Convex Optimization M2

Lecture 2
Convex Optimization Problems
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

\( f : \mathbb{R}^n \rightarrow \mathbb{R} \) 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 $\mathbb{R}_{++}$
affine functions are convex and concave; all norms are convex

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

- affine function $f(x) = a^T x + b$
- norms: $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$ 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

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

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

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

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

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

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**

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

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

(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

\( f \) is **differentiable** if \( \text{dom} f \) 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: \( f(x) = (1/2)x^T P x + q^T x + r \) (with \( P \in S^n \))

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

convex if \( P \succeq 0 \)

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

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

convex (for any \( A \))

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

\[
\nabla^2 f(x, y) = \frac{2}{y^3} \begin{bmatrix} y & y \end{bmatrix} \begin{bmatrix} y & -x \\ -x & -x \end{bmatrix}^T \succeq 0
\]

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

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

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

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

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

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

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

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

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

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

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

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

\(f\) is convex if and only if \(\text{epi} f\) is a convex set
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: \( f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i) \) 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:

\[
f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}
\]
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 $C$: $S_C(x) = \sup_{y \in C} y^T x$ is convex
- distance to farthest point in a set $C$:

$$f(x) = \sup_{y \in C} \|x - y\|$$

- maximum eigenvalue of symmetric matrix: for $X \in S^n$,

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

composition of \( g : \mathbb{R}^n \rightarrow \mathbb{R} \) and \( h : \mathbb{R} \rightarrow \mathbb{R} \):

\[
f(x) = h(g(x))
\]

\( f \) 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
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

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

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

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

\( g \) is convex, hence Schur complement \( A - B C^{-1} B^T \succeq 0 \)

- distance to a set: \( \text{dist}(x, S) = \inf_{y \in S} \|x - y\| \) is convex if \( S \) is convex
the **perspective** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the function $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R},$

$$g(x, t) = tf \left( \frac{x}{t} \right), \quad \text{dom } g = \{(x, t) \mid x/t \in \text{dom } f, \ t > 0\}$$

g is convex if $f$ is convex

**examples**

- $f(x) = x^T x$ is convex; hence $g(x, t) = x^T x / t$ is convex for $t > 0$

- negative logarithm $f(x) = -\log x$ is convex; hence relative entropy $g(x, t) = t \log t - t \log x$ is convex on $\mathbb{R}^2_{++}$

- if $f$ is convex, then

$$g(x) = (c^T x + d)f \left( \frac{(Ax + b)}{(c^T x + d)} \right)$$

is convex on $\{x \mid c^T x + d > 0, \ (Ax + b)/(c^T x + d) \in \text{dom } f \}$
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 \( f(x) = -\log x \)

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

- strictly convex quadratic \( f(x) = (1/2)x^TQx \) with \( Q \in S^n_+ \)

\[
\begin{align*}
\ast f(y) & = \sup_x (y^T x - (1/2)x^TQx) \\
& = \frac{1}{2} y^T Q^{-1} y
\end{align*}
\]
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
modified Jensen inequality: for quasiconvex f

\[ 0 \leq \theta \leq 1 \implies f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\} \]
Log-concave and log-convex functions

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

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

\[ f \text{ is log-convex if } \log f \text{ is convex} \]

- 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

\[f(x) = \int g(x + y)p(y) dy, \quad g(u) = \begin{cases} 1 & u \in C \\ 0 & u \not\in C, \end{cases}\]

$p$ is pdf of $y$
example: yield function

\[ Y(x) = \text{Prob}(x + w \in S) \]

- \( x \in \mathbb{R}^n \): nominal parameter values for product
- \( w \in \mathbb{R}^n \): random variations of parameters in manufactured product
- \( S \): set of acceptable values

If \( S \) is convex and \( w \) has a log-concave pdf, then

- \( Y \) is log-concave
- yield regions \( \{x \mid Y(x) \geq \alpha\} \) are convex
- Not necessarily tractable though
Convexity with respect to generalized inequalities

\[ f : \mathbb{R}^n \rightarrow \mathbb{R}^m \text{ is } K\text{-convex if } \text{dom } f \text{ is convex and } \]
\[ f(\theta x + (1 - \theta)y) \preceq_K \theta f(x) + (1 - \theta)f(y) \]

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

example \( f : S^m \rightarrow S^m \), \( f(X) = X^2 \) is \( S^m_+ \)-convex

proof: for fixed \( z \in \mathbb{R}^m \), \( z^T X^2 z = \|Xz\|^2 \) is convex in \( X \), i.e.,
\[ z^T(\theta X + (1 - \theta)Y)^2 z \leq \theta z^T X^2 z + (1 - \theta)z^T Y^2 z \]

for \( X, Y \in S^m \), \( 0 \leq \theta \leq 1 \)

therefore \( (\theta X + (1 - \theta)Y)^2 \preceq \theta X^2 + (1 - \theta)Y^2 \)
Convex Optimization Problems
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*}

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

**optimal value:**

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

\begin{itemize}
\item $p^* = \infty$ if problem is infeasible (no $x$ satisfies the constraints)
\item $p^* = -\infty$ if problem is unbounded below
\end{itemize}
Optimal and locally optimal points

\( x \) is \textbf{feasible} if \( x \in \text{dom} f_0 \) and it satisfies the constraints

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

\( x \) is \textbf{locally optimal} if there is an \( R > 0 \) such that \( x \) is optimal for

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

\textbf{examples} (with \( n = 1, \ m = p = 0 \))

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

the standard form optimization problem has an implicit constraint

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

- we call \( \mathcal{D} \) 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:

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

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

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

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

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

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

standard form convex optimization problem

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

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

often written as

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

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

\[
\begin{align*}
\text{minimize} & \quad f_0(\mathbf{x}) = x_1^2 + x_2^2 \\
\text{subject to} & \quad f_1(\mathbf{x}) = x_1 / (1 + x_2^2) \leq 0 \\
& \quad h_1(\mathbf{x}) = (x_1 + x_2)^2 = 0
\end{align*}
\]

- \(f_0\) 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

\[
\begin{align*}
\text{minimize} & \quad x_1^2 + x_2^2 \\
\text{subject to} & \quad x_1 \leq 0 \\
& \quad x_1 + x_2 = 0
\end{align*}
\]
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**

  \[
  \text{minimize } f_0(x) \quad \text{subject to } \quad Ax = b
  \]

  $x$ 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**

  \[
  \text{minimize } f_0(x) \quad \text{subject to } \quad x \succeq 0
  \]

  $x$ is optimal if and only if

  \[ x \in \text{dom } f_0, \quad x \succeq 0, \quad \begin{cases} 
  \nabla f_0(x)_i \geq 0 \quad x_i = 0 \\
  \nabla f_0(x)_i = 0 \quad x_i > 0 
  \end{cases} \]
Equivalent convex problems

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

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

minimizing over some variables

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

is equivalent to

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

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

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

with \( f_0 : \mathbb{R}^n \rightarrow \mathbb{R} \) 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, \( t \leq p^* \)

_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)

\[
\begin{align*}
\text{minimize} & \quad c^T x + d \\
\text{subject to} & \quad Gx \preceq h \\
& \quad Ax = b
\end{align*}
\]

- convex problem with affine objective and constraint functions
- feasible set is a polyhedron
Examples

diet problem: choose quantities $x_1, \ldots, x_n$ of $n$ foods

- one unit of food $j$ costs $c_j$, contains amount $a_{ij}$ of nutrient $i$
- healthy diet requires nutrient $i$ in quantity at least $b_i$

to find cheapest healthy diet,

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

piecewise-linear minimization

$$\text{minimize} \quad \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad a_i^T x + b_i \leq t, \quad i = 1, \ldots, m
\end{align*}$$
Chebyshev center of a polyhedron

Chebyshev center of

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

is center of largest inscribed ball

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

- \( a_i^T x \leq b_i \) for all \( x \in \mathcal{B} \) if and only if

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

- hence, \( x_c, r \) 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 \( c^T y + d z \)
subject to \( G y \preceq h z \)
\( A y = b z \)
\( e^T y + f z = 1 \)
\( z \geq 0 \)
**Quadratic program (QP)**

\[
\text{minimize} \quad \frac{1}{2} x^T P x + q^T x + r \\
\text{subject to} \quad Gx \preceq h \\
\quad \quad \quad \quad \quad Ax = b
\]

- \( P \in S_+^n \), so objective is convex quadratic
- minimize a convex quadratic function over a polyhedron
Examples

**least-squares**

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

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

**linear program with random cost**

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

- \( c \) is random vector with mean \( \bar{c} \) and covariance \( \Sigma \)
- hence, \( c^T x \) is random variable with mean \( c^T x \) 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)

minimize \( (1/2)x^T P_0 x + q_0^T x + r_0 \)
subject to \( (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \)
\( Ax = b \)

- \( P_i \in S^n_+; \) 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

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

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

- 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 \( a_i \in \mathcal{E}_i \)

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

- 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 $\mathcal{E}_i$:

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

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

- robust LP

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

is equivalent to the SOCP

$$\text{minimize } c^T x$$
subject to $a_i^T x + \|P_i^T x\|_2 \leq b_i, \quad i = 1, \ldots, m$

(follows from $\sup_{\|u\|_2 \leq 1}(a_i + P_i u)^T x = a_i^T x + \|P_i^T x\|_2$)
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

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

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

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

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

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

monomial function

\[ f(x) = cx_1^{a_1}x_2^{a_2} \cdots x_n^{a_n}, \quad \text{dom } f = \mathbb{R}_{++}^n \]

with \( c > 0 \); exponent \( \alpha_i \) can be any real number

posynomial function: sum of monomials

\[ f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}, \quad \text{dom } f = \mathbb{R}_{++}^n \]

gometric program (GP)

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

with \( f_i \) posynomial, \( h_i \) monomial
Geometric program in convex form

change variables to \( y_i = \log x_i \), and take logarithm of cost, constraints

- monomial \( f(x) = cx_1^{a_1} \cdots x_n^{a_n} \) transforms to
  \[
  \log f(e^{y_1}, \ldots, e^{y_n}) = a^T y + b \quad (b = \log c)
  \]

- posynomial \( f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}} \) transforms to
  \[
  \log f(e^{y_1}, \ldots, e^{y_n}) = \log \left( \sum_{k=1}^{K} e^{a_{ik}^T y + b_k} \right) \quad (b_k = \log c_k)
  \]

- geometric program transforms to convex problem

  minimize \( \log \left( \sum_{k=1}^{K} \exp\left( a_{0k}^T y + b_{0k} \right) \right) \)

  subject to \( \log \left( \sum_{k=1}^{K} \exp\left( a_{ik}^T y + b_{ik} \right) \right) \leq 0, \quad i = 1, \ldots, m \)

  \( Gy + d = 0 \)
Minimizing spectral radius of nonnegative matrix

Perron-Frobenius eigenvalue $\lambda_{pf}(A)$

- exists for (elementwise) positive $A \in \mathbb{R}^{n \times n}$
- a real, positive eigenvalue of $A$, equal to spectral radius $\max_i |\lambda_i(A)|$
- determines asymptotic growth (decay) rate of $A^k$: $A^k \sim \lambda_{pf}^k$ as $k \to \infty$
- alternative characterization: $\lambda_{pf}(A) = \inf\{\lambda \mid AV \preceq \lambda V \text{ for some } V \succ 0\}$

minimizing spectral radius of matrix of posynomials

- minimize $\lambda_{pf}(A(x))$, where the elements $A(x)_{ij}$ are posynomials of $x$
- equivalent geometric program:

$$\begin{align*}
\text{minimize} & \quad \lambda \\
\text{subject to} & \quad \sum_{j=1}^{n} A(x)_{ij} v_j / (\lambda v_i) \leq 1, \quad i = 1, \ldots, n
\end{align*}$$

variables $\lambda, v, x$
Generalized inequality constraints

convex 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 \( F_i, G \in S^k \)

- 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 $Ax \preceq b$

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

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

SOCP and equivalent SDP

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

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

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

where \(A(x) = A_0 + x_1A_1 + \cdots + x_nA_n\) (with given \(A_i \in S^k\))

equivalent SDP

\[
\text{minimize} \quad t \\
\text{subject to} \quad A(x) \preceq tl
\]

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

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

minimize $\|A(x)\|_2 = (\lambda_{\text{max}}(A(x)^TA(x)))^{1/2}$

where $A(x) = A_0 + x_1A_1 + \cdots + x_nA_n$ (with given $A_i \in S^{p \times q}$)

equivalent SDP

minimize $t$

subject to $\begin{bmatrix} tI & A(x) \\ A(x)^T & tl \end{bmatrix} \succeq 0$

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

$\|A\|_2 \leq t \iff A^TA \preceq t^2I$, $t \geq 0$

$\iff \begin{bmatrix} tl & A \\ A^T & tl \end{bmatrix} \succeq 0$