Introduction to Convex Optimization

Lecture 1: Linear Algebra Review

Silin DU

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

June 5, 2022

Contents

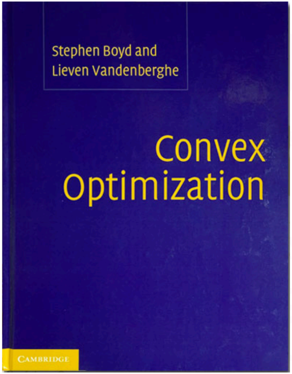
Outline of this Course

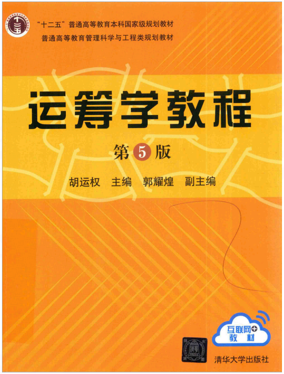
In this course, we focus on convex optimization, including three parts and the following modules:

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.





Lecture Overview

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. Functions of Vectors
7. Derivatives and Gradients

We put all proofs in appendix.

Contents

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 $\left(A^\top\right)^{-1} = \left(A^{-1}\right)^\top$.

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 ($a + bi \to a - bi$, 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: $\operatorname{tr}(A) = \sum_{i}^{n} \lambda_i$, 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 $AA^* = A^* A = I$.

Contents

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

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 $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A) \le \min(m, n)$, and when equality holds, A is full rank.
3. For any $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$, $\text{rank}(AB) \le \min\left(\text{rank}(A), \text{rank}(B)\right)$.
4. For any $A, B \in \mathbb{R}^{m\times n}$, $\text{rank}(A+B) \le \text{rank}(A) + \text{rank}(B)$.

Inverse

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 $\mathcal{R}\left(R^\top\right)$ 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

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 hermition 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.

Contents

Appendix 1: Norm Equivalence I

Appendix 2: Property of Dual Norm I