Optimization: Introduction

Instructor: Jin Zhang

Department of Mathematics
Southern University of Science and Technology
Spring 2024

What is mathematical optimization?

- Optimization models the goal of solving a problem in the "optimal way."
- Examples:
  - Running a business: to maximize profit, minimize loss, maximize efficiency, or minimize risk.
  - Design: minimize the weight of a bridge/truss, and maximize the strength, within the design constraints
  - Planning: select a flight route to minimize time or fuel consumption of an airplane
- Formal definition: to minimize (or maximize) a real function by deciding the values of free variables from within an allowed set.

- Optimization is an essential tool in life, business, and engineer.
- Examples:
  - Walmart pricing and logistics
  - Airplane engineering such as shape design and material selection
  - Packing millions of transistors in a computer chip in a functional way

  achieving these requires analyzing many related variables and possibilities, taking advantages of tiny opportunities.
- We will
  - cover some of the sophisticated mathematics for optimization
  - be closer to the reality than most other math courses

Status of optimization

- Last few decades: astonishing improvements in computer hardware and software, which motivated great leap in optimization modeling, algorithm designs, and implementations.
- Solving certain optimization problems has become standard techniques and everyday practice in business, science, and engineering. It is now possible to solve certain optimization problems with thousands, millions, and even thousands of millions of variables.
- Optimization has become part of undergrad curriculum. Optimization (along with statistics) has been the foundation of machine learning and big-data analytics. Matlab has two optimization toolboxes......

Ingredients of successful optimization

- modeling: turn a problem into one of the typical optimization formulations
- algorithms: an (iterative) procedure that leads you toward a solution (most optimization problems do not have a closed-form solution)
- software and, for some problems, hardware implementation: realize the algorithms and return numerical solutions

First examples

- Find two nonnegative numbers whose sum is up to 6 so that their product is a maximum.
- Find the largest area of a rectangular region provided that its perimeter is no greater than 100.
- Given a sequence of n numbers that are not all negative, find two indices so that the sum of those numbers between the two (including them) is a maximum.

Optimization formulation

$$
\begin{array}{rl}
\text{minimize} & f(x) \\
\text{subject to} & x \in \Omega
\end{array}
$$

- "minimize" is often abbreviated as "min"
- decision variable is typically stated under "minimize", unless obvious
- "subject to" is often shortened to "s.t."
- in linear and nonlinear optimization, feasible set Ω is represented by

$$
\begin{aligned}
h_i(x) = b_i,\ & i \in \mathcal{E} \quad \text{(equality constraints)} \\
g_j(x) \le b'_j,\ & j \in \mathcal{I} \quad \text{(inequality constraints).}
\end{aligned}
$$

Quadratic program–The Linear Least Squares Problem

- **Polynomial Fitting.** Suppose the observed data is $y_i \in \mathbb{R}$ for each time point t_i, $i = 1, 2, \cdots, N$, respectively. The underlying assumption it is that there is some function of time $f : \mathbb{R} \to \mathbb{R}$ such that $y_i = f(t_i),\ i = 1, 2, \cdots, N$.

  How to provide and estimate of the function f? $\Rightarrow$ One way is by a polynomial of a fixed degree, say n:

  $$p(t) = x_0 + x_1 t + x_2 t^2 + \cdots + x_n t^n.$$

  How to determine the coefficients that "best" fit the data?

  - If were possible to exactly fit the data, then there would exist a value for the coefficient.
  - If $n \ll N$, it cannot expect to fit the data perfectly and so there will be errors. It is wish to minimize the sum of the squares of the errors in the fit:

  $$\min_x \ \frac{1}{2} \sum_{i=1}^{N} (x_0 + x_1 t_i + x_2 t_i^2 + \cdots + x_n t_i^n - y_i)^2.$$

Quadratic program–The Linear Least Squares Problem

This minimization problem has the form

$$\min_{x} \ \frac{1}{2}\|Vx - y\|_2^2,$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^n \\ 1 & t_2 & t_2^2 & \cdots & t_2^n \\ \vdots & & & & \\ 1 & t_N & t_N^2 & \cdots & t_N^n \end{bmatrix}.$$

Quadratic program–The Linear Least Squares Problem

- **Linear Regression and Maximum Likelihood.** Considering a new drug therapy for reducing inflammation in a targeted population, and we have a relatively precise way of measuring inflammation for each member of this population. Trying to determine the dosing to achieve a target level of inflammation.

  Sampling a collection of N individuals from the target population, registar their dose z_{i0} and the values of their individual specific covariates $z_{i1}, z_{i2}, \cdots, z_{in}, i = 1, 2, \cdots, N$. After dosing we observe that the resultant inflammation for the ith subject to be $y_i, i = 1, 2, \cdots, N$. By saying that the "resultant level of inflammation is on average a linear function of the dose and other individual specific covariates", we mean that there exist coefficients $x_0, x_1, x_2, \cdots, x_n$ such that

  $$y_i = x_0 z_{i0} + x_1 z_{i1} + x_2 z_{i2} + \cdots + x_n z_{in} + v_i,$$

  where v_i is an instance of a random variable representing the individuals deviation from the linear model.

Quadratic program–The Linear Least Squares Problem

Assume that the random variables v_i are independently identically distributed $N(0, \sigma^2)$. The probability density function for the the normal distribution $N(0, \sigma^2)$ is

$$\frac{1}{\sigma\sqrt{2\pi}}\mathrm{EXP}[-v^2/(2\sigma^2)].$$

Given values for the coefficients x_i, the likelihood function for the sample $y_i, i = 1, 2, \cdots, N$ is the joint probability density function evaluated at this observation, which is given by

$$L(x, y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \mathrm{EXP}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_0 z_{i0} + \cdots + x_n z_{in} - y_i)^2\right].$$

To choose those values of the coefficients $x_0, x_1, \cdots, x_n$ that make the observation $y_1, y_2, \cdots, y_n$ most probable. One may maximize the likelihood function $L(x; y)$:

$$\max_{x \in \mathbb{R}^{n+1}} L(x; y).$$

Quadratic program–The Linear Least Squares Problem

It is equivalent to the problem

$$\min_{x\in\mathbb{R}^{n+1}} n\ln(\sigma\sqrt{2}\pi) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_0 z_{i0} + \cdots + x_n z_{in} - y_i)^2.$$

It is equivalent to the linear least squares problem

$$\min_{x\in\mathbb{R}^{n+1}} \frac{1}{2}\|Ax - y\|^2,$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, x = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ and } H = \begin{bmatrix} z_{10} & z_{11} & z_{12} & \cdots & z_{1n} \\ z_{20} & z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & & & & \\ z_{N0} & z_{N1} & z_{N2} & \cdots & z_{Nn} \end{bmatrix}.$$

A linearly constrained quadratic program (QP)

From Griva-Nash-Sofer § 1.2:

$$\begin{aligned} \text{min} \qquad & f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 \\ \text{subject to} \qquad & x_1 + x_2 = 3 \end{aligned}$$

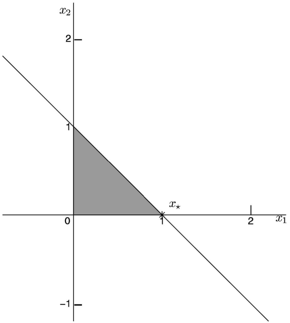
Elements: decision variables, parameters, constraint, feasible set, objective.

Linear program (LP)

From Griva-Nash-Sofer § 1.2:

$$\begin{array}{rl} \min & f(x) = -(x_1 + x_2) \\ \text{subject to} & x_1 + x_2 \le 1 \\ & 0 \le x_1,\ 0 \le x_2 \end{array}$$

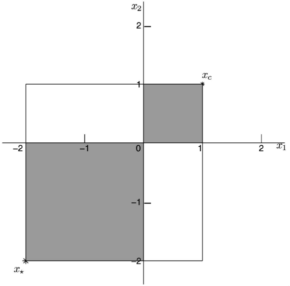
Elements: decision variables, constraints, feasible set, objective.



Nonlinear linear program (NLP)

$$\begin{aligned} \min \quad & f(x) = -(x_1 + x_2)^2 \\ \text{subject to} \quad & x_1 x_2 \geq 0 \\ & -2 \leq x_1 \leq 1 \\ & -2 \leq x_2 \leq 1 \end{aligned}$$

Elements: decision variables, feasible set, objective, (box) constraints, global minimizer vs local minimizer.



Global vs local solution

- "Solution" means "optimal solution"
- **Global solution** x^***:** $f(x^*) \le f(x)$ for all $x \in \Omega$
- **Local solution** x^***:** $\exists \delta > 0$ such that $f(x^*) \le f(x)$ for all $x \in \Omega$ and $\|x - x^*\| \le \delta$.
- a (global or local) solution x^* is unique if "$\le$" holds strictly as "$<$"
- In general, it is difficult to tell if a local solution is global because algorithms can only check "nearby points" and have not clue of behaviors "farther away." Hence, a "solution" may refer to a local solution.
- A local solution to a convex program is globally optimal. A LP is convex.
- A "stationary point" (where the derivative is zero) is also known as a solution, but it can be a maximization, minimization, or saddle point.

This course

- Most of this course focuses on finding local solutions and, for convex programs, global solutions. This is seemingly odd but there are good reasons:
- Asking for global solutions is computationally intractable, in general.
- Most global optimization algorithms (often takes long time to run) seeks the global solution by finding local solutions
- Many useful problems are convex, that is, a local solution is global
- In some applications, a local solution is an improvement from an existing point. Local solutions are OK.