Lecturer: Dabeen Lee

1 Outline

In this lecture, we study

- quadratic programming,
- semidefinite programming,
- conic programming,
- derivation of dual conic programs.

2 Quadratic programming

A quadratic program (QP) is an optimization problem of the following form.

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}x^\top Q x + p^\top x \\ \text{subject to} \quad & Ax \geq b \end{aligned} \tag{QP}$$

The quadratic program is convex only if Q is positive semidefinite.

2.1 Example: portfolio optimization

We studied the following formulation of portfolio optimization.

$$\begin{aligned} \text{maximize} \quad & \mu^\top x - \gamma x^\top \Sigma x \\ \text{subject to} \quad & 1^\top x = 1, \\ & x \in \mathbb{R}^n_+ \end{aligned}$$

where $\gamma > 0$ and Σ is a covariance matrix that is positive semidefinite. Note that

$$\max\{f(x):\ x \in C\} = -\min\{-f(x):\ x \in C\}$$

holds for any objective function f and any feasible set C. Thus, the formulation is equivalent to

$$\begin{aligned} \text{minimize} \quad & \gamma x^\top \Sigma x - \mu^\top x \\ \text{subject to} \quad & 1^\top x = 1 \\ & x \geq 0 \end{aligned}$$

which is a quadratic program because $\gamma > 0$ and Σ is positive semidefinite.

2.2 Example: support vector machine

The next example is the formulation of support vector machine.

$$\min_{w,b} \quad \lambda\|w\|_2^2 + \frac{1}{n}\sum_{i=1}^n \max\{0,\ 1 - y_i(w^\top x_i - b)\}.$$

Here, $\|w\|_2^2 = w^\top w = w^\top I w$ where I is the identity matrix, and therefore, $\|w\|_2^2$ is a convex quadratic function. Moreover, the max terms in the objective can be replaced by adding some auxiliary variables. Note that the formulation is equivalent to

$$\begin{aligned} \text{minimize} \quad & \lambda w^\top w + \frac{1}{n}\sum_{i=1}^n t_i \\ \text{subject to} \quad & t_i \geq \max\{0,\ 1 - y_i(w^\top x_i - b)\} \text{ for } i = 1, \ldots, n. \end{aligned}$$

Next, we can rewrite the constraints as linear constraints as the following.

$$\begin{aligned} \text{minimize} \quad & \lambda w^\top w + \frac{1}{n}\sum_{i=1}^n t_i \\ \text{subject to} \quad & t_i \geq 1 - y_i(w^\top x_i - b) \quad \text{for } i = 1, \ldots, n, \\ & t_i \geq 0 \quad \text{for } i = 1, \ldots, n. \end{aligned}$$

Therefore, it is a convex quadratic program with a quadratic objective and linear constraints.

2.3 Example: LASSO

Recall that LASSO can be formulated as

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

Note that

$$\|y - X\beta\|_2^2 = (y - X\beta)^\top(y - X\beta) = \beta^\top X^\top X\beta - 2y^\top X\beta + y^\top y$$

Here, $X^\top X$ is positive semidefinite because

$$u^\top X^\top X u = \|Xu\|_2^2 \geq 0$$

for any vector u. In addition, $y^\top y$ is a constant term which can be ignored from the objective. Moreover, we can replace the $\|\beta\|_1$ term by an auxiliary variable and a set of linear constraints. To be specific, the problem is equivalent to

$$\begin{aligned} \text{minimize} \quad & \frac{1}{n}\beta^\top X^\top X\beta - \frac{2}{n}y^\top X\beta + \lambda t \\ \text{subject to} \quad & t \geq \sum_{i=1}^d s_i, \\ & s_i \geq \beta \geq -s_i \text{ for } i = 1, \ldots, d. \end{aligned}$$

Hence, LASSO can be reformulated as a quadratic program.

3 Semidefinite programming

3.1 Motivation: max-cut

Semidefinite programming provides useful tools for solving difficult combinatorial optimization problems. For example, we consider the "max-cut problem" defined as follows. Given a graph $G = (V, E)$, find a partition the vertex set V so that the number of edges crossing the partition is maximized. Here, a partition (V_1, V_2) of V consists of two sets V_1, V_2 satisfying $V_1 \cup V_2 = V$ and $V_1 \cap V_2 = \emptyset$, and the set of edges crossing the partition is basically $\{uv \in E : u \in V_1, v \in V_2\}$. For example, in Figure 6.1, there is a graph of 5 vertices partitioned into red and black vertices, and the edges highlighted are the ones crossing the partition.

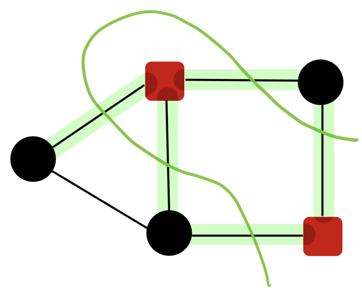

Figure 6.1: Edges crossing a partition

The problem can be formulated by the following (discrete) optimization problem:

$$\begin{aligned} \text{maximize} \quad & \sum_{ij \in E} \frac{1 - x_i x_j}{2} \\ \text{subject to} \quad & x_i \in \{-1, 1\} \text{ for } i \in V. \end{aligned}$$

As long as $x_i \in \mathbb{R}$, $x_i \in \{-1, 1\}$ is equivalent to $x_i^2 = 1$. Hence, the formulation is equivalent to

$$\begin{aligned} \text{maximize} \quad & \sum_{ij \in E} \frac{1 - x_i x_j}{2} \\ \text{subject to} \quad & x_i^2 = 1 \text{ for } i \in V. \end{aligned}$$

Let $d = |V|$. Then we consider a $d \times d$ matrix X whose entry at ith row and jth column, X_{ij}, is $x_i x_j$. Then we have that $X = xx^\top$, which is the outer product of vector x and itself. In fact, X is of the form $X = xx^\top$ if and only if X is positive semidefinite and the rank of X is precisey 1. What this implies is that, the max-cut formulation can be rewritten as

$$\begin{aligned} \text{maximize} \quad & \sum_{ij \in E} \frac{1 - X_{ij}}{2} \\ \text{subject to} \quad & X_{ii} = 1 \text{ for } i \in V, \\ & X \succeq 0, \\ & \text{rank}(X) = 1. \end{aligned}$$

Here, the constrsint $\text{rank}(X) = 1$ is nonconvex. A common approach is to take out the nonconvex constraint and consider

$$\begin{aligned} \text{maximize} \quad & \sum_{ij \in E} \frac{1 - X_{ij}}{2} \\ \text{subject to} \quad & X_{ii} = 1 \text{ for } i \in V, \\ & X \succeq 0. \end{aligned}$$

This is often called the *semidefinite programming (SDP) relaxation* of max-cut.

3.2 General form

More generally, a *semidefinite program* is an optimization problem of the following form. Let C and $A_1, \ldots, A_m$ be $d \times d$ matrices, and we have

$$\begin{aligned} \text{minimize} \quad & \text{tr}(C^\top X) \\ \text{subject to} \quad & \text{tr}(A_\ell^\top X) = b_\ell \text{ for } \ell = 1, \ldots, m \\ & X \succeq 0 \end{aligned} \tag{SDP}$$

where

$$\text{tr}(C^\top X) = \sum_{i=1}^{d} \sum_{j=1}^{d} C_{ij} X_{ij} \quad \text{and} \quad \text{tr}(A_\ell^\top X) = \sum_{i=1}^{d} \sum_{j=1}^{d} (A_\ell)_{ij} X_{ij}.$$

Here, if we view matrix X as a $(d \times d)$-dimensional vector, then the objective and the equality constraints are "linear" in X. Hence, (SDP) is analogous to linear programming. Recall that we defined the linear programming (LP) dual of a given linear program. Likewise, we may define the notion of semidefinite programming (SDP) dual. The *dual* of (SDP) is

$$\begin{aligned} \text{maximize} \quad & \sum_{\ell=1}^{m} b_\ell y_\ell \\ \text{subject to} \quad & \sum_{\ell=1}^{m} y_\ell A_\ell \preceq C \end{aligned} \tag{dual-SDP}$$

where $\sum_{\ell=1}^{m} y_\ell A_\ell \preceq C$ means $C - \sum_{\ell=1}^{m} y_\ell A_\ell$ is positive semidefinite. If an optimization is in either form, we say that it is a semidefinite program.

We will study more about duality later in this course. We have dicussed LP duality, and in particular, we covered how to derive the dual of a linear program and learned duality theorems. The notion of duality extends to more general classes of convex programming problems. We will learn how to derive the dual of a given optimization problem, and we will define the associated weak and strong duality statements.

3.3 Example: quadratic programming

(QP) can be rewritten as

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & Ax \geq b, \\ & x^\top Q x + 2p^\top x \leq 2t. \end{aligned}$$

In fact, this can be expressed as an instance of (dual-SDP) by rewriting $Ax \geq b$ and $x^\top Qx + 2p^\top x \leq 2t$ using some positive semidefinite matrices.

Note that $Ax - b$ is a vector and $Ax \geq b$ means that the entries of $Ax - b$ are nonnegative. $\mathrm{Diag}(Ax - b)$ is the diagonal matrix whose diagonal entries are the components of $Ax - b$. In fact, $Ax - b \geq 0$ holds if and only if

$$\mathrm{Diag}(Ax - b) \succeq 0$$

which means that $\mathrm{Diag}(Ax - b)$ is positive semidefinite.

Next we consider $x^\top Qx + 2p^\top x \leq 2t$ where Q is positive semidefinite.

Lemma 6.1. *For any positive semidefinite matrix Q, there exists a matrix P such that $Q = P^\top P$.*

Proof. By the eigendecomposition theorem for symmetric marices, Q can be written as $Q = U\Lambda U^\top$ where U is an orthonormal matrix and Λ is a diagonal matrix whose diagonal entries consist of the eigenvalues of Q. Since Q is positive semidefinite, all its eigenvalues are nonnegative, and therefore, all diagonal entries of Λ are nonnegative. Then $\Lambda^{1/2}$ can be properly defined by taking the square root of each diagonal entry of Λ. Then $\Lambda = (\Lambda^{1/2})^\top \Lambda^{1/2}$ as $\Lambda^{1/2}$ is symmetric as well. Then

$$Q = U\Lambda U^\top = U(\Lambda^{1/2})^\top \Lambda^{1/2} U^\top = (\Lambda^{1/2}U^\top)^\top (\Lambda^{1/2}U^\top).$$

Taking $P = \Lambda^{1/2}U^\top$, we have $Q = P^\top P$. □

By Lemma 6.1, $x^\top Qx + 2p^\top x \leq 2t$ is equivalent to

$$x^\top P^\top Px + 2p^\top x \leq 2t$$

for some matrix P. We also need the following result.

Lemma 6.2. *Let $y \in \mathbb{R}^d$. Then $y^\top y \leq s$ is equivalent to*

$$\begin{pmatrix} s & y^\top \\ y & I \end{pmatrix} \succeq 0$$

where I is the $d \times d$ identity matrix.

Proof. ($\Leftarrow$) Note that

$$(1, -y^\top) \begin{pmatrix} s & y^\top \\ y & I \end{pmatrix} \begin{pmatrix} 1 \\ -y \end{pmatrix} = s - y^\top y \geq 0.$$

($\Rightarrow$) Let $u \in \mathbb{R}$ and $v \in \mathbb{R}^d$. Then

$$\begin{aligned} (u, v^\top) \begin{pmatrix} s & y^\top \\ y & I \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} &= u^2 s + 2uy^\top v + v^\top v \\ &\geq u^2 y^\top y + 2uy^\top v + v^\top v \\ &= (uy + v)^\top (uy + v) \\ &\geq 0. \end{aligned}$$

Therefore, the matrix is positive semidefinite as required. □

By Lemma 6.2, $x^\top P^\top P x + 2p^\top x \leq 2t$ is equivalent to

$$\begin{pmatrix} 2t - 2p^\top x & (Px)^\top \\ Px & I \end{pmatrix} \succeq 0.$$

Finally, we have shown that (QP) is equivalent to the following optimization problem.

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \mathrm{Diag}(Ax - b) \succeq 0, \\ & \begin{pmatrix} 2t - 2p^\top x & (Px)^\top \\ Px & I \end{pmatrix} \succeq 0. \end{aligned}$$

4 Conic programming

Recall that a linear program (LP) is an optimization problem with a linear objective and a system of linear inequality constraints, as follows.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax \geq b. \end{aligned} \tag{LP}$$

Here, if the rows of A are $a_1^\top, \ldots, a_n^\top$ and the components of b are $b_1, \ldots, b_n$, then the linear system $Ax \geq b$ consists of linear inequality constraints $a_1^\top x \geq b_1, \ldots, a_n^\top x \geq b_n$. Note that Ax itself is a column vector whose components are $a_1^\top x, \ldots, a_n^\top x$. Basically, the arithmetic "$\geq$" compares two column vectors Ax and b coordinatewise.

$Ax \geq b$ is equivalent to $Ax - b \geq 0$, which means that each component of the column vector $Ax - b$ is nonnegative. We know that $\mathbb{R}_+^n$ is the nonnegative orthant, that is, the set of vectors all whose coordinates are nonnegative. Hence, $Ax - b \geq 0$ is equivalent to $\mathbb{R}_+^n$. Then the following is an equivalent expression for the above linear program.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax - b \in \mathbb{R}_+^n. \end{aligned}$$

Let us take a closer look at the nonnegative orthant $\mathbb{R}_+^n$. It satisfies the following properties.

1. $\mathbb{R}_+^n$ is a convex cone.
2. $\mathbb{R}_+^n$ is *pointed*, which means that if $v \in \mathbb{R}_+^n$ and $-v \in \mathbb{R}_+^n$, then it must be that $v = 0$.

In fact, $\mathbb{R}_+^n$ is not just a pointed convex cone. There are other important properties of $\mathbb{R}_+^n$.

3. $\mathbb{R}_+^n$ is *closed*, which means that for any convergent sequence $\{v^n\}_{n\in\mathbb{N}}$ contained in $\mathbb{R}_+^n$, its limit $\lim_{n\to\infty} v^n$ also belongs to $\mathbb{R}_+^n$.
4. $\mathbb{R}_+^n$ has a nonempty *interior*. Equivalently, $\mathbb{R}_+^n$ contains an *interior point*. A vector v is an interior point of a set K if there exists an open ball around v which is fully contained in K. Then the interior of a set K, denoted $\mathrm{int}(K)$, is defined as the set of all its interior points. The interior of $\mathbb{R}_+^n$ is $\mathbb{R}_{++}^n$, the positive orthant.

In summary, the nonnegative orthant $\mathbb{R}_+^n$ is a pointed and closed convex cone with a nonempty interior. In fact, there are other closed convex cones that are pointed and have a nonempty interior. For example,

- The *Lorentz cone*.

$$\{(x_1, \ldots, x_{n-1}, x_n)^\top \in \mathbb{R}^n : \ \|(x_1, \ldots, x_{n-1})^\top\|_2 \leq x_n\}.$$

  Other equivalent names include the *second-order cone*, the *ice-cream cone*, and the *ℓ_2-norm cone*. Its interior is given by

$$\{(x_1, \ldots, x_{n-1}, x_n)^\top \in \mathbb{R}^n : \ \|(x_1, \ldots, x_{n-1})^\top\|_2 < x_n\}.$$

- The positive semidefinite cone.

$$\{S \in \mathbb{S}^d : \ x^\top S x \geq 0 \quad \text{for all } x \in \mathbb{R}^d\}.$$

  Its interior is the positive definite cone, the set of all positive definite matrices.

A *conic program* is an optimization problem defined with a pointed and closed convex cone K with a nonempty interior, as follows.

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax - b \in K. \end{aligned} \tag{CP}$$

Again, when $K = \mathbb{R}^n_+$, the problem reduces to a linear program. As we use the arithmetic "$\geq$" to indicate that a vector belongs to $\mathbb{R}^n_+$, we use notation "$\geq_K$" to indicate that a vector belongs to cone K. Basically, $Ax - b \in K$ is equivalent to $Ax - b \geq_K 0$ and $Ax \geq_K b$.

Example 6.3. When K is the second-order cone, the conic program (CP) is referred to as a *second-order cone program*. When K is the positive semidefinite cone, (CP) is a semidefinite program.

5 Conic duality

We know that the dual of the linear program (LP) is given by

$$\begin{aligned} \text{maximize} \quad & b^\top y \\ \text{subject to} \quad & A^\top y = c \\ & y \geq 0. \end{aligned} \tag{dual-LP}$$

Let us see how to derive the dual! Note that for any $y \geq 0$ (or $y \in \mathbb{R}^n_+$) and system $Ax \geq b$, we have $y^\top(Ax - b) \geq 0$ because $y \geq 0$ and $Ax - b \geq 0$. Then it follows that

$$y^\top Ax \geq y^\top b.$$

If y further satisfies

$$A^\top y = c,$$

then we have

$$y^\top Ax = c^\top x \geq y^\top b = b^\top y.$$

In summary, if we take $x \in \mathbb{R}^d$ satisfying $Ax \geq b$ and $y \in \mathbb{R}^n$ with $y \geq 0$ and $A^\top y = c$, then $c^\top x$ is always lower bounded by $b^\top y$. Then we can try to find the best possible lower bound by maximizing the value of $b^\top y$, which is precisely what (dual-LP) does!

Following the basic idea behind obtaining the dual linear program, we may obtain and define the dual of the conic program (CP). The *dual cone* of $K \subseteq \mathbb{R}^n$ is defined as

$$K^* = \left\{ y \in \mathbb{R}^n : \ y^\top x \geq 0 \ \ \forall x \in K \right\}.$$

The dual cone of the nonnegative orthant $\mathbb{R}^d_+$ is $\mathbb{R}^d_+$ itself.

Example 6.4. The dual cone of the positive semidefinite cone $\mathbb{S}^d_+$ is given by

$$\left\{X \in \mathbb{R}^{d\times d}:\ \mathrm{tr}(X^\top S) = \sum_{i=1}^d \sum_{j=1}^d X_{ij}S_{ij} \geq 0 \ \ \forall S \in \mathbb{S}^d_+\right\}.$$

In fact, the positive semidefinite cone $\mathbb{S}^d_+$ is *self-dual*, meaning that its dual cone is itself.

Theorem 6.5 (See Theorem 2.3.1 in [BTN01])**.** *Let K be a pointed and closed convex cone with nonempty interior. Then its dual cone K^* is also a pointed and closed convex cone with nonempty interior. Moreover, $(K^*)^* = K$.*

Let us see how to derive and define the dual of the conic program!

(1) Take x such that $Ax - b \in K$ and $y \in K^*$. Then $y^\top(Ax - b) \geq 0$, and therefore,

$$y^\top Ax \geq y^\top b.$$

(2) If $y \in K^*$ further satisfies $A^\top y = c$, then

$$c^\top x = y^\top Ax \geq y^\top b = b^\top y.$$

(3) Then

$$\begin{aligned} \text{maximize} \quad & b^\top y \\ \text{subject to} \quad & A^\top y = c \\ & y \in K^* \end{aligned} \tag{dual-CP}$$

provides a lower bound on the value of (CP). Here, (dual-CP) is the dual conic program of (CP).

Taking the dual of a maximization problem is similar; the dual will give an upper bound on the problem.

Example 6.6. We consider the following semidefinite program.

$$\begin{aligned} \text{maximize} \quad & \sum_{\ell=1}^m b_\ell y_\ell \\ \text{subject to} \quad & \sum_{\ell=1}^m y_\ell A_\ell \preceq C \end{aligned}$$

To obtain its dual, we take a positive semidefinite matrix X. As the positive semidefinite cone $\mathbb{S}^d_+$ is self-dual, it follows that

$$\mathrm{tr}\left(X^\top\left(C - \sum_{\ell=1}^m y_\ell A_\ell\right)\right) = \mathrm{tr}(C^\top X) - \sum_{\ell=1}^m y_\ell \cdot \mathrm{tr}((A_\ell)^\top X) \geq 0.$$

If X satisfies

$$\mathrm{tr}((A_\ell)^\top X) = b_\ell \quad \text{for } \ell = 1, \ldots, m,$$

then

$$\operatorname{tr}(C^\top X) \geq \sum_{\ell=1}^{m} y_\ell \cdot \operatorname{tr}((A_\ell)^\top X) = \sum_{\ell=1}^{m} b_\ell y_\ell.$$

This means that

$$\begin{aligned} \text{minimize} \quad & \operatorname{tr}(C^\top X) \\ \text{subject to} \quad & \operatorname{tr}((A_\ell)^\top X) = b_\ell \quad \text{for } \ell = 1, \ldots, m \\ & X \succeq 0 \end{aligned}$$

provides an upper bound on the first semidefinite program.

References

[BTN01] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on Modern Convex Optimization*. Society for Industrial and Applied Mathematics, 2001. 6.5