Introduction

Dabeen Lee

Industrial and Systems Engineering, KAIST

IE 539: Convex Optimization

September 2, 2024

Prologue

- In today's fast-paced world driven by data, the ability to extract valuable insights and make informed decisions is more crucial than ever.

Prologue

- In today's fast-paced world driven by data, the ability to extract valuable insights and make informed decisions is more crucial than ever.
- Optimization, the process of finding the best solution among a set of alternatives, lies at the heart of this endeavor.

Prologue

- In today's fast-paced world driven by data, the ability to extract valuable insights and make informed decisions is more crucial than ever.
- Optimization, the process of finding the best solution among a set of alternatives, lies at the heart of this endeavor.

Examples

- Predicting customer behavior.

Prologue

- In today's fast-paced world driven by data, the ability to extract valuable insights and make informed decisions is more crucial than ever.
- Optimization, the process of finding the best solution among a set of alternatives, lies at the heart of this endeavor.

Examples

- Predicting customer behavior.
- Optimizing supply chains.

Prologue

- In today's fast-paced world driven by data, the ability to extract valuable insights and make informed decisions is more crucial than ever.
- Optimization, the process of finding the best solution among a set of alternatives, lies at the heart of this endeavor.

Examples

- Predicting customer behavior.
- Optimizing supply chains.
- Designing machine learning models.

Prologue

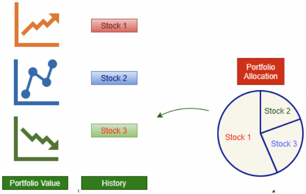
- In today's fast-paced world driven by data, the ability to extract valuable insights and make informed decisions is more crucial than ever.
- Optimization, the process of finding the best solution among a set of alternatives, lies at the heart of this endeavor.

Examples

- Predicting customer behavior.
- Optimizing supply chains.
- Designing machine learning models.
- Solving complex decision-making problems.

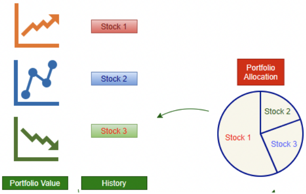
What is "Optimization"?

Portfolio optimization



What is “Optimization”?

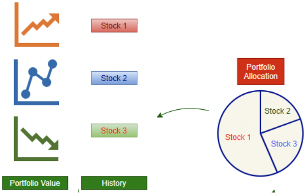
Portfolio optimization



- d financial assets (stocks, bonds, etc).

What is "Optimization"?

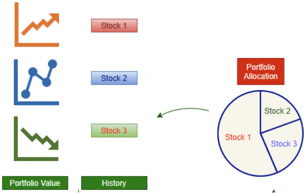
Portfolio optimization



- d financial assets (stocks, bonds, etc).
- Asset i that has return μ_i.

What is "Optimization"?

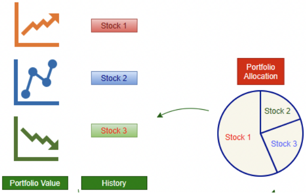
Portfolio optimization



- d financial assets (stocks, bonds, etc).
- Asset i that has return μ_i.
- σ_{ij} is the covariance of assets i and j.

What is "Optimization"?

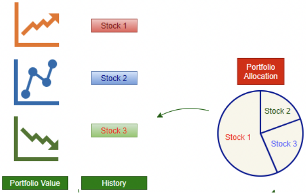
Portfolio optimization



- d financial assets (stocks, bonds, etc).
- Asset i that has return μ_i.
- σ_{ij} is the covariance of assets i and j.
- We allocate x_i fraction of our budget to asset i.

What is "Optimization"?

Portfolio optimization



- d financial assets (stocks, bonds, etc).
- Asset i that has return μ_i.
- σ_{ij} is the covariance of assets i and j.
- We allocate x_i fraction of our budget to asset i.

Goal: find a portfolio (allocation) maximizing return while minimizing risk (measured as a function of the covariance).

What is “Optimization”?

Facility location

▶ Solve

Solution summary

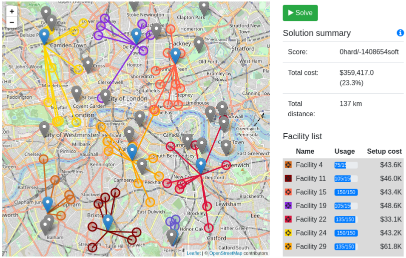
Score:	0hard/-1408654soft
Total cost:	$359,417.0 (23.3%)
Total distance:	137 km

Facility list

Name	Usage	Setup cost
Facility 4	75/15	$43.6K
Facility 11	105/15	$46.0K
Facility 15	150/150	$43.4K
Facility 19	105/15	$48.6K
Facility 22	135/150	$33.1K
Facility 24	150/150	$43.2K
Facility 29	135/150	$61.8K

What is “Optimization”?

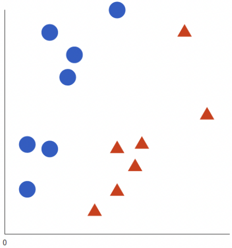
Facility location



Goal: build “fire stations” covering all households while minimizing the longest distance to a household.

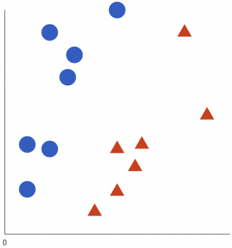
What is "Optimization"?

Support vector machine



What is “Optimization”?

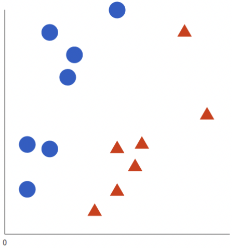
Support vector machine



- n data $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i \in \{-1, 1\}$ are labels.

What is "Optimization"?

Support vector machine



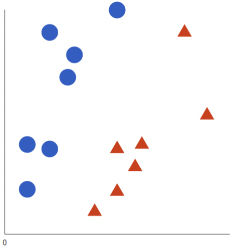
- n data $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i \in \{-1, 1\}$ are labels.
- We want to find a separating hyperplane

$$w^\top x = b$$

to classify data with $+1$ and data with -1.

What is "Optimization"?

Support vector machine



- n data $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i \in \{-1, 1\}$ are labels.
- We want to find a separating hyperplane

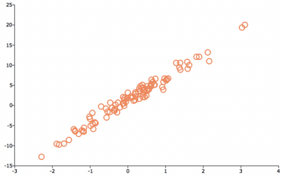
$$w^\top x = b$$

to classify data with $+1$ and data with -1.

Goal: find a separating hyperplane $w^\top x = b$ with the "gap" $(1/\|w\|)$ being maximized.

What is “Optimization”?

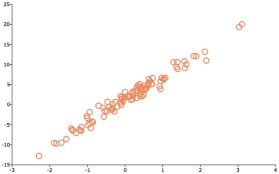
Linear regression



- n data points $(x_1, y_1), \ldots, (x_n, y_n)$.

What is "Optimization"?

Linear regression



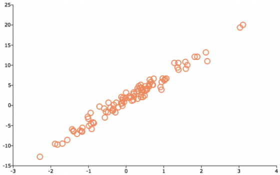
- n data points $(x_1, y_1), \ldots, (x_n, y_n)$.
- We want to find a linear rule

$$y = \beta^\top x$$

that best represents the relationship between x and y.

What is "Optimization"?

Linear regression



- n data points $(x_1, y_1), \ldots, (x_n, y_n)$.
- We want to find a linear rule

$$y = \beta^\top x$$

that best represents the relationship between x and y.

Goal: find β minimizing

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta^\top x_i)^2}_{\text{mean squared error}} + \underbrace{R(\beta)}_{\text{regularization}} \quad .$$

What is “Optimization”?

Neural networks

- The linear model is often too restrictive in practical applications.

What is “Optimization”?

Neural networks

- The linear model is often too restrictive in practical applications.
- In modern data science, neural networks are commonly used to solve a supervised learning task.

What is "Optimization"?

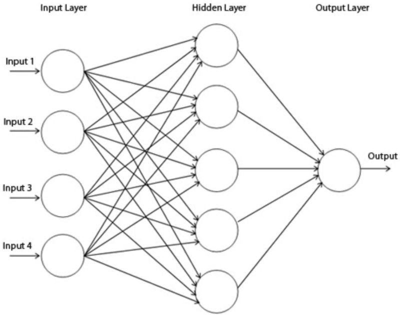
Neural networks

- The linear model is often too restrictive in practical applications.
- In modern data science, neural networks are commonly used to solve a supervised learning task.
- For simplicity, let us focus on a neural network with a single hidden layer.

What is "Optimization"?

Neural networks

- The linear model is often too restrictive in practical applications.
- In modern data science, neural networks are commonly used to solve a supervised learning task.
- For simplicity, let us focus on a neural network with a single hidden layer.



What is "Optimization"?

Neural networks

- Given the predictor variable vector $x \in \mathbb{R}^d$, our hypothesis is that the response variable $y \in \mathbb{R}$ satisfies

$$\mathbb{E}[y \mid x] = w_2^\top \sigma(W_1^\top x) \tag{1}$$

where

What is “Optimization”?

Neural networks

- Given the predictor variable vector $x \in \mathbb{R}^d$, our hypothesis is that the response variable $y \in \mathbb{R}$ satisfies

$$\mathbb{E}\left[y \mid x\right] = w_2^\top \sigma(W_1^\top x) \tag{1}$$

where

- $W_1^\top x$ is the output of the input layer,
- σ is an activation function,
- w_2 is the weight vector that the hidden layer applies.

What is "Optimization"?

Neural networks

- Given the predictor variable vector $x \in \mathbb{R}^d$, our hypothesis is that the response variable $y \in \mathbb{R}$ satisfies

$$\mathbb{E}[y \mid x] = w_2^\top \sigma(W_1^\top x) \tag{1}$$

  where
  - $W_1^\top x$ is the output of the input layer,
  - σ is an activation function,
  - w_2 is the weight vector that the hidden layer applies.

- As linear regression, we may consider the mean squared error,

$$\min_{W_1, w_2} \quad \frac{1}{n} \sum_{i=1}^{n} \left(y_i - w_2^\top \sigma(W_1^\top x_i) \right)^2 . \tag{2}$$

What is "Optimization"?

Neural networks

- Given the predictor variable vector $x \in \mathbb{R}^d$, our hypothesis is that the response variable $y \in \mathbb{R}$ satisfies

$$\mathbb{E}\left[y \mid x\right] = w_2^\top \sigma(W_1^\top x) \tag{1}$$

  where
  - $W_1^\top x$ is the output of the input layer,
  - σ is an activation function,
  - w_2 is the weight vector that the hidden layer applies.
- As linear regression, we may consider the mean squared error,

$$\min_{W_1, w_2} \quad \frac{1}{n} \sum_{i=1}^{n} \left(y_i - w_2^\top \sigma(W_1^\top x_i)\right)^2 . \tag{2}$$

- ReLU and the sigmoid function are common choices for σ.

Optimization basics

A mathematical optimization problem has the following canonical form.

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned}$$

where

Optimization basics

A mathematical optimization problem has the following canonical form.

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned}$$

where

- x is referred to as the decision vector, the vector of decision variables, or simply the decision variables,

Optimization basics

A mathematical optimization problem has the following canonical form.

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned}$$

where

- x is referred to as the decision vector, the vector of decision variables, or simply the decision variables,
- $f(x)$ is the objective function that we want to optimize,

Optimization basics

A mathematical optimization problem has the following canonical form.

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned}$$

where

- x is referred to as the decision vector, the vector of decision variables, or simply the decision variables,
- $f(x)$ is the objective function that we want to optimize,
- $\mathcal{X}$ is the domain from which we may take values of the decision variables,

Optimization basics

A mathematical optimization problem has the following canonical form.

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned}$$

where

- x is referred to as the decision vector, the vector of decision variables, or simply the decision variables,
- $f(x)$ is the objective function that we want to optimize,
- $\mathcal{X}$ is the domain from which we may take values of the decision variables,
- $\min_{x}$ indicates that the goal is to find and assigne values to the decision variables x minimizing the associated objective function value.

Optimization basics

A mathematical optimization problem has the following canonical form.

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & x \in \mathcal{X} \end{aligned}$$

where

- x is referred to as the decision vector, the vector of decision variables, or simply the decision variables,
- $f(x)$ is the objective function that we want to optimize,
- $\mathcal{X}$ is the domain from which we may take values of the decision variables,
- $\min_{x}$ indicates that the goal is to find and assigne values to the decision variables x minimizing the associated objective function value.

For $\max_{x}$ instead of $\min_{x}$, the goal is to maximize the objective function.

Who is Dabeen?

Office: E2-2 #2109

Email: dabeenl@kaist.ac.kr

Office hours: Tuesday 2:00 - 3:00 pm

Research interests:

- Optimization Theory (discrete, continuous, stochastic, online).
- Optimization for Machine Learning.
- Algorithm Design for Operations Research (resource allocation, scheduling, dynamic routing).

About this course

The following is a tentative list of topics covered in this course.

Theory

Convex Analysis (sets, functions, operations)

Optimality Conditions

Semidefinite Programming

Quadratic Programming

Algorithms

Gradient Descent (GD)

Proximal, Projected, Online, Stochastic GD

Frank-Wolfe

Proximal Point Algorithm and Augmented Lagrangian Method

Operator Splitting and ADMM

Newton's method and Quasi Newton methods

Applications

Machine Learning (SVM, LASSO, Ridge Regression, Policy Gradient, etc.)

Statistics (Uncertainty Quantification, Inverse Covariance Selection)

Operations Research (Advertisement Allocation, Facility Location, Portfolio Optimization)

- Many more applications will be discussed on the way.
- We might also cover other algorithms such as Mirror descent, and Interior Point Methods.

Logistics

Class times: Monday and Wednesday 2:30 - 3:45 pm.

Assessment:

- 5 assignments (50%)
- Course project (20%)
- Take-home final (30%)

No attendance check (but be responsible!)

Assignment

- Being comfortable with making mathematical arguments, writing proofs and programming is required throughout this course.
- **Typesetting in LaTeX is required for any submission.**
- Easiest option: Overleaf (https://www.overleaf.com)

```latex
\documentclass{article}
\usepackage{graphicx} % Required for inserting images

\title{IE331 Assignment 1}
\author{Dabeen Lee}
\date{\today}

\usepackage{amsmath,amssymb}

\begin{document}

\maketitle

\begin{enumerate}
    \item My answer to question 1 is ...

    $$\text{epi}(f) = \left\{(x,t)\in\mathbb{R}^d\times\mathbb{R}:\ \exists y\in\mathbb{R}^p\text{ s.t. } Ax + Dy + h t \leq r \right\}$$

    \item \begin{enumerate}
        \item My answer to question 2(a) is ...

        \begin{align*}
    \min\quad & f(x)\\
    \text{s.t.}\quad & g_i(x)\leq b_i,\quad i\in[m],\\
    &x\in\mathbb{R}^d.
\end{align*}
        \item My answer to question 2(b) is ...

            $$d(x)=\sum_{i\in[n]}\left\|x-v^i\right\|_\infty.$$
        \item My answer to question 2(c) is ...

            $$d(x)=\sum_{i\in[n]}\left\|x-v^i\right\|_1.$$
        \item My answer to question 2(d) is ...
    \end{enumerate}
\end{enumerate}

\end{document}

```

IE331 Assignment 1

Dabeen Lee

March 14, 2023

1. My answer to question 1 is ...

$$\text{epi}(f) = \left\{(x,t)\in\mathbb{R}^d\times\mathbb{R}:\ \exists y\in\mathbb{R}^p\text{ s.t. } Ax + Dy + h t \leq r \right\}$$

2. (a) My answer to question 2(a) is ...

$$\begin{aligned}\min\quad & f(x)\\ \text{s.t.}\quad & g_i(x)\leq b_i,\quad i\in[m],\\ &x\in\mathbb{R}^d.\end{aligned}$$

   (b) My answer to question 2(b) is ...

$$d(x)=\sum_{i\in[n]}\left\|x-v^i\right\|_\infty.$$

   (c) My answer to question 2(c) is ...

$$d(x)=\sum_{i\in[n]}\left\|x-v^i\right\|_1.$$

   (d) My answer to question 2(d) is ...

Project

There are two options.

1. Review of a research paper related to convex optimization.

   - Choose a paper published in a journal or announced at a coference.

2. Formulation and implementation of algorithms or methods for certain optimization problem.

   - Numerical implementation is required.

Submit (1) a proposal and (2) a final report.

Objectives

We formulate a decision-making problem as an optimization model

$$P \quad : \quad \min_{x \in D} f(x).$$

Objectives

We formulate a decision-making problem as an optimization model

$$P \;:\; \min_{x \in D} f(x).$$

Then

- We have to study the structure of the problem, f and D.

Objectives

We formulate a decision-making problem as an optimization model

$$P \ : \ \min_{x \in D} f(x).$$

Then

- We have to study the structure of the problem, f and D.
  - Is P convex? a linear program (LP)? a quadratic program (QP)? a semidefinite program (SDP)?

Objectives

We formulate a decision-making problem as an optimization model

$$P \; : \; \min_{x \in D} f(x).$$

Then

- We have to study the structure of the problem, f and D.
  - Is P convex? a linear program (LP)? a quadratic program (QP)? a semidefinite program (SDP)?
  - Is f smooth? strongly convex? both?

Objectives

We formulate a decision-making problem as an optimization model

$$P \quad : \quad \min_{x \in D} f(x).$$

Then

- We have to study the structure of the problem, f and D.
  - Is P convex? a linear program (LP)? a quadratic program (QP)? a semidefinite program (SDP)?
  - Is f smooth? strongly convex? both?
  - Is D convex? an affine subspace?

Objectives

We formulate a decision-making problem as an optimization model

$$P \ : \ \min_{x \in D} f(x).$$

Then

- We have to study the structure of the problem, f and D.
  - Is P convex? a linear program (LP)? a quadratic program (QP)? a semidefinite program (SDP)?
  - Is f smooth? strongly convex? both?
  - Is D convex? an affine subspace?
- We have to figure out and test candidate algorithms for solving P.

Objectives

We formulate a decision-making problem as an optimization model

$$P \ : \ \min_{x \in D} f(x).$$

Then

- We have to study the structure of the problem, f and D.
  - Is P convex? a linear program (LP)? a quadratic program (QP)? a semidefinite program (SDP)?
  - Is f smooth? strongly convex? both?
  - Is D convex? an affine subspace?
- We have to figure out and test candidate algorithms for solving P.
  - Gradient Descent, simply? Proximal Gradient Descent? Newton's method?

Objectives

We formulate a decision-making problem as an optimization model

$$P \; : \; \min_{x \in D} f(x).$$

Then

- We have to study the structure of the problem, f and D.
  - Is P convex? a linear program (LP)? a quadratic program (QP)? a semidefinite program (SDP)?
  - Is f smooth? strongly convex? both?
  - Is D convex? an affine subspace?
- We have to figure out and test candidate algorithms for solving P.
  - Gradient Descent, simply? Proximal Gradient Descent? Newton's method?

For this task, we need comprehensive knowledge in convex optimization.

Later, this knowledge will help you create a new optimization problem.

Example: dual methods

Consider

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + By = c \end{aligned}$$

where f, g are convex and A, B, c are matrices of appropriate dimension.

Example: dual methods

Consider

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + By = c \end{aligned}$$

where f, g are convex and A, B, c are matrices of appropriate dimension.

How do we solve the problem?

Example: dual methods

Consider

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + By = c \end{aligned}$$

where f, g are convex and A, B, c are matrices of appropriate dimension.

How do we solve the problem?

- If f and g are both strongly convex, then Gradient Ascent in the dual.

Example: dual methods

Consider

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + By = c \end{aligned}$$

where f, g are convex and A, B, c are matrices of appropriate dimension.

How do we solve the problem?

- If f and g are both strongly convex, then Gradient Ascent in the dual.
- If only f is strongly convex while g has an easy Prox, then Proximal Gradient in the dual.

Example: dual methods

Consider

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + By = c \end{aligned}$$

where f, g are convex and A, B, c are matrices of appropriate dimension.

How do we solve the problem?

- If f and g are both strongly convex, then Gradient Ascent in the dual.
- If only f is strongly convex while g has an easy Prox, then Proximal Gradient in the dual.
- If neither f nor g is strongly convex, then Proximal Point Algorithm in the dual or ADMM.