Optimisation Combinatoire et Convexe.

Introduction, convexité, dualité.
Today

- Convex optimization: introduction
- Course organization and other gory details...
- Convex optimization: basic concepts
Convex Optimization
Convex optimization

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_1(x) \leq 0, \ldots, f_m(x) \leq 0
\end{align*}
\]

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

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

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

- This template includes LS, LP, QP, and many others.
- \textbf{Good news:} convex problems (LP, QP, etc) are \textit{fundamentally tractable}.
- \textbf{Bad news:} this is an exception, most nonconvex are \textit{completely intractable}.
Convex optimization

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

- Historical view: nonlinear problems are hard, linear ones are easy.

- In reality: **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].
Convexity, complexity

- 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.
Linear Programming

- 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.
From LP to structured convex programs

- Nesterov and Nemirovskii [1994] show that the interior point methods (IPM) 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.
Large-scale convex programs

Interior point methods.

- IPM essentially solved once and for all a broad range of medium-scale convex programs.
- For large-scale problems, computing a single Newton step is often too expensive.

First order methods.

- Dependence on precision is polynomial $O(1/\epsilon^\alpha)$, not logarithmic $O(\log(1/\epsilon))$. This is OK in many applications (stats, etc).
- Run a much larger number of cheaper iterations. No Hessian means significantly lower memory and CPU costs per iteration.
- No unified analysis (self-concordance for IPM): large library of disparate methods.
- Algorithmic choices strictly constrained by problem structure.

Objective: classify these techniques, study their performance & complexity.
Symmetric cone programs

- An important particular case: linear programming on symmetric cones

  minimize \( c^T x \)
  subject to \( Ax - b \in \mathcal{K} \)

- 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: \( \{ X \in \mathbb{S}^n : X \succeq 0 \} \)

- Broad class of problems can be represented in this way.

- **Good news:** Fast, reliable, open-source solvers available (SDPT3, CVX, etc).

This course will describe some “exotic” applications of these programs.
A few “miracles”

Beyond convexity.

- **Hidden convexity.** Convex programs solving nonconvex problems (*S*-lemma).
- **Approximation results.** Approximating combinatorial problems by convex programs.
  - Approximate *S*-lemma.
  - Approximation ratio for MaxCut, etc.
- **Recovery results on *ℓ*₁ penalties.** Finding sparse solutions to optimization problems using convex penalties.
  - Sparse signal reconstruction.
  - Matrix completion (collaborative filtering, NETFLIX, etc.).
Course Organization
Course outline

- Fundamental definitions
  - A brief primer on convexity and duality theory

- Algorithmic complexity
  - Interior point methods, self-concordance.
  - First order algorithms: complexity and classification.

- Modern applications
  - Statistics
  - Geometrical problems, graphs.
  - ...

- Some “miracles”: approximation, asymptotic and hidden convexity results
  - Measure concentration results.
  - $S$-lemma, MaxCut, low rank SDP solutions, nonconvex QCQP, etc.
  - High dimensional geometry
  - $\ell_1$ recovery, matrix completion, convex deconvolution, etc.
Info

- Course website with lecture notes, homework, etc.
  
  http://www.di.ens.fr/~aspremon/

- A final exam.
Contact info on http://www.di.ens.fr/~aspremon/

Email: aspremon@ens.fr

Dual PhDs: Ecole Polytechnique & Stanford University

Interests: Optimization, machine learning, statistics & finance.
All lecture notes will be posted online, none of the books below are required.

- “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
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

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

**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 \( S \)
Hyperplanes and halfspaces

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

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

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

\[ a^T x = b \]

\[ a^T x \geq b \]

\[ a^T x \leq b \]
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 $P \in \mathbb{S}^{n}_{++}$ (i.e., $P$ symmetric positive definite)

  other representation: $\{ x_c + Au \mid \|u\|_2 \leq 1 \}$, with $A$ square and nonsingular.

- Representation impacts problem formulation & complexity.

- Idem for polytopes, with polynomial number of vertices, exponential number of facets, and vice-versa.
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 $t \in \mathbb{R}$
- $\|x + y\| \leq \|x\| + \|y\|$\n
notation: $\| \cdot \|$ is general (unspecified) norm; $\| \cdot \|_{\text{symb}}$ is particular norm

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

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

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 \( m = 2 \):
Affine function

suppose \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is affine \( (f(x) = Ax + b \text{ with } A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m) \)

- 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 \( \{x \mid x_1A_1 + \cdots + x_mA_m \preceq B\} \)
  (with \( A_i, B \in S^p \))

- hyperbolic cone \( \{x \mid x^TPx \leq (c^Tx)^2, c^Tx \geq 0\} \) (with \( P \in S_+^n \))
**Perspective and linear-fractional function**

**Perspective function** \( P : \mathbb{R}^{n+1} \to \mathbb{R}^n \):

\[
P(x, t) = x/t, \quad \text{dom} \ P = \{(x, t) \mid t > 0\}
\]

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.
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 $[0,1]$: 

\[
K = \{ x \in \mathbb{R}^n \mid x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1} \geq 0 \text{ for } t \in [0,1] \}
\]
generalized inequality defined by a proper cone $K$:

$$x \leq_K y \iff y - x \in K, \quad x <_K y \iff y - x \in \text{int } K$$

examples

- componentwise inequality ($K = \mathbb{R}^n_+$)

$$x \leq_{\mathbb{R}^n_+} y \iff x_i \leq y_i, \quad i = 1, \ldots, n$$

- matrix inequality ($K = \mathbf{S}^n_+$)

$$X \leq_{\mathbf{S}^n_+} Y \iff Y - X \text{ positive semidefinite}$$

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

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

$$x \leq_K y, \quad u \leq_K v \implies x + u \leq_K y + v$$
if \( C \) and \( D \) are disjoint convex sets, then there exists \( a \neq 0, b \) such that

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

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

Classical result. Proof relies on minimizing distance between set, and using the argmin to explicitly produce separating hyperplane.
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:

$$y \succeq_{K^*} 0 \iff y^T x \geq 0 \text{ for all } x \succeq_K 0$$
Convex Functions
Outline

- basic properties and examples
- operations that preserve convexity
- the conjugate function
- quasiconvex functions
- log-concave and log-convex functions
- convexity with respect to generalized inequalities
**Definition**

\( f : \mathbb{R}^n \to \mathbb{R} \) is convex if \( \text{dom} \, f \) is a convex set and

\[
 f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)
\]

for all \( x, y \in \text{dom} \, f, \ 0 \leq \theta \leq 1 \)

- \( f \) is concave if \(-f\) is convex

- \( f \) is strictly convex if \( \text{dom} \, f \) is convex and

\[
 f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)
\]

for \( x, y \in \text{dom} \, f, \ x \neq y, \ 0 < \theta < 1 \)
Examples on $\mathbb{R}$

convex:
- affine: $ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$
- exponential: $e^{ax}$, for any $a \in \mathbb{R}$
- powers: $x^\alpha$ on $\mathbb{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on $\mathbb{R}$, for $p \geq 1$
- negative entropy: $x \log x$ on $\mathbb{R}_{++}$

concave:
- affine: $ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$
- powers: $x^\alpha$ on $\mathbb{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbb{R}_{++}$
Examples on $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$

Affine functions are convex and concave; all norms are convex.

**Examples on $\mathbb{R}^n$**

- Affine function $f(x) = a^T x + b$
- Norms: $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_k |x_k|$.

**Examples on $\mathbb{R}^{m \times n}$ ($m \times n$ matrices)**

- Affine function

  $$f(X) = \text{Tr}(A^T X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$$

- Spectral (maximum singular value) norm

  $$f(X) = \|X\|_2 = \sigma_{\text{max}}(X) = (\lambda_{\text{max}}(X^T X))^{1/2}$$
Restriction of a convex function to a line

\[ f : \mathbb{R}^n \to \mathbb{R} \] is convex if and only if the function \( g : \mathbb{R} \to \mathbb{R}, \)

\[ g(t) = f(x + tv), \quad \text{dom } g = \{ t \mid x + tv \in \text{dom } f \} \]

is convex (in \( t \)) for any \( x \in \text{dom } f, \ v \in \mathbb{R}^n \)

can check convexity of \( f \) by checking convexity of functions of one variable

**example.** \( f : S^n \to \mathbb{R} \) with \( f(X) = \log \det X, \ \text{dom } X = S^n_{++} \)

\[ g(t) = \log \det(X + tV) = \log \det X + \log \det(I + tX^{-1/2}VX^{-1/2}) \]

\[ = \log \det X + \sum_{i=1}^{n} \log(1 + t\lambda_i) \]

where \( \lambda_i \) are the eigenvalues of \( X^{-1/2}VX^{-1/2} \)

\( g \) is concave in \( t \) (for any choice of \( X \succ 0, \ V \)); hence \( f \) is concave
extended-value extension $\tilde{f}$ of $f$ is

$$\tilde{f}(x) = f(x), \quad x \in \text{dom } f, \quad \tilde{f}(x) = \infty, \quad x \notin \text{dom } f$$

often simplifies notation; for example, the condition

$$0 \leq \theta \leq 1 \implies \tilde{f}(\theta x + (1 - \theta)y) \leq \theta \tilde{f}(x) + (1 - \theta)\tilde{f}(y)$$

(as an inequality in $\mathbb{R} \cup \{\infty\}$), means the same as the two conditions

- $\text{dom } f$ is convex
- for $x, y \in \text{dom } f$,

$$0 \leq \theta \leq 1 \implies f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$
First-order condition

\[ f \text{ is differentiable if } \text{dom } f \text{ is open and the gradient} \]

\[
\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)
\]

exists at each \( x \in \text{dom } f \)

1st-order condition: differentiable \( f \) with convex domain is convex iff

\[
f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \text{dom } f
\]

first-order approximation of \( f \) is global underestimator
Second-order conditions

$f$ is **twice differentiable** if $\text{dom } f$ is open and the Hessian $\nabla^2 f(x) \in S^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \ldots, n,$$

exists at each $x \in \text{dom } f$

**2nd-order conditions:** for twice differentiable $f$ with convex domain

- $f$ is convex if and only if
  $$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \text{dom } f$$

- if $\nabla^2 f(x) \succ 0$ for all $x \in \text{dom } f$, then $f$ is strictly convex
Examples

**quadratic function:** \( f(x) = (1/2)x^TPx + q^Tx + r \) (with \( P \in S^n \))

\[
\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P
\]

convex if \( P \succeq 0 \)

**least-squares objective:** \( f(x) = \|Ax - b\|_2^2 \)

\[
\nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^TA
\]

convex (for any \( A \))

**quadratic-over-linear:** \( f(x, y) = x^2/y \)

\[
\nabla^2 f(x, y) = \frac{2}{y^3} \left[ \begin{array}{c} y \\ -x \end{array} \right] \left[ \begin{array}{c} y \\ -x \end{array} \right]^T \succeq 0
\]

convex for \( y > 0 \)
log-sum-exp: \( f(x) = \log \sum_{k=1}^{n} \exp x_k \) is convex

\[
\nabla^2 f(x) = \frac{1}{1^T z} \text{diag}(z) - \frac{1}{(1^T z)^2} zz^T \quad (z_k = \exp x_k)
\]

to show \( \nabla^2 f(x) \succeq 0 \), we must verify that \( v^T \nabla^2 f(x) v \geq 0 \) for all \( v \):

\[
v^T \nabla^2 f(x) v = \frac{\left( \sum_k z_k v_k^2 \right) \left( \sum_k z_k \right) - \left( \sum_k v_k z_k \right)^2}{\left( \sum_k z_k \right)^2} \geq 0
\]

since \( \left( \sum_k v_k z_k \right)^2 \leq \left( \sum_k z_k v_k^2 \right) \left( \sum_k z_k \right) \) (from Cauchy-Schwarz inequality)

geometric mean: \( f(x) = \left( \prod_{k=1}^{n} x_k \right)^{1/n} \) on \( \mathbb{R}^{n}_{++} \) is concave

(similar proof as for log-sum-exp)
Epigraph and sublevel set

\( \alpha \)-sublevel set of \( f : \mathbb{R}^n \to \mathbb{R} \):

\[
C_\alpha = \{ x \in \text{dom } f \mid f(x) \leq \alpha \}
\]

sublevel sets of convex functions are convex (converse is false)

epigraph of \( f : \mathbb{R}^n \to \mathbb{R} \):

\[
\text{epi } f = \{ (x, t) \in \mathbb{R}^{n+1} \mid x \in \text{dom } f, \ f(x) \leq t \}
\]

\( f \) is convex if and only if \( \text{epi } f \) is a convex set

A. d’Aspremont. M1 ENS.
Jensen’s inequality

**basic inequality:** if \( f \) is convex, then for \( 0 \leq \theta \leq 1 \),

\[
f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)
\]

**extension:** if \( f \) is convex, then

\[
f(\mathbb{E} z) \leq \mathbb{E} f(z)
\]

for any random variable \( z \)

basic inequality is special case with discrete distribution

\[
\text{Prob}(z = x) = \theta, \quad \text{Prob}(z = y) = 1 - \theta
\]
Operations that preserve convexity

practical methods for establishing convexity of a function

1. verify definition (often simplified by restricting to a line)

2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$

3. show that $f$ is obtained from simple convex functions by operations that preserve convexity

- nonnegative weighted sum
- composition with affine function
- pointwise maximum and supremum
- composition
- minimization
- perspective
Positive weighted sum & composition with affine function

**nonnegative multiple:** \( \alpha f \) is convex if \( f \) is convex, \( \alpha \geq 0 \)

**sum:** \( f_1 + f_2 \) convex if \( f_1, f_2 \) convex (extends to infinite sums, integrals)

**composition with affine function:** \( f(Ax + b) \) is convex if \( f \) is convex

**examples**

- log barrier for linear inequalities

\[
f(x) = - \sum_{i=1}^{m} \log(b_i - a_i^T x), \quad \text{dom } f = \{ x \mid a_i^T x < b_i, i = 1, \ldots, m \}
\]

- (any) norm of affine function: \( f(x) = \|Ax + b\| \)
Pointwise maximum

if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

examples

- piecewise-linear function: $f(x) = \max_{i=1,\ldots,m}(a^T_i x + b_i)$ is convex

- sum of $r$ largest components of $x \in \mathbb{R}^n$:

  $$f(x) = x[1] + x[2] + \cdots + x[r]$$

  is convex ($x[i]$ is $i$th largest component of $x$)

proof:

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$$
Pointwise supremum

If \( f(x, y) \) is convex in \( x \) for each \( y \in A \), then

\[
g(x) = \sup_{y \in A} f(x, y)
\]
is convex

**Examples**

- Support function of a set \( C \): \( S_C(x) = \sup_{y \in C} y^Tx \) is convex
- Distance to farthest point in a set \( C \):

\[
f(x) = \sup_{y \in C} \|x - y\|
\]
- Maximum eigenvalue of symmetric matrix: for \( X \in \mathbb{S}^n \),

\[
\lambda_{\text{max}}(X) = \sup_{\|y\|_2=1} y^TXy
\]
Composition with scalar functions

composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$:

$$f(x) = h(g(x))$$

$f$ is convex if

- $g$ convex, $h$ convex, $\tilde{h}$ nondecreasing
- $g$ concave, $h$ convex, $\tilde{h}$ nonincreasing

proof (for $n = 1$, differentiable $g, h$)

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

note: monotonicity must hold for extended-value extension $\tilde{h}$

examples

- $\exp g(x)$ is convex if $g$ is convex
- $1/g(x)$ is convex if $g$ is concave and positive
Vector composition

composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

$f$ is convex if

- $g_i$ convex, $h$ convex, $\tilde{h}$ nondecreasing in each argument
- $g_i$ concave, $h$ convex, $\tilde{h}$ nonincreasing in each argument

proof (for $n = 1$, differentiable $g, h$)

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

examples

- $\sum_{i=1}^{m} \log g_i(x)$ is concave if $g_i$ are concave and positive
- $\log \sum_{i=1}^{m} \exp g_i(x)$ is convex if $g_i$ are convex
Minimization

if $f(x, y)$ is convex in $(x, y)$ and $C$ is a convex set, then

$$g(x) = \inf_{y \in C} f(x, y)$$

is convex

examples

- $f(x, y) = x^T A x + 2 x^T B y + y^T C y$ with

  $$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \quad C \succ 0$$

  minimizing over $y$ gives $g(x) = \inf_y f(x, y) = x^T (A - BC^{-1}B^T)x$

  $g$ is convex, hence Schur complement $A - BC^{-1}B^T \succeq 0$

- distance to a set: $\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if $S$ is convex
The conjugate function

the **conjugate** of a function $f$ is

$$f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$$

- $f^*$ is convex (even if $f$ is not)
- Used in regularization, duality results, . . .
examples

- negative logarithm $f(x) = -\log x$

\[
    f^*(y) = \sup_{x > 0} (xy + \log x) \\
    = \begin{cases} 
        -1 - \log(-y) & y < 0 \\
        \infty & \text{otherwise} 
    \end{cases}
\]

- strictly convex quadratic $f(x) = (1/2)x^T Q x$ with $Q \in \mathbf{S}_+^n$

\[
    f^*(y) = \sup_{x} (y^T x - (1/2)x^T Q x) \\
    = \frac{1}{2} y^T Q^{-1} y
\]
Quasiconvex functions

$f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex if $\text{dom } f$ is convex and the sublevel sets

$$S_\alpha = \{x \in \text{dom } f \mid f(x) \leq \alpha\}$$

are convex for all $\alpha$

- $f$ is quasiconcave if $-f$ is quasiconvex
- $f$ is quasilinear if it is quasiconvex and quasiconcave
Examples

- $\sqrt{|x|}$ is quasiconvex on $\mathbb{R}$
- $\text{ceil}(x) = \inf\{z \in \mathbb{Z} \mid z \geq x\}$ is quasilinear
- $\log x$ is quasilinear on $\mathbb{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbb{R}_{++}^2$
- linear-fractional function
  \[
  f(x) = \frac{a^T x + b}{c^T x + d}, \quad \text{dom } f = \{x \mid c^T x + d > 0\}
  \]
  is quasilinear
- distance ratio
  \[
  f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \quad \text{dom } f = \{x \mid \|x - a\|_2 \leq \|x - b\|_2\}
  \]
  is quasiconvex
Properties

modified Jensen inequality: for quasiconvex $f$

$$0 \leq \theta \leq 1 \implies f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\}$$

first-order condition: differentiable $f$ with cvx domain is quasiconvex iff

$$f(y) \leq f(x) \implies \nabla f(x)^T(y - x) \leq 0$$

sums of quasiconvex functions are not necessarily quasiconvex
Log-concave and log-convex functions

A positive function $f$ is log-concave if $\log f$ is concave:

$$f(\theta x + (1 - \theta)y) \geq f(x)^\theta f(y)^{1-\theta} \quad \text{for } 0 \leq \theta \leq 1$$

$f$ is log-convex if $\log f$ is convex

- powers: $x^a$ on $\mathbb{R}_{++}$ is log-convex for $a \leq 0$, log-concave for $a \geq 0$
- many common probability densities are log-concave, e.g., normal:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1} (x-\bar{x})}$$

- cumulative Gaussian distribution function $\Phi$ is log-concave

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} \, du$$
Properties of log-concave functions

- twice differentiable $f$ with convex domain is log-concave if and only if
  \[ f(x) \nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T \]
  for all $x \in \text{dom } f$

- product of log-concave functions is log-concave

- sum of log-concave functions is not always log-concave

- integration: if $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is log-concave, then
  \[ g(x) = \int f(x, y) \, dy \]
  is log-concave (not easy to show)
consequences of integration property

- convolution \( f \ast g \) of log-concave functions \( f, g \) is log-concave

\[
(f \ast g)(x) = \int f(x - y)g(y)dy
\]

- if \( C \subseteq \mathbb{R}^n \) convex and \( y \) is a random variable with log-concave pdf then

\[
f(x) = \text{Prob}(x + y \in C)
\]

is log-concave

proof: write \( f(x) \) as integral of product of log-concave functions

\[
f(x) = \int g(x + y)p(y) \, dy, \quad g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C \end{cases}, \quad p \text{ is pdf of } y
\]
Convex Optimization Problems
Outline

- optimization problem in standard form
- convex optimization problems
- quasiconvex optimization
- linear optimization
- quadratic optimization
- geometric programming
- generalized inequality constraints
- semidefinite programming
- vector optimization
Optimization problem in standard form

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

- \(x \in \mathbb{R}^n\) is the optimization variable
- \(f_0 : \mathbb{R}^n \to \mathbb{R}\) is the objective or cost function
- \(f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m\), are the inequality constraint functions
- \(h_i : \mathbb{R}^n \to \mathbb{R}\) are the equality constraint functions

**optimal value:**

\[
p^* = \inf \{ f_0(x) \mid f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad h_i(x) = 0, \quad i = 1, \ldots, p \}
\]

- \(p^* = \infty\) if problem is infeasible (no \(x\) satisfies the constraints)
- \(p^* = -\infty\) if problem is unbounded below
Optimal and locally optimal points

$x$ is **feasible** if $x \in \text{dom } f_0$ and it satisfies the constraints

A feasible $x$ is **optimal** if $f_0(x) = p^*$; $X_{\text{opt}}$ is the set of optimal points

$x$ is **locally optimal** if there is an $R > 0$ such that $x$ is optimal for

$$
\begin{align*}
\text{minimize (over } z) & \quad f_0(z) \\
\text{subject to} & \quad f_i(z) \leq 0, \quad i = 1, \ldots, m, \quad h_i(z) = 0, \quad i = 1, \ldots, p \\
& \quad \|z - x\|_2 \leq R
\end{align*}
$$

**examples** (with $n = 1, \ m = p = 0$)

- $f_0(x) = 1/x, \ \text{dom } f_0 = \mathbb{R}_{++}: \ p^* = 0, \ \text{no optimal point}$
- $f_0(x) = -\log x, \ \text{dom } f_0 = \mathbb{R}_{++}: \ p^* = -\infty$
- $f_0(x) = x \log x, \ \text{dom } f_0 = \mathbb{R}_{++}: \ p^* = -1/e, \ x = 1/e \ \text{is optimal}$
- $f_0(x) = x^3 - 3x, \ p^* = -\infty, \ \text{local optimum at } x = 1$
Implicit constraints

the standard form optimization problem has an implicit constraint

\[ x \in D = \bigcap_{i=0}^{m} \text{dom } f_i \cap \bigcap_{i=1}^{p} \text{dom } h_i, \]

- we call \( D \) the domain of the problem
- the constraints \( f_i(x) \leq 0, h_i(x) = 0 \) are the explicit constraints
- a problem is unconstrained if it has no explicit constraints \( (m = p = 0) \)

example:

\[
\text{minimize } f_0(x) = - \sum_{i=1}^{k} \log(b_i - a_i^T x)
\]

is an unconstrained problem with implicit constraints \( a_i^T x < b_i \)
Feasibility problem

find \( x \)
subject to \( f_i(x) \leq 0, \quad i = 1, \ldots, m \)
\( h_i(x) = 0, \quad i = 1, \ldots, p \)

\( \text{can be considered a special case of the general problem with } f_0(x) = 0: \)

minimize \( 0 \)
subject to \( f_i(x) \leq 0, \quad i = 1, \ldots, m \)
\( h_i(x) = 0, \quad i = 1, \ldots, p \)

- \( p^* = 0 \) if constraints are feasible; any feasible \( x \) is optimal
- \( p^* = \infty \) if constraints are infeasible
Convex optimization problem

standard form convex optimization problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad a^T_i x = b_i, \quad i = 1, \ldots, p
\end{align*}
\]

- $f_0, f_1, \ldots, f_m$ are convex; equality constraints are affine
- problem is quasiconvex if $f_0$ is quasiconvex (and $f_1, \ldots, f_m$ convex)

often written as

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

important property: feasible set of a convex optimization problem is convex
example

minimize \( f_0(x) = x_1^2 + x_2^2 \)
subject to \( f_1(x) = x_1/(1 + x_2^2) \leq 0 \)
\( h_1(x) = (x_1 + x_2)^2 = 0 \)

- \( f_0 \) is convex; feasible set \( \{(x_1, x_2) \mid x_1 = -x_2 \leq 0\} \) is convex
- not a convex problem (according to our definition): \( f_1 \) is not convex, \( h_1 \) is not affine
- equivalent (but not identical) to the convex problem

minimize \( x_1^2 + x_2^2 \)
subject to \( x_1 \leq 0 \)
\( x_1 + x_2 = 0 \)
Local and global optima

any locally optimal point of a convex problem is (globally) optimal

Proof: suppose $x$ is locally optimal and $y$ is optimal with $f_0(y) < f_0(x)$

$x$ locally optimal means there is an $R > 0$ such that

\[ z \text{ feasible, } \|z - x\|_2 \leq R \implies f_0(z) \geq f_0(x) \]

consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, so $0 < \theta < 1/2$
- $z$ is a convex combination of two feasible points, hence also feasible
- $\|z - x\|_2 = R/2$ and

\[ f_0(z) \leq \theta f_0(x) + (1 - \theta)f_0(y) < f_0(x) \]

which contradicts our assumption that $x$ is locally optimal
Optimality criterion for differentiable $f_0$

$x$ is optimal if and only if it is feasible and

$$\nabla f_0(x)^T(y - x) \geq 0 \quad \text{for all feasible } y$$

if nonzero, $\nabla f_0(x)$ defines a supporting hyperplane to feasible set $X$ at $x$
unconstrained problem: $x$ is optimal if and only if

$$x \in \text{dom} f_0, \quad \nabla f_0(x) = 0$$

equality constrained problem

minimize $f_0(x)$ subject to $Ax = b$

$x$ is optimal if and only if there exists a $\nu$ such that

$$x \in \text{dom} f_0, \quad Ax = b, \quad \nabla f_0(x) + A^T\nu = 0$$

minimization over nonnegative orthant

minimize $f_0(x)$ subject to $x \succeq 0$

$x$ is optimal if and only if

$$x \in \text{dom} f_0, \quad x \succeq 0, \quad \left\{ \begin{array}{l}
\nabla f_0(x)_i \geq 0 \quad x_i = 0 \\
\nabla f_0(x)_i = 0 \quad x_i > 0
\end{array} \right.$$
two problems are (informally) **equivalent** if the solution of one is readily obtained from the solution of the other, and vice-versa

some common transformations that preserve convexity:

- **eliminating equality constraints**

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

is equivalent to

\[
\begin{align*}
\text{minimize (over } z) & \quad f_0(Fz + x_0) \\
\text{subject to} & \quad f_i(Fz + x_0) \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

where \( F \) and \( x_0 \) are such that

\[
Ax = b \iff x = Fz + x_0 \text{ for some } z
\]
■ introducing equality constraints

\[
\begin{align*}
\text{minimize} & \quad f_0(A_0x + b_0) \\
\text{subject to} & \quad f_i(A_ix + b_i) \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

is equivalent to

\[
\begin{align*}
\text{minimize (over } x, y_i) & \quad f_0(y_0) \\
\text{subject to} & \quad f_i(y_i) \leq 0, \quad i = 1, \ldots, m \\
y_i = A_ix + b_i, \quad i = 0, 1, \ldots, m
\end{align*}
\]

■ introducing slack variables for linear inequalities

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad a_i^Tx \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]

is equivalent to

\[
\begin{align*}
\text{minimize (over } x, s) & \quad f_0(x) \\
\text{subject to} & \quad a_i^Tx + s_i = b_i, \quad i = 1, \ldots, m \\
s_i \geq 0, \quad i = 1, \ldots m
\end{align*}
\]
**epigraph form:** standard form convex problem is equivalent to

\[
\begin{align*}
\text{minimize (over } x, t \text{)} & \quad t \\
\text{subject to} & \quad f_0(x) - t \leq 0 \\
& \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

**minimizing over some variables**

\[
\begin{align*}
\text{minimize} & \quad f_0(x_1, x_2) \\
\text{subject to} & \quad f_i(x_1) \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

is equivalent to

\[
\begin{align*}
\text{minimize} & \quad \tilde{f}_0(x_1) \\
\text{subject to} & \quad f_i(x_1) \leq 0, \quad i = 1, \ldots, m
\end{align*}
\]

where \( \tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2) \)
Quasiconvex optimization

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

with \( f_0 : \mathbb{R}^n \to \mathbb{R} \) quasiconvex, \( f_1, \ldots, f_m \) convex

can have locally optimal points that are not (globally) optimal
quasiconvex optimization via convex feasibility problems

\[ f_0(x) \leq t, \quad f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad Ax = b \quad (1) \]

- for fixed \( t \), a convex feasibility problem in \( x \)
- if feasible, we can conclude that \( t \geq p^* \); if infeasible, \( t \leq p^* \)

\[ \text{Bisection method for quasiconvex optimization} \]

given \( l \leq p^*, u \geq p^*, \) tolerance \( \epsilon > 0 \).

repeat

1. \( t := (l + u)/2 \).
2. Solve the convex feasibility problem (1).
3. if (1) is feasible, \( u := t \); else \( l := t \).

until \( u - l \leq \epsilon \).

requires exactly \( \lceil \log_2 ((u - l)/\epsilon) \rceil \) iterations (where \( u, l \) are initial values)
Linear program (LP)

minimize \quad c^T x + d
subject to \quad Gx \preceq h
\quad Ax = b

- convex problem with affine objective and constraint functions
- feasible set is a polyhedron
Chebyshev center of a polyhedron

Chebyshev center of

\[ P = \{ x \mid a_i^T x \leq b_i, \ i = 1, \ldots, m \} \]

is center of largest inscribed ball

\[ B = \{ x_c + u \mid \|u\|_2 \leq r \} \]

\[ a_i^T x \leq b_i \text{ for all } x \in B \text{ if and only if } \]

\[ \sup \{ a_i^T(x_c + u) \mid \|u\|_2 \leq r \} = a_i^T x_c + r \|a_i\|_2 \leq b_i \]

\[ \text{hence, } x_c, r \text{ can be determined by solving the LP} \]

\[
\begin{align*}
\text{maximize} & \quad r \\
\text{subject to} & \quad a_i^T x_c + r \|a_i\|_2 \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]
(Generalized) linear-fractional program

minimize \( f_0(x) \)
subject to \( Gx \preceq h \)
\( Ax = b \)

linear-fractional program

\[
f_0(x) = \frac{c^T x + d}{e^T x + f}, \quad \text{dom } f_0(x) = \{ x \mid e^T x + f > 0 \}
\]

- a quasiconvex optimization problem; can be solved by bisection
- also equivalent to the LP (variables \( y, z \))

minimize \( c^T y + dz \)
subject to \( G y \preceq h z \)
\( A y = b z \)
\( e^T y + f z = 1 \)
\( z \geq 0 \)
Quadratic program (QP)

\[
\begin{align*}
\text{minimize} & \quad (1/2)x^TPx + q^Tx + r \\
\text{subject to} & \quad Gx \preceq h \\
& \quad Ax = b
\end{align*}
\]

- \( P \in \mathbb{S}_+^n \), so objective is convex quadratic
- minimize a convex quadratic function over a polyhedron
Examples

least-squares

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

- analytical solution $x^* = A^\dagger b$ ($A^\dagger$ is pseudo-inverse)
- can add linear constraints, e.g., $l \leq x \leq u$

linear program with random cost

$$\begin{align*}
\text{minimize} & \quad \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E} c^T x + \gamma \text{var}(c^T x) \\
\text{subject to} & \quad Gx \leq h, \quad Ax = b
\end{align*}$$

- $c$ is random vector with mean $\bar{c}$ and covariance $\Sigma$
- hence, $c^T x$ is random variable with mean $\bar{c}^T x$ and variance $x^T \Sigma x$
- $\gamma > 0$ is risk aversion parameter; controls the trade-off between expected cost and variance (risk)
Quadratically constrained quadratic program (QCQP)

\[ \begin{align*}
\text{minimize} & \quad \frac{1}{2} x^T P_0 x + q_0^T x + r_0 \\
\text{subject to} & \quad \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*} \]

- \( P_i \in S^n_+ \); objective and constraints are convex quadratic
- if \( P_1, \ldots, P_m \in S^n_{++} \), feasible region is intersection of \( m \) ellipsoids and an affine set
Second-order cone programming

\[
\begin{align*}
\text{minimize} & \quad f^T x \\
\text{subject to} & \quad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\
& \quad Fx = g
\end{align*}
\]

\((A_i \in \mathbb{R}^{n_i \times n}, \ F \in \mathbb{R}^{p \times n})\)

- inequalities are called second-order cone (SOC) constraints:
  \[(A_i x + b_i, c_i^T x + d_i) \in \text{second-order cone in } \mathbb{R}^{n_i+1}\]

- for \(n_i = 0\), reduces to an LP; if \(c_i = 0\), reduces to a QCQP
- more general than QCQP and LP
Robust linear programming

the parameters in optimization problems are often uncertain, e.g., in an LP

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

there can be uncertainty in $c, a_i, b_i$

two common approaches to handling uncertainty (in $a_i$, for simplicity)

- **deterministic model:** constraints must hold for all $a_i \in \mathcal{E}_i$

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

- **stochastic model:** $a_i$ is random variable; constraints must hold with probability $\eta$

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad \text{Prob}(a_i^T x \leq b_i) \geq \eta, \quad i = 1, \ldots, m
\end{align*}
\]
deterministic approach via SOCP

- choose an ellipsoid as $\mathcal{E}_i$:

  $$
  \mathcal{E}_i = \{ \bar{a}_i + P_i u \mid \|u\|_2 \leq 1 \} \quad (\bar{a}_i \in \mathbb{R}^n, \ P_i \in \mathbb{R}^{n \times n})
  $$

  center is $\bar{a}_i$, semi-axes determined by singular values/vectors of $P_i$

- robust LP

  $$
  \begin{align*}
  \text{minimize} \quad & \quad c^T x \\
  \text{subject to} \quad & \quad a_i^T x \leq b_i \quad \forall a_i \in \mathcal{E}_i, \ i = 1, \ldots, m
  \end{align*}
  $$

  is equivalent to the SOCP

  $$
  \begin{align*}
  \text{minimize} \quad & \quad c^T x \\
  \text{subject to} \quad & \quad \bar{a}_i^T x + \|P_i^T x\|_2 \leq b_i, \quad i = 1, \ldots, m
  \end{align*}
  $$

  (follows from $\sup_{\|u\|_2 \leq 1} (\bar{a}_i + P_i u)^T x = \bar{a}_i^T x + \|P_i^T x\|_2$)
stochastic approach via SOCP

- Assume \( a_i \) is Gaussian with mean \( \bar{a}_i \), covariance \( \Sigma_i \) (\( a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i) \)).

- \( a_i^T x \) is Gaussian r.v. with mean \( \bar{a}_i^T x \), variance \( x^T \Sigma_i x \); hence

\[
\text{Prob}(a_i^T x \leq b_i) = \Phi\left( \frac{b_i - \bar{a}_i^T x}{\|\Sigma_i^{1/2} x\|_2} \right)
\]

where \( \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \) is CDF of \( \mathcal{N}(0,1) \).

- Robust LP

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

with \( \eta \geq 1/2 \), is equivalent to the SOCP

\[
\begin{align*}
&\text{minimize} & c^T x \\
&\text{subject to} & \bar{a}_i^T x + \Phi^{-1}(\eta) \|\Sigma_i^{1/2} x\|_2 \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]
Impact of reliability

\[ \{ x \mid \text{Prob}(a_i^T x \leq b_i) \geq \eta, \ i = 1, \ldots, m \} \]

\( \eta = 10\% \)

\( \eta = 50\% \)

\( \eta = 90\% \)
Generalized inequality constraints

custom problem with generalized inequality constraints

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

- \( f_0 : \mathbb{R}^n \to \mathbb{R} \) convex; \( f_i : \mathbb{R}^n \to \mathbb{R}^{k_i} \) \( K_i \)-convex w.r.t. proper cone \( K_i \)
- same properties as standard convex problem (convex feasible set, local optimum is global, etc.)

**conic form problem**: special case with affine objective and constraints

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Fx + g \preceq_{K} 0 \\
& \quad Ax = b
\end{align*}
\]

extends linear programming (\( K = \mathbb{R}^m_+ \)) to nonpolyhedral cones
Semidefinite program (SDP)

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0 \\
& \quad Ax = b
\end{align*}
\]

with \( F_i, G \in \mathbb{S}^k \)

- inequality constraint is called linear matrix inequality (LMI)
- includes problems with multiple LMI constraints: for example,

\[
x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \quad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0
\]

is equivalent to single LMI

\[
x_1 \begin{bmatrix} \hat{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \hat{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \hat{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0
\]
LP and SOCP as SDP

LP and equivalent SDP

LP: minimize $c^T x$
subject to $Ax \leq b$

SDP: minimize $c^T x$
subject to $\text{diag}(Ax - b) \leq 0$

(note different interpretation of generalized inequality $\leq$)

SOCP and equivalent SDP

SOCP: minimize $f^T x$
subject to $\|A_ix + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m$

SDP: minimize $f^T x$
subject to $\begin{bmatrix} (c_i^T x + d_i)I & A_ix + b_i \\ (A_ix + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m$
Eigenvalue minimization

\[ \text{minimize} \quad \lambda_{\text{max}}(A(x)) \]

where \( A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n \) (with given \( A_i \in S^k \))

equivalent SDP

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad A(x) \preceq tI
\end{align*}
\]

- variables \( x \in \mathbb{R}^n, t \in \mathbb{R} \)
- follows from

\[ \lambda_{\text{max}}(A) \leq t \iff A \preceq tI \]
Matrix norm minimization

\[ \text{minimize} \quad \|A(x)\|_2 = \left( \lambda_{\text{max}}(A(x)^T A(x)) \right)^{1/2} \]

where \( A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n \) (with given \( A_i \in S^{p \times q} \))

equivalent SDP

\[ \text{minimize} \quad t \]
\[ \text{subject to} \quad \begin{bmatrix} tI & A(x) \\ A(x)^T & tI \end{bmatrix} \succeq 0 \]

- variables \( x \in \mathbb{R}^n, t \in \mathbb{R} \)
- constraint follows from

\[ \|A\|_2 \leq t \quad \iff \quad A^T A \leq t^2 I, \quad t \geq 0 \]
\[ \iff \quad \begin{bmatrix} tI & A \\ A^T & tI \end{bmatrix} \succeq 0 \]
Duality
Outline

- Lagrange dual problem
- weak and strong duality
- optimality conditions
- perturbation and sensitivity analysis
- examples
- generalized inequalities
Lagrangian

standard form problem (not necessarily convex)

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad h_i(x) = 0, \quad i = 1, \ldots, p
\end{align*}
\]

variable \( x \in \mathbb{R}^n \), domain \( D \), optimal value \( p^* \)

Lagrangian: \( L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \), with \( \text{dom } L = D \times \mathbb{R}^m \times \mathbb{R}^p \),

\[
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
\]

- weighted sum of objective and constraint functions
- \( \lambda_i \) is Lagrange multiplier associated with \( f_i(x) \leq 0 \)
- \( \nu_i \) is Lagrange multiplier associated with \( h_i(x) = 0 \)
Lagrange dual function

Lagrange dual function: $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$,

$$g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)$$

$$= \inf_{x \in D} \left( f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

$g$ is concave, can be $-\infty$ for some $\lambda, \nu$

lower bound property: if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^*$

proof: if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^* \geq g(\lambda, \nu)$
Least-norm solution of linear equations

\[
\begin{align*}
\text{minimize} & \quad x^T x \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

dual function

- Lagrangian is \( L(x, \nu) = x^T x + \nu^T (Ax - b) \)
- to minimize \( L \) over \( x \), set gradient equal to zero:

\[
\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \implies \quad x = -(1/2)A^T \nu
\]

- plug in in \( L \) to obtain \( g \):

\[
g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4} \nu^T AA^T \nu - b^T \nu
\]

a concave function of \( \nu \)

lower bound property: \( p^* \geq -(1/4)\nu^T AA^T \nu - b^T \nu \) for all \( \nu \)
Standard form LP

\[
\begin{align*}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & Ax = b, \quad x \succeq 0
\end{align*}
\]

dual function

- Lagrangian is

\[
L(x, \lambda, \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x
\]

\[
= -b^T \nu + (c + A^T \nu - \lambda)^T x
\]

- \( L \) is linear in \( x \), hence

\[
g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} 
-b^T \nu & A^T \nu - \lambda + c = 0 \\
-\infty & \text{otherwise}
\end{cases}
\]

\( g \) is linear on affine domain \( \{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\} \), hence concave

**lower bound property:** \( p^* \geq -b^T \nu \) if \( A^T \nu + c \succeq 0 \)
Equality constrained norm minimization

\[
\begin{align*}
\text{minimize} & \quad \|x\| \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

dual function

\[
g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} 
\nu^T \nu & \|A^T \nu\|_* \leq 1 \\
-\infty & \text{otherwise}
\end{cases}
\]

where \(\|\nu\|_* = \sup_{\|u\| \leq 1} u^T \nu\) is dual norm of \(\| \cdot \|\).

proof: follows from \(\inf_x (\|x\| - y^T x) = 0\) if \(\|y\|_* \leq 1\), \(-\infty\) otherwise

- if \(\|y\|_* \leq 1\), then \(\|x\| - y^T x \geq 0\) for all \(x\), with equality if \(x = 0\)
- if \(\|y\|_* > 1\), choose \(x = tu\) where \(\|u\| \leq 1\), \(u^T y = \|y\|_* > 1\):

\[
\|x\| - y^T x = t(\|u\| - \|y\|_*) \rightarrow -\infty \quad \text{as } t \rightarrow \infty
\]

lower bound property: \(p^* \geq b^T \nu\) if \(\|A^T \nu\|_* \leq 1\)
Two-way partitioning

minimize $x^T W x$
subject to $x_i^2 = 1, \quad i = 1, \ldots, n$

- a nonconvex problem; feasible set contains $2^n$ discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; $W_{ij}$ is cost of assigning $i, j$ to the same set; $-W_{ij}$ is cost of assigning to different sets

dual function

$$g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) = \inf_x x^T (W + \text{diag}(\nu)) x - 1^T \nu$$

$$= \begin{cases} -1^T \nu & W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

lower bound property: $p^* \geq -1^T \nu$ if $W + \text{diag}(\nu) \succeq 0$

example: $\nu = -\lambda_{\min}(W) 1$ gives bound $p^* \geq n \lambda_{\min}(W)$
The dual problem

Lagrange dual problem

maximize \( g(\lambda, \nu) \)
subject to \( \lambda \succeq 0 \)

- finds best lower bound on \( p^* \), obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted \( d^* \)
- \( \lambda, \nu \) are dual feasible if \( \lambda \succeq 0, (\lambda, \nu) \in \text{dom} \ g \)
- often simplified by making implicit constraint \( (\lambda, \nu) \in \text{dom} \ g \) explicit

example: standard form LP and its dual (page 101)

minimize \( c^T x \)
subject to \( Ax = b \)
\( x \succeq 0 \)

maximize \( -b^T \nu \)
subject to \( A^T \nu + c \succeq 0 \)
Weak and strong duality

**weak duality:** \( d^* \leq p^* \)

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems
  for example, solving the SDP

\[
\begin{align*}
\text{maximize} & \quad -1^T \nu \\
\text{subject to} & \quad W + \text{diag}(\nu) \succeq 0
\end{align*}
\]

\( W \) gives a lower bound for the two-way partitioning problem on page 103

**strong duality:** \( d^* = p^* \)

- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called **constraint qualifications**
Slater’s constraint qualification

strong duality holds for a convex problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

if it is strictly feasible, i.e.,

\[
\exists x \in \text{int } D : \quad f_i(x) < 0, \quad i = 1, \ldots, m, \quad Ax = b
\]

- also guarantees that the dual optimum is attained (if \( p^* > -\infty \))
- can be sharpened: \( e.g. \), can replace \( \text{int } D \) with \( \text{relint } D \) (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, . . .
- there exist many other types of constraint qualifications
Feasibility problems

feasibility problem A in \( x \in \mathbb{R}^n \).

\[
f_i(x) < 0, \quad i = 1, \ldots, m, \quad h_i(x) = 0, \quad i = 1, \ldots, p
\]

feasibility problem B in \( \lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p \).

\[
\lambda \succeq 0, \quad \lambda \neq 0, \quad g(\lambda, \nu) \geq 0
\]

where \( g(\lambda, \nu) = \inf_x (\sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)) \)

- feasibility problem B is convex (\( g \) is concave), even if problem A is not
- A and B are always **weak alternatives**: at most one is feasible
  proof: assume \( \tilde{x} \) satisfies A, \( \lambda, \nu \) satisfy B

\[
0 \leq g(\lambda, \nu) \leq \sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) < 0
\]

- A and B are **strong alternatives** if exactly one of the two is feasible (can prove infeasibility of A by producing solution of B and vice-versa).
Inequality form LP

primal problem

\[
\text{minimize } \quad c^T x \\
\text{subject to } \quad Ax \preceq b
\]

dual function

\[
g(\lambda) = \inf_x \left( (c + A^T \lambda)^T x - b^T \lambda \right) = \begin{cases} 
-b^T \lambda & A^T \lambda + c = 0 \\
-\infty & \text{otherwise}
\end{cases}
\]

dual problem

\[
\text{maximize } \quad -b^T \lambda \\
\text{subject to } \quad A^T \lambda + c = 0, \quad \lambda \succeq 0
\]

- from Slater’s condition: \( p^* = d^* \) if \( A\tilde{x} \prec b \) for some \( \tilde{x} \)
- in fact, \( p^* = d^* \) except when primal and dual are infeasible
Quadratic program

primal problem (assume $P \in S_{++}^n$)

\[
\begin{align*}
\text{minimize} & \quad x^T P x \\
\text{subject to} & \quad A x \preceq b
\end{align*}
\]

dual function

\[
g(\lambda) = \inf_x \left( x^T P x + \lambda^T (A x - b) \right) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda
\]

dual problem

\[
\begin{align*}
\text{maximize} & \quad -(1/4) \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\
\text{subject to} & \quad \lambda \succeq 0
\end{align*}
\]

- from Slater’s condition: $p^* = d^*$ if $A \tilde{x} \prec b$ for some $\tilde{x}$
- in fact, $p^* = d^*$ always
A nonconvex problem with strong duality

minimize \( x^T Ax + 2b^T x \)
subject to \( x^T x \leq 1 \)

nonconvex if \( A \not\succeq 0 \)

dual function: \( g(\lambda) = \inf_x (x^T (A + \lambda I)x + 2b^T x - \lambda) \)

- unbounded below if \( A + \lambda I \not\succeq 0 \) or if \( A + \lambda I \succeq 0 \) and \( b \not\in \mathcal{R}(A + \lambda I) \)
- minimized by \( x = -(A + \lambda I)^\dagger b \) otherwise: \( g(\lambda) = -b^T (A + \lambda I)^\dagger b - \lambda \)

dual problem and equivalent SDP:

maximize \( -b^T (A + \lambda I)^\dagger b - \lambda \)
subject to \( A + \lambda I \succeq 0 \)
\( b \in \mathcal{R}(A + \lambda I) \)

maximize \( -t - \lambda \)
subject to \( \begin{bmatrix} A + \lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \)

strong duality although primal problem is not convex (more later)
Complementary slackness

Assume strong duality holds, $x^*$ is primal optimal, $(\lambda^*, \nu^*)$ is dual optimal

$$f_0(x^*) = g(\lambda^*, \nu^*) = \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right)$$

$$\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*)$$

$$\leq f_0(x^*)$$

hence, the two inequalities hold with equality

- $x^*$ minimizes $L(x, \lambda^*, \nu^*)$
- $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$\lambda_i^* > 0 \implies f_i(x^*) = 0, \quad f_i(x^*) < 0 \implies \lambda_i^* = 0$$
Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable $f_i$, $h_i$):

1. **Primal feasibility**: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$

2. **Dual feasibility**: $\lambda \succeq 0$

3. **Complementary slackness**: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$

4. **Gradient of Lagrangian with respect to $x$ vanishes** (first order condition):

   $$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

If strong duality holds and $x$, $\lambda$, $\nu$ are optimal, then they must satisfy the KKT conditions.
KKT conditions for convex problem

If \( \tilde{x}, \tilde{\lambda}, \tilde{\nu} \) satisfy KKT for a convex problem, then they are optimal:

- from complementary slackness: \( f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \)
- from 4th condition (and convexity): \( g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \)

hence, \( f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu}) \)

If Slater’s condition is satisfied, \( x \) is optimal if and only if there exist \( \lambda, \nu \) that satisfy KKT conditions

- recall that Slater implies strong duality, and dual optimum is attained
- generalizes optimality condition \( \nabla f_0(x) = 0 \) for unconstrained problem

Summary:

- When strong duality holds, the KKT conditions are necessary conditions for optimality
- If the problem is convex, they are also sufficient
example: water-filling (assume $\alpha_i > 0$)

$$\begin{align*}
\text{minimize} & \quad - \sum_{i=1}^{n} \log(x_i + \alpha_i) \\
\text{subject to} & \quad x \succeq 0, \quad 1^T x = 1
\end{align*}$$

$x$ is optimal iff $x \succeq 0$, $1^T x = 1$, and there exist $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}$ such that

$$\lambda \succeq 0, \quad \lambda_i x_i = 0, \quad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine $\nu$ from $1^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$

interpretation

- $n$ patches; level of patch $i$ is at height $\alpha_i$
- flood area with unit amount of water
- resulting level is $1/\nu^*$
Duality and problem reformulations

- equivalent formulations of a problem can lead to very different duals
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting

**common reformulations**

- introduce new variables and equality constraints
- make explicit constraints implicit or vice-versa
- transform objective or constraint functions

\[ \text{e.g., replace } f_0(x) \text{ by } \phi(f_0(x)) \text{ with } \phi \text{ convex, increasing} \]
Introducing new variables and equality constraints

\[
\begin{align*}
\text{minimize} & \quad f_0(Ax + b) \\
\text{dual function is constant:} & \quad g = \inf_x L(x) = \inf_x f_0(Ax + b) = p^* \\
\text{we have strong duality, but dual is quite useless}
\end{align*}
\]

reformulated problem and its dual

\[
\begin{align*}
\text{minimize} & \quad f_0(y) & \text{maximize} & \quad b^T \nu - f_0^*(\nu) \\
\text{subject to} & \quad Ax + b - y = 0 & \text{subject to} & \quad A^T \nu = 0
\end{align*}
\]

dual function follows from

\[
g(\nu) = \inf_{x, y} (f_0(y) - \nu^T y + \nu^T Ax + b^T \nu) = \begin{cases} 
-f_0^*(\nu) + b^T \nu & A^T \nu = 0 \\
-\infty & \text{otherwise}
\end{cases}
\]
norm approximation problem: minimize $\|Ax - b\|$

minimize $\|y\|$
subject to $y = Ax - b$

can look up conjugate of $\| \cdot \|$, or derive dual directly

$$g(\nu) = \inf_{x,y} (\|y\| + \nu^T y - \nu^T Ax + b^T \nu)$$

$$= \begin{cases} 
  b^T \nu + \inf_y (\|y\| + \nu^T y) & A^T \nu = 0 \\
  -\infty & \text{otherwise}
\end{cases}$$

$$= \begin{cases} 
  b^T \nu & A^T \nu = 0, \quad \|\nu\|_* \leq 1 \\
  -\infty & \text{otherwise}
\end{cases}$$

(see page 100)

dual of norm approximation problem

maximize $b^T \nu$
subject to $A^T \nu = 0, \quad \|\nu\|_* \leq 1$
Implicit constraints

**LP with box constraints:** primal and dual problem

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad -1 \leq x \leq 1
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad -b^T \nu - 1^T \lambda_1 - 1^T \lambda_2 \\
\text{subject to} & \quad c + A^T \nu + \lambda_1 - \lambda_2 = 0 \\
& \quad \lambda_1 \geq 0, \quad \lambda_2 \geq 0
\end{align*}
\]

**reformulation with box constraints made implicit**

\[
\begin{align*}
\text{minimize} & \quad f_0(x) = \begin{cases} 
  c^T x & -1 \leq x \leq 1 \\
  \infty & \text{otherwise}
\end{cases} \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

dual function

\[
g(\nu) = \inf_{-1 \leq x \leq 1} (c^T x + \nu^T (Ax - b)) = -b^T \nu - \|A^T \nu + c\|_1
\]

**dual problem:** maximize \(-b^T \nu - \|A^T \nu + c\|_1\)
Problems with generalized inequalities

minimize \( f_0(x) \)
subject to \( f_i(x) \preceq K_0 \), \( i = 1, \ldots, m \)
\( h_i(x) = 0, \quad i = 1, \ldots, p \)

\( \preceq K_i \) is generalized inequality on \( \mathbb{R}^{k_i} \)

definitions are parallel to scalar case:

- Lagrange multiplier for \( f_i(x) \preceq K_i 0 \) is vector \( \lambda_i \in \mathbb{R}^{k_i} \)
- Lagrangian \( L : \mathbb{R}^n \times \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R} \), is defined as

\[
L(x, \lambda_1, \cdots, \lambda_m, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
\]

- dual function \( g : \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m} \times \mathbb{R}^p \to \mathbb{R} \), is defined as

\[
g(\lambda_1, \ldots, \lambda_m, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda_1, \cdots, \lambda_m, \nu)
\]
**lower bound property:** if $\lambda_i \succeq_{K_i^*} 0$, then $g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^*$

**proof:** if $\tilde{x}$ is feasible and $\lambda \succeq_{K_i^*} 0$, then

$$f_0(\tilde{x}) \geq f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x})$$

$$\geq \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu)$$

$$= g(\lambda_1, \ldots, \lambda_m, \nu)$$

minimizing over all feasible $\tilde{x}$ gives $p^* \geq g(\lambda_1, \ldots, \lambda_m, \nu)$

**dual problem**

maximize $g(\lambda_1, \ldots, \lambda_m, \nu)$

subject to $\lambda_i \succeq_{K_i^*} 0$, $i = 1, \ldots, m$

- **weak duality:** $p^* \geq d^*$ always
- **strong duality:** $p^* = d^*$ for convex problem with constraint qualification (for example, Slater’s: primal problem is strictly feasible)
Semidefinite program

primal SDP \((F_i, G \in S^k)\)

- minimize \(c^T x\)
- subject to \(x_1 F_1 + \cdots + x_n F_n \preceq G\)

- Lagrange multiplier is matrix \(Z \in S^k\)
- Lagrangian \(L(x, Z) = c^T x + \text{Tr}(Z(x_1 F_1 + \cdots + x_n F_n - G))\)
- dual function

\[
g(Z) = \inf_x L(x, Z) = \begin{cases} 
-\text{Tr}(GZ) & \text{Tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\
-\infty & \text{otherwise}
\end{cases}
\]

dual SDP

- maximize \(-\text{Tr}(GZ)\)
- subject to \(Z \succeq 0, \quad \text{Tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n\)

\(p^* = d^*\) if primal SDP is strictly feasible \((\exists x \text{ with } x_1 F_1 + \cdots + x_n F_n < G)\)
Let’s consider the following Second Order Cone Program (SOCP):

\[
\begin{align*}
\text{minimize} & \quad f^T x \\
\text{subject to} & \quad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m,
\end{align*}
\]

with variable \(x \in \mathbb{R}^n\). Let’s show that the dual can be expressed as

\[
\begin{align*}
\text{maximize} & \quad \sum_{i=1}^m (b_i^T u_i + d_i v_i) \\
\text{subject to} & \quad \sum_{i=1}^m (A_i^T u_i + c_i v_i) + f = 0 \\
& \quad \|u_i\|_2 \leq v_i, \quad i = 1, \ldots, m,
\end{align*}
\]

with variables \(u_i \in \mathbb{R}^{n_i}, v_i \in \mathbb{R}, i = 1, \ldots, m\) and problem data given by \(f \in \mathbb{R}^n\), \(A_i \in \mathbb{R}^{n_i \times n}, b_i \in \mathbb{R}^{n_i}, c_i \in \mathbb{R}\) and \(d_i \in \mathbb{R}\).
We can derive the dual in the following two ways:

1. Introduce new variables $y_i \in \mathbb{R}^{n_i}$ and $t_i \in \mathbb{R}$ and equalities $y_i = A_i x + b_i, t_i = c_i^T x + d_i$, and derive the Lagrange dual.

2. Start from the conic formulation of the SOCP and use the conic dual. Use the fact that the second-order cone is self-dual:

$$ t \geq \|x\| \iff tv + x^T y \geq 0, \text{ for all } v, y \text{ such that } v \geq \|y\| $$

The condition $x^T y \leq tv$ is a simple Cauchy-Schwarz inequality.
We introduce new variables, and write the problem as

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad \|y_i\|_2 \leq t_i, \quad i = 1, \ldots, m \\
& \quad y_i = A_i x + b_i, \quad t_i = c_i^T x + d_i, \quad i = 1, \ldots, m
\end{align*}
\]

The Lagrangian is

\[
L(x, y, t, \lambda, \nu, \mu) = c^T x + \sum_{i=1}^{m} \lambda_i (\|y_i\|_2 - t_i) + \sum_{i=1}^{m} \nu_i^T (y_i - A_i x - b_i) + \sum_{i=1}^{m} \mu_i (t_i - c_i^T x - d_i)
\]

\[
= (c - \sum_{i=1}^{m} A_i^T \nu_i - \sum_{i=1}^{m} \mu_i c_i)^T x + \sum_{i=1}^{m} (\lambda_i \|y_i\|_2 + \nu_i^T y_i) + \sum_{i=1}^{m} (-\lambda_i + \mu_i) t_i
\]

\[
- \sum_{i=1}^{n} (b_i^T \nu_i + d_i \mu_i).
\]
The minimum over $x$ is bounded below if and only if

$$\sum_{i=1}^{m}(A_i^T \nu_i + \mu_i c_i) = c.$$ 

To minimize over $y_i$, we note that

$$\inf_{y_i}(\lambda_i \|y_i\|_2 + \nu_i^T y_i) = \begin{cases} 0 & \|\nu_i\|_2 \leq \lambda_i \\ -\infty & \text{otherwise.} \end{cases}$$

The minimum over $t_i$ is bounded below if and only if $\lambda_i = \mu_i$. 
The Lagrange dual function is

\[ g(\lambda, \nu, \mu) = \begin{cases} 
- \sum_{i=1}^{n} (b_i^T \nu_i + d_i \mu_i) & \text{if } \sum_{i=1}^{m} (A_i^T \nu_i + \mu_i c_i) = c, \\
\|\nu_i\|_2 \leq \lambda_i, & \mu = \lambda \\
-\infty & \text{otherwise}
\end{cases} \]

which leads to the dual problem

\[
\text{maximize } - \sum_{i=1}^{n} (b_i^T \nu_i + d_i \lambda_i) \\
\text{subject to } \sum_{i=1}^{m} (A_i^T \nu_i + \lambda_i c_i) = c \\
\|\nu_i\|_2 \leq \lambda_i, \quad i = 1, \ldots, m.
\]

which is again an SOCP
We can also express the SOCP as a **conic form** problem

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad -(c_i^T x + d_i, A_i x + b_i) \preceq_{K_i} 0, \quad i = 1, \ldots, m.
\end{align*}
\]

The Lagrangian is given by:

\[
L(x, u_i, v_i) = c^T x - \sum_i (A_i x + b_i)^T u_i - \sum_i (c_i^T x + d_i) v_i
\]

\[
= (c - \sum_i (A_i^T u_i + c_i v_i))^T x - \sum_i (b_i^T u_i + d_i v_i)
\]

for \((v_i, u_i) \succeq_{K_i} 0\) (which is also \(v_i \geq \|u_i\|\))
Duality: SOCP

With

\[ L(x, u_i, v_i) = \left( c - \sum_i (A_i^T u_i + c_i v_i) \right)^T x - \sum_i (b_i^T u_i + d_i v_i) \]

the dual function is given by:

\[ g(\lambda, \nu, \mu) = \begin{cases} 
- \sum_{i=1}^n (b_i^T \nu_i + d_i \mu_i) & \text{if } \sum_{i=1}^m (A_i^T \nu_i + \mu_i c_i) = c, \\
-\infty & \text{otherwise}
\end{cases} \]

The conic dual is then:

\[
\begin{align*}
\text{maximize} & \quad - \sum_{i=1}^n (b_i^T u_i + d_i v_i) \\
\text{subject to} & \quad \sum_{i=1}^m (A_i^T u_i + v_i c_i) = c \\
& \quad (v_i, u_i) \succeq K_i^0, \quad i = 1, \ldots, m.
\end{align*}
\]
References


