Fundamentals of Computational Mathematics

Optimization

Introduction

Prof. Manuel F. Aprile

- PhD in Discrete Optimization
- Post-doc in Bruxelles
- Researcher in Padova

EPFL, Lausanne

- PhD in Discrete Optimization
- Post-doc in Bruxelles
- Researcher in Padova

Sources

- Optimization methods in Finance, Cornuejols & Tütüncü (1st and 2nd edition)
- Numerical Optimization, Nocedal & Wright
- Convex Optimization, Boyd & Vandenberghe
- Specific sources indicated at the beginning of the slides
- Various material shared via Moodle

What is optimization?

> "True optimization is the revolutionary contribution of modern research to decision processes."
>
> George Dantzig

What is optimization?

Optimization problem: Minimizing (or maximizing) a given function over a given domain.

$$\min f(x) : x \in S$$

- $x \in \mathbb{R}^n$: vector of decision variables
- f: objective function (cost, utility...)
- S: feasible region (choice space) = constraints on x (budget constraints, integrality...)

Note: solving $\max f(x)$ is equivalent to solving $\min -f(x)$, hence we will only deal with minimization problems.

Unconstrained optimization

No constraints on x: $S = \mathbb{R}^n$.

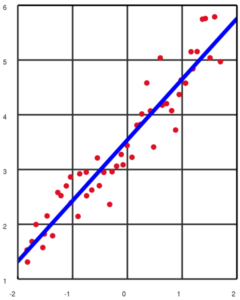
Example:

$$\min x_1 x_2 - x_1^2 + x_2, \; x \in \mathbb{R}^2$$

Very useful as foundation for more sophisticated methods.

Example: least squares

Fitting a data set with a line: given samples (x_i, y_i), $i = 1, \dots, N$ find linear function β that minimizes: $\sum_i (\beta x_i - y_i)^2$



Wikipedia

Usually reformulated as minimizing the squared norm of a vector: $\min_x ||Ax - b||_2^2$.

Example: least squares

Fitting a data set with a line: given samples (x_i, y_i), $i = 1, \dots, N$ find linear function β that minimizes: $\sum_i (\beta x_i - y_i)^2$

Usually reformulated as minimizing the squared norm of a vector: $\min_x ||Ax - b||_2^2$.

For this there is an analytical solution: $x = (A^T A)^{-1} A^T b$.

But we will mostly deal with problems where an analytical solution is hard / too expensive to obtain, and we will see algorithms to obtain approximate solutions.

Constrained optimization

Constraints can be:

- Linear: $a^\top x \leq b$, $a^\top x = b$
- Convex: $x \in S$ where S is a convex set: e.g. $x_1^2 + x_2^2 \leq 1$.
- Non-convex: S is a general set, for instance $S = \mathbb{Z}^n$ (integrality constraint)

Example: portfolio optimization

- Variables: amounts invested in different assets (e.g. stocks)
- Constraints: budget, max / min investment per asset, minimum return, ...
- Objective: minimize risk, or maximize expected return, ...

Assume we are given, for each asset i, its expected return μ_i.

Example: portfolio optimization

$$\max \sum_{i=1}^{n} \mu_i x_i$$
$$\mathbb{1}^\top x = 1$$
$$x \geq \mathbb{0}$$

For $i = 1, \dots, n$:

- x_i = percentage of budget to be invested in stock i.
- $\mathbb{1}$ is the all-ones vector of appropriate size.
- $\mu_i = \mathbb{E}(R_i)$ expected value of random variable R_i for the return of stock i

Example: portfolio optimization

$$\max \sum_{i=1}^{n} \mu_i x_i$$

$$\mathbb{1}^\top x = 1$$

$$x \geq \mathbb{0}$$

For $i = 1, \ldots, n$:

- x_i = percentage of budget to be invested in stock i.
- $\mathbb{1}$ is the all-ones vector of appropriate size.
- $\mu_i = \mathbb{E}(R_i)$ expected value of random variable R_i for the return of stock i

Problem: the model does not take into account diversification and risk, and the optimal solution just allocates all the budget to the stock with highest expected return μ_i.

Example: portfolio optimization

Markowitz' mean-variance model (Nobel Prize in Economics in 1990)

$$
\begin{aligned}
\min\ & x^\top Q\, x \\
& \mathbb{1}^\top x = 1 \\
& \mu^\top x \geq R \\
& x \geq \mathbb{0}
\end{aligned}
$$

Example: portfolio optimization

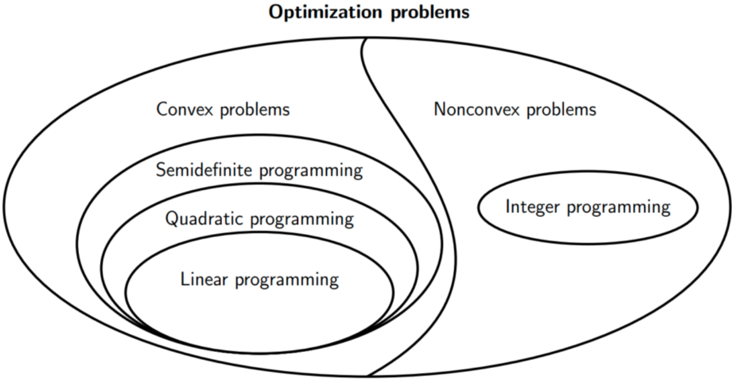
Markowitz' mean-variance model (Nobel Prize in Economics in 1990)

$$\begin{aligned} \min\ & x^\top Q\, x \\ & \mathbb{1}^\top x = 1 \\ & \mu^\top x \geq R \\ & x \geq \mathbb{0} \end{aligned}$$

For $i = 1, \dots, n$:

- x_i = percentage of budget to be invested in stock i.
- μ_i= expected return of stock i
- Q_{ij}= correlation between the variance of returns of stocks i and j

Goal: minimize the variance of the investment, while ensuring a high expected return.



http://alphonsusadubredu.com/optimization/

Overview of the course

- Unconstrained optimization
- Linear programming
- (General) constrained optimization
- Quadratic programming
- Integer programming
- ...

Two main challenges in optimization:

- Modeling our problem as an optimization problem
  - choosing the right decision variables
  - choosing the right constraints
  - choosing the right objective function
- Solving our problem:
  - exact techniques
  - iterative approaches
  - heuristics

Two main challenges in optimization:

- Modeling our problem as an optimization problem
  - choosing the right decision variables
  - choosing the right constraints
  - choosing the right objective function
- Solving our problem:
  - exact techniques
  - iterative approaches
  - heuristics

No free lunch theorem (Wolpert and Macready)

No algorithm performs better than any other when averaged over all possible inputs.

There is no single best optimization algorithm: problem-specific knowledge is essential for choosing the right algorithm.

Local vs. Global optima

Given the optimization problem $\min f(x) : x \in S$, a point $x^* \in S$ is a **global optimum** (minimizer) for the problem if:

$$f(x^*) \leq f(x) \, \forall \, x \in S.$$

Local vs. Global optima

Given the optimization problem $\min f(x) : x \in S$, a point $x^* \in S$ is a **global optimum** (minimizer) for the problem if:

$$f(x^*) \leq f(x) \, \forall \, x \in S.$$

A point $x^* \in S$ is a **local optimum** (minimizer) for the problem if, for some $\varepsilon > 0$:

$$f(x^*) \leq f(x) \, \forall \, x \in S \cap B_\varepsilon(x^*)$$

where $B_\varepsilon(x^*)$ is the open ball with radius ε centered at x^*.

$$B_\varepsilon(x^*) = \{x \in \mathbb{R}^n : ||x - x^*|| < \varepsilon\}$$

Local vs. Global optima

Given the optimization problem $\min f(x) : x \in S$, a point $x^* \in S$ is a **global optimum** (minimizer) for the problem if:

$$f(x^*) \leq f(x) \, \forall \, x \in S.$$

A point $x^* \in S$ is a **local optimum** (minimizer) for the problem if, for some $\varepsilon > 0$:

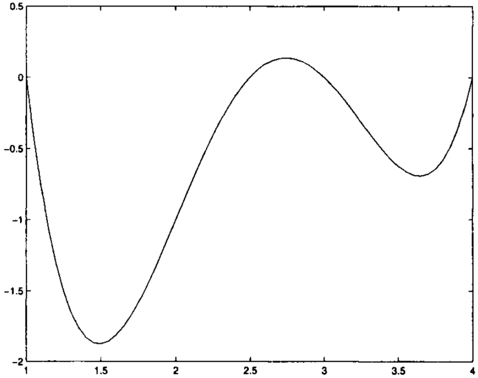
$$f(x^*) \leq f(x) \, \forall \, x \in S \cap B_\varepsilon(x^*)$$

where $B_\varepsilon(x^*)$ is the open ball with radius ε centered at x^*.

$$B_\varepsilon(x^*) = \{x \in \mathbb{R}^n : ||x - x^*|| < \varepsilon\}$$

Global optimum $\implies$ Local optimum

Local vs. Global optima

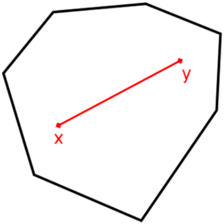


Convexity

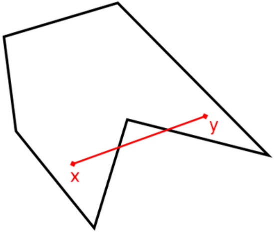
A set S is **convex** if, for any $x, y \in S$ and $\lambda \in [0, 1]$, we have

$$\lambda x + (1 - \lambda)y \in S.$$

I.e.: the line segment between any two points of S is contained in S.



Convex region



Non-convex region

Convexity

A set S is **convex** if, for any $x, y \in S$ and $\lambda \in [0,1]$, we have

$$\lambda x + (1-\lambda)y \in S.$$

I.e.: the line segment between any two points of S is contained in S.

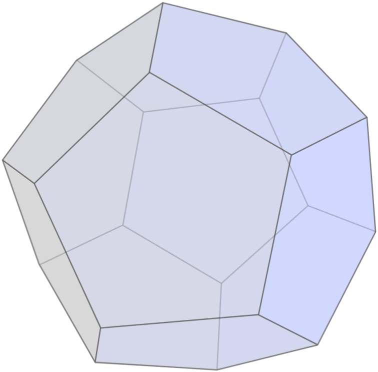
Exercises:

1. Halfspaces ($\{x \in \mathbb{R}^n : a^\top x \leq b\}$) and hyperplanes $\{x \in \mathbb{R}^n : a^\top x = b\}$ are convex.
2. The intersection of two convex sets is convex.

Polyhedra

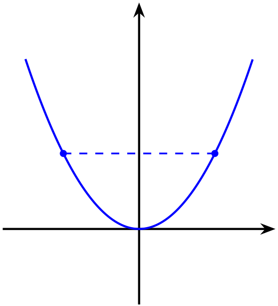
A **polyhedron** is a set $P = \{x \in \mathbb{R}^n : Ax \leq b\}$, i.e. an intersection of halfspaces.

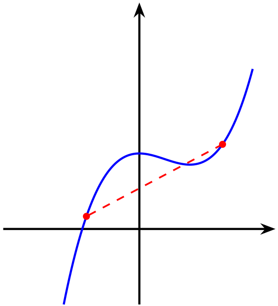
Polyhedra are convex.



Let S be convex. A function $f : S \to \mathbb{R}$ is **convex** if for any $x, y \in S$ and $\lambda \in [0, 1]$, we have

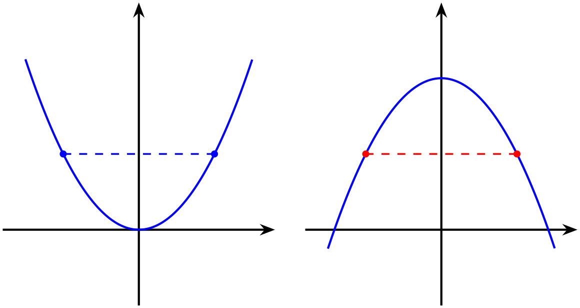
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$





Let S be convex. A function $f : S \to \mathbb{R}$ is **convex** if for any $x, y \in S$ and $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$



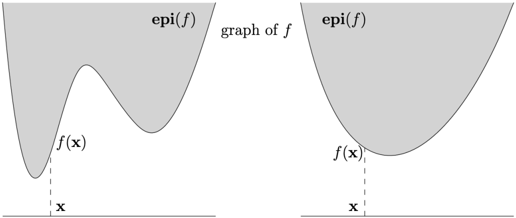
Let S be convex. A function $f : S \to \mathbb{R}$ is **concave** if $-f$ is convex. Maximizing a concave function is the same as minimizing a convex function.

The **epigraph** of f is the set

$$\mathrm{epi}(f) = \{(x, r) \in S \times \mathbb{R} : f(x) \leq r\}$$

Exercise

f is a convex function if and only if $\mathrm{epi}(f)$ is a convex set.

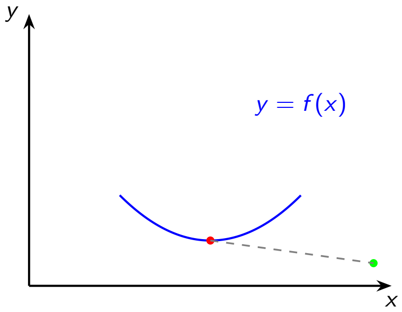


Convex optimization

Problem $\min f(x) : x \in S$ is a **convex optimization problem** if both S and $f : S \to \mathbb{R}$ are convex.

Theorem

In a convex optimization problem, all local optima are also global optima.



Convex optimization

Problem $\min f(x) : x \in S$ is a **convex optimization problem** if both S and $f : S \to \mathbb{R}$ are convex.

Theorem

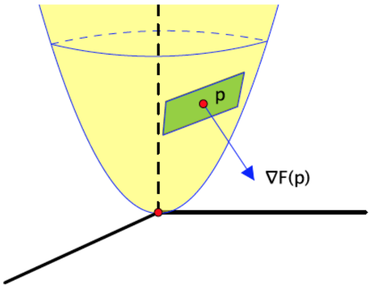
In a convex optimization problem, all local optima are also global optima.

Convex optimization problems can be solved efficiently (i.e. in polynomial time), while problems that are not convex are usually harder.

Gradient and stationary points

Gradient of f: $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n}\right)$.

First order approximation of f at x_0: $f(x) \approx f(x_0) + \nabla f(x_0)(x - x_0)$.



$\nabla f(x_0)$ is the normal vector of the hyperplane tangent to f in x_0, which has equation $f(x_0) + \nabla f(x_0)(x - x_0)$.

Gradient and stationary points

Gradient of f: $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n}\right)$.

First order approximation of f at x_0: $f(x) \approx f(x_0) + \nabla f(x_0)(x - x_0)$.

A point x_0 is critical or stationary for f if $\nabla f(x_0) = (0, \dots, 0)$.

Lemma

If $f : S \to \mathbb{R}$ is smooth (= all derivatives exist and are continuous), S is open, and x_0 is a local maximum or minimum for f, then x_0 is a stationary point of f.

Taylor approximation (2nd order)

Single variable: $x, x_0 \in \mathbb{R}$

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$$

Hessian of f:

$$[\nabla^2 f(x)]_{i,j} = \frac{\partial f(x)}{\partial x_i \partial x_j}$$

General: $x, x_0 \in \mathbb{R}^n$

$$f(x) \approx f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^\top \nabla^2 f(x_0)(x - x_0)$$

Idea: this polynomial is a good approximation for $f(x)$ when x is close to x_0.

Excellent youtube video by **3Blue1Brown**.

Taylor approximation (2^{nd} order)

Single variable: $x, x_0 \in \mathbb{R}$

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$$

Hessian of f:

$$[\nabla^2 f(x)]_{i,j} = \frac{\partial f(x)}{\partial x_i \partial x_j}$$

General: $x, x_0 \in \mathbb{R}^n$

$$f(x) \approx f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^\top \nabla^2 f(x_0)(x - x_0)$$

Alternative but equivalent form (x close to 0):

$$f(x + x_0) \approx f(x_0) + \nabla f(x_0)x + \frac{1}{2}x^\top \nabla^2 f(x_0)x$$

PSD matrices

A square, symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semidefinite (PSD)** if $x^\top A x \geq 0$ for all $x \in \mathbb{R}^n$.

Proposition

Given a square, symmetric matrix $A \in \mathbb{R}^{n \times n}$, the following are equivalent:

1. A is PSD;
2. all eigenvalues of A are non-negative;
3. $A = B^\top B$ where $B \in \mathbb{R}^{n \times n}$ is invertible.

PSD matrices

A square, symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semidefinite (PSD)** if $x^\top A x \geq 0$ for all $x \in \mathbb{R}^n$.

Proposition

Given a square, symmetric matrix $A \in \mathbb{R}^{n \times n}$, the following are equivalent:

1. A is PSD;
2. all eigenvalues of A are non-negative;
3. $A = B^\top B$ where $B \in \mathbb{R}^{n \times n}$ is invertible.

A square, symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive definite (PD)** if $x^\top A x > 0$ for all $x \in \mathbb{R}^n$, $x \neq 0$.

A is PD $\iff$ all its eigenvalues are positive.

The **covariance matrix** is a symmetric matrix that summarizes the covariances between random variables $X_1, \dots, X_n$.

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

- Symmetric: $\Sigma = \Sigma^T$
- Diagonal elements are variances: $\sigma_{ii} = \mathrm{Var}(X_i)$
- Off-diagonal elements are covariances: $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$

Proposition

The covariance matrix is positive semidefinite.

Proof: Can write $\Sigma = \mathbb{E}\left[(x-\mu)(x-\mu)^T\right]$ where μ is the vector of expected values of $x = (X_1, \ldots, X_n)$.

Then for $y \in \mathbb{R}^n$,

$$
\begin{aligned}
y^T \Sigma y &= y^T \mathbb{E}\left[(x-\mu)(x-\mu)^T\right] y \\
&= \mathbb{E}\left[(y^T(x-\mu))^2\right] \geq 0
\end{aligned}
$$

$\blacksquare$

Conditions for convexity

If $S \subseteq \mathbb{R}$ and $f : S \to \mathbb{R}$ is twice-continuously differentiable, then f is convex if and only if $f''(x) \geq 0 \; \forall \, x \in S$.

If $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$ is twice-continuously differentiable, then f is convex if and only if its Hessian $H = \nabla^2 f(x)$ is PSD for every $x \in S$.

Conditions for convexity

If $S \subseteq \mathbb{R}$ and $f : S \to \mathbb{R}$ is twice-continuously differentiable, then f is convex if and only if $f''(x) \geq 0 \; \forall \, x \in S$.

If $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$ is twice-continuously differentiable, then f is convex if and only if its Hessian $H = \nabla^2 f(x)$ is PSD for every $x \in S$.

Corollary

The quadratic function $x^T Q x$ is convex if and only if Q is PSD.

This implies that the mean-variance portfolio optimization problems that we will study are convex problems.

Exercise

Compute the gradient and Hessian of the Rosenbrock function

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Show that $x^* = (1, 1)$ is the only local optimum of this function, and that the Hessian matrix at that point is positive definite.

Exercise

Let a be a given n-vector, and A be a given $n \times n$ symmetric matrix. Compute the gradient and Hessian of $f_1(x) = a^T x$ and $f_2(x) = x^T A x$.

Exercise

Show that, if f, g are convex functions on the same domain, $f + g$ is a convex function.