Convex optimization

What is it
and
what is it good for...

Significance of convex optimization

- The set of all feasible solutions is always convex.
- Every local minimum is also a global minimum.
- Efficient algorithms based on interior point methods - already implemented in various solvers or software packages.
- Difficulty: identifying and formulating the optimization problem as a convex optimization problem $\longrightarrow$ CONVEX ANALYSIS

Topic 1: What is it and what is it good for...

- Surprisingly many problems can be formulated as convex optimization problems: data fitting, portfolio optimization, optimal control, statistical estimation, combinatorial optimization, topology / experimental design, geometric problems,.....
- Nonconvex problems - many algorithms are based on solving a sequence of convex optimization problems
- Convex relaxation of nonconvex problems: bounds on optimal solution

"In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity"

T. Rockafellar

Topic 1: What is it and what is it good for...

Main themes:

- Convex analysis
  - Convex sets
  - Convex functions
- Convex optimization problems in general
- Duality
- Optimality conditions
- Applications of convex optimization and CVX modeling system
- Interior point methods

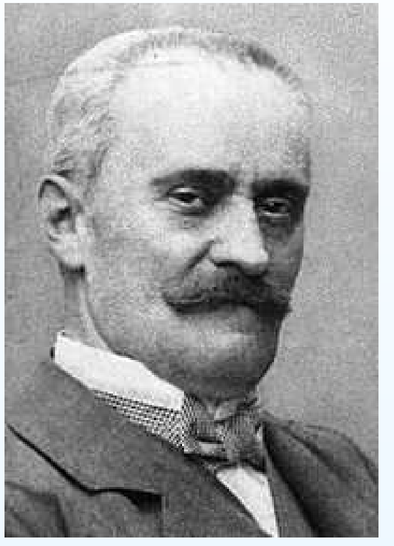
History of (convex) optimization

- 1900- WW2 - convex analysis - properties of convex sets and convex functions (Carathéodory, Minkowski, Farkas, Steinitz)
- 1928 - John von Neumann - existence of the minmax saddle point (game theory)
- 1930 - Wassily Leontief - the first linear program formulation
- 1947 - George Dantzig - simplex algorithm
- 1950' - quadratic programming

Topic 1: What is it and what is it good for...

C. Carathéodory

H. Minkowski



G. Farkas

E. Steinitz

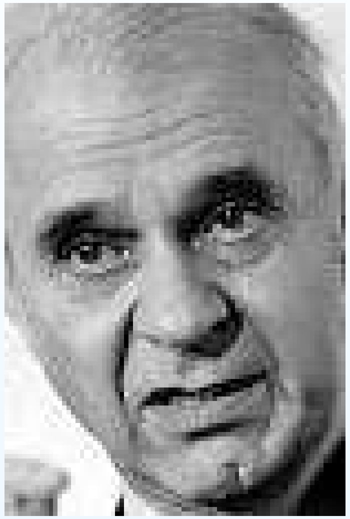
Constantin Carathéodory	(1873-1950)
Hermann Minkowski	(1864-1909)
Gyula Farkas	(1847-1930)
Ernst Steinitz	(1871-1928)

Topic 1: What is it and what is it good for...

- 1960' - geometric programming (Duffin, Peterson, Zener)
- WW2-1980 - duality theory, optimizality conditions (Kuhn, Tucker, Fenchel, Rockafellar)
- 1958 - Sion - generalization of von Neumann's theorem
- 1960-1970 - new scientific branch: complexity theory
- 1968 - Fiacco, McCormick - barrier methods
- 1972 - Klee, Minty - simplex method is NOT polynomial

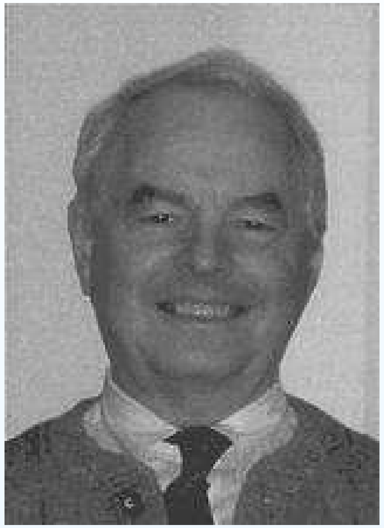
Topic 1: What is it and what is it good for...

J. von Neumann



W. Leontief

G. Dantzig

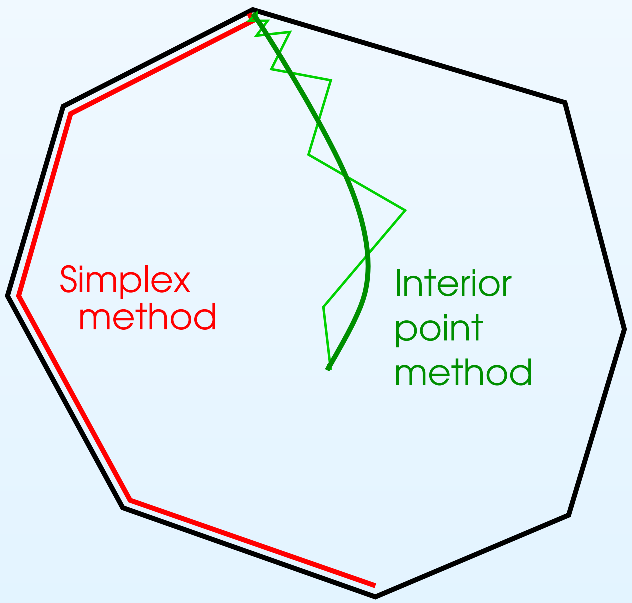


R. T. Rockafellar

John von Neumann	(1903-1957)
Wassily Leontief	(1905-1999)
George Dantzig	(1914-2005)
R. Tyrrell Rockafellar	http://www.math.washington.edu/ rtr/mypage.html

Topic 1: What is it and what is it good for...

- 1979 - Leonid Khachian - ellipsoid method - the 1st polynomial algorithm for linear programming, slow in practice
- 1984 - Narendra Karmarkar - "projective" algorithm for linear programming - polynomial AND fast, beginning of the modern interior point methods (IPM)



Topic 1: What is it and what is it good for...

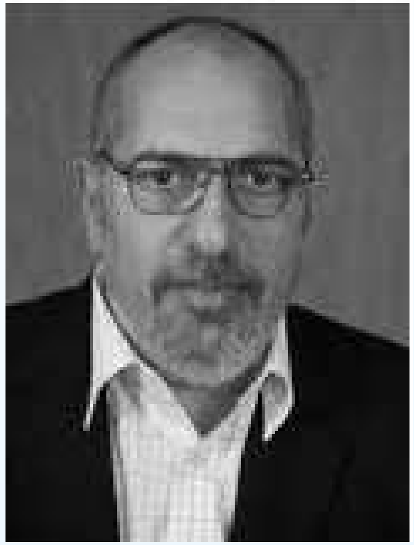
- 1988 - Nesterov, Nemirovski - generalization of IMP for convex optimization problems
- 1992 - semidefinite programming - first efficient algorithms
- 2000-.... - general cone programming, applications in engineering, finance, etc. , methods for non smooth convex optimization problems

Topic 1: What is it and what is it good for...

H. W. Kuhn

A. W. Tucker

Y. Nesterov



A. Nemirovski

Harold W. Kuhn	http://www.math.princeton.edu/directory/harold-w-kuhn
Albert W. Tucker	(1905-1995)
Yurii Nesterov	http://www.core.ucl.ac.be/ nesterov/
Arkadi Nemirovski	http://www.isye.gatech.edu/faculty-staff/profile.php?entry=an63

Convex optimization problem formulation:

$$\begin{array}{ll} Min & f_0(x) \\ & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array}$$

$f_i(x) : \mathbb{R}^n \to \mathbb{R},\ i = 0, 1, \ldots, m$ - convex functions
$h_i(x) : \mathbb{R}^n \to \mathbb{R},\ i = 1, \ldots, p$ - affine functions

Feasible set:

$$\mathcal{P} = \{x \mid f_i(x) \leq 0, i = 1, \ldots, m, h_i(x) = 0, i = 1, \ldots, p.\}$$

Optimal value:

$$p^* = \inf\{f_0(x) \mid x \in \mathcal{P}\} \in \mathbb{R} \cup \pm\infty, \quad p^* = +\infty \ \Leftrightarrow \ \mathcal{P} = \emptyset.$$

Widely known subclasses of convex optimization:

1. **Least squares problem**

$$Min \quad f_0(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{m} (a_i^T x - b_i)^2$$

2. **Linear programming**

$$\begin{aligned} Min \quad & c^T x \\ & a_i^T x \le b_i, \quad i = 1, 2, \ldots, m \end{aligned}$$

Example: Aproximation problem:

$$Min \quad f_0(x) = \|Ax - b\|$$

$A \in \mathbb{R}^{m \times n},\ b \in \mathbb{R}^m,\ x \in \mathbb{R}^n$, $\|\cdot\|$ is a norm in $\mathbb{R}^m$.

- $Ax \approx b$, residual $r = Ax - b$
- If $b \in \mathcal{S}(A)$, then clearly $p^* = 0$.

l_2	l_∞	l_1
$Min \quad \|Ax - b\|_2$	$Min \quad \|Ax - b\|_\infty$	$Min \quad \|Ax - b\|_1$
$Min \quad \left(\sum_{i=1}^m r_i^2\right)^{\frac{1}{2}}$	$Min \quad \max_i \|r_i\|$	$Min \quad \sum_{i=1}^m \|r_i\|$

l_2 - Least-squares

- equivalent with convex quadratic programming problem

$$Min \quad f_0(x) = x^T A^T Ax - 2b^T Ax + b^T b$$

- $\hat{x}$ is the optimal solution $\Leftrightarrow A^T Ax = A^T b$
- Assuming:
  1. columns of A are linearly independent;
  2. $b \notin \mathcal{S}(A)$ ($\Rightarrow \ m > n$),

  - the optimal solution can be expressed as

$$\hat{x} = (A^T A)^{-1} A^T b$$

- puts large penalty on large residuals

l_∞ - Chebyshev (minimax) approximation problem

- Chebyshev approximation is equivalent to linear programming problem:

$$\begin{aligned} Min \quad & t \\ & -t\mathbf{1} \leq Ax - b \leq t\mathbf{1} \end{aligned}$$

where $\mathbf{1} = (1, 1, \ldots, 1)^T \in \mathbb{R}^m$, $t \in \mathbb{R}$.

l_1 - approximation problem

- l_1 - approximation is equivalent to linear programming problem:

$$\begin{aligned} Min \quad & \mathbf{1}^T t \\ & -t \leq Ax - b \leq t \end{aligned}$$

where $t \in \mathbb{R}^m$

- large number of zero or very small residuals

Generalized convex programming problem

Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex set.

$$\begin{array}{ll} Min & f_0(x) \\ & x \in \mathcal{C} \end{array}$$

The set $\mathcal{C}$ should be properly described in order to analyze the problem and to design an efficient algorithm for solving it.

Example: $\mathcal{S}_+^2 = \{X \in \mathbb{R}^{2\times 2} \mid X = X^T, X \text{ is p.s.d}\}$ - convex set

$$X = \begin{pmatrix} x & y \\ y & z \end{pmatrix} \in \mathcal{S}_+^2 \;\Leftrightarrow\; x \geq 0,\; z \geq 0,\; y^2 - xz \leq 0$$

- concovex constraint

Topic 1: What is it and what is it good for...

- $\mathcal{K}$ - proper cone (closed, convex, pointed, $int(\mathcal{K}) \neq \emptyset$)
- $x \preceq_{\mathcal{K}} y \;\Leftrightarrow\; y - x \in \mathcal{K}$
- $\preceq_{\mathcal{K}}$ - partial ordering (reflexive, antisymmetric, transitive), invariant w.r.t. addition, nonnegative scaling, limit; it is NOT a linear ordering.

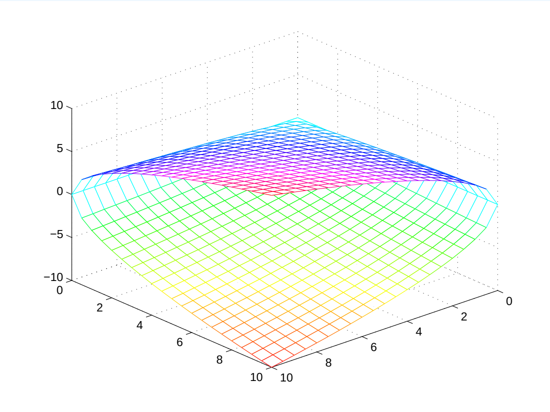
\- $\mathcal{K} \subseteq \mathbb{R}^m$ - proper cone
\- $\preceq_{\mathcal{K}}$ - generalized inequality

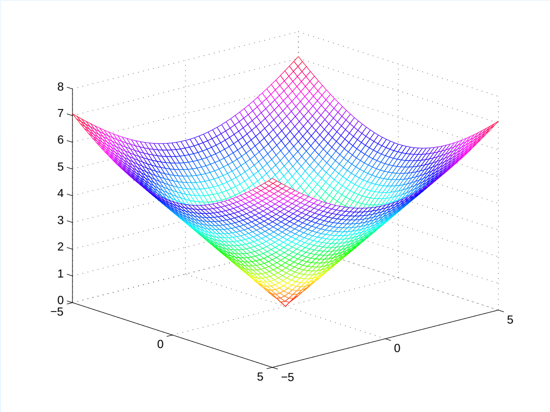
A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called $\mathcal{K}$-convex $\Leftrightarrow \forall x, y \in \mathbb{R}^n$, $\forall \lambda \in [0, 1]$:

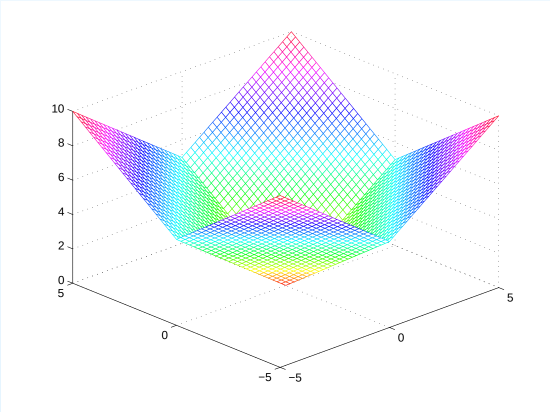
$$f(\lambda x + (1-\lambda)y) \preceq_{\mathcal{K}} \lambda f(x) + (1-\lambda) f(y).$$

Examples of proper cones:

- $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$
- $\mathcal{S}^n_+ = \{X \in \mathbb{R}^{n\times n} \mid X = X^T, X \text{ je k.s.d.}\}$
- $\mathcal{K}_{norm} = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\| \leq t\}$
- $\mathcal{K}_{pol} = \{c \in \mathbb{R}^n \mid c_1 + c_2 t + \cdots + c_n t^{n-1} \geq 0 \; \forall t \in [0,1]\}$







Generalized convex programming problem:

$$\begin{array}{ll} Min & f_0(x) \\ & f_i(x) \preceq_{\mathcal{K}_i} 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{array}$$

$\mathcal{K}_0 = \mathbb{R}^n_+$, $\mathcal{K}_i, (i = 1, \ldots, m)$ are (possibly different) proper cones and the functions $f_i, (i = 0, 1, \ldots, m)$ are $\mathcal{K}_i$-convex.

If the functions $f_0, f_1, \ldots, f_m$ are linear $\Rightarrow$

$\mathcal{K}$	$\mathbb{R}^n_+$	$\mathcal{S}^n_+$	$\mathcal{K}_{\|\cdot\|_2}$
Problem:	LP	SDP	SOCP

Semidefinite programming

- $C, A_1, \dots, A_m \in \mathcal{S}^n$, $b \in \mathbb{R}^m$

- SDP in standard form:

$$\begin{array}{ll} Min & \mathrm{tr}(CX) \\ & \mathrm{tr}(A_i X) = b_i \quad i = 1, \dots, m \\ & X \succeq 0 \end{array}$$

- SDP with LMI constraints

$$\begin{array}{ll} Min & b^T x \\ & x_1 A_1 + \cdots + x_m A_m \preceq C \end{array}$$

Example: Markowitz portfolio optimization

n	number of assets
x_i	(relative) amount of the i-th asset ($i = 1, 2, \ldots, n$)
p_i	return of the i-th asset
$\bar{p} = E(p)$	expected return
$\Sigma = E((p - \bar{p})(p - \bar{p})^T)$	covariance matrix
$x^T \Sigma x$	portfolio variance x - risk meassure

Classical portfolio problem:

$$\begin{array}{ll} Min_x & x^T \Sigma x \\ & \bar{p}^T x \geq r_{min} \\ & \mathbf{1}^T x = 1, \qquad x \geq 0 \end{array}$$

- convex quadratic problem with linear constraints

Alternative problem:
Given portfolio x and only partial information about the covariance matrix Σ - for example

- $L_{ij} \leq \Sigma_{ij} \leq U_{ij}$,
- Σ_{ii} - are known, partial / no information about $\Sigma_{ij}, i \neq j$
- $l_{ij} \leq \frac{\Sigma_{ij}}{\sqrt{(\Sigma_{ii}\Sigma_{jj})}} \leq u_{ij}$

We are looking for the worst-case risk withing the given constraints:

$$\begin{array}{ll} Max_{\Sigma} & x^T \Sigma x \\ & L_{ij} \leq \Sigma_{ij} \leq U_{ij} \\ & \Sigma \succeq 0 \end{array}$$

- semidefinite programming problem, since $x^T \Sigma x = \mathrm{tr}(x^T \Sigma x) = \mathrm{tr}(\Sigma x x^T)$.