = c \cdot \operatorname{tr}(A), c \in \mathbb{R}$.
4. $\operatorname{tr}\left(A^\top B\right) = \operatorname{tr}\left(AB^\top\right) = \operatorname{tr}\left(B^\top A\right) = \operatorname{tr}\left(BA^\top\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$, where $A, B \in \mathbb{R}^{m \times n}$. When $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m}, \operatorname{tr}(AB) = \operatorname{tr}(BA)$.
5.* Cyclic Property: $\operatorname{tr}(ABCD) = \operatorname{tr}(BCDA) = \operatorname{tr}(CDAB) = \operatorname{tr}(DABC)$. In general $\operatorname{tr}(ABC) \neq \operatorname{tr}(BCA)$.
6.* Trace and Eigenvalue: $\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i$, where λ_i is the i-th eigenvalue of A.

Orthogonal Matrix

Definition 5 (Orthogonal)

Two vectors are orthogonal if $x^\top y = 0$.

Definition 6 (Orthogonal Matrix)

A square matrix $A \in \mathbb{R}^{n \times n}$ is orthogonal if $AA^\top = A^\top A = I$.

Definition 7 (Unitary Matrix)

A square matrix $A \in \mathbb{C}^{n \times n}$ is orthogonal if $AA^* = A^* A = I$.

A is orthogonal means $A^\top = A^{-1}$.

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix
- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

Norm

Definition 8 (Norm)

A norm is a function $f : \mathbb{R}^n \to \mathbb{R}$, which satisfies four conditions:

1. Non-negativity: $\forall x \in \mathbb{R}^n, f(x) \geq 0$.
2. Definiteness: $f(x) = 0$ iff $x = 0$.
3. Homogeneity: $\forall x \in \mathbb{R}^n, t \in \mathbb{R}, f(tx) = |t| f(x)$.
4. Triangle-inequality: $\forall x, t \in \mathbb{R}^n, f(x + y) \leq f(x) + f(y)$.

A well-defined norm is a measure of "distance" or "length". Here are samples of different norms.

1. l_2-norm (Euclidean Norm/Distance): $\|x\|_2 = \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}$.
2. l_1-norm: $\|x\|_1 = \sum_{i=1}^n |x_i|$.
3. l_∞-norm: $\|x\|_\infty = \max_i |x|_i$.
4. l_p-norm: $\|x\|_p = \left(\sum_{i=1}^n x_i^p\right)^{\frac{1}{p}}, p \in \mathbb{R}, p \geq 1$.
5. Frobenius-norm (A Matrix Norm):

$$\forall A \in \mathbb{R}^{m \times n}, \|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2} = \sqrt{\operatorname{tr}\left(A^\top A\right)}$$

Norm Equivalence

Theorem 1 (Norm Equivalence)

For any two norms $\|\cdot\|_a$ and $\|\cdot\|_b$, there exists $0 < C_1 \leq C_2$, such that

$$C_1\|x\|_b \leq \|x\|_a \leq C_2\|x\|_b$$

We provide detailed proof in Appendix 1.

Dual Norm

Definition 9 (Dual Norm)

Let $\|\cdot\|$ be a norm in $\mathbb{R}^n$, the associated dual norm $\|\cdot\|_*$ is defined as $\|z\|_* = \sup\{z^\top x \mid \|x\| \leq 1\}$.

Property:

1. The dual of dual-norm is the original norm itself.$(\|\cdot\|_*) = \|\cdot\|$.
2. The dual of a l_2-norm is l_2-norm.
3. The dual of a l_p-norm is l_q-norm, where $\frac{1}{p} + \frac{1}{q} = 1, p \geq 1, q \geq 1$.

For the third property, we provide detailed proof in Appendix 2.

Linear Independence

Definition 10 (Linear Independence)

A set of vectors $\{x_1, x_2, ..., x_k\} \in \mathbb{R}^n$ is said to be linear independence if no vector can be represented as the linear combination of remaining ones. Mathematically, $\{x_1, x_2, ..., x_k\} \in \mathbb{R}^n$ are said to be linear independence if $\sum_{i=1}^{k} a_i x_i = 0$ can be only satisfied by $a_i = 0, \forall i \in 1, 2, ..., k$.

Definition 11 (Affine Independence)

A set of vectors $\{x_0, x_1, x_2, ..., x_k\} \in \mathbb{R}^n$ is said to be affine independence, if there $\nexists \sum_{i=1}^{k} |a_i| > 0$, s.t. $\sum_{i=0}^{k} a_i x_i = 0$ and $\sum_{i=0}^{k} a_i = 0$.

Vectors $\{x_0, x_1, x_2, ..., x_k\}$ are affine independent iff $\{x_i - x_0\}, i = 1, ..., k$ are linear independent.

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix

- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

Column Rank

Definition 12 (Column Rank)

The column rank of a matrix $A \in \mathbb{R}^{m \times n}$ is the size/cardinality of the largest subset of linear independent columns of A.

Property:

1. For any $A \in \mathbb{R}^{m \times n}$, the row rank equals to the column rank.
2. For any $A \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A) \leq \min(m, n)$, and when equality holds, A is full rank.
3. For any $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$, $\operatorname{rank}(AB) \leq \min\left(\operatorname{rank}(A), \operatorname{rank}(B)\right)$.
4. For any $A, B \in \mathbb{R}^{m \times n}$, $\operatorname{rank}(A+B) \leq \operatorname{rank}(A) + \operatorname{rank}(B)$.

Definition 13 (Inverse)

The inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted by A^{-1}, which satisfies $AA^{-1} = I$.

- Non-square matrix A doesn't have inverse but has pseudo inverse, which is corresponding to the singular decomposition.
- A square matrix is invertible iff it's full rank.

Null Space

Definition 14 (Null Space)

The null space of $A \in \mathbb{R}^{m\times n}$ is the following subspace of $\mathbb{R}^n$,

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

Definition 15 (Range)

The range of $A \in \mathbb{R}^{m\times n}$ is defined as

$$\mathcal{R}(A) = \{Ax, x \in \mathbb{R}^n\}$$

simply the linear space spanned by the column vectors of matrix A. So $\mathcal{R}\left(A^\top\right)$ is the linear combination of rows of A.

Theorem 2

For any $x \in \mathbb{R}^n$, we can split x into two parts, namely $x = y + z$, where $y \in \mathcal{N}(A)$ and $z \in \mathcal{R}\left(A^\top\right)$ and $\mathcal{N}(A) \cap \mathcal{R}\left(A^\top\right) = \emptyset$.

Determinant

Recall that in many linear algebra textbooks, we can only learn how to calculate the determinant of a matrix, without a clear definition of it.

Definition 16 (Determinant)

The function $|\cdot| : \mathbb{R}^{n\times n} \to \mathbb{R}$, which satisfies the following conditions and is unique, is termed as "determinant".

1. $|I_n| = 1$.
2. Given A, if multiply a row of A by $t \in \mathbb{R}$, then the determinant of the new matrix is $t|A|$.
3. Exchange two rows of A, the determinant of the new matrix is $-|A|$.

Property

1. $\left|A^\top\right| = |A|$.
2. $A, B \in \mathbb{R}^{m\times n}, |AB| = |A| \cdot |B|$.
3. $A \in \mathbb{R}^{m\times n}, |A| = 0$ iff A is singular.
4. $A \in \mathbb{R}^{n\times n}$ and A is non-singular, $\left|A^{-1}\right| = \frac{1}{|A|}$.

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix

- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

Definition 17 (Eigenvalue and Eigenvector)

For a square matrix $A \in \mathbb{R}^{n\times n}$, if there exists $x \in \mathbb{R}$ and $\lambda \in \mathbb{R}$ such that $Ax = \lambda x$, then λ is an eigenvalue of A and x is an eigenvector corresponding to λ.

Property

1. Trace and Eigenvalue: $\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$.
2. Determinant and Eigenvalue: $|A| = \prod_{i=1}^{n} \lambda_i$.
3. Rank and Eigenvalue: $\mathrm{rank}(A)$ is equal to the number of non-zero eigenvalues of A.
4. If A is non-singular and λ_i is the eigenvalue of A, then $\frac{1}{\lambda_i}$ is the eigenvalue of A^{-1} and they have the same eigenvector.
5. If A is hermitian and full rank, the basis of eigenvectors may be chosen to be mutually orthogonal and the eigenvalues are real.

Eigenvalue and Eigenvector II

Definition 18 (Similarity Transform)

For a given matrix A, pre and post multiplying A by another square matrix V and its inverse V^{-1} gives a similarity transform, i.e., VAV^{-1}.

Similarity Transform preserves the eigenvalue of a matrix, i.e., if λ and u are an eigenpair of A, then λ and Vu are the eigenpair of VAV^{-1}.

Definition 19 (Diagonalizable)

Matrix $A \in \mathbb{R}^{n\times n}$ is called diagonalizable if it's similar to a diagonal matrix B, i.e., $B = VAV^{-1}$.

Matrix $A \in \mathbb{R}^{n\times n}$ is diagonalizable if it has n linearly independent eigenvectors.

Lemma 1

Eigenvectors of distinct eigenvalues are linearly independent.

Positive Semi-Definite Matrix

Definition 20 (Positive Semi-Definite Matrix)

A symmetric matrix is positive semi-definite iff all its eigenvalues are non-negative. Mathematically, $A \in \mathbb{R}^{n \times n}$ is positive semi-definite iff $x^\top Ax \geq 0, \forall x \neq 0, x \in \mathbb{R}^{n \times n}$.

Definition 21 (Positive Definite Matrix)

A symmetric matrix is positive definite iff all its eigenvalues are positive. Mathematically, $A \in \mathbb{R}^{n \times n}$ is positive semi-definite iff $x^\top Ax > 0, \forall x \neq 0, x \in \mathbb{R}^{n \times n}$.

Definition 22 (Negative Definite Matrix)

A symmetric matrix is positive definite iff all its eigenvalues are negative. Mathematically, $A \in \mathbb{R}^{n \times n}$ is positive semi-definite iff $x^\top Ax < 0, \forall x \neq 0, x \in \mathbb{R}^{n \times n}$.

Positive/Negative definite matrix is invertible.

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix

- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

LU Decomposition I

There are several common methods to decompose a matrix, including LU, QR, Cholesky, singular value decomposition and eigendecomposition.

Definition 23 (LU Decomposition)

Let A be a square matrix. An LU decomposition refers to the factorization of A, with proper row and/or column orderings or permutations, into two factors —— a lower triangular matrix L and an upper triangular matrix U:

$$A = LU$$

$$\begin{bmatrix} 4 & 3 \\ 6 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1.5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 0 & -1.5 \end{bmatrix}$$

LU Decomposition II

If A is invertible, then it admits an LU decomposition if and only if all its leading principal minors are nonzero. If A is a singular matrix of rank k, then it admits an LU decomposition if the first k leading principal minors are nonzero, although the converse is not true.

The following matrix doesn't have an LU decomposition.

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

Definition 24 (Minors)

Let A be an $m \times n$ matrix and k an integer with $0 < k \leq m, k \leq n$. A $k \times k$ minor of A, also called minor determinant of order k of A or, if $m = n$, $(n-k)$th minor determinant of A is the determinant of a $k \times k$ matrix obtained from A by deleting $m - k$ rows and $n - k$ columns.

LU Decomposition III

Definition 25 (LDU Decomposition)

A lower-diagonal-upper (LDU) decomposition of A is a decomposition of the from

$$A = LDU$$

where D is a diagonal matrix and L and U are lower and upper triangular matrices. The diagonal entries of L and U are 1.

If A is invertible, then it admits an LDU factorization if and only if all its leading principal minors are nonzero.

QR Decomposition

Definition 26 (QR Decomposition)

Let A be a square matrix. A QR decomposition refers to the factorization of A into two factors —— an orthogonal matrix Q ($Q^\top Q = I$) and an upper triangular matrix U:

$$A = QR$$

If A is invertible, then the factorization is unique and the diagonal elements of R are positive.

Cholesky Decomposition

Definition 27 (Cholesky Decomposition)

The Cholesky decomposition of a symmetric positive-definite matrix A is a decomposition of the form

$$A = LL^\top$$

where L is a lower triangle matrix with real and positive diagonal entries.

- Every symmetric positive-definite matrix has a unique Cholesky decomposition.
- If A can be written as $LL^\top$ for some invertible L, then A is a symmetric positive-definite matrix.

Definition 28 (Eigendecomposition)

Let $A \in \mathbb{R}^{n\times n}$ be a matrix with n linearly independent eigenvectors $q_i (i = 1, 2, ..., n)$. Then A can be factorized as

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

where the i-th column of $Q \in \mathbb{R}^{n\times n}$ is q_i and Λ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues, $\Lambda_{ii} = \lambda_i$.

1. Note that only diagonalizable matrices can be factorized in this way. (See for Definition 19) The following matrix is a counterexample.

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

Eigendecomposition II

2. The decomposition can be derived by the property of eigenvectors.

$$Aq_i = \lambda_i q_i, \quad AQ = Q\Lambda, \quad A = Q\Lambda Q^{-1}$$

3. Generally, the n eigenvectors are orthonormal ($q_i q_i^\top = 1, q_i^\top q_j = 0, \forall i \neq j$) but they need not to be.

4. If all eigenvalues of A are non zero, then A is invertible and its inverse is given by

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

where $\left[\Lambda^{-1}\right]_{ii} = \frac{1}{\lambda_i}$.

Eigendecomposition III

Theorem 3 (Eigendecomposition of Hermitian Matrix)

If A is Hermitian (symmetric), then the eigenvalues of A are real, and it can be decomposed as $A = UDU^*$ ($A = UDU^\top$), where D is a diagonal matrix with $D_{ii} = \lambda_i$ and U forms an orthogonal basis for $\mathbb{C}^n$ ($\mathbb{R}^n$).

- If v is an eigenvector of A, then $cv, c \in \mathbb{R}$ is also an eigenvector of A. When c is complex, we get a complex eigenvector. By Theorem 3, when A is Hermitian, the eigenvalues of A are real. Then we can always pick the eigenvector with real entries, e.g., if $v = a + bi, a, b \in \mathbb{R}^n$, then $A(a + bi) = \lambda(a + bi) \to Aa = \lambda a$.

Definition 29 (Singular Value Decomposition)

For any matrix $A \in \mathbb{R}^{m \times n} (m > n)$ with rank r, A can be decomposed as

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

where U and V are orthogonal matrices with size m and n respectively. And

$$\Sigma = \operatorname{diag}\left(\sqrt{\lambda_1}, ..., \sqrt{\lambda_r}\right)$$

where λ_i are the non-zero eigenvalues of the matrix $AA^\top$, and $\sqrt{\lambda_i}$ are the singular values of A. Matrices U and V satisfy

$$u_i = \frac{1}{\lambda_i} A v_i$$

where u_i and v_i are the i-th columns of matrices U and V, while V is the eigenvector matrix of $AA^\top$.

Singular Value Decomposition II

1. A is invertible $\Leftrightarrow$ A is non-singular.
2. If $A \in \mathbb{C}^{m\times n}$, U and V are unitary matrices and $AA^\top$ should be replaced by AA^*.
3. λ_i are the singular value of A and $\sqrt{\lambda_i}$ are the eigenvalues of the squared matrix $AA^\top$.
4. The number of non-zero singular values of A is equal to the rank of A.
5. The Frobenius norm of a matrix $\|A\|_F$ is equal to the Euclidean norm of the vectors of its singular values, i.e.,

$$\|A\|_F = \sqrt{\sum_{i=1}^{n} \lambda_i^2} = \sqrt{\operatorname{tr}\left(AA^\top\right)}$$

Singular Value Decomposition III

Definition 30 (Pseudo Inverse)

Let $A \in \mathbb{R}^{m\times n}, n \leq m$ be a full-rank matrix. The pseudo inverse of A is $A^{+} = \left(A^\top A\right)^{-1} A^\top$.

1. It's easy to show that $A^{+}A = \left(A^\top A\right)^{-1} A^\top A = I$.
2. Since $A = U\bar{\Sigma}V^\top$, we have

$$
\begin{aligned}
\left(A^\top A\right)^{-1} A^\top &= \left(V\bar{\Sigma}^\top U^\top U\bar{\Sigma}V^\top\right)^{-1} V\bar{\Sigma}^\top U^\top \\
&= \left(V\bar{\Sigma}^\top \bar{\Sigma}V^\top\right)^{-1} V\bar{\Sigma}^\top U^\top && (U \text{ is orthogonal}) \\
&= V\left(\bar{\Sigma}^\top \bar{\Sigma}\right)^{-1} V^\top V\bar{\Sigma}U^\top && (V \text{ is orthogonal}) \\
&= V\left(\bar{\Sigma}^\top \bar{\Sigma}\right)^{-1} \bar{\Sigma}U^\top \\
&= V\bar{\Sigma}^{+}U^\top
\end{aligned}
$$

3. If $n \geq m$, then $A^{+} = A^\top \left(AA^\top\right)^{-1}$ and $AA^{+} = I$.
4. When A is not full-rank, A^{+} still exists. (See for Moore–Penrose inverse)

Schur Decomposition

Definition 31 (Schur Decomposition)

For any matrix $A \in \mathbb{C}^{n \times n}$, A can be factorized as

$$A = UTU^*$$

where $U \in \mathbb{C}^{n \times n}$ is a unitary matrix, and $T \in \mathbb{C}^{n \times n}$ is an upper triangular matrix. When A is a Hermitian matrix (i.e., $A = A^*$), it can be shown that $T^* = T$ and T is a diagonal matrix.

1. Since U is a unitary matrix, we have $A = UTU^{-1}$. A is similar to T (see for Definition 18) and the diagonal entries of T is the eigenvalues of A.

Schur Complement I

Definition 32 (Schur Complement)

Let $M \in \mathbb{R}^{(p+q)\times(p+q)}$ be a block matrix

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

where $A \in \mathbb{R}^{p\times p}, B \in \mathbb{R}^{p\times q}, C \in \mathbb{R}^{q\times p}, D \in \mathbb{R}^{q\times q}$. The Schur complements of M is defined as

$$M_{/A} = D - CA^{-1}B$$
$$M_{/D} = A - BD^{-1}C$$

1. $\det(M) = \det(A) \cdot \det\left(M_{/A}\right) = \det(D) \cdot \det\left(M_{/D}\right)$.

Schur Complement II

2. For a symmetric M,

$$M = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix}$$

   M is positive definite iff A and $M_{/A}$ or D and $M_{/D}$ are positive definite. M is positive semi-definite if and only if A or D is positive definite (needs to be invertible), and $M_{/A}$ or $M_{/D}$ is positive semi-definite.

3. $\operatorname{rank}(M) = \operatorname{rank}(A) + \operatorname{rank}(M_{/A}) = \operatorname{rank}(D) + \operatorname{rank}(M_{/D})$.

Application 1

Let $I_p \in \mathbb{R}^{p\times p}$ be an identity matrix. Matrix L is a lower triangular matrix.

$$
\begin{aligned}
ML &= \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I_p & 0 \\ -D^{-1}C & I_q \end{bmatrix} \\
&= \begin{bmatrix} A - BD^{-1}C & B \\ 0 & D \end{bmatrix} \\
&= \begin{bmatrix} I_p & BD^{-1} \\ 0 & I_q \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}
\end{aligned}
$$

In analogous to LDU decomposition:

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

Application 2

To solve the equations:

$$M \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A & B^\top \\ B & D \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

Suppose A is invertible, then upper rows left multiply BA^{-1} yields:

$$BA^{-1}Ax + BA^{-1}B^\top\lambda = BA^{-1}f$$
$$Bx = BA^{-1}f - BA^{-1}B\lambda$$

Plug into lower row:

$$\lambda = \left(BA^{-1}B^\top - C\right)^{-1}\left(BA^{-1}f - g\right)$$

$BA^{-1}B^\top - C$ is the Schur complement of M on A. When computing M^{-1} is rather difficult, computing A^{-1} and $BA^{-1}B^\top$ can be easier.

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix

- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

Vector Functions I

A function $f : \mathbb{R}^n \to \mathbb{R}^m, n \geq 1, m \geq 1$ takes a vector as inputs and returns a real value or a vector. For example,

$$
\begin{aligned}
f(x) &= \|x\|_2^2 = x^\top x, && x \in \mathbb{R}^n \\
f(x) &= ax + b, && x \in \mathbb{R}^n, a \in \mathbb{R}, b \in \mathbb{R}^n \\
f(x) &= \frac{c^\top x + d}{e^\top x + f}, && x \in \mathbb{R}^n, c, e \in \mathbb{R}^n, e, f \in \mathbb{R} \\
f(x) &= \begin{bmatrix} \cos(x_1) & \sin(x_2) & x_3^2 \end{bmatrix}, && x \in \mathbb{R}^3
\end{aligned}
$$

Vector Functions II

We can apply the rules in calculus to get the gradient of vector functions.

Example 1: l_2-norm

$$f(x) = \|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}, \quad x \in \mathbb{R}^n$$

$$\frac{\partial f}{\partial x_k} = \frac{1}{2}\left(\sum_{i=1}^{n} x_i^p\right)^{-\frac{1}{2}} \cdot 2x_k = \left(\sum_{i=1}^{n} x_i^2\right)^{-\frac{1}{2}} x_k$$

We can also write it as

$$f(x) = \sqrt{x^\top x}, \quad \nabla f(x) = \frac{x}{\sqrt{x^\top x}}$$

Note that $\nabla f(x)$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$.

Example 2: l_p-norm

$$f(x) = \|x\|_p = \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1}{p}}, \quad x \in \mathbb{R}^n, p \geq 1$$

$$\frac{\partial f}{\partial x_k} = \frac{1}{p}\left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1-p}{p}} \cdot p x_k^{p-1}$$

$$= \left(\sum_{i=1}^{n} x_i^p\right)^{\frac{1-p}{p}} x_k^{p-1}$$

Example 3:

$$f(x) = \begin{bmatrix} \cos(x_1) & \sin(x_2) + x_3^2 \end{bmatrix}, \quad x \in \mathbb{R}^3$$

$$\frac{\partial f}{\partial x_1} = \begin{bmatrix} -\sin(x_1) & 0 \end{bmatrix}$$

$$\frac{\partial f}{\partial x_2} = \begin{bmatrix} 0 & \cos(x_2) \end{bmatrix}$$

$$\frac{\partial f}{\partial x_3} = \begin{bmatrix} 0 & 2x_3 \end{bmatrix}$$

Vector Functions IV

Example 4:

$$f(x) = \frac{1}{c^\top x + b}, \quad x, c \in \mathbb{R}^n, b \in \mathbb{R}$$
$$\nabla f(x) = \frac{-c}{\left(c^\top x + b\right)^2}$$

Example 5:

$$f(x) = x^\top A x + b^\top x + c, \quad x, b \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}, c \in \mathbb{R}$$
$$\nabla f(x) = 2Ax + b, \quad \nabla^2 f(x) = 2A$$

Note that $\nabla^2 f(x) : \mathbb{R}^{n\times n} \to \mathbb{R}$.

Contents

Matrix

Norm

Rank

Eigenvalue and Eigenvector

Matrix Decomposition

Vector Functions

Appendix

- Appendix 1: Norm Equivalence
- Appendix 2: Property of Dual Norm

Appendix 1: Norm Equivalence I

Appendix 2: Property of Dual Norm I