Introduction to optimization

15.093: Optimization

Dimitris Bertsimas
Alexandre Jacquillat

Sloan School of Management
Massachusetts Institute of Technology

Scope of optimization

From data to decisions

Descriptive analytics

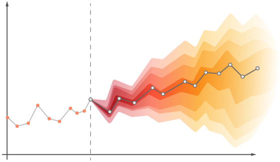
- Data management
- Data visualization

→ What happened?

Predictive analytics

- Machine learning
- Forecasting

→ What will happen?



Prescriptive analytics

- Optimization
- Simulation

→ What should I do?

→ Optimization deals with complex decision-making problems that involve many interdependent decisions and several interrelated constraints

Optimization problems

- Fermat, Newton, 1600s:

$$\min\ f(x) \quad \rightarrow \quad \frac{df(x)}{dx} = 0$$

- Euler, 1755:

$$\min\ f(x_1, \cdots, x_n) \quad \rightarrow \quad \nabla f(\boldsymbol{x}) = 0$$

- Lagrange, 1797:

$$\min\ f(x_1, \cdots, x_n) \quad \text{s.t. } g_i(x_1, \cdots, x_n) = 0,\ \forall i = 1, \cdots, m$$

$$\rightarrow \quad \mathcal{L}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i g_i(\boldsymbol{x})$$

Extended-form notation

$$\begin{aligned} \min \quad & f(x_1, \cdots, x_n) \\ \text{s.t.} \quad & g_1(x_1, \cdots, x_n) \leq 0 \\ & \vdots \\ & g_m(x_1, \cdots, x_n) \leq 0 \end{aligned}$$

Vector notation

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_1(\boldsymbol{x}) \leq 0 \\ & \vdots \\ & g_m(\boldsymbol{x}) \leq 0 \end{aligned}$$

Compact notation

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0,\ \forall i \end{aligned}$$

Optimization terminology

Formulation (General optimization problem)

$$\begin{aligned} \min \quad & f(x_1, \cdots, x_n) \\ \textit{s.t.} \quad & g_1(x_1, \cdots, x_n) \leq 0 \\ & \vdots \\ & g_m(x_1, \cdots, x_n) \leq 0 \end{aligned} \qquad \begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \textit{s.t.} \quad & g_1(\boldsymbol{x}) \leq 0 \\ & \vdots \\ & g_m(\boldsymbol{x}) \leq 0 \end{aligned} \qquad \begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \textit{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0, \ \forall i \end{aligned}$$

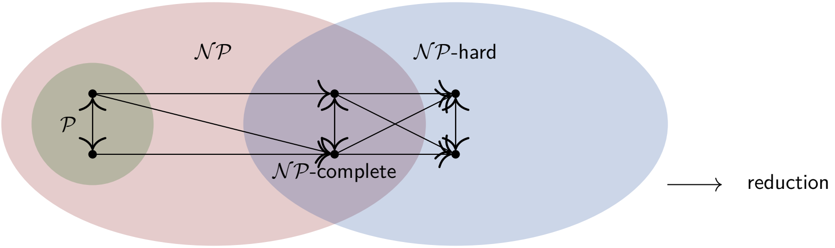
- **Objective function:** $f : \mathbb{R}^n \to \mathbb{R}$
- **Constraint functions:** $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \cdots, m$
- **Feasible region:**

$$\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n : g_i(x) \leq 0, \quad \forall i = 1, \cdots, m\}$$

- **Decision variable:** $x \in \mathcal{F}$
- **Optimal solution:**

$$\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x} \in \mathcal{F}} f(\boldsymbol{x})$$

Remark on complexity theory and tractability

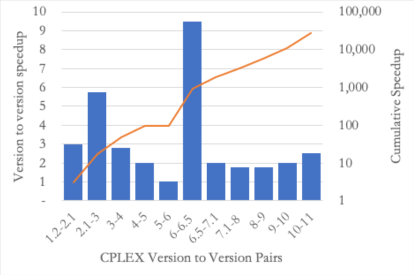


- Objective: developing practically tractable optimization methods
  - The equation `efficient` = `polynomially solvable` has been accepted as the theoretical translation of a "practical algorithm"
  - However, "inefficient" does not mean intractable: $\mathcal{NP}$-hard problems are routinely solved in practice
  - → Practically tractable methods to solve problems of the size encountered in practice, and in appropriate computational times for the application

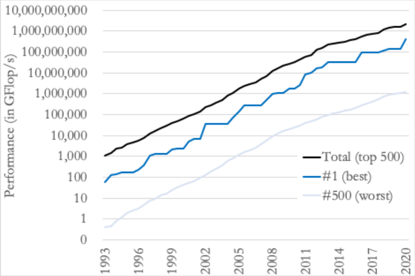
Improvements in optimization technologies

- Software: CPLEX 29,000 times faster from 1991 to 2007 and Gurobi 53 times faster from 2009 to 2019, a cumulative speedup of 1,500,000
- Hardware: computing speedup of another factor of 1,500,000

→ From problems with dozens of variables/constraints in hours to problems with millions of variables/constraints in seconds or minutes



Improvements in integer optimization

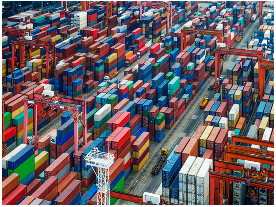


Improvements in computing power

Optimization across industries


Transportation


Energy


Supply chain


Health care


Finance

Sport

→ Modeling and algorithmic tools at the core of data science

Classes of optimization problems

Linear optimization (LO)

- Optimization with continuous variables, linear objective function and linear constraints
- Very efficient solution algorithms, which scale to problems with millions of variables and constraints

Formulation (Linear optimization)

$$\begin{aligned} \min \quad & c_1x_1 + \cdots + c_nx_n \\ \textit{s.t.} \quad & a_{11}x_1 + \cdots + a_{1n}x_n \geq b_1 \\ & a_{21}x_1 + \cdots + a_{2n}x_n \geq b_2 \\ & \cdots \\ & a_{m1}x_1 + \cdots + a_{mn}x_n \geq b_m \\ & x_1, \cdots, x_n \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & \sum_{j=1}^n c_jx_j \\ \textit{s.t.} \quad & \sum_{j=1}^n a_{1j}x_j \geq b_1 \\ & \cdots \\ & \sum_{j=1}^n a_{mj}x_j \geq b_m \\ & x_j \geq 0, \ \forall j \end{aligned}$$

$$\begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \textit{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

Non-linear optimization (NLO)

Formulation (Non-linear optimization)

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) && \textit{non-linear objective: } f:\mathbb{R}^n \to \mathbb{R} \\
\textit{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, \ \forall i = 1,\ldots,m && \textit{non-linear constraints: } g_i:\mathbb{R}^n \to \mathbb{R} \\
& \boldsymbol{x} \ge \boldsymbol{0}
\end{aligned}
$$

- Important sub-classes of problems
  - Convex optimization: $f, g_1, \cdots, g_m$ convex functions
  - Convex quadratic optimization

$$g_i(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q}_i \boldsymbol{x} + \boldsymbol{p}_i^\top \boldsymbol{x} + q_i, \qquad \boldsymbol{Q}_i \text{ positive semi-definite}$$

  - Non-convex quadratic optimization

$$g_i(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q}_i \boldsymbol{x} + \boldsymbol{p}_i^\top \boldsymbol{x} + q_i, \qquad \boldsymbol{Q}_i \text{ non positive semi-definite}$$

  - Second-order conic optimization

$$g_i(\boldsymbol{x}) = \|\boldsymbol{a}_i^\top \boldsymbol{x} + b_i\|_2 + \boldsymbol{p}_i^\top \boldsymbol{x} + q_i$$

Mixed-integer optimization (MIO)

- Linear optimization with discrete variables
  - Mixed-integer optimization: $\boldsymbol{x} \in \mathbb{Z}^p \times \mathbb{R}^{n-p}$
  - Integer optimization: $\boldsymbol{x} \in \mathbb{Z}^n$
  - Binary optimization: $\boldsymbol{x} \in \{0,1\}^n$
- Astonishing progress in mixed-integer optimization algorithms, scaling to problems with thousands to millions of variables and constraints

Formulation (Mixed-integer optimization)

$$
\begin{aligned}
\min \quad & c_1x_1 + \cdots + c_nx_n \\
\textit{s.t.} \quad & a_{11}x_1 + \cdots + a_{1n}x_n \geq b_1 \\
& a_{21}x_1 + \cdots + a_{2n}x_n \geq b_2 \\
& \ldots \\
& a_{m1}x_1 + \cdots + a_{mn}x_n \geq b_m \\
& x_1, \cdots, x_n \geq 0 \\
& x_1, \cdots, x_p \in \mathbb{Z}
\end{aligned}
$$

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\textit{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \\
& \boldsymbol{x} \in \mathbb{Z}_+^p \times \mathbb{R}_+^{n-p}
\end{aligned}
$$

Optimization under uncertainty

- Decisions oftentimes need to be made without perfect information about current or future operating conditions
  - Data uncertainty: data inaccuracy, missing data, measurement errors
  - Core uncertainty: decisions made before learning critical information
- Stochastic optimization (SO): description of uncertainty via probabilistic scenarios, and minimization of expected cost
- Robust optimization (RO): description of uncertainty via uncertainty sets, and minimization of worst-stage cost
- Tailored solution algorithms, scaling to large-scale problems

Formulation (Stochastic optimization)

$$\begin{aligned} \min \quad & \sum_{s \in \mathcal{S}} p_s \cdot \left(\boldsymbol{c}_s^\top \boldsymbol{x}_s \right) \\ \textit{s.t.} \quad & \boldsymbol{A}_s \boldsymbol{x}_s = \boldsymbol{b}_s, \ \forall s \in \mathcal{S} \\ & \boldsymbol{x}_s \geq \boldsymbol{0}, \ \forall s \in \mathcal{S} \end{aligned}$$

Formulation (Robust optimization)

$$\begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \textit{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \ \forall \boldsymbol{A} \in \mathcal{U} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

Course syllabus

Structure of the course

Lectures	Topic
1	Introduction and overview
2–6	Linear optimization
7–9	Duality theory
10–16	Non-linear optimization
17–19	Integer optimization
20–22	Optimization under uncertainty
23–24	Decomposition methods

- Canvas page: https://canvas.mit.edu/courses/20858
- Piazza page: http://piazza.com/mit/fall2023/15093
- Overall course objectives:
  - **Modeling:** translating practical problems into mathematical models
  - **Theory:** mathematical and geometrical structure for algorithmic insights
  - **Algorithms:** solver technologies, algorithm design, scalability
  - **Practice:** problem modeling, decision support tools, managerial insights
- Use of state-of-the-art solvers through Julia/JuMP (see here)

Deliverables

- Assignments (25%)
  - 7 short individual assignments, schedule on the Syllabus
  - Assignments will combine theory and computations
- Two midterm examinations
  - Tuesday, October 17 (25%)
  - Tuesday, November 21 (25%)
- Term project (25%)
  - Self-formed teams of two (three, with justification)
  - Decision-making problem, optimization methods, computational results, and practical recommendations
  - October 6: Team registration
  - November 10: One-page interim report (20% of grade)
  - December 08: Final report (40%)
  - December 08: Poster presentations (40%)
- Class attendance and engagement: an important tie-breaker

The week at a glance

Section	Day	Time	Location
Section A	Tuesday, Thursday	1:00 pm – 2:30 am	E51-345
Section B	Tuesday, Thursday	8:30 am – 10:00 am	E62-276
Recitation A	Friday	1:00 pm – 2:00 pm	1-190
Recitation B	Friday	2:00 pm – 3:00 pm	1-190
Office Hours (AJ)	Monday	11:00 am – 12:00 pm	E62-687
Office Hours (TK)	Monday	3:00 – 4:00 pm	E62-687
Office Hours (DB)	Tuesday	2:30 pm – 3:30 pm	E62-574
Office Hours (FC)	Tuesday	2:30 – 3:30 pm	E62-529
Office Hours (MH)	Wednesday	10:00 – 11:00 am	E62-575
Office Hours (AS)	Thursday	10:00 – 11:00 am	E62-529

- Textbooks: *Introduction to Linear Optimization*, Bertsimas & Tsitsiklis and *Convex Optimization*, Boyd & Vandenberghe
- Attendance required and participation encouraged, in class & recitation
- Generative AI is encouraged but end products should be yours
- See syllabus on class conduct, special needs, and academic integrity
- Take care of yourselves: get enough sleep, enough rest, avoid stress

Linear optimization

History of linear optimization

- Pre-algorithmic period
  - Chinese mathematicians (≈ 300 BC), Newton (1707): linear equalities
  - Fourier, 1826: system of linear inequalities
  - de la Vallée Poussin: optimization of absolute values
  - Kantorovich, Koopmans, 1930s: formulations and solution method
  - von Neumann, 1928: game theory, duality
  - Farkas, Minkowski, Carathéodory, 1870-1930: foundations
- Algorithmic period
  - George Dantzig, 1947: simplex method
  - 1950s–70s: applications, large scale optimization, optimization under uncertainty, complexity theory
  - Leonid Khachiyan, 1979: ellipsoid algorithm
  - Narendra Karmarkar, 1984: interior point algorithms
  - 2000s: robust optimization
  - 2010s: huge-scale optimization
- "If one would take statistics about which mathematical problem is using up most of the computer time in the world, then the answer would probably be linear optimization." – Laszlo Lovasz

The knapsack problem

- You are planning a trip and are trying to select what to pack
  - You have access to n items, indexed by $j = 1, \cdots, n$
  - Each item j carries a value c_j and occupies a volume a_j
  - You can carry up to a total volume of b
- Decision variables

$$x_j = \text{amount of item } j \text{ to carry}$$

- Optimization formulation: maximize total value, while complying with capacity and ensuring that all quantities are non-negative

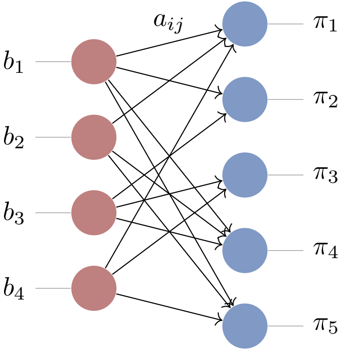
Formulation (Knapsack problem)

$$\begin{aligned} \max \ & \sum_{j=1}^{n} c_j x_j \\ \textit{s.t.} \ & \sum_{j=1}^{n} a_j x_j \leq b \\ & x_j \geq 0, \ \forall j = 1, \cdots, n \end{aligned}$$

Manufacturing: data

- A facility manufactures several products, each with specific characteristics and production requirements
- The facility is limited in terms of raw materials and labor hours
- Problem: how to plan production to maximize profits?

- Available data:
  - n products $j = 1, \cdots, n$
  - m raw materials $i = 1, \cdots, m$
  - π_j: profit of product j
  - b_i: available units of material i
  - a_{ij}: # units of material i needed to manufacture product j



Manufacturing: formulation

- Decision variable:

$$x_j : \text{ amount of product } j \text{ manufactured}$$

- Optimization formulation
  - Objective: maximizing total profit
  - Constraints: complying with availability of all raw materials
  - Production quantities are non-negative

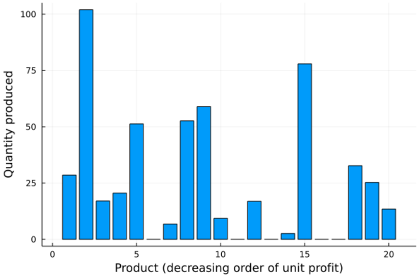
Formulation (Manufacturing problem)

$$\begin{aligned} \max \quad & \pi_1 x_1 + \cdots + \pi_n x_n \\ \textit{s.t.} \quad & a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\ & a_{21}x_1 + \cdots + a_{2n}x_n \leq b_2 \\ & \ldots \\ & a_{m1}x_1 + \cdots + a_{mn}x_n \leq b_m \\ & x_1, \cdots, x_n \geq 0 \end{aligned}$$

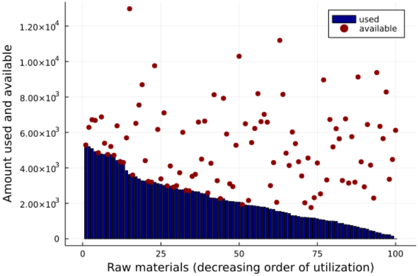
$$\begin{aligned} \max \quad & \sum_{j=1}^{n} \pi_j x_j \\ \textit{s.t.} \quad & \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \ \forall i = 1, \cdots, m \\ & x_j \geq 0, \ \forall j \end{aligned}$$

Manufacturing: solution

Product manufacturing decisions



Resource utilization decisions



- Which products are manufactured? Why don't we produce more?
- How much materials are utilized? Why don't we use more?
- What are the drivers of these decisions?

$\rightarrow$ Coupling constraints across products and across raw materials

$\rightarrow$ Many resource constraints are not binding

Manufacturing: edge of optimization

- Let us consider a greedy benchmark
  - Sort products by decreasing order of unit profit
  - Maximize quantity of first product until some resource gets depleted
  - Maximize quantity of second product until some resource gets depleted
  - $\cdots$ [repeat until last product]
- The optimal solution increases profitability by 12%

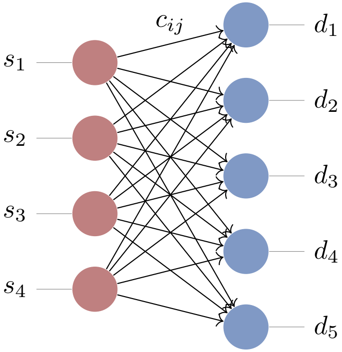
$\rightarrow$ Edge of optimization lies in thinking globally as opposed to locally

Method	Metrics	Prod. 1	Prod. 2	Prod. 10	Prod. 20	Total
Heuristic	Unit profit	$48	$37	$19.4	$11.5	—
	Quantity	56.5	104.3	0	0	423.4
	Total profit	$2,717	$3,891	0	0	$11,389
Optimization	Unit profit	$48	$37	$19.4	$11.5	—
	Quantity	28.5	102.0	9.3	13.5	515.9
	Total profit	$1,373	$3,803	$181	$154	$12,771

Transportation: data

- A logistics provider needs to transport products from its plants (where products are manufactured) to its warehouses (where they are stored)
- Each plant has limited product availability
- Each warehouse has specific demand requirements
- Problem: how to satisfy demand from all warehouses at minimal cost?

- Available data:
  - n warehouses $j = 1, \cdots, n$
  - m plants $i = 1, \cdots, m$
  - s_i capacity of plant i
  - d_j demand of warehouse j
  - c_{ij}: unit cost of transportation from plant i to warehouse j



Transportation: formulation

- Decision variable:

  x_{ij} : number of units to ship from plant i to warehouse j

- Optimization formulation
  - Objective: minimizing total transportation costs
  - Constraints: warehouse demand and plant capacity

Formulation (Transportation problem)

$$
\begin{aligned}
\min \quad & c_{11}x_{11} + \cdots + c_{mn}x_{mn} \\
\textit{s.t.} \quad & x_{11} + \cdots + x_{m1} \geq d_1 \\
& \cdots \\
& x_{1n} + \cdots + x_{mn} \geq d_n \\
& x_{11} + \cdots + x_{1n} \leq s_1 \\
& \cdots \\
& x_{m1} + \cdots + x_{mn} \leq s_m \\
& x_{11}, \cdots, x_{mn} \geq 0
\end{aligned}
$$

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \\
\textit{s.t.} \quad & \sum_{i=1}^{m} x_{ij} \geq d_j, \qquad \forall j = 1, \cdots, n \\
& \sum_{j=1}^{n} x_{ij} \leq s_i, \qquad \forall i = 1, \cdots, m \\
& x_{ij} \geq 0, \qquad \forall i, j
\end{aligned}
$$

Structure of linear optimization models

Linear optimization: standard form

Formulation (General form of linear optimization)

$$\begin{aligned}
\min / \max \quad & \boldsymbol{c}^\top \boldsymbol{x} & \\
\textit{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{x} \geq b_i & \forall i \in \mathcal{M}_1 \\
& \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i & \forall i \in \mathcal{M}_2 \\
& \boldsymbol{a}_i^T \boldsymbol{x} = b_i & \forall i \in \mathcal{M}_3 \\
& x_j \geq 0 & \forall j \in \mathcal{N}_1 \\
& x_j \leq 0 & \forall j \in \mathcal{N}_2 \\
& x_j \text{ free} & \forall j \in \mathcal{N}_3
\end{aligned}$$

Formulation (Standard form of linear optimization)

$$\begin{aligned}
\min \quad & \boldsymbol{c}^\top \boldsymbol{x} \\
\textit{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& \boldsymbol{x} \geq \boldsymbol{0}
\end{aligned}$$

Equivalent linear optimization formulations

Proposition

Every LO problem can be written in standard form.

- Maximization vs. minimization

$$\max \quad \boldsymbol{c}^\top \boldsymbol{x} \quad \longrightarrow \quad -\min \quad (-\boldsymbol{c})^\top \boldsymbol{x}$$

- Inequality vs. equality: introduction of *slack variables*

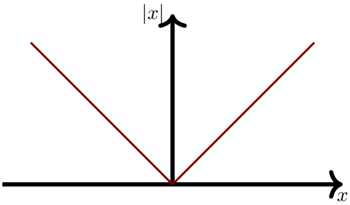
$$\begin{cases} \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i & \forall i \in \mathcal{M}_1 \\ \boldsymbol{a}_i^T \boldsymbol{x} \geq b_i & \forall i \in \mathcal{M}_2 \end{cases} \quad \longrightarrow \quad \begin{cases} \boldsymbol{a}_i^T \boldsymbol{x} + s_i = b_i, \ s_i \geq 0 & \forall i \in \mathcal{M}_1 \\ \boldsymbol{a}_i^T \boldsymbol{x} - s_i = b_i, \ s_i \geq 0 & \forall i \in \mathcal{M}_2 \end{cases}$$

- Free vs. positive variables: introduction of *positive part* & *negative part*

$$\boldsymbol{x} \text{ free} \quad \longrightarrow \quad \begin{cases} \boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^- \\ \boldsymbol{x}^+, \boldsymbol{x}^- \geq \boldsymbol{0} \end{cases}$$

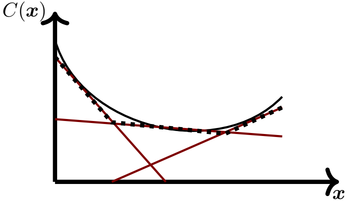
- In practice, a formulation is typically given in general form but optimization solvers start by converting it into standard form

On the power of linear optimization modeling



$$\min \quad \sum_{j=1}^{n} \underbrace{c_j}_{\geq 0} |x_j| \qquad \text{s.t.} \quad \boldsymbol{Ax} \geq \boldsymbol{b}$$

$$\begin{aligned} \min \quad & \sum_{j=1}^{n} c_j z_j \\ \text{s.t.} \quad & \boldsymbol{Ax} \geq \boldsymbol{b} \\ & z_j \geq x_j, \ \forall j \\ & z_j \geq -x_j, \ \forall j \end{aligned}$$



$$\min \quad \sum_{j=1}^{n} C(x_j) \qquad \text{s.t.} \quad \boldsymbol{Ax} \geq \boldsymbol{b}$$

$$\begin{aligned} \min \quad & \sum_{j=1}^{n} z_j \\ \text{s.t.} \quad & \boldsymbol{Ax} \geq \boldsymbol{b} \\ & z_j \geq \boldsymbol{d}_{kj}^{\top} \boldsymbol{x}_j + g_{kj}, \ \forall j, k \end{aligned}$$

→ Linear optimization enables to model piece-wise linear convex function and to approximate convex functions

Linear optimization formulations

- Structure your data consistently
  - Define categorical sets (or indices) at the outset
  - All numerical parameters need to be defined as a function of the sets
- Define your decision variables clearly and consistently
  - Again, all decision variables need to be defined as a function of the sets
  - Example: decision variable is a vector in the manufacturing problem, and a matrix in the transportation problem
- Write down the (linear) objective function and the (linear) constraints
  - Physical requirements, managerial requirements, system dynamics, etc.
  - Interdependencies across decision variables, requiring "global" optimization as opposed to "local" decision-making
  - A single objective function can capture multi-objective structures to trade off competing goals
- Importance of defining the scope of the problem: comprehensive model with solution in hours vs. restricted model with solution in seconds?
- No systematic method available—an art and a science in optimization model to capture the problem at hand

Conclusion

Summary

Takeaway

Optimization broadly used across industries to turn data into decisions.

Takeaway

Modeling components: inputs, decision variables, objective, constraints.

Takeaway

A science and an art to optimization modeling in practice.

Takeaway

Optimization classes: linear optimization, mixed-integer optimization, optimization under uncertainty, and non-linear optimization.

Takeaway

Linear optimization: critical technique to model large-scale and important problems faced in practice.