

Lecture 3: Canonical Problem Forms

Lecturer: Yu-Xiang Wang

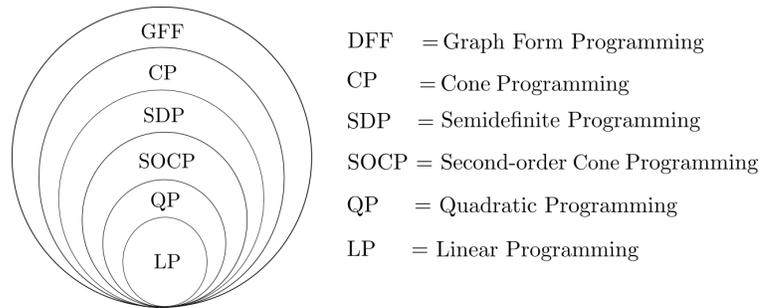
Scribes: Krishna Gudipati

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

3.1 Introduction

Optimization is a huge class of problems. There is a hierarchy of convex optimization problems. We'll talk about linear programming, quadratic programming, second-order cone programming, and semidefinite programming problems. These 4 canonical problems described in these notes relate to each other according to the high level picture shown below:



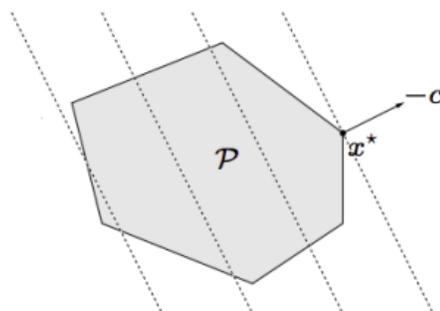
3.2 Linear Program

A linear program or LP is a convex optimization problem with affine objective and constraint functions.

Basic form	Standard form
$\min_x \quad c^\top x$	$\min_x \quad c^\top x$
$\text{subject to } \quad Gx \preceq h$	$\text{subject to } \quad Ax = b$
$\quad \quad \quad Ax = b$	$\quad \quad \quad x \succeq 0$

The transition from the general form to the standard form is obtained by adding slack variables s so that the constraints $Dx \leq d$ become $Dx + s = d, s \geq 0$.

The feasible set of LPs always forms a polyhedron P shown in the figure below. The objective function can be visualized as isolines with constant cost, visualized by the dotted lines. The optimal solution is at the boundary point that touches the isoline of least cost.



LP is the simplest type of convex optimization problem first introduced by Kantorovich in the late 1930s and Dantzig in the 1940s.

The interest in *Linear Programming* began during World War II in order to deal with problems of transportation, scheduling, and allocation of resources subject to certain restrictions such as costs and availability.

The first LP solver was developed in the late 1940s (Dantzig's "simplex algorithm"), and now LP solvers are considered a mature technology.

3.2.1 Examples

Example 3.1. Diet : Find cheapest combination of foods that satisfies some nutritional requirements.

$$\begin{array}{ll} \min_x & c^\top x \\ \text{subject to} & Ax \succeq b \\ & x \succeq 0 \end{array}$$

- c_j : per-unit cost of food j
- b_i : minimum required intake of nutrient i
- A_{ij} : content of nutrient i per unit of food j
- x_j : units of food j in the diet

Example 3.2. Transportation problem : Ship commodities from given sources to destinations at min cost.

$$\begin{array}{ll} \min_x & \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \\ \text{subject to} & \sum_{j=1}^n x_{ij} \leq s_i, \quad i=1, \dots, m \\ & \sum_{i=1}^m x_{ij} \leq d_j, \quad j=1, \dots, n \\ & x \succeq 0 \end{array}$$

- 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 3.3. Basis Pursuit :

Given $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$, where $p > n$. Suppose we want to find a solution to the linear system $X\beta = y$. Because $p > n$, this system is undetermined so there are many solutions. One option is to seek the sparsest possible solution

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

This is a non-convex problem. We will instead consider the convex relaxation of this problem known as Basis Pursuit.

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

It can be shown that this is a linear program by reformulation.

$$\begin{aligned} \min_{\beta, z} \quad & \mathbf{1}^T z \\ \text{subject to} \quad & z \geq \beta \\ & z \geq -\beta \\ & X\beta = y \end{aligned}$$

3.3 Quadratic Program

A convex optimization problem is called a quadratic program (QP) if the objective function is (convex) quadratic, and the constraint functions are affine.

Basic form	Standard form
$\begin{aligned} \min_x \quad & \frac{1}{2}x^T Q^T x + c^T x \\ \text{subject to} \quad & Gx \preceq h \\ & Ax = b \end{aligned}$	$\begin{aligned} \min_x \quad & \frac{1}{2}x^T Q^T x + c^T x \\ \text{subject to} \quad & Ax = b \\ & x \succeq 0 \end{aligned}$

The problem is convex iff the matrix Q is positive semidefinite. i.e $Q \succeq 0$.

In a quadratic program, we minimize a convex quadratic function over a polyhedron.

Quadratic programs include linear programs as a special case, by taking $Q = 0$.

3.3.1 Examples

Example 3.4. Portfolio optimization : Construct financial portfolio with optimal performance/risk tradeoff.

$\begin{aligned} \max_x \quad & \mu^T x - \frac{\gamma}{2} x^T Q x \\ \text{subject to} \quad & \mathbf{1}^T x = 1 \\ & x \succeq 0 \end{aligned}$	<ul style="list-style-type: none"> • μ : expected assets' returns • Q: covariance matrix of assets' returns • γ : risk aversion • x : portfolio holdings (percentages)
--	---

Example 3.5. Support Vector Machines : Given $y \in \{-1, 1\}^n$, $X \in R^{n \times p}$ having rows x_1, \dots, x_n , the SVM problem:

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

Example 3.6. Lasso : Given $y \in \{R\}^n$, $X \in R^{n \times p}$, we can write the constrained form of the Lasso as follows:

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

We can also write this in the called Lagrangian or penalized form with $\lambda \geq 0$ as the tuning parameter:

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

3.3.2 Quadratically constrained quadratic program (QCQP)

If the objective as well as the inequality constraint functions are (convex) quadratic then the problem is called *quadratically constrained quadratic program*.

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T Q^T x + c^T x \\ \text{subject to} \quad & \frac{1}{2}x^T P_i^T x + q_i^T x + r_i \leq 0 \quad i=1, \dots, m \\ & Ax = b \end{aligned}$$

In a QCQP, we minimize a convex quadratic function over a feasible region that is the intersection of ellipsoids (when $P_i \prec 0$).

Quadratically constrained quadratic programs include quadratic programs (and therefore also linear programs) as a special case, by taking $P_i = 0$.

3.4 Semidefinite Program

Semi-definite programs are a much bigger subset of convex optimization problems than convex quadratic programs. We can extend from linear programs to semi-definite programs by changing the order (\leq) involved in the inequality constraint $Gx \leq h$ to a different kind of order in some vector space. We work with the vector-space S^n .

3.4.1 Notes about Symmetric Matrices

- S^n is space of $n \times n$ symmetric matrices
- S_+^n is the space of positive semidefinite matrices, i.e., $S_+^n = \{X \in S^n: u^T X u \geq 0 \text{ for all } u \in R^n\}$
- S_{++}^n is the space of positive definite matrices, i.e., $S_{++}^n = \{X \in S^n: u^T X u > 0 \text{ for all } u \in R^n\}$
- The canonical inner product in S^n : for $X, Y \in S^n$, $X \bullet Y = \text{tr}(XY) = \sum_{i=1}^n \sum_{j=1}^n X_{ij} Y_{ij}$.
- Trace satisfies the property that if the product ABC is well-defined and the result is a square matrix, then $\text{tr}(ABC) = \text{tr}(BCA)$.
- Loewner ordering: Given $X, Y \in S^n$, $X \succeq Y \iff X - Y \in S_+^n$
- For $x, y \in R^n$, $\text{diag}(x) \succeq \text{diag}(y) \iff x \geq y$ (elementwise)

3.4.2 Optimization Problem

Semidefinite programming can be regarded as an extension of linear programming where the componentwise inequalities between vectors are replaced by matrix inequalities. It is of the form:

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

where $c \in R^n$, $F_0, F_1, \dots, F_n \in S^d$, $A \in R^{m \times n}$, $b \in R^m$

The inequality constraint is called Linear Matrix Inequality (LMI).

SDPs can also be written in standard form, which is derived from LPs with our analogous inner product and partial ordering.

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

where $C, A_1, \dots, A_m \in S^n$

To convert any SDP to standard form, we make use of slack variables. In particular, we can first split x into positive and negative parts, i.e., $x = x^+ - x^-$, such that $x^+, x^- \geq 0$. Next, the inequality can be cast into equality by introducing a slack variable $Y \succeq 0$. Then the problem now becomes

$$\begin{aligned} \min_{x^+, x^-, Y} \quad & c^T x^+ - c^T x^- \\ \text{subject to} \quad & (x_1^+ - x_1^-)F_1 + \dots + (x_n^+ - x_n^-)F_n + Y = F_0 \\ & Ax^+ - Ax^- = b \\ & x^+ \geq 0, x^- \geq 0, Y \succeq 0 \end{aligned}$$

Now, the standard form can be realized by constructing block matrices out of x^+ , x^- , and Y and rearranging coefficient matrices.

Every linear program with optimization variable x is also a semidefinite program. To see this, consider the standard form SDP where $X = \text{diag}(x)$.

3.4.3 Examples

Example 3.7. Lovasz Theta Function

Let $G = (N, E)$, N is the node and E is the edge.

$\omega(G)$ is the clique number of G i.e the largest set of nodes that are completely connected.

$\chi(G)$ is the chromatic number of G i.e the minimum number of colors that suffice to color the nodes of graph.

The Lovasz theta function $\mathcal{V}(G) =$

$$\begin{aligned} \max_X \quad & \mathbf{1}\mathbf{1}^T \bullet X \\ \text{subject to} \quad & I \bullet X = 1 \\ & X_{ij} = 0, (i, j) \notin E \\ & X \succeq 0 \end{aligned}$$

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

Example 3.8. Trace Norm Minimization

Let $A : R^{m \times n} \rightarrow R^p$ be a linear map,

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

Trace norm $\|X\|_{tr} = \|\sigma(X)\|_1$, the sum of the singular values of X .

The dual of the nuclear norm is operator norm: $\|X\|_{op} = \|\sigma(X)\|_\infty = \max \|Xu\|_2 : \|u\|_2 \leq 1$

$$\begin{aligned} \min_X \quad & \mathbf{1}\mathbf{1}^T \bullet X \\ \text{subject to} \quad & I \bullet X = 1 \\ & X_{ij} = 0, (i, j) \notin E \\ & X \succeq 0 \end{aligned}$$

Consider finding the minimum-rank solution of the underdetermined linear system:

$$\begin{aligned} \min_X \quad & \text{rank}(X) \\ \text{subject to} \quad & A(X) = b \end{aligned}$$

This problem is non-convex and hard to solve. Therefore, a popular approximation is to instead minimize the trace of the matrix:

$$\begin{aligned} \min_X \quad & \|X\|_{tr} \\ \text{subject to} \quad & A(X) = b \end{aligned}$$

The intuition behind this relaxation is that $\text{rank}(X) = \sum_i \mathbf{1}\{\sigma_i(x) \neq 0\}$, i.e. the L_0 -norm of the singular values, so as we saw with the basis pursuit problem, we use an L_1 -norm relaxation (where $\|X\|_{tr} = \sum_i \sigma_i(x)$).

3.5 Conic Program

A conic program is an optimization problem of the form:

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

where $c, x \in R^n$, $A \in R^{m \times n}$, $b \in R^m$; $D : R^n \rightarrow Y$ is a linear map and $d \in Y$ for a Euclidean space Y ; and $K \subseteq Y$ is a closed convex cone.

Conic inequality: a constraint $x \in K$ where K is a convex cone in R^m .

$$x \succeq_K y \iff x - y \in K$$

$$x \succ_K y \iff x - y \in \text{int } K \text{ (interior of } K)$$

Conic program is again very similar to LP, the only distinction is the set of linear inequalities are replaced with conic inequalities, i.e. $D(x) + d \leq_K 0$.

If $K = R_+^n$ the nonnegative orthant then the inequality has the form $D(x) + d \leq 0$, we recover the LP. Similarly if $K = S_+^n$, we recover SDP.

3.6 Second Order Cone Program

A second order cone program (SOCP) is defined as the following special case of a cone program:

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

The inequalities are called second order cone constraints. Indeed, they form a special class of convex sets called *second order cones*.

We can see that every LP is a second order cone program.

Theorem 3.9. *SOCP's are Conic Programs*

Proof. Recall the second order cone:

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

Consider the constraint

$$\|D_i x + d_i\|_2 \leq e_i^T x + f_i \implies (\|D_i x + d_i\|_2, e_i^T x + f_i) \in Q_i \quad \forall i = 1, \dots, p$$

Now, taking the cone as the cartesian product of all the previous cones gives us the conic program, i.e $K = Q_1 \times \dots \times Q_p$ □

Theorem 3.10. *Every QP is an SOCP*

Proof. We can rewrite the canonical QP program in the following form by introducing a variable t :

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

Consider the constraint $\frac{1}{2} x^T Q x \leq t$. We can show that

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

$$\begin{aligned} \left\| \left(\frac{1}{\sqrt{2}} Q^{\frac{1}{2}} x, \frac{1}{2}(1-t) \right) \right\|_2 &\leq \frac{1}{2}(1+t) \\ \left\| \left(\frac{1}{\sqrt{2}} Q^{\frac{1}{2}} x, \frac{1}{2}(1-t) \right) \right\|_2^2 &\leq \frac{1}{4}(1+t)^2 \\ \frac{1}{2} x^T Q x + \frac{1}{4}(1-t)^2 &\leq \frac{1}{4}(1+t)^2 \\ \frac{1}{2} x^T Q x &\leq t \end{aligned}$$

□

Theorem 3.11. *Every SOCP is a Semidefinite Program(SDP)*

Proof. Schur Complement theorem: If A,C are symmetric and C \succ 0

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \implies A - BC^{-1}B^T \succeq 0$$

Apply this theorem to the following matrix

$$\begin{bmatrix} tI & x \\ x^T & t \end{bmatrix} \succeq 0 \implies tI - \frac{xx^T}{t} \succeq 0 \implies \|x\|_2 \leq t$$

Therefore, every SOCP can be transformed into an SDP. □

Thus we arrive at the convex programming hierarchy

$$LP \subset QP \subset SOCP \subset SDP \subset CP$$

3.7 Approximation Algorithm for Max Cut

Max Cut: Given an undirected graph $G=(V, E)$, find a partition of V into two subsets S, \bar{S} so as to maximize the number of edges having one endpoint in S and the other in \bar{S} .

Weighted Max Cut: Given an undirected graph $G=(V, E)$ and a positive weight w_e for each edge, find a partition of V into two subsets S, \bar{S} so as to maximize the combined weight of the edges having one endpoint in S and the other in \bar{S} .

Max Cut:

- is NP-hard
- is the same as finding maximum bipartite subgraph of G
- can be thought of a variant of the 2-coloring problem in which we try to maximize the number of edges consisting of two different colors.

Let $x_j = 1$ if $j \in S$ and $x_j = -1$ if $j \in \bar{S}$

$$\begin{aligned} \max_x & \quad \frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n w_{ij}(1 - x_i x_j) \\ \text{such that} & \quad x_j \in \{-1, 1\}, j=1, \dots, n \end{aligned}$$

Reformulation

$$\begin{aligned} \max_{Y \in R^{n \times n}, x \in R^n} & \quad \frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n w_{ij}(1 - Y_{i,j}) \\ \text{such that} & \quad Y_{j,j} = 1, j=1, \dots, n \\ & \quad Y = xx^T \end{aligned}$$

Convex relaxation

$$\begin{aligned} \max_{Y \in R^{n \times n}} & \quad \frac{1}{4} \sum_{i=1}^n \sum_{j=1}^n w_{ij}(1 - Y_{i,j}) \\ \text{such that} & \quad Y_{j,j} = 1, j=1, \dots, n \\ & \quad Y \succeq 0 \end{aligned}$$

Goemans and Williamson algorithm :

1. Convex relaxation: solve an SDP instead
2. Randomized rounding: Sample v uniformly from the unit sphere in R^n , decompose $Y = UU^T$, output $\text{sign}(Uv)$.

References

- [1] S. BOYD and L. VANDENBERGHE, "Convex Optimization," Chapter 4
- [2] A. NEMIROVSKI and A. BEN-TAL, "Lectures on modern convex optimization," Chapter 1-4

Also, the lecture notes from similar course offerings at various universities and reference material from internet were referred to while making this scribe notes.