Convex Optimization Basics

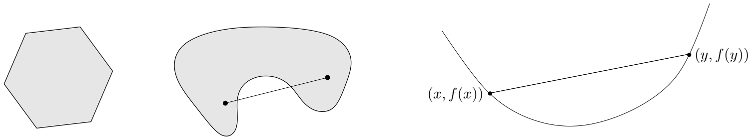
Yu-Xiang Wang
CS292A

(Based on Ryan Tibshirani's 10-725)

Last time: convex sets and functions

"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?

Outline

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

$$\begin{aligned} \min_{x \in D} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \ i = 1, \dots m \\ & Ax = b \end{aligned}$$

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

- f is called criterion or objective function
- g_i is called inequality constraint function
- If $x \in D$, $g_i(x) \le 0$, $i = 1, \dots m$, 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 $f^\star$

- If x is feasible and $f(x) = f^\star$, then x is called optimal; also called a solution, or a minimizer[1]
- If x is feasible and $f(x) \le f^\star + \epsilon$, then x is called ϵ-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

$$
\begin{array}{ll}
\min\limits_x & f(x) \\
\text{subject to} & g_i(x) \le 0, \\
& \quad i = 1, \ldots m \\
& Ax = b
\end{array}
\quad \Longleftrightarrow \quad
\begin{array}{ll}
\max\limits_x & -f(x) \\
\text{subject to} & g_i(x) \le 0, \\
& \quad i = 1, \ldots m \\
& Ax = b
\end{array}
$$

Both are called convex optimization problems

[1] Note: a convex optimization problem need not have solutions, i.e., need not attain its minimum, but we will not be careful about this

Solution set

Let X_{opt} be the set of all solutions of convex problem, written

$$\begin{array}{rcll} X_{\text{opt}} & = & \text{argmin} & f(x) \\ & & \text{subject to} & g_i(x) \leq 0, \ i = 1, \ldots m \\ & & & Ax = b \end{array}$$

Key property: X_{opt} is a convex set

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

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

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_{opt} contains one element

Example: lasso

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

$$\begin{aligned} \min_{\beta} \quad & \|y - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \leq s \end{aligned}$$

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:

$$\sum_{i=1}^n \rho(y_i - x_i^T \beta), \quad \rho(z) = \begin{cases} \frac{1}{2} z^2 & |z| \leq \delta \\ \delta |z| - \frac{1}{2}\delta^2 & \text{else} \end{cases} \quad ?$$

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:

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

Is this convex? What is the criterion, constraints, feasible set? Is the solution (β, β_0, ξ) 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 β component, at the solution?

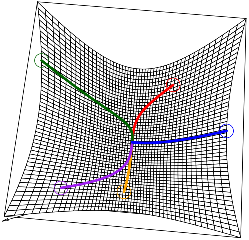
Local minima are global minima

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

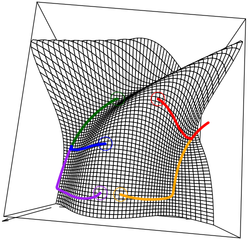
$$f(x) \le f(y) \quad \text{for all feasible } y \text{ such that } \|x - y\|_2 \le R$$

Reminder: for convex optimization problems, local optima are global optima

Proof simply follows from definitions



Convex



Nonconvex

Rewriting constraints

The optimization problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \ i = 1, \dots m \\ & Ax = b \end{aligned}$$

can be rewritten as

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

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

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

$$\min_x \ f(x) + I_C(x)$$

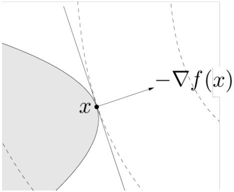
First-order optimality condition

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

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



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 $b \notin \mathrm{col}(Q)$, then there is no solution (i.e., $\min_x \ f(x) = -\infty$)
- if Q is singular and $b \in \mathrm{col}(Q)$, then there are infinitely many solutions

  $$x = -Q^{+}b + z, \quad z \in \mathrm{null}(Q)$$

  where Q^{+} is the pseudoinverse of Q

Example: equality-constrained minimization

Consider the equality-constrained convex problem:

$$\min_{x} \; f(x) \;\; \text{subject to} \;\; Ax = b$$

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

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

Result follows because $\mathrm{null}(A)^\perp = \mathrm{row}(A)$

Example: projection onto a convex set

Consider projection onto convex set C:

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

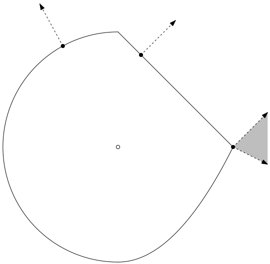
First-order optimality condition says that the solution x satisfies

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

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

$$\begin{array}{ll} \min\limits_{x_1, x_2} & f(x_1, x_2) \\ \text{subject to} & g_1(x_1) \le 0 \\ & g_2(x_2) \le 0 \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \min\limits_{x_1} & \tilde{f}(x_1) \\ \text{subject to} & g_1(x_1) \le 0 \end{array}$$

where $\tilde{f}(x_1) = \min\{f(x_1, x_2) : g_2(x_2) \le 0\}$. The right problem is convex if the left problem is

Example: hinge form of SVMs

Recall the SVM problem

$$\begin{aligned} \min_{\beta,\beta_0,\xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{subject to} \quad & \xi_i \geq 0,\ y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i,\ i = 1,\dots n \end{aligned}$$

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 ξ 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

$$\begin{aligned} &\min_x \; f(x) \;\; \text{subject to} \;\; x \in C \\ \iff &\min_x \; h(f(x)) \;\; \text{subject to} \;\; x \in C \end{aligned}$$

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:

$$\begin{aligned} &\min_x \; f(x) \;\; \text{subject to} \;\; x \in C \\ \iff &\min_y \; f(\phi(y)) \;\; \text{subject to} \;\; \phi(y) \in C \end{aligned}$$

Example: geometric programming

A monomial is a function $f : \mathbb{R}^n_{++} \to \mathbb{R}$ 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

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

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

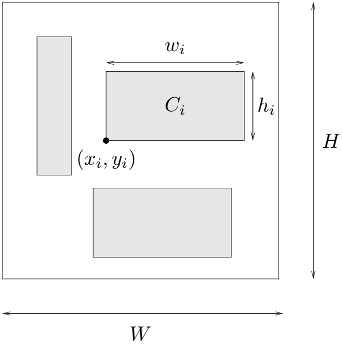
$$\gamma (e^{y_1})^{a_1} (e^{y_2})^{a_2} \cdots (e^{y_n})^{a_n} = e^{a^T y + b}$$

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

$$\begin{array}{ll} \min\limits_{x} & \log \left(\displaystyle\sum_{k=1}^{p_0} e^{a_{0k}^T y + b_{0k}} \right) \\ \text{subject to} & \log \left(\displaystyle\sum_{k=1}^{p_i} e^{a_{ik}^T y + b_{ik}} \right) \le 0, \ i = 1, \dots m \\ & c_j^T y + d_j = 0, \ j = 1, \dots r \end{array}$$

This is convex, recalling the convexity of soft max functions

Several interesting problems are geometric programs, e.g., floor planning:



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

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \ i = 1, \ldots m \\ & Ax = b \end{aligned}$$

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

$$\begin{aligned} \min_y \quad & f(My + x_0) \\ \text{subject to} \quad & g_i(My + x_0) \le 0, \ i = 1, \ldots m \end{aligned}$$

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

Introducing slack variables

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

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0, \; i = 1, \ldots m \\ & Ax = b \end{aligned}$$

we can transform the inequality constraints via

$$\begin{aligned} \min_{x,s} \quad & f(x) \\ \text{subject to} \quad & s_i \ge 0, \; i = 1, \ldots m \\ & g_i(x) + s_i = 0, \; i = 1, \ldots m \\ & Ax = b \end{aligned}$$

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:

$$\begin{aligned} \max_{x,b} \quad & \sum_{t=1}^{T} \alpha_t u(x_t) \\ \text{subject to} \quad & b_{t+1} = b_t + f(b_t) - x_t, \ t = 1, \dots T \\ & 0 \le x_t \le b_t, \ t = 1, \dots T \end{aligned}$$

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:

$$b_{t+1} \le b_t + f(b_t) - x_t, \ t = 1, \dots T?$$

Example: principal components analysis

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

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

Here $\|A\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p A_{ij}^2$, the entrywise squared ℓ_2 norm, and $\operatorname{rank}(A)$ denotes the rank of A

Also called principal components analysis or PCA problem. Given $X = UDV^T$, 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

$$\min_{Z \in \mathbb{S}^p} \|X - XZ\|_F^2 \;\; \text{subject to} \;\; \mathrm{rank}(Z) = k, \;\; Z \text{ is a projection}$$
$$\iff \max_{Z \in \mathbb{S}^p} \mathrm{tr}(SZ) \;\; \text{subject to} \;\; \mathrm{rank}(Z) = k, \;\; Z \text{ is a projection}$$

where $S = X^T X$. Hence constraint set is the nonconvex set

$$C = \Big\{ Z \in \mathbb{S}^p : \lambda_i(Z) \in \{0, 1\}, \; i = 1, \ldots p, \; \mathrm{tr}(Z) = k \Big\}$$

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 $\mathcal{F}_k = \mathrm{conv}(C)$, its convex hull. Note

$$\begin{aligned} \mathcal{F}_k &= \{Z \in \mathbb{S}^p : \lambda_i(Z) \in [0,1],\ i = 1, \ldots p,\ \mathrm{tr}(Z) = k\} \\ &= \{Z \in \mathbb{S}^p : 0 \preceq Z \preceq I,\ \mathrm{tr}(Z) = k\} \end{aligned}$$

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

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

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 conerning 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

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

- This is the algorithm for the sparse PCA problem that achieves the minimax rate. (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 $j \in \bar{S}$.

$$\begin{array}{ll} \max\limits_{x} & \dfrac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(1 - x_i x_j) \\ \text{s.t.} & x_j \in \{-1, 1\}, j = 1, ..., n \end{array}$$

- Goemans and Williamson algorithm:
  1. Convex relaxation: solve an SDP instead.
  2. Randomized rounding.
- You get a 0.87856 approximation of an NP-complete problem.

Approximation Algorithm for MaxCut

Reformulation (without changing the problem):

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

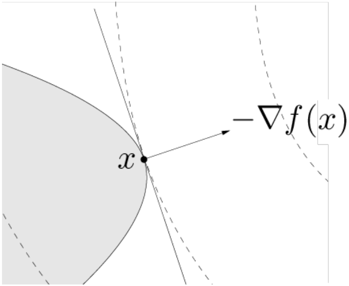
The convex relaxation:

$$\begin{aligned} \max_{Y \in \mathbb{R}^{n\times n}} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(1 - Y_{i,j}) \\ \text{s.t.} \quad & Y_{i,i} = 1 \quad \forall j = 1, ..., n \\ & Y \succeq 0. \end{aligned}$$

Goemans and Williamson: Sample v uniformly from the unit sphere in $\mathbb{R}^n$, output $\mathrm{sign}(Yv)$.

Quick Summary

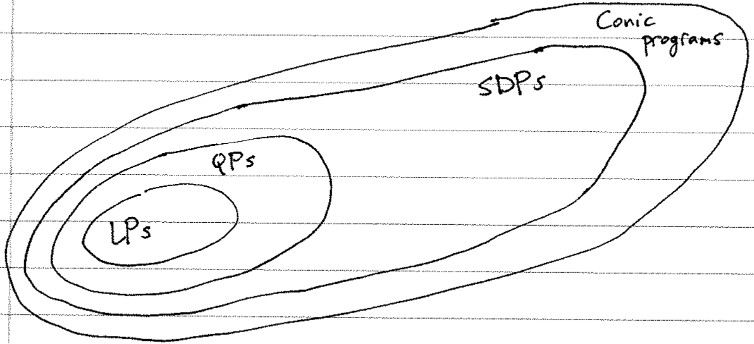
- Optimization terminology (e.g., criterion, constraints, feasible points, solutions)
- Properties and first-order optimality



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

Hierarchy of Canonical Optimizations

- Linear programs
- Quadratic programs
- Semidefinite programs
- Cone programs


Conic programs
SDPs
QPs
LPs

Linear program

A linear program or LP is an optimization problem of the form

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Dx \le d \\ & Ax = b \end{aligned}$$

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!)

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & Dx \geq d \\ & x \geq 0 \end{aligned}$$

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

$$
\begin{array}{ll}
\min\limits_{x} & \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} c_{ij} x_{ij} \\
\text{subject to} & \sum\limits_{j=1}^{n} x_{ij} \le s_i,\ i = 1, \ldots, m \\
& \sum\limits_{i=1}^{m} x_{ij} \ge d_j,\ j = 1, \ldots, n,\ \ x \ge 0
\end{array}
$$

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:

$$\begin{array}{ll} \min_{\beta} & \|\beta\|_0 \\ \text{subject to} & X\beta = y \end{array}$$

where recall $\|\beta\|_0 = \sum_{j=1}^p 1\{\beta_j \neq 0\}$, the ℓ_0 "norm"

The ℓ_1 approximation, often called basis pursuit:

$$\begin{array}{ll} \min_{\beta} & \|\beta\|_1 \\ \text{subject to} & X\beta = y \end{array}$$

Basis pursuit is a linear program. Reformulation:

$$
\begin{array}{ll}
\displaystyle\min_{\beta} & \|\beta\|_1 \\
\text{subject to} & X\beta = y
\end{array}
\quad \iff \quad
\begin{array}{ll}
\displaystyle\min_{\beta, z} & 1^T z \\
\text{subject to} & z \geq \beta \\
& z \geq -\beta \\
& X\beta = y
\end{array}
$$

(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]

$$\begin{aligned} \min_{\beta} \quad & \|\beta\|_1 \\ \text{subject to} \quad & \|X^T(y - X\beta)\|_\infty \le \lambda \end{aligned}$$

Here $\lambda \ge 0$ 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"

Standard form

A linear program is said to be in standard form when it is written as

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Any linear program can be rewritten in standard form (check this!)

Convex quadratic program

A convex quadratic program or QP is an optimization problem of the form

$$\begin{aligned} \min_x \quad & c^T x + \frac{1}{2} x^T Q x \\ \text{subject to} \quad & Dx \le d \\ & Ax = b \end{aligned}$$

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:

$$\begin{aligned} \max_x \quad & \mu^T x - \frac{\gamma}{2} x^T Q x \\ \text{subject to} \quad & 1^T x = 1 \\ & x \geq 0 \end{aligned}$$

Interpretation:

- μ : expected assets' returns
- Q : covariance matrix of assets' returns
- γ : 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:

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

This is a quadratic program

Example: lasso

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

$$\begin{aligned} \min_{\beta} \quad & \|y - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \le s \end{aligned}$$

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

Alternative parametrization (called Lagrange, or penalized form):

$$\min_{\beta} \ \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

Now $\lambda \ge 0$ is a tuning parameter. And again, this can be rewritten as a quadratic program (check this!)

Standard form

A quadratic program is in standard form if it is written as

$$\begin{aligned} \min_x \quad & c^T x + \frac{1}{2} x^T Q x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Any quadratic program can be rewritten in standard form

Motivation for semidefinite programs

Consider linear programming again:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Dx \le d \\ & Ax = b \end{aligned}$$

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

- $\mathbb{S}^n$ is space of $n \times n$ symmetric matrices
- $\mathbb{S}^n_+$ is the space of positive semidefinite matrices, i.e.,

$$\mathbb{S}^n_+ = \{X \in \mathbb{S}^n : u^T X u \ge 0 \text{ for all } u \in \mathbb{R}^n\}$$

- $\mathbb{S}^n_{++}$ is the space of positive definite matrices, i.e.,

$$\mathbb{S}^n_{++} = \left\{X \in \mathbb{S}^n : u^T X u > 0 \text{ for all } u \in \mathbb{R}^n \setminus \{0\}\right\}$$

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 $X, Y \in \mathbb{S}^n$,

$$X \bullet Y = \mathrm{tr}(XY)$$

- 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 $x, y \in \mathbb{R}^n$, $\mathrm{diag}(x) \succeq \mathrm{diag}(y) \iff x \geq y$ (recall, the latter is interpreted elementwise)

Semidefinite program

A semidefinite program or SDP is an optimization problem of the form

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & x_1 F_1 + \ldots + x_n F_n \preceq F_0 \\ & Ax = b \end{aligned}$$

Here $F_j \in \mathbb{S}^d$, 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!)

Standard form

A semidefinite program is in standard form if it is written as

$$\begin{aligned} \min_{X} \quad & C \bullet X \\ \text{subject to} \quad & A_i \bullet X = b_i, \; i = 1, \ldots m \\ & X \succeq 0 \end{aligned}$$

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]

$$
\begin{array}{rcll}
\vartheta(G) & = & \max_X & 11^T \bullet X \\
 & & \text{subject to} & I \bullet X = 1 \\
 & & & X_{ij} = 0,\ (i,j) \notin E \\
 & & & X \succeq 0
\end{array}
$$

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,

$$A(X) = \begin{pmatrix} A_1 \bullet X \\ \ldots \\ A_p \bullet X \end{pmatrix}$$

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

$$\begin{array}{ll} \min\limits_X & \mathrm{rank}(X) \\ \text{subject to} & A(X) = b \end{array}$$

Trace norm approximation:

$$\begin{array}{ll} \min\limits_X & \|X\|_{\mathrm{tr}} \\ \text{subject to} & A(X) = b \end{array}$$

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

Conic program

A conic program is an optimization problem of the form:

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

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, $K = \mathbb{S}^n_+$

Second-order cone program

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

$$\begin{aligned} \min_{x} \quad & c^T x \\ \text{subject to} \quad & \|D_i x + d_i\|_2 \le e_i^T x + f_i, \; i = 1, \dots p \\ & Ax = b \end{aligned}$$

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

$$Q = \{(x,t) : \|x\|_2 \le t\}$$

So we have

$$\|D_i x + d_i\|_2 \le e_i^T x + f_i \iff (D_i x + d_i, e_i^T x + f_i) \in Q_i$$

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

$$\|x\|_2 \le t \iff \begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0$$

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

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

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

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

LPs $\subseteq$ QPs $\subseteq$ SOCPs $\subseteq$ SDPs $\subseteq$ Conic programs

completing the picture we saw at the start

References and further reading

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapter 4
- O. Guler (2010), "Foundations of optimization", Chapter 4
- D. Bertsimas and J. Tsitsiklis (1997), "Introduction to linear optimization," Chapters 1, 2
- A. Nemirovski and A. Ben-Tal (2001), "Lectures on modern convex optimization," Chapters 1–4