Optimization Methods and Applications

Lecture 1. Introduction to Optimization Problems, and Convexity

Dr. Huiqi Guan

Fudan University

September 9, 2025

Roadmap

This lecture

- Syllabus
- Introduction to optimization problems
- Convexity
  - convex sets
  - convex functions
  - convex programming

General Information

- Instructor: Dr. Huiqi Guan
  - Education
    - Ph.D., University of Miami
    - M.S. and B.S., Tsinghua University
  - Research Interests
    - Supply chain contract, Interface issues with marketing and information economics, Sharing economy, Economics models in supply chain management
- Office: Room 517, Siyuan Building
- Office hour: before or after class or by appointment
- Email: guanhuiqi@fudan.edu.cn
- TA: Yizhou Lu
- TA's Email: yzlu21@m.fudan.edu.cn

Grading

• Class Participation (at least one class presentation)	20%
• Assignments (once every two weeks)	15%
• Project	5%
• Midterm Exam	20%
• Final Exam	40%

- Open book, and Non-cumulative Exams

Introduction to Optimization

- Optimization often appears in any scenario in which you are trying to make certain decisions and reach the best possible outcome.
- Optimization is concerned with the study of maximization and minimization of mathematical functions. Very often the arguments of (i.e., variables or unknowns in) these functions are subject to side conditions or constraints.

Introduction to Optimization

- By virtue of its great utility in such diverse areas as applied science, engineering, economics, finance, medicine, and statistics, optimization holds an important place in the practical world and the scientific world.
- Indeed, as far back as the Eighteenth Century, the famous Swiss mathematician and physicist Leonhard Euler (1707-1783) proclaimed[1] that
  . . . nothing at all takes place in the Universe in which some rule of maximum or minimum does not appear.

[1] See Leonhardo Eulero, Methodus Inviendi Lineas Curvas Maximi Minimive Proprietate Gaudentes, Lausanne & Geneva, 1744, p. 245.

Quantitative or Mathematical Models

The class of optimization problems considered in this course can all be expressed in the form

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{x} \in \mathcal{X} \end{aligned}$$

where $\mathcal{X}$ usually specified by constraints:

$$\begin{aligned} c_i(\mathbf{x}) &= 0, i \in \mathcal{E} \\ c_i(\mathbf{x}) &\leq 0, i \in \mathcal{I}. \end{aligned}$$

- $\mathbf{x}$: decision variable/activity, data/parameter
- $f(\mathbf{x})$: objective/goal/target
- $\mathcal{X}$: constraint/limitation/requirement
- $\mathcal{E}/\mathcal{I}$:equality/inequality constraint
- $c_i(\mathbf{x})$: constraint function/the right-hand side

Model Classifications

Optimization problems are generally divided into Linear and Nonlinear problems based upon the objective and constraints of the problem

- Linear Optimization: If both the objective and the constraints are linear functions
- Linearly Constrained Optimization: If the constraints are linear functions
- Nonlinear Optimization: If either the objective or the constraints contain nonlinear functions
- There are various sub-classifications among nonlinear problems e.g. quadratic, convex, etc.

The Optimization Process

- Formulate real life problems into mathematical models
  - Study the environment and clearly understand the problem
  - Formulate the problem using verbal description
  - Define notations for parameters and decision variables
  - Construct a model using mathematical expressions
  - Collect necessary data; Transform the raw data to parameter values
- Implement the model and solution algorithms using a computer: analyze the models and develop efficient procedures to obtain best solutions
- Interpret computer solutions and perform sensitivity analysis
- Implementation: put the knowledge gained from the solution to work
- Monitor the validity and effectiveness of the model and update it when necessary

What do You Learn?

- Models–the art: How we choose to represent real problems
- Theory–the science: What we know about different classes of models; e.g. necessary and sufficient conditions for optimality
- Algorithms–the tools: How we apply the theory to robustly and efficiently solve powerful models

Course Description

- Linear Programming
  - Convex analysis, Simplex method, Duality and sensitivity analysis, Network simplex method, Integer programming
- Nonlinear Programming
  - Newton method, Conjugate gradient method, Quadratic programming, Lagrangian duality, Penalty functions
- Algorithm

References

1. Leonard D. Berkovitz. Convexity and Optimization in $\mathbb{R}^n$. John Wiley & Sons, 2001.
2. Bazaraa, Mokhtar S., Hanif D. Sherali, and Chitharanjan M. Shetty. Nonlinear programming: theory and algorithms. John Wiley & Sons, 2006.
3. Michael Kupferschmid. Introduction to Mathematical Programming: Theory and Algorithms of Linear and Nonlinear Optimization.
4. MIT Open courseware, Course No. 15.053.
5. Ignizio, James P., and Tom M. Cavalier. Linear programming. Prentice-Hall, Inc., 1994.
6. Luenberger, David G. and Yinyu Ye. Linear and Nonlinear Programming. Springer, 2010.
7. 孙文瑜等, 最优化方法, 高等教育出版社，2010.

Optimization Problems

- Resource Allocation Problem
- Transportation problem
- Minimum-cost network flow problem
- Facility location problem
- Air Traffic Control
- Travelling salesman problem
- Minimize max-TSP-tour
- Stochastic resource allocation problem
- Supporting Vector Machine

The Art of Modeling

Objective to distill the real-world as accurately and succinctly as possible into a quantitative model

- Don't want models to be too generalized. Models that are overly generalized might not draw much real world value from your results. Ex. Trying to analyze traffic flows by modeling every single individual using different assumptions
- Don't want models to be too specific. Models that are overly specific might lose the ability to solve problems or gain insights. Ex. Analyzing traffic flows assuming every person has the same characteristics

Formulation of Optimization Models: Four-Step Rule

- Sort out data and parameters from the verbal description
- Define the set of decision variables
- Formulate the objective function of data and decision variables
- Set up equality and/or inequality constraints

Resource Allocation Problem

The Wyndor Glass Co. produces high-quality glass products, including wood-framed windows and aluminum-framed glass doors. It has three plants. Aluminum frames and hardware are made in Plant 1, wood frames are made in Plant 2, and Plant 3 is used to produce glass and assemble the products. Wyndor produces to products which require the resources of the 3 plants as follows:

	Production Time per Batch		
Plant	Aluminum	Wood	Production Time
1	1	0	4
2	0	2	12
3	3	2	18
Profit per batch	$3000	$5000	

Find the optimal production plan for Wyndor.

Resource Allocation Problem continued

Suppose Wyndor will produce x_1 wood-frame windows, and x_2 aluminum-frame windows. The optimization problem is formulated as follows:

$$
\begin{aligned}
\max \quad & 3000x_1 + 5000x_2 \\
\text{s.t.} \quad & x_1 \leq 4 \\
& 2x_2 \leq 12 \\
& 3x_1 + 2x_2 \leq 18 \\
& x_1, \; x_2 \geq 0.
\end{aligned}
$$

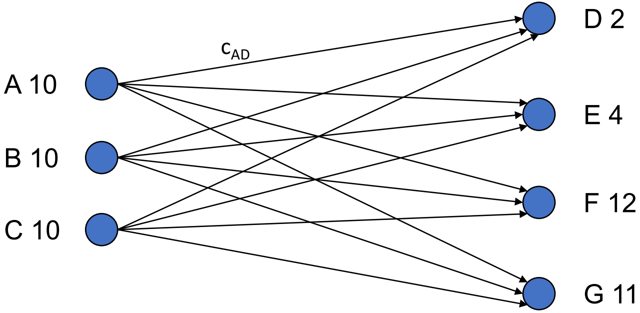
Resource Allocation Problem continued

$$\max \sum_{j=1}^{m} c_j x_j$$
$$\text{s.t.} \sum_{j=1}^{m} a_{ij} x_j \le b_i, \text{for } i = 1, \ldots, n$$
$$x_j \ge 0, \text{for } j = 1, \ldots, m$$

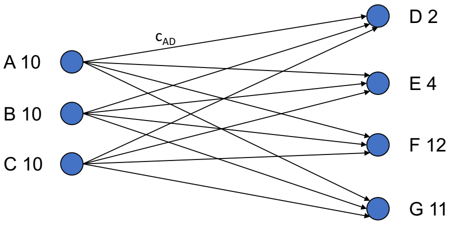
- m= the number of products
- n= the number of resources
- c_j=profit margin of product j
- b_i=amount of resource i
- a_{ij}=consumption of resource i for per unit product j
- x_j=units of product j

Transportation Problem

You have 3 distribution centers (DCs), and need to deliver product to 4 customers. Find cheapest way to satisfy all demand.



Transportation Problem continued



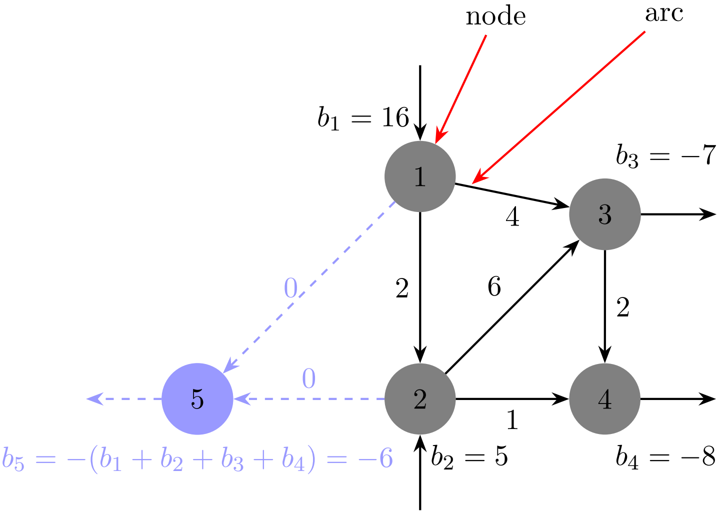
- Parameter
  - M=the number of locations for distribution centers
  - N=the number of locations for customers
  - S_i=supply at location i, for $i = 1, \dots, M$
  - d_j=demand at location j, for $j = 1, \dots, N$
  - c_{ij}=the cost to ship one unit from i to j
- Decision variables
  - x_{ij}=units shipped from DC i to customer j

Transportation Problem Formulation

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{M} \sum_{j=1}^{N} c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{j=1}^{N} x_{ij} \leq S_i, \text{for } i = 1, \ldots, M \\
& \sum_{i=1}^{M} x_{ij} \geq d_j, \text{for } j = 1, \ldots, N \\
& x_{ij} \geq 0, \text{for all } i, j
\end{aligned}
$$

Minimum-cost Network Flow Problem

Find the minimum cost that matches supply and demand.



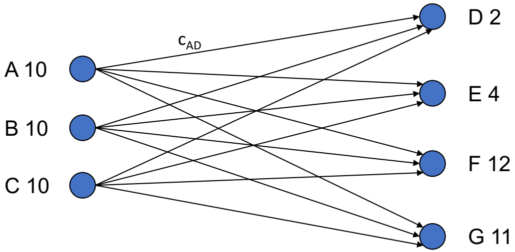
- N=the number of nodes
- A=the set of all arcs
- b_i=supply ($b_i > 0$) or demand ($b_i < 0$) at node i
- c_{ij}=unit cost associated with arc $(i, j) \in A$
- x_{ij}=quantity of flow from node i to node j along arc $(i, j) \in A$

Minimum-cost Network Flow Problem Formulation

$$
\begin{aligned}
\min \quad & \sum_{(i,j)\in A} c_{ij}x_{ij} \\
\text{s.t.} \quad & \sum_{j:(i,j)\in A} x_{ij} - \sum_{j:(j,i)\in A} x_{ji} = b_i, \text{for } i = 1, \ldots, N \\
& x_{ij} \geq 0, \text{for all } (i,j) \in A
\end{aligned}
$$

Facility Location Problem

You will build 3 DCs, and need to deliver product to 4 markets.



Find the locations of DCs with minimized total distance weighted by the shipment from the DCs to the markets.

Facility Location Problem continued

- c_i=capacity at DC i
- (a_j, b_j)=known location of market j for $j = 1, , n$
- r_j=demand at location j
- (x_i, y_i)=unknown location of DC i for $i = 1, , m$
- d_{ij}=distance from DC i to market area j
- x_{ij} =units shipped from DC i to market j

Facility Location Problem Formulation

$$
\begin{aligned}
\min \ & \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} x_{ij} \\
\text{s.t.} \ & \sum_{j=1}^{n} x_{ij} \le c_i, \text{for } i = 1, \ldots, m \\
& \sum_{i=1}^{m} x_{ij} \ge r_j, \text{for } j = 1, \ldots, n \\
& d_{ij} = [(x_i - a_j)^2 + (y_i - b_j)^2]^{1/2}, \text{for } i = 1, \ldots, m; j = 1, \ldots, n \\
& x_{ij} \ge 0, \text{for all } i, j
\end{aligned}
$$

Air Traffic Control

Air plane $j, j = 1, \ldots, n$ arrives at the airport within the time interval $[a_j, b_j]$ in the order of $1, 2, \ldots, n$. The airport wants to find the arrival time for each air plane such that the narrowest metering time (inter-arrival time between two consecutive airplanes) is the greatest.

Air Traffic Control Formulation

Let t_j be the arrival time of plane j. Then

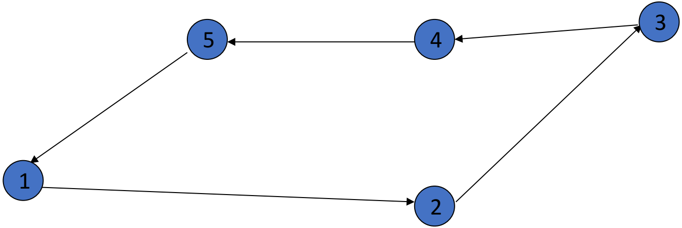
$$\begin{aligned} \max \quad & \min_{j=1,\ldots,n-1} (t_{j+1} - t_j) \\ \text{s.t. } & a_j \le t_j \le b_j, \quad j = 1, 2, \ldots, n. \end{aligned}$$

Air Traffic Control continued

Equivalent smooth formulation:

$$
\begin{aligned}
\max \ & \Delta \\
\text{s.t.} \ & t_2 - t_1 - \Delta \geq 0, \\
& t_3 - t_2 - \Delta \geq 0, \\
& \dots, \\
& t_n - t_{n-1} - \Delta \geq 0, \\
& a_j \leq t_j \leq b_j, \quad j = 1, 2, \dots, n.
\end{aligned}
$$

A Salesman wishes to travel around a given set of cities, and return to the beginning, covering the smallest total distance.



Travelling Salesman Problem continued

- d_{ij}= the distance from city i to city $j, i, j = 1, \ldots, n$
- $x_{ij} = \begin{cases} 1, & \text{if city } j \text{ is visited immediately following city } i \\ 0, & \text{otherwise} \end{cases}$

Travelling Salesman Problem Formulation

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_{ij} \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_{ij} = 1, \text{for } i = 1, \ldots, n \\
& \sum_{i=1}^{n} x_{ij} = 1, \text{for } j = 1, \ldots, n \\
& t_i - t_j + n x_{ij} \leq n - 1, \text{for } i, j = 2, \ldots, n \\
& x_{ij} \in \{0, 1\}, \text{for } i, j = 1, \ldots, n \\
& t_i \in \mathbb{R}, \text{for } i = 1, \ldots, n
\end{aligned}
$$

- There is no subtour without location 1.
- There is tour with location 1.
- Suppose the tour is $1, 2, \ldots, n, 1$. Let $t_i = i, i \geq 2$. We have $t_i - t_j + n x_{ij} \leq n - 1, \text{for } i, j = 2, \ldots, n.$

Minimize max-TSP-tour

Given two-dimension sensor points $\mathbf{a}_j, j = 1, \ldots, n$, and the vehicle locations $b_i, i = 1, \ldots, m$; find the best m clusters assigned to each vehicle such that the maximum of the TSP (Traveling Salesman Problem) tour length is minimized.

Minimize max-TSP-tour continued

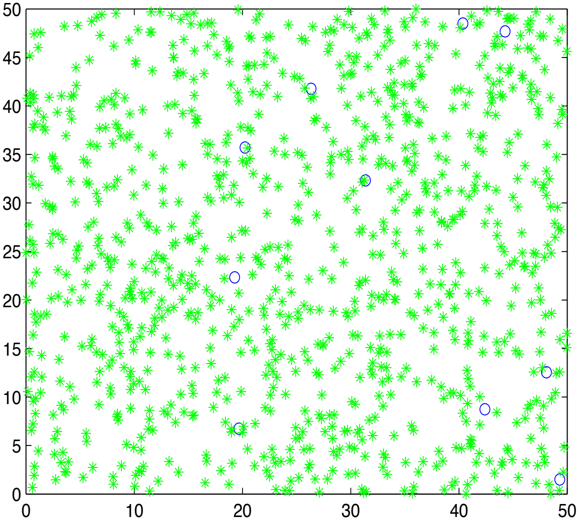


Figure 1: Base-Station Location.

Minimize max-TSP-tour Formulation

$$\begin{aligned} \min \ & \lambda \\ \text{s.t. } & TSP(S_i) \leq \lambda, \quad \forall i \\ & \cup S_i = \mathcal{N}, \end{aligned}$$

where $\mathcal{N}$ is the set of all customers, $S_i \subset \mathcal{N}$ is the subset of points assigned to vehicle i, and $TSP(S_i)$ is the minimal TSP tour-length to visit all points in set S_i by vehicle i.

Stochastic Resource Allocation Problem

$$
\begin{aligned}
\max \quad & \sum_{j=1}^{m} c_j x_j \\
\text{s.t.} \quad & \sum_{j=1}^{m} a_{ij} x_j \le b_i, \text{for } i = 1, \ldots, n \\
& x_j \ge 0, \text{for } j = 1, \ldots, m
\end{aligned}
$$

- c_j=profit margin of product j
- b_i=amount of resource i
- a_{ij}=consumption of resource i for per unit product j
- x_j=units of product j

Stochastic Resource Allocation continued

$$\begin{aligned} \max \quad & \sum_{j=1}^{m} \bar{c}_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^{m} a_{ij} x_j \leq b_i, \text{for } i = 1, \ldots, n \\ & x_j \geq 0, \text{for } j = 1, \ldots, m \end{aligned}$$

- c_j is random with mean $\bar{c}_j$ and the covariance matrix is V.

Stochastic Resource Allocation continued

$$\begin{aligned} \min \ & \mathbf{x^t V x} \\ \text{s.t. } & \sum_{j=1}^{m} a_{ij} x_j \le b_i, \text{for } i = 1, \ldots, n \\ & x_j \ge 0, \text{for } j = 1, \ldots, m \end{aligned}$$

- c_j is random with mean $\bar{c}_j$ and the covariance matrix is V.
- $\mathbf{x^t} = (x_1, \ldots, x_m)$

Stochastic Resource Allocation continued

$$
\begin{aligned}
\min \ & \mathbf{x^t V x} \\
\text{s.t. } & \sum_{j=1}^{m} a_{ij} x_j \leq b_i, \text{for } i = 1, \ldots, n \\
& \bar{\mathbf{c}}^{\mathbf{t}} \mathbf{x} \geq \bar{z} \\
& x_j \geq 0, \text{for } j = 1, \ldots, m
\end{aligned}
$$

- c_j is random with mean $\bar{c}_j$ and the covariance matrix is V.
- $\mathbf{x^t} = (x_1, \ldots, x_m), \bar{\mathbf{c}}^{\mathbf{t}} = (c_1, \ldots, c_m)$
- $\bar{z}$= the required expected payoff

Stochastic Resource Allocation continued

$$
\begin{aligned}
\max \quad & P(z \geq \bar{z}) \\
\text{s.t.} \quad & \sum_{j=1}^{m} a_{ij} x_j \leq b_i, \text{for } i = 1, \ldots, n \\
& \mathbf{c}^{\mathbf{t}} \mathbf{x} = z \\
& x_j \geq 0, \text{for } j = 1, \ldots, m
\end{aligned}
$$

- c_j is random with mean $\bar{c}_j$ and the covariance matrix is V.
- $\mathbf{x}^{\mathbf{t}} = (x_1, \ldots, x_m), \bar{\mathbf{c}}^{\mathbf{t}} = (c_1, \ldots, c_m)$
- $\bar{z}$= the required expected payoff

Stochastic Resource Allocation continued

$$
\begin{aligned}
\max\ & E(u(z)) \\
\text{s.t. } & \sum_{j=1}^{m} a_{ij} x_j \leq b_i, \text{for } i = 1, \ldots, n \\
& \mathbf{c}^{\mathbf{t}} \mathbf{x} = z \\
& x_j \geq 0, \text{for } j = 1, \ldots, m
\end{aligned}
$$

- c_j is random with mean $\bar{c}_j$ and the covariance matrix is V.
- $\mathbf{x}^{\mathbf{t}} = (x_1, \ldots, x_m), \overline{\mathbf{c}}^{\mathbf{t}} = (c_1, \ldots, c_m)$
- $\bar{z}$= the required expected payoff

Stochastic Resource Allocation continued

$$\begin{aligned} \max \ & k\bar{\mathbf{c}}^{\mathbf{t}}\mathbf{x} - \frac{1}{2}k^2\mathbf{x}^{\mathbf{t}}\mathbf{V}\mathbf{x} \\ \text{s.t.} \ & \sum_{j=1}^{m} a_{ij}x_j \leq b_i, \text{for } i = 1, \ldots, n \\ & x_j \geq 0, \text{for } j = 1, \ldots, m \end{aligned}$$

- $\mathbf{x}^{\mathbf{t}} = (x_1, \ldots, x_m), \bar{\mathbf{c}}^{\mathbf{t}} = (c_1, \ldots, c_m)$
- $\mathbf{c} \sim N(\bar{\mathbf{c}}, V), u(z) = 1 - e^{-kz}$

Supporting Vector Machine

Suppose we have two-class discrimination data. We assign the first class with 1 and the second with -1. A powerful discrimination method is the Supporting Vector Machine (SVM).

Supporting Vector Machine continued

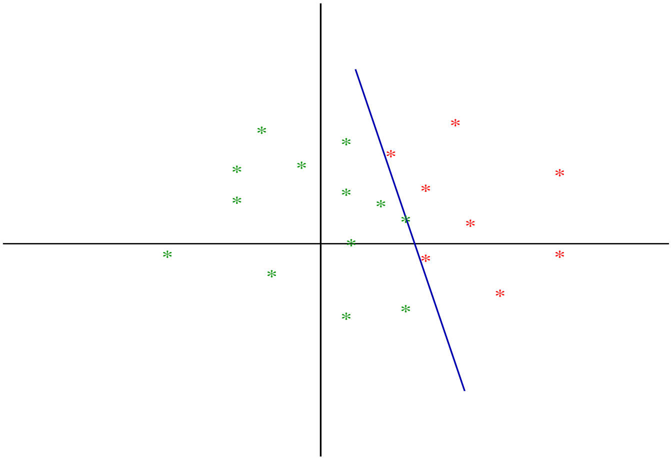


Figure 2: Linear Support Vector Machine

Supporting Vector Machine continued

Let the data point i be given by $\mathbf{a}_i \in \mathbb{R}^d, i = 1, \dots, n$. With this data set, we have some $\bar{y}_i = 1$ (in the first class) and the rest $\bar{y}_i = -1$ (in the second class).

We wish to find a hyperplane in $\mathbb{R}^d$ to separate $\mathbf{a}_i$s with red from $\mathbf{a}_j$s with green. Mathematically, we wish to find $\omega \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ such that

$$\mathbf{a}_i^T \omega + \beta > 0 \quad \forall \{i, \bar{y}_i = 1\}$$

and

$$\mathbf{a}_i^T \omega + \beta < 0 \quad \forall \{i, \bar{y}_i = -1\}.$$

Since the distance between a point $\mathbf{x}_i$ and a hyperplane $\mathbf{x}_i^T \omega + \beta = 0$ is $\frac{|\mathbf{x}_i^T \omega + \beta|}{\|\omega\|}$, then the distance between the data point i and the hyperplane $\mathbf{x}_i^T \omega + \beta = 0$ is $\frac{y_i(\mathbf{a}_i^T \omega + \beta)}{\|\omega\|}$.

The "optimal" hyperplane will maximize the smallest distance between the data points and the separating plane. Thus, we will maximize λ such that $\lambda \leq \frac{y_i(\mathbf{a}_i^T \omega + \beta)}{\|\omega\|}, i = 1, \dots n$.

Supporting Vector Machine continued

Let $\hat{\omega} = \frac{\omega}{\lambda\|\omega\|}$ and $\hat{\beta} = \frac{\beta}{\lambda\|\omega\|}$. Then $\lambda = \frac{1}{\|\hat{\omega}\|}$ and the optimization problem is

$$\begin{aligned} \max \ & \frac{1}{\|\hat{\omega}\|} \\ \text{s.t. } & \bar{y}_i(\mathbf{a}_i^T\hat{\omega} + \hat{\beta}) \geq 1, \quad \forall i. \end{aligned}$$

If a clean separation is possible, we can formulate the problem as an optimization problem:

$$\begin{aligned} \min \ & \|\hat{\omega}\|^2 \\ \text{s.t. } & \bar{y}_i(\mathbf{a}_i^T\hat{\omega} + \hat{\beta}) \geq 1, \quad \forall i. \end{aligned}$$

A clean separation may not be possible for noisy data. Another formulation of the problem is a minimization problem:

$$\begin{aligned} \min \ & \|\hat{\omega}\|^2 + \gamma \sum_{i=1}^{n} (\xi_i)^2 \\ \text{s.t. } & \bar{y}_i(\mathbf{a}_i^T \hat{\omega} + \hat{\beta}) \geq 1 + \xi_i, \quad \forall i. \end{aligned}$$

Convexity

Convexity is the basic concept that must be involved in the optimization theory. The convex nonlinear programming model is a special kind of important model, which plays a very important role in the theoretical proof and algorithm research of optimization.

- Convex Set
- Convex Function
- Convex Programming

Notations

- n-dimensional vector: $\mathbf{x} = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$
  - $\mathbf{e} = (1, \ldots, 1)^t$, all elements equal to 1.
  - $\mathbf{e}_j = \underbrace{(0, \ldots, 1, \ldots, 0)^t}_{jth}$, only the jth element equals to 1.
- Inner product: $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$
- $\|\mathbf{x}\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2} = (\langle \mathbf{x}, \mathbf{x} \rangle)^{1/2}$
- Cauchy–Schwarz inequality: $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$
- Triangle inequality: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- m×n matrix:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} = (\mathbf{a}_1, \mathbf{a}_2, , \ldots, \mathbf{a}_n)$$

  - $\mathbf{I}$: identity matrix, n×n square matrix with ones on the main diagonal and zeros elsewhere.
  - rank($\mathbf{A}$): The rank of $\mathbf{A}$ is the maximum number of linearly independent rows or, equivalently, the maximum number of linearly independent columns of the matrix $\mathbf{A}$.

Definition 1.1

Let D be an arbitrary set in $\mathbb{R}^n$.

- $\text{int}(D)$ (interior of D): $\forall \mathbf{x} \in \text{int}(D), \exists \epsilon > 0$, such that $B(\mathbf{x}, \epsilon) = \{\mathbf{y} :\| \mathbf{y} - \mathbf{x} \| < \epsilon\} \subseteq D$.
- D is called open if $D = \text{int}(D)$.
- ∂D (boundary of D): $\forall \mathbf{x} \in \partial D, \forall \epsilon > 0$, such that $B(\mathbf{x}, \epsilon) \cap D \neq \phi$ and $B(\mathbf{x}, \epsilon) \cap (\mathbb{R}^n \setminus D) \neq \phi$.
- D'(limit points of D): $\mathbf{x} \in D'$ if for every $\epsilon > 0$ there exists a point $\mathbf{x}_\epsilon \neq \mathbf{x}$ such that $\mathbf{x}_\epsilon \in B(\mathbf{x}, \epsilon) \cap D$.
- $\bar{D}$ (closure of D): $\forall \mathbf{x} \in \bar{D}, \forall \epsilon > 0$, such that $B_\epsilon(\mathbf{x}) \cap D \neq \phi$, i.e. $\bar{D} = \text{int}(D) \cup \partial D = D \cup D'$.
- D is called closed if $D = \bar{D}$. If D is closed and $\mathbf{x_n} \to \bar{\mathbf{x}}, \{\mathbf{x_n}\}_{n=1}^{\infty} \in D$, then $\bar{\mathbf{x}} \in D$.
- D is said to be bounded if there exists an $M > 0$ such that $\|\mathbf{x}\| < M$ for all $\mathbf{x} \in D$.

Compact Set

Definition 1.2 (Compact Set)

Let S be a set in $\mathbb{R}^n$. A collection of open sets $\{O_\alpha\}_{\alpha \in A}$ is said to be an open cover of S if every $\mathbf{x} \in S$ is contained in some O_α. A set S in $\mathbb{R}^n$ is said to be compact if for every open cover $\{O_\alpha\}_{\alpha \in A}$ of S there is a finite collection of sets $O_1, \ldots, O_m$ from the original collection $\{O_\alpha\}_{\alpha \in A}$ such that the finite collection $O_1, \ldots, O_m$ is also an open cover of S.

Theorem 1.3

In $\mathbb{R}^n$ a set S is compact if and only if every sequence $\{\mathbf{x}_k\}$ of points in S has a subsequence $\{\mathbf{x}_{k_j}\}$ that converges to a point in S.

Theorem 1.4

In $\mathbb{R}^n$ a set S is compact if and only if S is closed and bounded.

Definitions

Definition 1.5

- A real-valued function f defined on an open set D in $\mathbb{R}^n$ is said to be of class $C^{(k)}$ on D if all of the partial derivatives up to and including those of order k exist and are continuous on D.
- A function $\mathbf{f} = (f_1, f_2, \ldots, f_m)$ defined on D with range in $\mathbb{R}^m, m > 1$, is said to be of class $C^{(k)}$ on D if each of its component functions $f_i, i = 1, \ldots, m$, is of class $C^{(k)}$ on D.
- $$\nabla f = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$$
- Hessian matrix
$$\nabla^2 f = H(\mathbf{x}) = (\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}))$$

Taylor's Theorem

Theorem 1.6

If φ is of class $C^{(k)}$ in an open interval D, then for t_0 in D and h such that $t_0 + h$ is in D,

$$\varphi(t_0 + h) - \varphi(t_0) = \sum_{i=1}^{k-1} \frac{\varphi^{(i)}(t_0)}{i!} h^i + \frac{\varphi^{(k)}(\bar{t})}{k!} h^k,$$

where $\varphi^{(i)}$ denotes the ith derivative of φ and $\bar{t}$ is a point lying between t_0 and $t_0 + h$.

Corollary 1.7

Let D be an open set in $\mathbb{R}^n$ and let f be a real-valued function defined in D. If f is of class $C^{(1)}$ in D, then for each $\mathbf{x}_0$ in D and $\mathbf{h}$ such that $\mathbf{x}_0 + \mathbf{h}$ is in D

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \langle \nabla f(\mathbf{x}_0 + \bar{t}\mathbf{h}), \mathbf{h} \rangle,$$

where $0 < \bar{t} < 1$. If f is of class $C^{(2)}$, then

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \langle \nabla f(\mathbf{x}_0), \mathbf{h} \rangle + \frac{1}{2} \langle \mathbf{h}, H(\mathbf{x}_0 + \bar{t}\mathbf{h})\mathbf{h} \rangle,$$

where $0 < \bar{t} < 1$.

Line

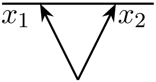
Definition 1.8 (Line)

The line in $\mathbb{R}^n$ through two points $\mathbf{x}_1$ and $\mathbf{x}_2$ is defined to be the set of points $\mathbf{x}$ such that $\mathbf{x} = \mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1)$, where t is any real number, or in set notation

$$\{\mathbf{x} : \mathbf{x} = \mathbf{x}_1 + t(\mathbf{x}_2 - \mathbf{x}_1), -\infty < t < \infty\}$$

or equivalently

$$\{\mathbf{x} : \mathbf{x} = (1-t)\mathbf{x}_1 + t\mathbf{x}_2, -\infty < t < \infty\}$$



Line Segment

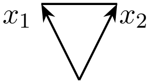
Definition 1.9 (Line Segment)

The closed line segment joining the points $\mathbf{x}_1$ and $\mathbf{x}_2$ is denoted by $[\mathbf{x}_1, \mathbf{x}_2]$ and is defined by

$$[\mathbf{x}_1, \mathbf{x}_2] = \{\mathbf{x} : \mathbf{x} = (1-t)\mathbf{x}_1 + t\mathbf{x}_2, 0 \le t \le 1\}$$

or equivalently

$$[\mathbf{x}_1, \mathbf{x}_2] = \{\mathbf{x} : \mathbf{x} = \alpha\mathbf{x}_1 + \beta\mathbf{x}_2, \alpha \ge 0, \beta \ge 0, \alpha + \beta = 1\}.$$



Note: open line segment: $(\mathbf{x}_1, \mathbf{x}_2) = \{\mathbf{x} : \mathbf{x} = (1-t)\mathbf{x}_1 + t\mathbf{x}_2, 0 < t < 1\}$; half-open line segment which includes $\mathbf{x}_1$ but not $\mathbf{x}_2$: $[\mathbf{x}_1, \mathbf{x}_2) = \{\mathbf{x} : \mathbf{x} = (1-t)\mathbf{x}_1 + t\mathbf{x}_2, 0 \le t < 1\}$; half-open line segment which includes $\mathbf{x}_2$ but not $\mathbf{x}_1$: $(\mathbf{x}_1, \mathbf{x}_2] = \{\mathbf{x} : \mathbf{x} = (1-t)\mathbf{x}_1 + t\mathbf{x}_2, 0 < t \le 1\}$.

Convex Set

Definition 1.10 (Convex Set)

A subset C of $\mathbb{R}^n$ is convex if for every pair of points $\mathbf{x}_1, \mathbf{x}_2$ in C the line segment

$$[\mathbf{x}_1, \mathbf{x}_2] = \{\mathbf{x} : \mathbf{x} = \alpha \mathbf{x}_1 + \beta \mathbf{x}_2, \alpha \geq 0, \beta \geq 0, \alpha + \beta = 1\}$$

belongs to C.

When speaking of a convex set, we shall always assume that the convex set is not empty.

Example 1.11

Convex sets: $\{\mathbf{x}\}, \mathbb{R}^n$.

Example

Example 1.12

The ball $\|\mathbf{x}\| \le r$ is a convex set.

Proof.

For any pair of points $\mathbf{x}, \mathbf{y}$ in the ball and $0 \le \alpha \le 1$, we have

$$\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\| \le \alpha\|\mathbf{x}\| + (1-\alpha)\|\mathbf{y}\| \le \alpha r + (1-\alpha)r = r,$$

which means that $\alpha\mathbf{x} + (1-\alpha)\mathbf{y}$ is in the ball. Thus, the ball is a convex set. $\square$

Combination of Points

- Linear Combination: $\sum_{i=1}^{m} \lambda_i \mathbf{x}_i, \lambda_i \in \mathbb{R}, \mathbf{x}_i \in \mathbb{R}^n, i = 1, \ldots, m.$
- Affine Combination: $\sum_{i=1}^{m} \lambda_i \mathbf{x}_i, \lambda_i \in \mathbb{R}, \mathbf{x}_i \in \mathbb{R}^n, i = 1, \ldots, m$, and $\sum_{i=1}^{m} \lambda_i = 1.$
- Convex Combination: $\sum_{i=1}^{m} \lambda_i \mathbf{x}_i, \lambda_i \geq 0, \mathbf{x}_i \in \mathbb{R}^n, i = 1, \ldots, m$, and $\sum_{i=1}^{m} \lambda_i = 1.$
- Convex Cone Combination: $\sum_{i=1}^{m} \lambda_i \mathbf{x}_i, \lambda_i \geq 0, \mathbf{x}_i \in \mathbb{R}^n, i = 1, \ldots, m.$

Example 1.13

For any pair of points $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2, \mathbf{x}_1, \mathbf{x}_2 \neq \mathbf{0}, \mathbf{x}_1 \neq \lambda \mathbf{x}_2, \forall \lambda \in \mathbb{R}$

- Linear Combination: $\mathbb{R}^2$
- Affine Combination: The line through these two points
- Convex Combination: The line segment with these two points as extreme points
- Convex Cone Combination: The cone with vertex at original and passing through these two points

Examples: Figures

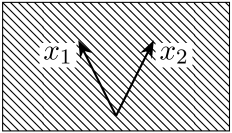


Figure 3: Linear combination.

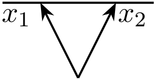


Figure 4: Affine combination.

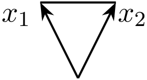


Figure 5: Convex combination.

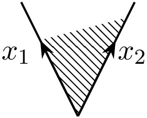


Figure 6: Convex cone combination.

Properties

Lemma 1.14

Let $\{C_\alpha\}$ be a collection of convex sets such that $C = \bigcap_\alpha C_\alpha$ is not empty. Then C is convex.

Note that if A and B are convex sets, then $A \bigcup B$ need not be convex. If A and B are two sets in $\mathbb{R}^n$ and if λ and μ are scalars, we define

$$\lambda A + \mu B = \{\mathbf{x} : \mathbf{x} = \lambda \mathbf{a} + \mu \mathbf{b}, \mathbf{a} \in A, \mathbf{b} \in B\}.$$

Lemma 1.15

If A and B are convex sets in $\mathbb{R}^n$ and λ and μ are scalars, then $\lambda A + \mu B$ is convex.

Lemma 1.16

Let $A_1 \subseteq \mathbb{R}^{n_1}, A_2 \subseteq \mathbb{R}^{n_2}, \ldots, A_k \subseteq \mathbb{R}^{n_k}$ and let $A_1, \ldots, A_k$ be convex. Then $A_1 \times \cdots \times A_k = \{(\mathbf{a}_1, \ldots, \mathbf{a}_n) | \mathbf{a}_i \in A_i, i = 1, \ldots, n\}$ is a convex set in $\mathbb{R}^{n_1 + \cdots + n_k}$.

Examples

Definition 1.17 (Hyperplane)

A hyperplane $H_{\mathbf{a}}^{b}$ is the set of points $\mathbf{x} = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ that satisfy $\langle \mathbf{a}, \mathbf{x} \rangle = \mathbf{a}^{\mathbf{t}}\mathbf{x} = b$, where $\mathbf{a} = (a_1, \ldots, a_n)^t \in \mathbb{R}^n, \mathbf{a} \neq \mathbf{0}$ and $b \in \mathbb{R}$. The vector $\mathbf{a}$ is said to be a normal to the hyperplane.

Definition 1.18 (Half space)

A closed half space corresponding to the hyperplane $\langle \mathbf{a}, \mathbf{x} \rangle = b$, is either of the sets $H_{\mathbf{a}}^{b+} = \{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle \geq b\}$ or $H_{\mathbf{a}}^{b-} = \{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle \leq b\}$. When these half spaces are defined as $\{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle > b\}$ or $\{\mathbf{x} : \langle \mathbf{a}, \mathbf{x} \rangle < b\}$, they are called open half spaces.

Definition 1.19 (Polyhedral set)

A polyhedral set is the intersection of a finite number of half spaces. Thus, the constraint set $S = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ is a polyhedral set because it is the intersection of m half spaces corresponding to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ and n half spaces corresponding to $\mathbf{x} \geq \mathbf{0}$.

Properties

Lemma 1.20

A set C in $\mathbb{R}^n$ is convex if and only if
<u>every convex combination of points</u> in C is also in C.

Lemma 1.21

Let C be a convex set with nonempty interior. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two points with $\mathbf{x}_1 \in \text{int}(C)$ and $\mathbf{x}_2 \in \bar{C}$. Then the line segment $[\mathbf{x}_1, \mathbf{x}_2)$ is contained in the interior of C.

Corollary 1.22

If C is convex, then $\text{int}(C)$ is either empty or convex.

Corollary 1.23

Let C be convex and let $\text{int}(C) \neq \phi$. Then (i) $\overline{\text{int}(C)} = \bar{C}$ and (ii) $\text{int}(C) = \text{int}(\bar{C})$.

Properties

Proof of Lemma 1.21.

First, we assume $\mathbf{x}_2 \in C$. Since $\mathbf{x}_1 \in \text{int}(C)$, there exists $\epsilon > 0$ such that $B(\mathbf{x}_1, \epsilon) \subseteq C$. Consider $\mathbf{z} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, 0 < \alpha \le 1$. For any $\mathbf{y} \in B(\mathbf{z}, \alpha\epsilon)$, let $\mathbf{y}_1 = \frac{\mathbf{y}-(1-\alpha)\mathbf{x}_2}{\alpha}$. Then $\|\mathbf{y}_1 - \mathbf{x}_1\| = \frac{\|\mathbf{y}-\alpha\mathbf{x}_1-(1-\alpha)\mathbf{x}_2\|}{\alpha} = \frac{\|\mathbf{y}-\mathbf{z}\|}{\alpha} < \epsilon$, which yields $\mathbf{y}_1 \in B(\mathbf{x}_1, \epsilon)$. Note that $\mathbf{y} = \alpha\mathbf{y}_1 + (1-\alpha)\mathbf{x}_2$ and $\mathbf{y}_1, \mathbf{x}_2$ belong to C. Since C is convex, then $\mathbf{y} \in C$. Therefore, $B(\mathbf{z}, \alpha\epsilon) \subseteq C$ and then $\mathbf{z} \in \text{int}(C)$.

Second, we assume $\mathbf{x}_2 \in C'$. Since $\mathbf{x}_1 \in \text{int}(C)$, there exists $\epsilon > 0$ such that $B(\mathbf{x}_1, \epsilon) \in C$. Consider $\mathbf{z} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2, 0 < \alpha \le 1$. Since $\mathbf{x}_2 \in C'$, then there exists $\tilde{\mathbf{x}}_2 \in C$ such that $\|\tilde{\mathbf{x}}_2 - \mathbf{x}_2\| < \frac{\alpha\epsilon}{1-\alpha}$. Let $\tilde{\mathbf{x}}_1 = \frac{\mathbf{z}-(1-\alpha)\tilde{\mathbf{x}}_2}{\alpha}$. Then $\|\tilde{\mathbf{x}}_1 - \mathbf{x}_1\| = \frac{\|\mathbf{z}-\alpha\mathbf{x}_1-(1-\alpha)\tilde{\mathbf{x}}_2\|}{\alpha} = \frac{\|(1-\alpha)(\mathbf{x}_2-\tilde{\mathbf{x}}_2)\|}{\alpha} < \epsilon$, which yields $\tilde{\mathbf{x}}_1 \in B(\mathbf{x}_1, \epsilon)$ and then $\tilde{\mathbf{x}}_1 \in \text{int}(C)$. Since $\tilde{\mathbf{x}}_1 \in \text{int}(C), \tilde{\mathbf{x}}_2 \in C$ and $\mathbf{z} = \alpha\tilde{\mathbf{x}}_1 + (1-\alpha)\tilde{\mathbf{x}}_2$, by the first part, we know $\mathbf{z} \in \text{int}(C)$. $\square$

Convex Hull

Definition 1.24

The convex hull of a set A, denoted by $\text{co}(A)$, is the intersection of all convex sets containing A.

For a given set A, let $K(A)$ denote the set of all convex combinations of points in A, i.e.

$$K(A) = \{\sum_{i=1}^{m} \lambda_i \mathbf{x}_i | \mathbf{x}_i \in A, \lambda_i \geq 0, i = 1, 2, \ldots, m, \sum_{i=1}^{m} \lambda_i = 1, m \in \mathbb{N}\}.$$

It is easy to verify that $K(A)$ is convex. Clearly, $K(A) \supseteq A$.

Theorem 1.25

The convex hull of a set A is the set of all convex combinations of points in A; that is, $\text{co}(A) = K(A)$.

Corollary 1.26

A set A is convex if and only if $A = \text{co}(A)$.

Convex Hull

Definition 1.27

For each positive integer n, define

$$P_n = \{\mathbf{p} = (p_1, \ldots, p_n) : p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}.$$

Theorem 1.28

Let A be a subset of $\mathbb{R}^n$ and let $\mathbf{x} \in \text{co}(A)$. Then there exist $n+1$ points $\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}$ in A and a point $\mathbf{p}$ in P_{n+1} such that

$$\mathbf{x} = p_1\mathbf{x}_1 + \cdots + p_{n+1}\mathbf{x}_{n+1},$$

which indicates $\text{co}(A) = P_{n+1} \times \underbrace{A \times \cdots \times A}_{\text{n+1}}$

Lemma 1.29

If $O \neq \phi$ is an open subset of $\mathbb{R}^n$, then $\text{co}(O)$ is also open, i.e. $\text{co}(O) = \text{int}(\text{co }(O))$.

Note: Since $\text{int}(\text{co }(O))$ is convex and contains O, then $\text{co}(O) \subseteq \text{int}(\text{co }(O))$.

Convex Function

Definition 1.30 (Convex Function)

A real-valued function f defined on a convex set C in $\mathbb{R}^n$ is said to be convex if for every $\mathbf{x}_1$ and $\mathbf{x}_2$ in C and every $\alpha > 0, \beta > 0, \alpha + \beta = 1$,

$$f(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) \leq \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2). \tag{1}$$

The function f is said to be strictly convex if strict inequality holds in (1).

Note: The function value of the convex combination of the variables is no greater than the convex combination of the function values of the variables.

Definition 1.31

A function f defined on a convex set C is said to be concave if $-f$ is convex.

Geometric Interpretation

The geometric interpretation of the Definition 1.30 is that a convex function is a function such that if $\mathbf{z}_1$ and $\mathbf{z}_2$ are any two points in $\mathbb{R}^{n+1}$ on the graph of f, then points of the line segment $[\mathbf{z}_1, \mathbf{z}_2]$ joining $\mathbf{z}_1$ and $\mathbf{z}_2$ lie on or above the graph of f.

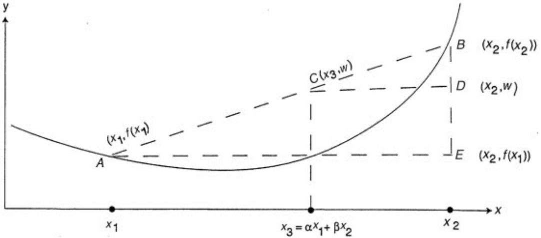


Figure 7: Geometric interpretation of convex function. Note: $w = \alpha f(\mathbf{x}_1) + \beta f(\mathbf{x}_2)$.

Examples

Example 1.32

Show $f(x) = (x-1)^2$ is a strictly convex function on $\mathbb{R}$.

Proof.

For any $x, y \in \mathbb{R}$ and $x \neq y, \alpha \in (0,1)$, we have

$$\begin{aligned} & f(\alpha x + (1-\alpha)y) - (\alpha f(x) + (1-\alpha) f(y)) \\ = & (\alpha x + (1-\alpha)y - 1)^2 - \alpha(x-1)^2 - (1-\alpha)(y-1)^2 \\ = & -\alpha(1-\alpha)(x-y)^2 < 0, \end{aligned}$$

which indicates that $f(x)$ is a strictly convex function on $\mathbb{R}$. □

Example 1.33

Show the linear function $f(\mathbf{x}) = \mathbf{c}^{\mathbf{t}}\mathbf{x} = c_1x_1 + c_2x_2 + \cdots + c_nx_n$ is a convex function on $\mathbb{R}^n$.

Jensen's Inequality

Theorem 1.34 (Jensen's Inequality)

Let f be a real-valued function defined on a convex set C in $\mathbb{R}^n$. Then f is convex if and only if for every finite set of points $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in C and every $\lambda = (\lambda_1, \ldots, \lambda_k)$ in P_k

$$f(\lambda_1\mathbf{x}_1 + \cdots + \lambda_k\mathbf{x}_k) \leq \lambda_1 f(\mathbf{x}_1) + \cdots + \lambda_k f(\mathbf{x}_k).$$

Proof.

Suppose f is convex and $\lambda_1 + \lambda_2 + \lambda_3 = 1, \lambda_i \geq 0$ and $\lambda_1 < 1$. Then

$$\begin{aligned}
f(\lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2 + \lambda_3\mathbf{x}_3) &= f(\lambda_1\mathbf{x}_1 + (1-\lambda_1)(\frac{\lambda_2}{1-\lambda_1}\mathbf{x}_2 + \frac{\lambda_3}{1-\lambda_1}\mathbf{x}_3) \\
&\leq \lambda_1 f(\mathbf{x}_1) + (1-\lambda_1) f(\frac{\lambda_2}{1-\lambda_1}\mathbf{x}_2 + \frac{\lambda_3}{1-\lambda_1}\mathbf{x}_3) \\
&\leq \lambda_1 f(\mathbf{x}_1) + \lambda_2 f(\mathbf{x}_2) + \lambda_3 f(\mathbf{x}_3).
\end{aligned}$$

$\square$

Epigraph

Definition 1.35

Let f be a real-valued function defined on a set A. The epigraph of f, denoted by $\text{epi}(f)$, is the set

$$\text{epi}(f) = \{(\mathbf{x}, y) : \mathbf{x} \in A, y \in \mathbb{R}, y \geq f(\mathbf{x})\}.$$

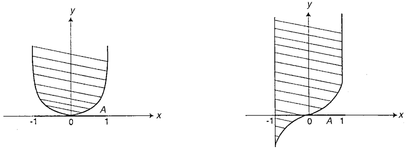


Figure 8: Illustration of epigraph.

Lemma 1.36

A function f defined on a convex set C is convex if and only if $\text{epi}(f)$ is convex.

Level Set

Definition 1.37

Let f be a real-valued function defined on a set A. The level set of f on A associated to a real value γ, denoted by $\mathrm{s}(f, \gamma)$, is the set

$$\mathrm{s}(f, \gamma) = \{\mathbf{x} : \mathbf{x} \in A, f(\mathbf{x}) \leq \gamma\}.$$

Lemma 1.38

If f is a convex function defined on a convex set C in $\mathbb{R}^n$, then for any real γ the set $\{\mathbf{x} : f(\mathbf{x}) \leq \gamma\}$ is either empty or convex.

Note: The converse to this lemma is false, as can be seen from the function $f(x) = x^3$.

Properties

Lemma 1.39

Let $\{f_\alpha\}$ be a family of convex functions defined on a convex set C in $\mathbb{R}^n$ and let there exist a real-valued function M with domain C such that, for each α, $f_\alpha(\mathbf{x}) \leq M(\mathbf{x})$ for all $\mathbf{x}$ in C. Then the function f defined by $f(\mathbf{x}) = \sup\{f_\alpha(\mathbf{x}) : \alpha\}$ is convex.

Hint: $y \geq f(\mathbf{x}) = \sup\{f_\alpha(\mathbf{x}) : \alpha\} \Leftrightarrow y \geq f_\alpha(\mathbf{x}), \forall \alpha$.

Lemma 1.40

Let $f_1, f_2, \ldots, f_k$ be convex functions defined on a convex set C in $\mathbb{R}^n$ and then $\varphi(\mathbf{x}) = \sum_{i=1}^k \lambda_i f_i(\mathbf{x}), \forall \lambda_i \geq 0, i = 1, 2, \ldots, k$ is convex.

Lemma 1.41

Let f be a convex function defined on a convex set C in $\mathbb{R}^n$. Let g be a nondecreasing convex function defined on an interval I in $\mathbb{R}$. Let $f(C) \subseteq I$. Then the composite function $g \circ f$ defined by $(g \circ f)(x) = g(f(x))$ is convex on C.

Three-chord Property

Lemma 1.42 (Three-chord Property)

Let g be a convex function defined on an interval $I \subseteq \mathbb{R}$. Then if $x < y < z$ are three points in I,

$$\frac{g(y) - g(x)}{y - x} \leq \frac{g(z) - g(x)}{z - x} \leq \frac{g(z) - g(y)}{z - y}.$$

If g is strictly convex, then strict inequality holds.

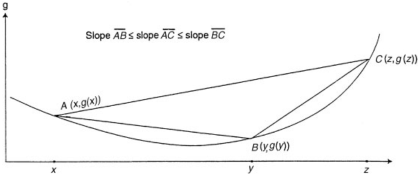


Figure 9: Illustration of three-chord property.

One-dimension Convex Function

Corollary 1.43

Let f be a convex function defined on a convex set C in $\mathbb{R}$. If x is an interior point of C, then f is continuous at x.

Proof.

For each interior point x in C, there exists an $\epsilon > 0$ such that $I = (x - \epsilon, x + \epsilon) \in C$. If $x - \epsilon < x_1 < y < x < z < x_2 < x + \epsilon$. By Lemma 1.42,

$$\frac{f(x) - f(x_1)}{x - x_1} \leq \frac{f(x) - f(y)}{x - y} \leq \frac{f(x_2) - f(x)}{x_2 - x},$$

$$\frac{f(x) - f(x_1)}{x - x_1} \leq \frac{f(z) - f(x)}{z - x} \leq \frac{f(x_2) - f(x)}{x_2 - x}.$$

Then when y, z are close enough to x, $f(y)$ and $f(z)$ will also be close enough to $f(x)$. Hence, f is continuous at x. □

One-dimension Convex Function

Theorem 1.44

Let f be of class $C^{(1)}$ on an open interval $I \subseteq \mathbb{R}$.

(1) Then f is convex on I if and only if for each x_0 in C $f(x) \geq f(x_0) + f'(x_0)(x - x_0)$ for all x in C. Also, f is strictly convex if and only if for each x_0 in C and all $x \neq x_0$ in C, $f(x) > f(x_0) + f'(x_0)(x - x_0)$.

(2) Let f be of class $C^{(2)}$ on an open interval $I \subseteq \mathbb{R}$. Then f is convex on I if and only if $f^{''}(x) \geq 0$ for all x in I. If $f^{''}(x) > 0$ for all x in I, then f is strictly convex.

Note: If f is strictly convex, then we cannot conclude that $f^{''}(x) > 0$ for all x, as the example $f(x) = x^4$ shows.

Proof.

(1) By Lemma 1.42.

(2) By the preceding result and Taylor's theorem, $f(x) - f(x_0) - f'(x_0)(x - x_0) = \frac{1}{2} f^{''}(x_*)(x - x_0)^2 \geq 0$, where x_* is a point between x_0 and x.

□

Applications of Convex Function

Example 1.45

(1) (Young's Inequality) $x^{\frac{1}{p}} y^{\frac{1}{q}} \le \frac{x}{p} + \frac{y}{q}$, if $x, y \ge 0, p, q > 0, \frac{1}{p} + \frac{1}{q} = 1$.

(2) (Hölder's Inequality) $\sum_{k=1}^{n} x_k y_k \le (\sum_{k=1}^{n} x_k^p)^{\frac{1}{p}} (\sum_{k=1}^{n} y_k^q)^{\frac{1}{q}}$ if $x_i, y_i \ge 0, p, q > 0, \frac{1}{p} + \frac{1}{q} = 1$.

Proof.

(1) Consider $f(x) = -\ln(x)$. For $x > 0$, $f^{'}(x) = -\frac{1}{x}$ and $f^{''}(x) = \frac{1}{x^2} > 0$. Therefore, $f(x)$ is a convex function.

(2) Let $x = \frac{x_k^p}{\sum_{k=1}^{n} x_k^p}$ and $y = \frac{y_k^q}{\sum_{k=1}^{n} y_k^q}$ in Young's Inequality, and then take sum from $k = 1$ to n.

□

Application of Jensen's Inequality

Example 1.46

(1) $\frac{x_1+x_2+\cdots+x_n}{n} \leq (\frac{x_1^p+x_2^p+\cdots+x_n^p}{n})^{\frac{1}{p}}$, if $x_i \geq 0, p \geq 1$.

(2) $\frac{x_1+x_2+\cdots+x_n}{n} \geq (\frac{x_1^p+x_2^p+\cdots+x_n^p}{n})^{\frac{1}{p}}$, if $x_i \geq 0, 0 < p \leq 1$.

Proof.

Consider $f(x) = x^p$. For $x > 0$, $f^{'}(x) = px^{p-1}$ and $f^{''}(x) = p(p-1)x^{p-2}$. Therefore, $f(x)$ is a convex function if $p \geq 1$ and a concave function if $0 < p \leq 1$. □

Discriminant Theorem of Convex Function

Theorem 1.47

Let f be a convex function defined on a convex set C in $\mathbb{R}^n$, and for any $\mathbf{x}, \mathbf{y}$, define $\varphi(t) = f((1-t)\mathbf{x} + t\mathbf{y}), t \in [0,1]$, then

(1) f is convex if and only if for any $\mathbf{x}, \mathbf{y} \in C$, the univariate function $\varphi(t)$ is a convex function on $[0,1]$.

(2) For any $\mathbf{x}, \mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}$, if $\varphi(t)$ is a strictly convex function on $[0,1]$, then $f(\mathbf{x})$ is a strictly convex function on C.

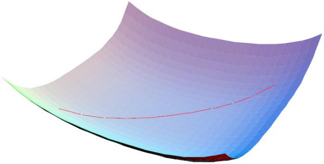


Figure 10: One-dimensional section of the convex function is convex.

Differentiable Convex Function

Theorem 1.48

Let C be an open convex set in $\mathbb{R}^n$ and let f be real valued and differentiable on C. Then f is convex if and only if for each $\mathbf{x}_0$ in C

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \langle \nabla f(\mathbf{x}_0), \mathbf{x} - \mathbf{x}_0 \rangle$$

for all $\mathbf{x}$ in C. Also, f is strictly convex if and only if for each $\mathbf{x}$ in C and all $\mathbf{x} \neq \mathbf{x}_0$ in C

$$f(\mathbf{x}) > f(\mathbf{x}_0) + \langle \nabla f(\mathbf{x}_0), \mathbf{x} - \mathbf{x}_0 \rangle.$$

Proof.

Consider $\varphi(t) = f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0)), t \in [-\epsilon, \epsilon]$($\epsilon$ small enough), and note $\varphi'(t) = \langle \nabla f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0)), \mathbf{x} - \mathbf{x}_0 \rangle$. □

Differentiable Convex Function

Theorem 1.49

Let f be of class $C^{(2)}$ on an open convex set D. Then f is convex on D if and only if the Hessian matrix

$$H(\mathbf{x}) = (\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}))$$

is positive semidefinite at each point $\mathbf{x}$ in D. If $H(\mathbf{x})$ is positive definite at each $\mathbf{x}$, then f is strictly convex.

Proof.

Consider $\varphi(t) = f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0)), t \in [-\epsilon, \epsilon]$($\epsilon$ small enough), and note $\varphi'(t) = \langle \nabla f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0)), \mathbf{x} - \mathbf{x}_0 \rangle$, and
$\varphi''(t) = \langle \mathbf{x} - \mathbf{x}_0, H(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0))(\mathbf{x} - \mathbf{x}_0) \rangle$. □

Convex Programming

Definition 1.50 (Convex Programming)

Let f be a convex function defined on a convex set C in $\mathbb{R}^n$. We define the problem $\min_{\mathbf{x}\in C} f(\mathbf{x})$ as a Convex Programming Problem.

Example 1.51

The linear programming:

$$\begin{aligned} \min\ & \mathbf{c}^t\mathbf{x} \\ \text{s.t. } & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned}$$

is a Convex Programming Problem.

Example

Example 1.52

Let C be an open convex set in $\mathbb{R}^n$, f be a convex function on C, $g_i, i = 1, \ldots, m$ be concave functions and $h_j, j = 1, \ldots, l$ be linear functions on $\mathbb{R}^n$. Then the following three programming problems are Convex Programming Problems:

$$(1)\ \begin{cases} \min & f(\mathbf{x}) \\ \text{s.t.} & h_j(\mathbf{x}) = 0, j = 1, \ldots, l, \end{cases}$$

$$(2)\ \begin{cases} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0, i = 1, \ldots, m, \end{cases}$$

$$(3)\ \begin{cases} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \geq 0, i = 1, \ldots, m, \\ & h_j(\mathbf{x}) = 0, j = 1, \ldots, l, \end{cases}$$

Example

Example 1.53

Is the following programming convex?

$$\begin{cases} \min & f(\mathbf{x}) = x_1^2 + x_2^2 - 4x_1 + 4 \\ s.t. & g_1(\mathbf{x}) = -x_1 + x_2 - 2 \le 0, \\ & g_2(\mathbf{x}) = x_1^2 - x_2 + 1 \le 0, \\ & x_1, x_2 \ge 0. \end{cases}$$

Proof.

Since $\nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\nabla^2 g_1(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 g_1}{\partial x_1^2} & \frac{\partial^2 g_1}{\partial x_2 \partial x_1} \\ \frac{\partial^2 g_1}{\partial x_1 \partial x_2} & \frac{\partial^2 g_1}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\nabla^2 g_2(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 g_2}{\partial x_1^2} & \frac{\partial^2 g_2}{\partial x_2 \partial x_1} \\ \frac{\partial^2 g_2}{\partial x_1 \partial x_2} & \frac{\partial^2 g_2}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$ are all positive semidefinite, the programming in the example is convex. □

Convex Programming Property

Theorem 1.54

(1) Every local minimizer of the convex programming problem is also global, and all the global minimizers consist a convex set.

(2) Let f be a strictly convex function defined on a convex set C in $\mathbb{R}^n$. The convex programming problem $\min_{\mathbf{x}\in C} f(\mathbf{x})$ has a local minimizer $\mathbf{x}^*$. Then $\mathbf{x}^*$ is the unique global minimizer.

Proof.

Suppose $\mathbf{x}^*$ is a local minimizer of the convex programming problem $\min_{\mathbf{x}\in C} f(\mathbf{x})$. Then there exists a $\delta > 0$ such that $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in B(\mathbf{x}^*, \delta) \cap C$. If $\mathbf{x}^*$ is not the global minimizer, then there exists an $\bar{\mathbf{x}} \in C$ such that $f(\mathbf{x}^*) > f(\bar{\mathbf{x}})$. Since f is a convex function, then $f(\alpha\bar{\mathbf{x}} + (1-\alpha)\mathbf{x}^*) \leq \alpha f(\bar{\mathbf{x}}) + (1-\alpha) f(\mathbf{x}^*) < \alpha f(\mathbf{x}^*) + (1-\alpha) f(\mathbf{x}^*) = f(\mathbf{x}^*), \forall \alpha \in (0,1]$. But for small enough α, we have $\alpha\bar{\mathbf{x}} + (1-\alpha)\mathbf{x}^* \in B(\mathbf{x}^*, \delta) \cap C$ and then $f(\mathbf{x}^*) \leq f(\alpha\bar{\mathbf{x}} + (1-\alpha)\mathbf{x}^*)$, which yields a contradiction. $\square$

Convexification

Claim: any optimization problem can be transformed into a convex programming problem.

Claim 1: any optimization problem with a nonlinear objective function can be transformed into another problem with a linear objective.

Let's start with an unconstrained problem:

$$\begin{array}{c}\text{minimize } f(\boldsymbol{x}) \\ \text{subject to } \boldsymbol{x} \in \mathbb{R}^n\end{array} \Longleftrightarrow \begin{array}{c}\text{minimize } z \\ \text{subject to } \boldsymbol{x} \in \mathbb{R}^n, z \in \mathbb{R} \\ f(\boldsymbol{x}) \le z\end{array}$$

Note that the new objective function in terms of the auxiliary variable z is Linear.

Convexification

In general, we have

$$
\begin{array}{rl}
\text{minimize} & f(\boldsymbol{x}) \\
\text{subject to} & \boldsymbol{x} \in \mathbb{R}^n \\
& \mathbf{h}(\mathbf{x}) = \mathbf{0} \\
& \mathbf{g}(\mathbf{x}) \leq \mathbf{0}
\end{array}
\iff
\begin{array}{rl}
\text{minimize} & z \\
\text{subject to} & \boldsymbol{x} \in \mathbb{R}^n, z \in \mathbb{R} \\
& \mathbf{h}(\mathbf{x}) = \mathbf{0} \\
& \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\
& f(\boldsymbol{x}) \leq z
\end{array}
$$

Convexification: Given a noncovex problem, how to reformulate it as a convex problem?

Convexification

Consider the problem:

$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}) \\ &\text{subject to } \boldsymbol{x} \in \mathbb{R}^n \\ &\qquad \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ &\qquad \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \end{aligned} \iff \begin{aligned} &\text{minimize } f(\boldsymbol{x}) \\ &\text{subject to } \boldsymbol{x} \in \mathbb{R}^n \\ &\qquad \mathbf{x} \in S(\text{ feasible set}) \end{aligned}$$

Theorem 1.55

If the objective function $f(\boldsymbol{x})$ is linear, then the following problems are equivalent:

$$\begin{aligned} &(1) \text{ minimize } f(\boldsymbol{x}) \\ &\quad \text{subject to } \boldsymbol{x} \in \mathbb{R}^n \\ &\qquad \boldsymbol{x} \in S(\text{feasible set}) \end{aligned} \iff \begin{aligned} &(2) \text{ minimize } f(\boldsymbol{x}) \\ &\quad \text{subject to } \boldsymbol{x} \in \mathbb{R}^n \\ &\qquad \boldsymbol{x} \in \mathrm{co}(S)(\text{convex hull of} S) \end{aligned}$$

Convexification

Note that

- (1) may or may not be convex, but (2) is always convex.
- The above theorem implies that:
  - (1) and (2) have the same (globally optimal) objective value
  - Every global solution of (1) is solution of (2) (but not the other way around)