Convex Optimization M2

Lecture 1
Today

- Convex optimization: introduction
- Course organization and other gory details...
- Convex sets, basic definitions.
Convex Optimization
How do we identify easy and hard problems?

Convexity: why is it so important?

Modeling: how do we recognize easy problems in real applications?

Algorithms: how do we solve these problems in practice?
Least squares (LS)

\[ \text{minimize} \quad \| Ax - b \|_2^2 \]

\[ A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m \text{ are parameters}; \quad x \in \mathbb{R}^n \text{ is variable} \]

- Complete theory (existence & uniqueness, sensitivity analysis . . . )
- Several algorithms compute (global) solution reliably
- We can solve dense problems with \( n = 1000 \) vbles, \( m = 10000 \) terms
- By exploiting structure (e.g., sparsity) can solve far larger problems

... LS is a (widely used) technology
Linear program (LP)

\[
\begin{align*}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & a_i^T x \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]

\(c, a_i \in \mathbb{R}^n\) are parameters; \(x \in \mathbb{R}^n\) is variable

- Nearly complete theory
  (existence & uniqueness, sensitivity analysis . . . )
- Several algorithms compute (global) solution reliably
- Can solve dense problems with \(n = 1000\) vsles, \(m = 10000\) constraints
- By exploiting structure (e.g., sparsity) can solve far larger problems

. . . LP is a (widely used) technology
Quadratic program (QP)

\[
\begin{align*}
\text{minimize} \quad & \|Fx - g\|_2^2 \\
\text{subject to} \quad & a_i^T x \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]

- Combination of LS & LP
- Same story . . . QP is a technology
- Reliability: Programmed on chips to solve real-time problems
- Classic application: portfolio optimization
The bad news

- LS, LP, and QP are exceptions
- Most optimization problems, even some very simple looking ones, are intractable
- The objective of this class is to show you how to recognize the nice ones...
- Many, many applications across all fields...
Polynomial minimization

\[
\text{minimize } p(x)
\]

\(p\) is polynomial of degree \(d\); \(x \in \mathbb{R}^n\) is variable

- Except for special cases (e.g., \(d = 2\)) this is a very difficult problem
- Even sparse problems with size \(n = 20, d = 10\) are essentially intractable
- All algorithms known to solve this problem require effort exponential in \(n\)
What makes a problem easy or hard?

Classical view:

- **linear** is easy
- **nonlinear** is hard(er)
What makes a problem easy or hard?

Emerging (and correct) view:

... the great watershed in optimization isn’t between linearity and nonlinearity, but convexity and nonconvexity.

— R. Rockafellar, SIAM Review 1993
**Convex optimization**

minimize \( f_0(x) \)
subject to \( f_1(x) \leq 0, \ldots, f_m(x) \leq 0 \)

\( x \in \mathbb{R}^n \) is optimization variable; \( f_i : \mathbb{R}^n \to \mathbb{R} \) are convex:

\[
f_i(\lambda x + (1 - \lambda)y) \leq \lambda f_i(x) + (1 - \lambda)f_i(y)
\]

for all \( x, y, 0 \leq \lambda \leq 1 \)

- includes LS, LP, QP, and **many others**
- like LS, LP, and QP, convex problems are **fundamentally tractable**
Example: Stochastic LP

Consider the following stochastic LP:

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad \Pr(\bar{a}_i^T x \leq b_i) \geq \eta, \quad i = 1, \ldots, m
\end{align*}
\]

coefficient vectors \( a_i \) IID, \( \mathcal{N}(\bar{a}_i, \Sigma_i) \); \( \eta \) is required reliability

- for fixed \( x \), \( a_i^T x \) is \( \mathcal{N}(\bar{a}_i^T x, x^T \Sigma_i x) \)
- so for \( \eta = 50\% \), stochastic LP reduces to LP

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad \bar{a}_i^T x \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]

and so is easily solved

- what about other values of \( \eta \), e.g., \( \eta = 10\% \? \); \( \eta = 90\% \)?
$\{x \mid \text{Prob}(a_i^T x \leq b_i) \geq \eta, \ i = 1, \ldots, m\}$

- $\eta = 10\%$
- $\eta = 50\%$
- $\eta = 90\%$
stochastic LP with reliability $\eta = 90\%$ is convex, and very easily solved

stochastic LP with reliability $\eta = 10\%$ is not convex, and extremely difficult

moral: very difficult and very easy problems can look quite similar (to the untrained eye)
A brief history. . .

- The field is about 50 years old.
- Starts with the work of Von Neumann, Kuhn and Tucker, etc.
- Explodes in the 60’s with the advent of “relatively” cheap and efficient computers. . .
- Key to all this: fast linear algebra
- Some of the theory developed before computers even existed. . .
Convex optimization: history

- Convexity $\implies$ low complexity:

  "... In fact the great watershed in optimization isn’t between linearity and nonlinearity, but convexity and nonconvexity."  
  T. Rockafellar.

- True: Nemirovskii and Yudin [1979].

- Very true: Karmarkar [1984].

- Seriously true: convex programming, Nesterov and Nemirovskii [1994].
Standard convex complexity analysis

- All convex minimization problems with: a first order oracle (returning $f(x)$ and a subgradient) can be solved in polynomial time in size and number of precision digits.

- Proved using the **ellipsoid method** by Nemirovskii and Yudin [1979].

- Very slow convergence in practice.
Simplex algorithm by Dantzig (1949): exponential worst-case complexity, very efficient in most cases.

Khachiyan [1979] then used the ellipsoid method to show the polynomial complexity of LP.

Karmarkar [1984] describes the first efficient polynomial time algorithm for LP, using interior point methods.
Nesterov and Nemirovskii [1994] show that the interior point methods used for LPs can be applied to a larger class of structured convex problems.

The **self-concordance** analysis that they introduce extends the polynomial time complexity proof for LPs.

Most operations that preserve convexity also preserve self-concordance.

The complexity of a certain number of elementary problems can be directly extended to a much wider class.
Symmetric cone programs

- An important particular case: linear programming on symmetric cones

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax - b \in \mathcal{K}
\end{align*}
\]

- These include the LP, second-order (Lorentz) and semidefinite cone:

  LP: \( \{ x \in \mathbb{R}^n : x \geq 0 \} \)
  Second order: \( \{ (x, y) \in \mathbb{R}^n \times \mathbb{R} : \|x\| \leq y \} \)
  Semidefinite: \( \{ X \in \mathbb{S}^n : X \succeq 0 \} \)

- Again, the class of problems that can be represented using these cones is extremely vast.
Course Organization
Course Plan

- Convex analysis & modeling
- Duality
- Algorithms: interior point methods, first order methods.
- Applications
Grading

Course website with lecture notes, homework, etc.

http://di.ens.fr/~aspremon/OptConvexeM2.html

- A few homeworks, will be posted online.
  
  Email your homeworks to dm.daspremont@gmail.com

  you will get an automatic reply to your message if it has been received.

- A final exam.
Short blurb

- Contact info on http://di.ens.fr/~aspremon
- Email: aspremon@ens.fr
- Dual PhDs: Ecole Polytechnique & Stanford University
- Interests: Optimization, machine learning, statistics & finance.
References

- All lecture notes will be posted online
- Textbook: **Convex Optimization** by Lieven Vandenberghe and Stephen Boyd, available online at:
  
  http://www.stanford.edu/~boyd/cvxbook/

- See also Ben-Tal and Nemirovski [2001], “Lectures On Modern Convex Optimization: Analysis, Algorithms, And Engineering Applications”, SIAM.
  
  http://www2.isye.gatech.edu/~nemirovs/

Convex Sets
Convex Sets

- affine and convex sets
- some important examples
- operations that preserve convexity
- generalized inequalities
- separating and supporting hyperplanes
- dual cones and generalized inequalities
**Affine set**

**line** through $x_1$, $x_2$: all points

$$x = \theta x_1 + (1 - \theta) x_2 \quad (\theta \in \mathbb{R})$$

**affine set**: contains the line through any two distinct points in the set

**example**: solution set of linear equations $\{x \mid Ax = b\}$
Convex set

**Line segment** between $x_1$ and $x_2$: all points

$$x = \theta x_1 + (1 - \theta)x_2$$

with $0 \leq \theta \leq 1$

**Convex set:** contains line segment between any two points in the set

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \implies \theta x_1 + (1 - \theta)x_2 \in C$$

**Examples** (one convex, two nonconvex sets)
**Convex combination and convex hull**

**convex combination** of $x_1, \ldots, x_k$: any point $x$ of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1$, $\theta_i \geq 0$

**convex hull** $\text{Co}S$: set of all convex combinations of points in $S$
Conic (nonnegative) combination of $x_1$ and $x_2$: any point of the form

$$x = \theta_1 x_1 + \theta_2 x_2$$

with $\theta_1 \geq 0$, $\theta_2 \geq 0$

Convex cone: set that contains all conic combinations of points in the set
Hyperplanes and halfspaces

**hyperplane**: set of the form \( \{ x \mid a^T x = b \} \) \((a \neq 0)\)

\[ a^T x = b \]

**halfspace**: set of the form \( \{ x \mid a^T x \leq b \} \) \((a \neq 0)\)

\[ a^T x \leq b \]

- \(a\) is the normal vector
- hyperplanes are affine and convex; halfspaces are convex
Euclidean balls and ellipsoids

(Euclidean) ball with center $x_c$ and radius $r$:

$$B(x_c, r) = \{ x \mid \| x - x_c \|_2 \leq r \} = \{ x_c + ru \mid \| u \|_2 \leq 1 \}$$

ellipsoid: set of the form

$$\{ x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1 \}$$

with $P \in S_+^n$ (i.e., $P$ symmetric positive definite)

other representation: $\{ x_c + Au \mid \| u \|_2 \leq 1 \}$ with $A$ square and nonsingular
Norm balls and norm cones

**norm:** a function $\| \cdot \|$ that satisfies

- $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$
- $\|tx\| = |t| \|x\|$ for $t \in \mathbb{R}$
- $\|x + y\| \leq \|x\| + \|y\|$

notation: $\| \cdot \|$ is general (unspecified) norm; $\| \cdot \|_{\text{symb}}$ is particular norm

**norm ball** with center $x_c$ and radius $r$: $\{ x \mid \|x - x_c\| \leq r \}$

**norm cone:** $\{(x, t) \mid \|x\| \leq t\}$

Euclidean norm cone is called second-order cone

norm balls and cones are convex
Polyhedra

solution set of finitely many linear inequalities and equalities

\[ Ax \preceq b, \quad Cx = d \]

\((A \in \mathbb{R}^{m \times n}, \ C \in \mathbb{R}^{p \times n}, \ \preceq \text{ is componentwise inequality})\)

polyhedron is intersection of finite number of halfspaces and hyperplanes
Positive semidefinite cone

**notation:**

- \( S^n \) is set of symmetric \( n \times n \) matrices

- \( S^n_+ = \{X \in S^n \mid X \succeq 0\} \): positive semidefinite \( n \times n \) matrices

\[ X \in S^n_+ \iff z^T X z \geq 0 \text{ for all } z \]

\( S^n_+ \) is a convex cone

- \( S^{n++} = \{X \in S^n \mid X \succ 0\} \): positive definite \( n \times n \) matrices

**example:** \[ \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in S^2_+ \]
Operations that preserve convexity

practical methods for establishing convexity of a set $C$

1. apply definition

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \implies \quad \theta x_1 + (1 - \theta)x_2 \in C$$

2. show that $C$ is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, . . . ) by operations that preserve convexity

- intersection
- affine functions
- perspective function
- linear-fractional functions
the intersection of (any number of) convex sets is convex

example:

\[ S = \{ x \in \mathbb{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3 \} \]

where \( p(t) = x_1 \cos t + x_2 \cos 2t + \cdots + x_m \cos mt \)

for \( m = 2 \):

\[ S \]
suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine ($f(x) = Ax + b$ with $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$)

- the image of a convex set under $f$ is convex

\[ S \subseteq \mathbb{R}^n \text{ convex} \implies f(S) = \{ f(x) \mid x \in S \} \text{ convex} \]

- the inverse image $f^{-1}(C)$ of a convex set under $f$ is convex

\[ C \subseteq \mathbb{R}^m \text{ convex} \implies f^{-1}(C) = \{ x \in \mathbb{R}^n \mid f(x) \in C \} \text{ convex} \]

examples

- scaling, translation, projection

- solution set of linear matrix inequality \( \{ x \mid x_1 A_1 + \cdots + x_m A_m \preceq B \} \) (with $A_i, B \in \mathcal{S}^p$)

- hyperbolic cone \( \{ x \mid x^T P x \leq (c^T x)^2, c^T x \geq 0 \} \) (with $P \in \mathcal{S}_+^n$)
Perspective and linear-fractional function

**perspective function** $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$:

$$P(x, t) = \frac{x}{t}, \quad \text{dom} \ P = \{(x, t) \mid t > 0\}$$

images and inverse images of convex sets under perspective are convex

**linear-fractional function** $f : \mathbb{R}^n \to \mathbb{R}^m$:

$$f(x) = \frac{Ax + b}{c^T x + d}, \quad \text{dom} \ f = \{x \mid c^T x + d > 0\}$$

images and inverse images of convex sets under linear-fractional functions are convex
example of a linear-fractional function

\[ f(x) = \frac{1}{x_1 + x_2 + 1} \]
Generalized inequalities

A convex cone $K \subseteq \mathbb{R}^n$ is a **proper cone** if

- $K$ is closed (contains its boundary)
- $K$ is solid (has nonempty interior)
- $K$ is pointed (contains no line)

**Examples**

- Nonnegative orthant $K = \mathbb{R}^n_+ = \{ x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \ldots, n \}$
- Positive semidefinite cone $K = \mathbb{S}^n_+$
- Nonnegative polynomials on $[0, 1]$:  
  \[
  K = \{ x \in \mathbb{R}^n \mid x_1 + x_2t + x_3t^2 + \cdots + x_n t^{n-1} \geq 0 \text{ for } t \in [0, 1] \} \]
generalized inequality defined by a proper cone $K$:

$$x \preceq_K y \iff y - x \in K, \quad x \prec_K y \iff y - x \in \text{int} K$$

elements

- componentwise inequality ($K = \mathbb{R}^n_+$)
  
  $$x \preceq_{\mathbb{R}^n_+} y \iff x_i \leq y_i, \quad i = 1, \ldots, n$$

- matrix inequality ($K = S^n_+$)
  
  $$X \preceq_{S^n_+} Y \iff Y - X \text{ positive semidefinite}$$

these two types are so common that we drop the subscript in $\preceq_K$

properties: many properties of $\preceq_K$ are similar to $\leq$ on $\mathbb{R}$, e.g.,

$$x \preceq_K y, \quad u \preceq_K v \implies x + u \preceq_K y + v$$
Minimum and minimal elements

\( \preceq_K \) is not in general a linear ordering: we can have \( x \not\preceq_K y \) and \( y \not\preceq_K x \)

\( x \in S \) is the **minimum element** of \( S \) with respect to \( \preceq_K \) if

\[
y \in S \implies x \preceq_K y
\]

\( x \in S \) is a **minimal element** of \( S \) with respect to \( \preceq_K \) if

\[
y \in S, \quad y \preceq_K x \implies y = x
\]

**example** (\( K = \mathbb{R}^2_+ \))

\( x_1 \) is the minimum element of \( S_1 \)

\( x_2 \) is a minimal element of \( S_2 \)
if $C$ and $D$ are disjoint convex sets, then there exists $a \neq 0$, $b$ such that

$$a^T x \leq b \text{ for } x \in C, \quad a^T x \geq b \text{ for } x \in D$$

the hyperplane $\{x \mid a^T x = b\}$ separates $C$ and $D$

strict separation requires additional assumptions (e.g., $C$ is closed, $D$ is a singleton)
supporting hyperplane theorem

**Supporting hyperplane** to set $C$ at boundary point $x_0$:

$$\{ x \mid a^T x = a^T x_0 \}$$

where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$

**supporting hyperplane theorem**: if $C$ is convex, then there exists a supporting hyperplane at every boundary point of $C$
Dual cones and generalized inequalities

dual cone of a cone $K$:

\[ K^* = \{ y \mid y^T x \geq 0 \text{ for all } x \in K \} \]

examples

- $K = \mathbb{R}_+^n$: $K^* = \mathbb{R}_+^n$
- $K = \mathbb{S}_+^n$: $K^* = \mathbb{S}_+^n$
- $K = \{(x, t) \mid \|x\|_2 \leq t\}$: $K^* = \{(x, t) \mid \|x\|_2 \leq t\}$
- $K = \{(x, t) \mid \|x\|_1 \leq t\}$: $K^* = \{(x, t) \mid \|x\|_\infty \leq t\}$

first three examples are self-dual cones

dual cones of proper cones are proper, hence define generalized inequalities:

\[ y \succeq_{K^*} 0 \iff y^T x \geq 0 \text{ for all } x \succeq_K 0 \]
Minimum and minimal elements via dual inequalities

**Minimum element** w.r.t. \( \preceq_K \)

\( x \) is minimum element of \( S \) iff for all \( \lambda \succ_K \ast 0 \), \( x \) is the unique minimizer of \( \lambda^T z \) over \( S \)

**Minimal element** w.r.t. \( \preceq_K \)

- if \( x \) minimizes \( \lambda^T z \) over \( S \) for some \( \lambda \succ_K \ast 0 \), then \( x \) is minimal

- if \( x \) is a minimal element of a *convex* set \( S \), then there exists a nonzero \( \lambda \succeq_K \ast 0 \) such that \( x \) minimizes \( \lambda^T z \) over \( S \)
References


