Convex Optimization Basics

Yu-Xiang Wang
CS292A

(Based on Ryan Tibshirani’s 10-725)
“Convex calculus” makes it easy to check convexity. Tools:

- Definitions of convex sets and functions, classic examples
- Key properties (e.g., first- and second-order characterizations for functions)
- Operations that preserve convexity (e.g., affine composition)

E.g., is \( \max \left\{ \log(1 + e^{a^T x}), \|Ax + b\|_1^5 \right\} \) convex?
Today:

- Optimization terminology
- Properties and first-order optimality
- Equivalent transformations
- Hierarchies of Canonical Problems
- Many examples!
Optimization terminology

Reminder: a convex optimization problem (or program) is

\[
\begin{align*}
\min_{x \in D} & \quad f(x) \\
\text{subject to} & \quad g_i(x) \leq 0, \ i = 1, \ldots , m \\
& \quad Ax = b
\end{align*}
\]

where \( f \) and \( g_i, i = 1, \ldots m \) are all convex, and the optimization domain is \( D = \text{dom}(f) \cap \bigcap_{i=1}^{m} \text{dom}(g_i) \) (often we do not write \( D \))

- \( f \) is called criterion or objective function
- \( g_i \) is called inequality constraint function
- If \( x \in D, g_i(x) \leq 0, i = 1, \ldots m, \) and \( Ax = b \) then \( x \) is called a feasible point
- The minimum of \( f(x) \) over all feasible points \( x \) is called the optimal value, written \( f^* \)
• If \( x \) is feasible and \( f(x) = f^* \), then \( x \) is called optimal; also called a solution, or a minimizer

• If \( x \) is feasible and \( f(x) \leq f^* + \epsilon \), then \( x \) is called \( \epsilon \)-suboptimal

• If \( x \) is feasible and \( g_i(x) = 0 \), then we say \( g_i \) is active at \( x \)

• Convex minimization can be reposed as concave maximization

\[
\begin{align*}
\min_x & \quad f(x) \\
\text{subject to} & \quad g_i(x) \leq 0, \quad i = 1, \ldots, m \\
\end{align*}
\]

\[
\begin{align*}
\max_x & \quad -f(x) \\
\text{subject to} & \quad g_i(x) \leq 0, \quad i = 1, \ldots, m \\
\end{align*}
\]

\( A x = b \)

Both are called convex optimization problems

\[\text{\footnotesize{1Note: a convex optimization problem need not have solutions, i.e., need not attain its minimum, but we will not be careful about this}}\]
Let $X_{\text{opt}}$ be the set of all solutions of convex problem, written

$$X_{\text{opt}} = \arg\min f(x)$$

subject to

$$g_i(x) \leq 0, \ i = 1, \ldots, m$$

$$Ax = b$$

Key property: $X_{\text{opt}}$ is a convex set

Proof: use definitions. If $x, y$ are solutions, then for $0 \leq t \leq 1$,

- $g_i(tx + (1 - t)y) \leq tg_i(x) + (1 - t)g_i(y) \leq 0$
- $A(tx + (1 - t)y) = tAx + (1 - t)Ay = b$
- $f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) = f^*$

Therefore $tx + (1 - t)y$ is also a solution

Another key property: if $f$ is strictly convex, then the solution is unique, i.e., $X_{\text{opt}}$ contains one element
**Example: lasso**

Given \( y \in \mathbb{R}^n, \; X \in \mathbb{R}^{n \times p} \), consider the lasso problem:

\[
\min_{\beta} \|y - X\beta\|_2^2
\]

subject to \( \|\beta\|_1 \leq s \)

Is this convex? What is the criterion function? The inequality and equality constraints? Feasible set? Is the solution unique, when:

- \( n \geq p \) and \( X \) has full column rank?
- \( p > n \) ("high-dimensional" case)?

How do our answers change if we changed criterion to **Huber loss**:

\[
\sum_{i=1}^{n} \rho(y_i - x_i^T \beta), \quad \rho(z) = \begin{cases} 
\frac{1}{2}z^2 & |z| \leq \delta \\
\delta|z| - \frac{1}{2}\delta^2 & \text{else}
\end{cases}
\]
Example: support vector machines

Given $y \in \{-1, 1\}^n$, $X \in \mathbb{R}^{n \times p}$ with rows $x_1, \ldots, x_n$, consider the support vector machine or SVM problem:

$$\min_{\beta, \beta_0, \xi} \quad \frac{1}{2} \|\beta\|^2_2 + C \sum_{i=1}^n \xi_i$$

subject to $\xi_i \geq 0$, $i = 1, \ldots, n$

$$y_i (x_i^T \beta + \beta_0) \geq 1 - \xi_i, \quad i = 1, \ldots, n$$

Is this convex? What is the criterion, constraints, feasible set? Is the solution $(\beta, \beta_0, \xi)$ unique? What if changed the criterion to

$$\frac{1}{2} \|\beta\|^2_2 + \frac{1}{2} \beta_0^2 + C \sum_{i=1}^n \xi_i^{1.01}?$$

For original criterion, what about $\beta$ component, at the solution?
Local minima are global minima

For a convex problem, a feasible point $x$ is called \textit{locally optimal} is there is some $R > 0$ such that

$$f(x) \leq f(y) \quad \text{for all feasible } y \text{ such that } \|x - y\|_2 \leq R$$

Reminder: for convex optimization problems, \textit{local optima are global optima}

Proof simply follows from definitions
Rewriting constraints

The optimization problem

\[
\min_x f(x) \\
\text{subject to } g_i(x) \leq 0, \ i = 1, \ldots m \\
Ax = b
\]

can be rewritten as

\[
\min_x f(x) \quad \text{subject to } x \in C
\]

where \( C = \{x : g_i(x) \leq 0, \ i = 1, \ldots m, \ Ax = b\} \), the feasible set. Hence the latter formulation is completely general.

With \( I_C \) the indicator of \( C \), we can write this in unconstrained form

\[
\min_x f(x) + I_C(x)
\]
First-order optimality condition

For a convex problem

$$\min_x f(x) \text{ subject to } x \in C$$

and differentiable $f$, a feasible point $x$ is optimal if and only if

$$\nabla f(x)^T (y - x) \geq 0 \text{ for all } y \in C$$

This is called the first-order condition for optimality

In words: all feasible directions from $x$ are aligned with gradient $\nabla f(x)$

Important special case: if $C = \mathbb{R}^n$ (unconstrained optimization), then optimality condition reduces to familiar $\nabla f(x) = 0$
Example: quadratic minimization

Consider minimizing the quadratic function

\[ f(x) = \frac{1}{2} x^T Q x + b^T x + c \]

where \( Q \succeq 0 \). The first-order condition says that solution satisfies

\[ \nabla f(x) = Qx + b = 0 \]

- if \( Q \succ 0 \), then there is a unique solution \( x = -Q^{-1}b \)
- if \( Q \) is singular and \( b \notin \text{col}(Q) \), then there is no solution (i.e., \( \min_x f(x) = -\infty \))
- if \( Q \) is singular and \( b \in \text{col}(Q) \), then there are infinitely many solutions

\[ x = -Q^+ b + z, \quad z \in \text{null}(Q) \]

where \( Q^+ \) is the pseudoinverse of \( Q \).
Example: equality-constrained minimization

Consider the equality-constrained convex problem:

$$\min_x f(x) \quad \text{subject to} \quad Ax = b$$

with $f$ differentiable. Let’s prove Lagrange multiplier optimality condition

$$\nabla f(x) + A^T u = 0 \quad \text{for some } u$$

According to first-order optimality, solution $x$ satisfies $Ax = b$ and

$$\nabla f(x)^T (y - x) \geq 0 \quad \text{for all } y \text{ such that } Ay = b$$

This is equivalent to

$$\nabla f(x)^T v = 0 \quad \text{for all } v \in \text{null}(A)$$

Result follows because $\text{null}(A)^\perp = \text{row}(A)$
Example: projection onto a convex set

Consider projection onto convex set $C$:

$$\min_x \|a - x\|_2^2 \quad \text{subject to} \quad x \in C$$

First-order optimality condition says that the solution $x$ satisfies

$$\nabla f(x)^T(y - x) = (x - a)^T(y - x) \geq 0 \quad \text{for all} \quad y \in C$$

Equivalently, this says that

$$a - x \in \mathcal{N}_C(x)$$

where recall $\mathcal{N}_C(x)$ is the normal cone to $C$ at $x$
Partial optimization

Reminder: \( g(x) = \min_{y \in C} f(x, y) \) is convex in \( x \), provided that \( f \) is convex in \((x, y)\) and \( C \) is a convex set.

Therefore we can always partially optimize a convex problem and retain convexity.

E.g., if we decompose \( x = (x_1, x_2) \in \mathbb{R}^{n_1+n_2} \), then

\[
\begin{align*}
\min_{x_1,x_2} & \quad f(x_1, x_2) \\
\text{subject to} & \quad g_1(x_1) \leq 0, \quad g_2(x_2) \leq 0
\end{align*}
\]

\( \iff \)

\[
\begin{align*}
\min_{x_1} & \quad \tilde{f}(x_1) \\
\text{subject to} & \quad g_1(x_1) \leq 0
\end{align*}
\]

where \( \tilde{f}(x_1) = \min \{ f(x_1, x_2) : g_2(x_2) \leq 0 \} \). The right problem is convex if the left problem is
Example: hinge form of SVMs

Recall the SVM problem

$$\min_{\beta, \beta_0, \xi} \frac{1}{2}||\beta||_2^2 + C \sum_{i=1}^{n} \xi_i$$

subject to \( \xi_i \geq 0, \ y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i, \ i = 1, \ldots n \)

Rewrite the constraints as \( \xi_i \geq \max\{0, 1 - y_i(x_i^T \beta + \beta_0)\} \). Indeed we can argue that we have \( = \) at solution

Therefore plugging in for optimal \( \xi \) gives the hinge form of SVMs:

$$\min_{\beta, \beta_0} \frac{1}{2}||\beta||_2^2 + C \sum_{i=1}^{n} \left[1 - y_i(x_i^T \beta + \beta_0)\right]_+$$

where \( a_+ = \max\{0, a\} \) is called the hinge function
Transformations and change of variables

If $h : \mathbb{R} \to \mathbb{R}$ is a **monotone increasing transformation**, then

$$
\min_x f(x) \text{ subject to } x \in C \\
\iff \min_x h(f(x)) \text{ subject to } x \in C
$$

Similarly, inequality or equality constraints can be transformed and yield equivalent optimization problems. Can use this to reveal the “hidden convexity” of a problem

If $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one, and its image covers feasible set $C$, then we can **change variables** in an optimization problem:

$$
\min_x f(x) \text{ subject to } x \in C \\
\iff \min_y f(\phi(y)) \text{ subject to } \phi(y) \in C
$$
Example: geometric programming

A **monomial** is a function $f : \mathbb{R}^n_{++} \rightarrow \mathbb{R}$ of the form

$$f(x) = \gamma x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

for $\gamma > 0$, $a_1, \ldots, a_n \in \mathbb{R}$. A **posynomial** is a sum of monomials,

$$f(x) = \sum_{k=1}^{p} \gamma_k x_1^{a_{k1}} x_2^{a_{k2}} \cdots x_n^{a_{kn}}$$

A **geometric program** is of the form

$$\begin{align*}
\min_x & \quad f(x) \\
\text{subject to} & \quad g_i(x) \leq 1, \ i = 1, \ldots, m \\
& \quad h_j(x) = 1, \ j = 1, \ldots, r
\end{align*}$$

where $f, g_i, i = 1, \ldots, m$ are posynomials and $h_j, j = 1, \ldots, r$ are monomials. This is nonconvex.
Given \( f(x) = \gamma x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n} \), let \( y_i = \log x_i \) and rewrite this as

\[
\gamma (e^{y_1})^{a_1} (e^{y_2})^{a_2} \cdots (e^{y_n})^{a_n} = e^{a^T y + b}
\]

for \( b = \log \gamma \). Also, a posynomial can be written as \( \sum_{k=1}^{p} e^{a_k^T y + b_k} \).

With this variable substitution, and after taking logs, a geometric program is equivalent to

\[
\min_x \quad \log \left( \sum_{k=1}^{p_0} e^{a_{0k}^T y + b_{0k}} \right)
\]

subject to

\[
\log \left( \sum_{k=1}^{p_i} e^{a_{ik}^T y + b_{ik}} \right) \leq 0, \quad i = 1, \ldots, m
\]

\[
c_j^T y + d_j = 0, \quad j = 1, \ldots, r
\]

This is convex, recalling the convexity of soft max functions.
Several interesting problems are geometric programs, e.g., floor planning:

![Diagram of floor planning problem](image)

See Boyd et al. (2007), “A tutorial on geometric programming”, and also Chapter 8.8 of B & V book
Eliminating equality constraints

Important special case of change of variables: eliminating equality constraints. Given the problem

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

we can always express any feasible point as \( x = My + x_0 \), where \( Ax_0 = b \) and \( \text{col}(M) = \text{null}(A) \). Hence the above is equivalent to

\[
\begin{align*}
\min_y & \quad f(My + x_0) \\
\text{subject to} & \quad g_i(My + x_0) \leq 0, \ i = 1, \ldots, m
\end{align*}
\]

Note: this is fully general but not always a good idea (practically)
Introducing slack variables

Essentially opposite to eliminating equality constraints: introducing slack variables. Given the problem

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

we can transform the inequality constraints via

\[
\begin{align*}
\min_{x,s} & \quad f(x) \\
\text{subject to} & \quad s_i \geq 0, \ i = 1, \ldots, m \\
& \quad g_i(x) + s_i = 0, \ i = 1, \ldots, m \\
& \quad Ax = b
\end{align*}
\]

Note: this is no longer convex unless \(g_i, \ i = 1, \ldots, n\) are affine
Relaxing nonaffine equalities

Given an optimization problem

$$\min_x f(x) \text{ subject to } x \in C$$

we can always take an enlarged constraint set $\tilde{C} \supseteq C$ and consider

$$\min_x f(x) \text{ subject to } x \in \tilde{C}$$

This is called a relaxation and its optimal value is always smaller or equal to that of the original problem

Important special case: relaxing nonaffine equality constraints, i.e.,

$$h_j(x) = 0, \ j = 1, \ldots r$$

where $h_j, \ j = 1, \ldots r$ are convex but nonaffine, are replaced with

$$h_j(x) \leq 0, \ j = 1, \ldots r$$
Example: maximum utility problem

The maximum utility problem models investment/consumption:

$$\max_{x,b} \sum_{t=1}^{T} \alpha_t u(x_t)$$
subject to
$$b_{t+1} = b_t + f(b_t) - x_t, \ t = 1, \ldots, T$$
$$0 \leq x_t \leq b_t, \ t = 1, \ldots, T$$

Here $b_t$ is the budget and $x_t$ is the amount consumed at time $t$; $f$ is an investment return function, $u$ utility function, both concave and increasing.

Is this a convex problem? What if we replace equality constraints with inequalities:

$$b_{t+1} \leq b_t + f(b_t) - x_t, \ t = 1, \ldots, T?$$
Example: principal components analysis

Given $X \in \mathbb{R}^{n \times p}$, consider the low rank approximation problem:

$$\min_{R} \| X - R \|_F^2 \quad \text{subject to} \quad \text{rank}(R) = k$$

Here $\| A \|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} A_{ij}^2$, the entrywise squared $\ell_2$ norm, and $\text{rank}(A)$ denotes the rank of $A$

Also called principal components analysis or PCA problem. Given $X = U D V^T$, singular value decomposition or SVD, the solution is

$$R = U_k D_k V_k^T$$

where $U_k, V_k$ are the first $k$ columns of $U, V$ and $D_k$ is the first $k$ diagonal elements of $D$. I.e., $R$ is reconstruction of $X$ from its first $k$ principal components
The PCA problem is not convex. Let’s recast it. First rewrite as

\[
\min_{Z \in S^p} \|X - XZ\|_F^2 \quad \text{subject to} \quad \text{rank}(Z) = k, \ Z \text{ is a projection}
\]

\[\iff \quad \max_{Z \in S^p} \text{tr}(SZ) \quad \text{subject to} \quad \text{rank}(Z) = k, \ Z \text{ is a projection} \]

where \( S = X^T X \). Hence constraint set is the nonconvex set

\[ C = \left\{ Z \in S^p : \lambda_i(Z) \in \{0, 1\}, \ i = 1, \ldots p, \ \text{tr}(Z) = k \right\} \]

where \( \lambda_i(Z), \ i = 1, \ldots n \) are the eigenvalues of \( Z \). Solution in this formulation is

\[ Z = V_k V_k^T \]

where \( V_k \) gives first \( k \) columns of \( V \)
Now consider relaxing constraint set to $\mathcal{F}_k = \text{conv}(C)$, its convex hull. Note

$$
\mathcal{F}_k = \{Z \in \mathbb{S}^p : \lambda_i(Z) \in [0, 1], \ i = 1, \ldots, p, \ \text{tr}(Z) = k\}
$$

$$
= \{Z \in \mathbb{S}^p : 0 \preceq Z \preceq I, \ \text{tr}(Z) = k\}
$$

This set is called the Fantope of order $k$. It is convex. Hence, the linear maximization over the Fantope, namely

$$
\max_{Z \in \mathcal{F}_k} \text{tr}(SZ)
$$

is a convex problem. Remarkably, this is equivalent to the original nonconvex PCA problem (admits the same solution)!

(Famous result: Fan (1949), “On a theorem of Weyl concerning eigenvalues of linear transformations”)


Sparse PCA with Fantope Projection and Selection

- Having an optimization formulation allows us to add additional problem specific considerations.
- Suppose we want the recovered principle components to be sparse

\[
\max_{Z \in \mathcal{F}_k} \text{tr}(SZ) - \lambda \sum_{i,j} |Z_{i,j}| \quad \text{subject to} \quad \text{rank}(R) = k
\]

- This is the algorithm for the sparse PCA problem that achieves the minimax rate. \text{(Vu and Lei, NIPS 2013).}
Approximation Algorithm for MaxCut

• Given a graph with nodes and edges and edge weights. Find a subset \( S \) of the nodes such that the sum of the weights \( w_{ij} \) of the edges between \( S \) and its complement \( \bar{S} \) is maximizes.

• Let \( x_j = 1 \) if \( j \in S \) and \( x_j = -1 \) if \( j \in \bar{S} \).

\[
\max_x \quad \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (1 - x_i x_j)
\]

s.t. \( x_j \in \{-1, 1\}, j = 1, ..., n \)

• Goemans and Williamson algorithm:
  1. **convex relaxation**: get the optimal solution \( \hat{x} \)
  2. **Randomized rounding**: Sample \( v \) uniformly from the unit sphere in \( \mathbb{R}^n \).
  3. Output: \( \text{sign}(v^T \hat{x}) \)

• You get a 0.87856 approximation of an NP-complete problem.
Approximation Algorithm for MaxCut

Reformulation:

\[
\max_{Y \in \mathbb{R}^{n \times n}, x \in \mathbb{R}^n} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (1 - Y_{i,j}) \\
\text{s.t.} \quad Y_{i,i} = 1 \quad \forall j = 1, \ldots, n \\
Y = xx^T.
\]

The convex relaxation:

\[
\max_{Y \in \mathbb{R}^{n \times n}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (1 - Y_{i,j}) \\
\text{s.t.} \quad Y_{i,i} = 1 \quad \forall j = 1, \ldots, n \\
Y \succeq 0.
\]
Quick Summary

- Optimization terminology (e.g., criterion, constraints, feasible points, solutions)
- Properties and first-order optimality

- Equivalent transformations (e.g., partial optimization, change of variables, eliminating equality constraints)
Hierarchy of Canonical Optimizations

• Linear programs
• Quadratic programs
• Semidefinite programs
• Cone programs
A linear program or LP is an optimization problem of the form

$$\min_x c^T x$$
subject to

$$Dx \leq d$$
$$Ax = b$$

Observe that this is always a convex optimization problem.

- First introduced by Kantorovich in the late 1930s and Dantzig in the 1940s.
- Dantzig’s simplex algorithm gives a direct (noniterative) solver for LPs (later in the course we’ll see interior point methods).
- Fundamental problem in convex optimization. Many diverse applications, rich history.
Example: diet problem

Find cheapest combination of foods that satisfies some nutritional requirements (useful for graduate students!)

$$\begin{align*}
\min_x & \quad c^T x \\
\text{subject to } & \quad Dx \geq d \\
& \quad x \geq 0
\end{align*}$$

Interpretation:

- \(c_j\): per-unit cost of food \(j\)
- \(d_i\): minimum required intake of nutrient \(i\)
- \(D_{ij}\): content of nutrient \(i\) per unit of food \(j\)
- \(x_j\): units of food \(j\) in the diet
Example: transportation problem

Ship commodities from given sources to destinations at min cost

\[
\begin{align*}
\min_x & \quad \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \\
\text{subject to} & \quad \sum_{j=1}^{n} x_{ij} \leq s_i, \quad i = 1, \ldots, m \\
& \quad \sum_{i=1}^{m} x_{ij} \geq d_j, \quad j = 1, \ldots, n, \quad x \geq 0
\end{align*}
\]

Interpretation:

- \( s_i \): supply at source \( i \)
- \( d_j \): demand at destination \( j \)
- \( c_{ij} \): per-unit shipping cost from \( i \) to \( j \)
- \( x_{ij} \): units shipped from \( i \) to \( j \)
Example: basis pursuit

Given $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$, where $p > n$. Suppose that we seek the sparsest solution to underdetermined linear system $X \beta = y$

Nonconvex formulation:

$$\min_{\beta} \quad \|\beta\|_0$$
subject to $X \beta = y$

where recall $\|\beta\|_0 = \sum_{j=1}^{p} 1\{\beta_j \neq 0\}$, the $\ell_0$ “norm”

The $\ell_1$ approximation, often called basis pursuit:

$$\min_{\beta} \quad \|\beta\|_1$$
subject to $X \beta = y$
Basis pursuit is a linear program. Reformulation:

\[
\min_{\beta} \| \beta \|_1 \quad \iff \quad \min_{\beta, z} 1^T z
\]

subject to \( X \beta = y \)

subject to \( z \geq \beta \)

\( z \geq -\beta \)

\( X \beta = y \)

(Check that this makes sense to you)
Example: Dantzig selector

Modification of previous problem, where we allow for \( X\beta \approx y \) (we don’t require exact equality), the Dantzig selector:\(^2\)

\[
\begin{align*}
\min_{\beta} & \quad \| \beta \|_1 \\
\text{subject to} & \quad \| X^T (y - X\beta) \|_\infty \leq \lambda
\end{align*}
\]

Here \( \lambda \geq 0 \) is a tuning parameter

Again, this can be reformulated as a linear program (check this!)

---

\(^2\)Candes and Tao (2007), “The Dantzig selector: statistical estimation when \( p \) is much larger than \( n \)”
A linear program is said to be in **standard form** when it is written as

$$\min_x c^T x$$

subject to

$$Ax = b$$

$$x \geq 0$$

Any linear program can be rewritten in standard form (check this!)
A convex **quadratic program** or QP is an optimization problem of the form

\[
\min_x \quad c^T x + \frac{1}{2} x^T Q x
\]

subject to

\[
Dx \leq d \quad Ax = b
\]

where \( Q \succeq 0 \), i.e., positive semidefinite

Note that this problem is not convex when \( Q \not\succeq 0 \)

From now on, when we say quadratic program or QP, we implicitly assume that \( Q \succeq 0 \) (so the problem is convex)
Example: portfolio optimization

Construct a financial portfolio, trading off performance and risk:

\[
\max_x \mu^T x - \frac{\gamma}{2} x^T Q x
\]

subject to

\[
1^T x = 1
\]

\[
x \geq 0
\]

Interpretation:

- \( \mu \): expected assets’ returns
- \( Q \): covariance matrix of assets’ returns
- \( \gamma \): risk aversion
- \( x \): portfolio holdings (percentages)
Example: support vector machines

Given \( y \in \{-1, 1\}^n \), \( X \in \mathbb{R}^{n \times p} \) having rows \( x_1, \ldots x_n \), recall the support vector machine or SVM problem:

\[
\min_{\beta, \beta_0, \xi} \quad \frac{1}{2} \|\beta\|_2^2 + C \sum_{i=1}^{n} \xi_i \\
\text{subject to} \quad \xi_i \geq 0, \ i = 1, \ldots n \\
y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i, \ i = 1, \ldots n
\]

This is a quadratic program
Example: lasso

Given \( y \in \mathbb{R}^n, \ X \in \mathbb{R}^{n \times p} \), recall the lasso problem:

\[
\begin{align*}
\min_{\beta} & \quad \| y - X \beta \|^2_2 \\
\text{subject to} & \quad \| \beta \|_1 \leq s
\end{align*}
\]

Here \( s \geq 0 \) is a tuning parameter. Indeed, this can be reformulated as a quadratic program (check this!)

Alternative parametrization (called Lagrange, or penalized form):

\[
\begin{align*}
\min_{\beta} & \quad \frac{1}{2}\| y - X \beta \|^2_2 + \lambda \| \beta \|_1 \\
\end{align*}
\]

Now \( \lambda \geq 0 \) is a tuning parameter. And again, this can be rewritten as a quadratic program (check this!)
A quadratic program is in standard form if it is written as

$$\min_x \quad c^T x + \frac{1}{2} x^T Q x$$

subject to \quad Ax = b

\quad x \geq 0

Any quadratic program can be rewritten in standard form
Motivation for semidefinite programs

Consider linear programming again:

\[
\begin{align*}
\min_x & \quad c^T x \\
\text{subject to} & \quad Dx \leq d \\
& \quad Ax = b
\end{align*}
\]

Can generalize by changing \( \leq \) to different (partial) order. Recall:

- \( S^n \) is space of \( n \times n \) symmetric matrices
- \( S^n_+ \) is the space of positive semidefinite matrices, i.e.,
  \[
  S^n_+ = \{ X \in S^n : u^T X u \geq 0 \text{ for all } u \in \mathbb{R}^n \}
  \]
- \( S^n_{++} \) is the space of positive definite matrices, i.e.,
  \[
  S^n_{++} = \{ X \in S^n : u^T X u > 0 \text{ for all } u \in \mathbb{R}^n \setminus \{0\} \} \]
Facts about $\mathbb{S}^n$, $\mathbb{S}^n_+$, $\mathbb{S}^n_{++}$

• Basic linear algebra facts, here $\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))$:

\[
X \in \mathbb{S}^n \implies \lambda(X) \in \mathbb{R}^n
\]
\[
X \in \mathbb{S}^n_+ \iff \lambda(X) \in \mathbb{R}^n_+
\]
\[
X \in \mathbb{S}^n_{++} \iff \lambda(X) \in \mathbb{R}^n_{++}
\]

• We can define an inner product over $\mathbb{S}^n$: given $X, Y \in \mathbb{S}^n$,

\[
X \bullet Y = \text{tr}(XY)
\]

• We can define a partial ordering over $\mathbb{S}^n$: given $X, Y \in \mathbb{S}^n$,

\[
X \succeq Y \iff X - Y \in \mathbb{S}^n_+
\]

Note: for $x, y \in \mathbb{R}^n$, $\text{diag}(x) \succeq \text{diag}(y) \iff x \succeq y$ (recall, the latter is interpreted elementwise)
A *semidefinite program* or SDP is an optimization problem of the form

$$\min_{x} \quad c^T x$$

subject to

$$x_1 F_1 + \ldots + x_n F_n \preceq F_0$$

$$Ax = b$$

Here $F_j \in \mathcal{S}^d$, for $j = 0, 1, \ldots n$, and $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. Observe that this is always a convex optimization problem.

Also, any linear program is a semidefinite program (check this!)
A semidefinite program is in **standard form** if it is written as

\[
\begin{align*}
\min_X & \quad C \bullet X \\
\text{subject to} & \quad A_i \bullet X = b_i, \quad i = 1, \ldots, m \\
& \quad X \succeq 0
\end{align*}
\]

Any semidefinite program can be written in standard form (for a challenge, check this!)
Example: theta function

Let $G = (N, E)$ be an undirected graph, $N = \{1, \ldots, n\}$, and

- $\omega(G)$ : clique number of $G$
- $\chi(G)$ : chromatic number of $G$

The Lovasz theta function:\(^3\)

$$\vartheta(G) = \max_X \quad 11^T \cdot X$$
subject to
$$I \cdot X = 1$$
$$X_{ij} = 0, \ (i, j) \notin E$$
$$X \succeq 0$$

The Lovasz sandwich theorem: $\omega(G) \leq \vartheta(\bar{G}) \leq \chi(G)$, where $\bar{G}$ is the complement graph of $G$

\(^3\)Lovasz (1979), “On the Shannon capacity of a graph”
Example: trace norm minimization

Let \( A : \mathbb{R}^{m \times n} \to \mathbb{R}^p \) be a linear map,

\[
A(X) = \begin{pmatrix}
A_1 \bullet X \\
\vdots \\
A_p \bullet X
\end{pmatrix}
\]

for \( A_1, \ldots A_p \in \mathbb{R}^{m \times n} \) (and where \( A_i \bullet X = \text{tr}(A_i^T X) \)). Finding lowest-rank solution to an underdetermined system, nonconvex:

\[
\min_X \text{rank}(X) \quad \text{subject to} \quad A(X) = b
\]

Trace norm approximation:

\[
\min_X \| X \|_{\text{tr}} \quad \text{subject to} \quad A(X) = b
\]

This is indeed an SDP (but harder to show, requires duality ...)
Conic program

A conic program is an optimization problem of the form:

\[
\begin{align*}
\min_x & \quad c^T x \\
\text{subject to} & \quad Ax = b \\
& \quad D(x) + d \in K
\end{align*}
\]

Here:

- \( c, x \in \mathbb{R}^n \), and \( A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \)
- \( D : \mathbb{R}^n \to Y \) is a linear map, \( d \in Y \), for Euclidean space \( Y \)
- \( K \subseteq Y \) is a closed convex cone

Both LPs and SDPs are special cases of conic programming. For LPs, \( K = \mathbb{R}^n_+ \); for SDPs, \( K = S^n_+ \)
Second-order cone program

A second-order cone program or SOCP is an optimization problem of the form:

\[
\begin{align*}
\min_{x} & \quad c^T x \\
\text{subject to} & \quad \|D_i x + d_i\|_2 \leq e_i^T x + f_i, \quad i = 1, \ldots, p \\
& \quad Ax = b
\end{align*}
\]

This is indeed a cone program. Why? Recall the second-order cone

\[ Q = \{ (x, t) : \|x\|_2 \leq t \} \]

So we have

\[
\|D_i x + d_i\|_2 \leq e_i^T x + f_i \iff (D_i x + d_i, e_i^T x + f_i) \in Q_i
\]

for second-order cone \( Q_i \) of appropriate dimensions. Now take

\[ K = Q_1 \times \ldots \times Q_p \]
Observe that every LP is an SOCP. Further, every SOCP is an SDP

Why? Turns out that

\[ \|x\|_2 \leq t \iff \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0 \]

Hence we can write any SOCP constraint as an SDP constraint

The above is a special case of the Schur complement theorem:

\[ \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \iff A - BC^{-1}B^T \succeq 0 \]

for \( A, C \) symmetric and \( C \succ 0 \)
Hey, what about QPs?

Finally, our old friend QPs “sneak” into the hierarchy. Turns out QPs are SOCPs, which we can see by rewriting a QP as

$$\begin{align*}
\min_{x,t} & \quad c^T x + t \\
\text{subject to} & \quad Dx \leq d, \quad \frac{1}{2} x^T Q x \leq t \\
& \quad Ax = b
\end{align*}$$

Now write $$\frac{1}{2} x^T Q x \leq t \iff \|(\frac{1}{\sqrt{2}} Q^{1/2} x, \frac{1}{2} (1 - t))\|_2 \leq \frac{1}{2} (1 + t)$$

Take a breath (phew!). Thus we have established the hierarchy

$$\text{LPs} \subseteq \text{QPs} \subseteq \text{SOCPs} \subseteq \text{SDPs} \subseteq \text{Conic programs}$$

completing the picture we saw at the start.
References and further reading

- D. Bertsimas and J. Tsitsiklis (1997), “Introduction to linear optimization,” Chapters 1, 2