Chapter 6

Semidefinite Programming

Semidefinite programming is a form of convex optimization that generalizes linear programming, and also provides a unified framework for several standard problems, including quadratic programming. The notes below follow closely [6].

Definition 6.1. An $n \times n$ real matrix X is *positive semidefinite* if

$$v^T X v \geq 0,\ \forall v \in \mathbb{R}^n.$$

We write $X \succeq 0$. We say that X is positive definite if

$$v^T X v > 0,\ \forall v \in \mathbb{R}^n.$$

We write $X \succ 0$.

We denote by S^n the set of $n \times n$ real symmetric matrices, that is

$$S^n = \{X \in M_n(\mathbb{R}),\ X^T = X\}.$$

Then

$$\begin{aligned} S^n_+ &= \{X \in S^n,\ X \succeq 0\}, \\ S^n_{++} &= \{X \in S^n,\ X \succ 0\}. \end{aligned}$$

Note that $X \succeq Y$ means $X - Y \succeq 0$.

Lemma 6.1. *The set*

$$S^n_+ = \{X \in S^n,\ X \succeq 0\}$$

is convex.

Proof. Pick $\lambda \in]0,1[$, and $X, W \in S^n_+$, we want to see that $\lambda X + (1-\lambda)W \in S^n_+$. So take any $v \in \mathbb{R}^n$, then

$$v^T(\lambda X + (1-\lambda)W)v = \lambda v^T X v + (1-\lambda)v^T W v \geq 0$$

as needed. □

We recall a few facts about symmetric matrices.

- If $X \in S^n$, we have a decomposition of X into $X = QDQ^T$ for Q orthonormal (that is $Q^{-1} = Q^T$) and D diagonal. The columns of Q form a set of n orthonormal eigenvectors of X, whose eigenvalues are the corresponding diagonal entries of D.
- If $X \in S^n$ and $X \succeq 0$, then $v^T QDQ^T v^T = (v^T Q)D(v^T Q)^T \geq 0$ for all v, so pick v^T to be in turn each of the row of $Q^{-1} = Q^T$, then $v^T Q$ will range through all the unit vectors, and $(v^T Q)D(v^T Q)^T$ will range through all the eigenvalues of X, so all of them are are non-negative.
- If $X \in S^n$, $X \succeq 0$ and if $x_{ii} = 0$, then $x_{ij} = x_{ji} = 0$ for all $j = 1, \dots, n$ (see Exercise 49).
- Consider the matrix
$$M = \begin{bmatrix} P & v \\ v^T & d \end{bmatrix}$$
for $P \succ 0$, $P \in S^n$, v a vector, d scalar. Then $M \succ 0 \iff d - v^T P^{-1} v > 0$. This is saying that a symmetric matrix is positive definite if and only if its Schur complement is.

Definition 6.2. A *semidefinite program (SDP)* is an optimization problem of the form

$$\begin{aligned} \min \quad & C \bullet X \\ s.t. \quad & A_i \bullet X = b_i, \ i = 1, \dots, m \\ & X \succeq 0 \end{aligned}$$

where we will assume that X, C and $A_i, i = 1, \dots, m$, are symmetric, and we use the notation:

$$C \bullet X = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} = \text{trace}(C^T X).$$

Recall that by definition, $\text{trace}(M) = \sum_{j=1}^{n} m_{jj}$.

Example 6.1. Take

$$A_1 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 7 \\ 1 & 7 & 5 \end{bmatrix}, \ A_2 = \begin{bmatrix} 0 & 2 & 8 \\ 2 & 6 & 0 \\ 8 & 0 & 4 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 0 \\ 3 & 0 & 7 \end{bmatrix}, \ b_1 = 11, \ b_2 = 19.$$

Then we have

$$\begin{aligned} \min \quad & C \bullet X = x_{11} + 4x_{12} + 6x_{13} + 9x_{22} + 7x_{33} \\ s.t. \quad & x_{11} + 2x_{13} + 3x_{22} + 14x_{23} + 5x_{33} = 11 \\ & 4x_{12} + 16x_{13} + 6x_{22} + 4x_{33} = 19 \\ & X \succeq 0. \end{aligned}$$

The above example almost looks like an LP, but for the constraint $X \succeq 0$. We will show next that in fact, every LP can be written as an SDP (the above example serves as a counter-example that the converse is not true).

Suppose that we have the LP:

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

(constraints of the form $Ax \leq b$ can be rewritten with equality using slack variables).

Then define

$$A_i = \mathrm{diag}(a_{i1}, \ldots, a_{in}),\ i = 1, \ldots, m,\ C = \mathrm{diag}(c_1, \ldots, c_n),\ X = \mathrm{diag}(x_1, \ldots, x_n)$$

and the above LP can be written as the following SDP:

$$\begin{aligned} \min \quad & C \bullet X = c^T x \\ s.t. \quad & A_i \bullet X = b_i,\ i = 1, \ldots, m \\ & X \succeq 0 \end{aligned}$$

where the eigenvalues of X are $x_1, \ldots, x_n$ so $x_1, \ldots, x_n \geq 0$ can be expressed in terms of $X \succeq 0$.

6.1 Duality

Given an SDP:

$$\begin{aligned} (SDP) : \min \quad & C \bullet X \\ s.t. \quad & A_i \bullet X = b_i,\ i = 1, \ldots, m \\ & X \succeq 0, \end{aligned}$$

define its dual as

$$\begin{aligned} (SDD) : \max \quad & \textstyle\sum_{i=1}^m y_i b_i \\ s.t. \quad & \textstyle\sum_{i=1}^m y_i A_i + S = C \\ & S \succeq 0. \end{aligned}$$

In particular, we get from the dual that $C - \sum_{i=1}^m y_i A_i \succeq 0$. Note that S has to be symmetric since C and A_i are.

Example 6.2. For our previous example, we get

$$\begin{aligned} (SDD)\max \quad & 11y_1 + 19y_2 \\ s.t. \quad & y_1 \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 7 \\ 1 & 7 & 5 \end{bmatrix} + y_2 \begin{bmatrix} 0 & 2 & 8 \\ 2 & 6 & 0 \\ 8 & 0 & 4 \end{bmatrix} + S = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 0 \\ 3 & 0 & 7 \end{bmatrix} \\ & S \succeq 0. \end{aligned}$$

Alternatively:

$$(SDD)\max \quad 11y_1 + 19y_2$$
$$s.t. \quad \begin{bmatrix} 1-y_1 & 2-2y_2 & 3-y_1-8y_2 \\ 2-2y_2 & 9-3y_1-6y_2 & -7y_1 \\ 3-y_1-8y_2 & -7y_1 & 7-5y_1-4y_2 \end{bmatrix} \succeq 0.$$

The above SDP-SDD formulation is the equivalent of the primal-dual formulation for LP:

$$\begin{aligned} (P):\min \quad & c^T x \\ s.t. \quad & A_i x = b_i,\ i = 1,\ldots,m \\ & x \geq 0, \end{aligned}$$

where A_i are the rows of A, and

$$\begin{aligned} (D):\max \quad & \textstyle\sum_{i=1}^m y_i b_i \\ s.t. \quad & \textstyle\sum_{i=1}^m y_i A_i^T + s = c \\ & s \geq 0. \end{aligned}$$

We express the duality here using equalities instead of inequalities, but as we saw in the chapter on linear programming, one can always move from one formulation to another, and express the dual programs accordingly.

Let us start by recalling what we know about weak and strong duality for LP, then we shall see what can be generalized to SDP.

Given a feasible solution x of (P), and a feasible solution (y, s) of its dual (D), we have, using that $A_i x = b_i$

$$c^T x - \sum_{i=1}^m y_i b_i = \left(c^T - \sum_{i=1}^m y_i A_i\right) x = s^T x \geq 0$$

since $x, s \geq 0$. This is the Weak Duality Theorem, which says that if x is feasible for (P) and y is feasible for (D), then $c^T x \leq b^T y$.

The difference $c^T x - \sum_{i=1}^m y_i b_i$ is called the *duality gap*, it is the gap between the two objective functions. We know from the Strong Duality Theorem that as long as the primal LP is feasible and bounded, then the primal and the dual both attain their optima with no duality gap. That is, there exist x^* and (y^*, s^*) feasible for the primal and the dual respectively, such that

$$c^T x^* - \sum_{i=1}^m y_i^* b_i = (s^*)^T x^* = 0.$$

It turns out that weak duality holds for SDP:

Theorem 6.2. *Given a feasible solution X of SDP, a feasible solution (Y, S) of SDD, the duality gap is*

$$C \bullet X - \sum_{i=1}^m y_i b_i = S \bullet X \geq 0.$$

If $C \bullet X - \sum_{i=1}^{m} y_i b_i = 0$, then X and (Y,S) are each optimal solutions to the SDP and SDD respectively, and furthermore $SX = 0$.

Proof. To start with, we want to prove that $S \bullet X \geq 0$. Since S and X are symmetric and positive definite, write

$$S = PDP^T, \; X = QEQ^T$$

for P, Q orthornomal and D, E diagonal matrices whose diagonal entries are non-negative. Then since S is symmetric:

$$\begin{aligned} S \bullet X &= \text{trace}(S^T X) = \text{trace}(SX) \\ &= \text{trace}(P(DP^T QEQ^T)) = \text{trace}(DP^T QEQ^T P) \end{aligned}$$

since $\text{trace}(MN) = \text{trace}(NM)$. Now multiplying $P^T QEQ^T P$ by $D = \text{diag}(d_1, \ldots, d_n)$ means that row j of $P^T QEQ^T P$ is multiplied by d_j, and

$$\text{trace}(DP^T QEQ^T P) = \sum_{j=1}^{n} d_j (P^T QEQ^T P)_{jj} \geq 0$$

because $d_j \geq 0$ (S is positive semidefinite) and $(P^T QEQ^T P)_{jj}$ are the diagonal coefficients of the matrix $(P^T QEQ^T P)$ which is positive semidefinite (write $v^T (P^T Q) E (P^T Q)^T v = w^T E w$ with $w = v^T (P^T Q)$), thus they are non-negative (see Exercise 50). This completes the proof that $S \bullet X \geq 0$.

Next we want to prove that if $C \bullet X - \sum_{i=1}^{m} y_i b_i = 0$, then X and (Y,S) are optimal and $SX = 0$. Since $C \bullet X - \sum_{i=1}^{m} y_i b_i = S \bullet X$, we have that $S \bullet X = 0$. Optimality is clear from this, because $C \bullet X \geq \sum_{i=1}^{m} y_i b_i$, for all X and (Y,S), so from the view point of X, it is true for every X that $C \bullet X \geq \sum_{i=1}^{m} y_i^* b_i$ where y^* gives the largest value, and thus since we want to minimize $C \bullet X$, optimality for X is reached with equality. From the view point of Y, for every Y, $C \bullet X^* \geq \sum_{i=1}^{m} y_i b_i$ for X^* which minimizes $C \bullet X$, and so optimality for Y is reached with equality.

We are left to show that $SX = 0$. We just showed above that $S \bullet X = \sum_{j=1}^{n} d_j (P^T QEQ^T P)_{jj}$ where every term of the sum is non-negative, so

$$\sum_{j=1}^{n} d_j (P^T QEQ^T P)_{jj} = 0$$

and for each j, either $d_j = 0$ or $(P^T QEQ^T P)_{jj} = 0$.

- If $d_j = 0$, then the jth row and column of D are zero, so S and SX have their jth row and column zero as well.
- If $(P^T QEQ^T P)_{jj} = 0$, then $(P^T QEQ^T P)_{ij} = (P^T QEQ^T P)_{ji} = 0$ (see one of the facts recalled above about symmetric positive definite matrices), and so the jth row and column of $(P^T QEQ^T P)$ are zero. Multiplying this matrix by PD will give a new matrix whose jth row and columns are still zero, which is SX.

□

Strong duality on the other hand does not hold. Here is a classical example given by Lovász.

Example 6.3.

$$\begin{aligned} \min \quad & y_1 \\ s.t. \quad & \begin{bmatrix} 0 & y_1 & 0 \\ y_1 & y_2 & 0 \\ 0 & 0 & y_1+1 \end{bmatrix} \succeq 0. \end{aligned}$$

We remember from the facts about symmetric positive semidefinite matrices listed earlier that a zero on the diagonal means the corresponding row and column are zero. Thus $y_1 = 0$ in any feasible solution for this SDP. Once $y_1 = 0$, we must have $y_2 \geq 0$, and so the minimum for this SDP is 0.

One can compute (see Exercise 51) that the dual of this SDP is:

$$\begin{aligned} \max \quad & -x_{33} \\ s.t. \quad & x_{12} + x_{21} + x_{33} = 1 \\ & x_{22} = 0 \\ & X \succeq 0. \end{aligned}$$

Since $x_{22} = 0$ and $X \succeq 0$, again, a zero on the diagonal means the corresponding row and column are zero, so $x_{12} = x_{21} = 0$ and $x_{23} = x_{32} = 0$. But $x_{12} + x_{21} + x_{33} = 1$, so $x_{33} = 1$ and the optimum is -1. So strong duality does not hold.

For SDP, strong duality holds under the so-called Slater conditions.

Theorem 6.3. *Let z_P^* and z_D^* denote the optimal values of the objective functions for the SDP and its dual respectively. Suppose there exists a feasible solution X^* of the SDP such that $X^* \succ 0$, and there exists a feasible solution (Y^*, S^*) of its dual SDD such that $S^* \succ 0$. Then both the SDP and the SDD attain their optimal value, and $z_P^* = z_D^*$.*

See [8] for a proof.

Weak duality extends from LP to SDP. Strong duality, under some stronger conditions for SDP than for LDP, extends as well. However, there is no direct analog of a basic feasible solution for SDP. There is no finite algorithm for solving SDP, but while this sounds discouraging at first, SDP is an actually very powerful optimization framework, and there are extremely efficient algorithms to solve semidefinite programs.

SDP has wide applications in combinatorial optimization.

- A number of NP-hard combinatorial optimization problems have convex relaxations that are semidefinite programs (we will give an example of such a relaxation in the next section). SDP relaxations are often tight in practice, in certains cases, the optimal solution fo the SDP relaxation can be converted to a feasible solution for the original problem with provably good objective value.

- SDP can be used to model constraints that include linear inequalities, convex quadratic inequalities, lower bounds on matrix norms, lower bounds on determinants of symmetric positive semidefinite matrices.
- SDP can be used to solve linear programs, to optimize a convex quadratic form under convex quadratic inequality constraints among many other applications (see Exercise 52 for an eigenvalue optimization example).

6.2 An SDP Relaxation of the Max Cut Problem

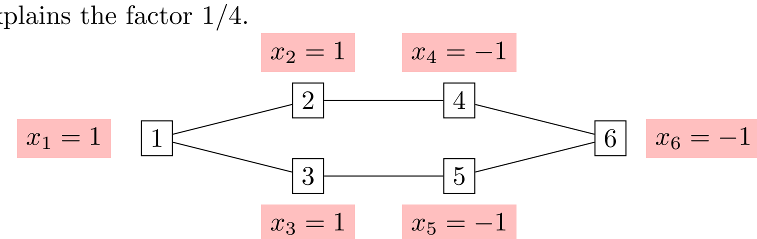
Let $G = (V, E)$ be an undirected graph. Let $w_{ij} = w_{ji} \geq 0$ be the weight on the edge $(i, j) \in E$.

Problem 6. The *max cut* problem consists of determining a subset S of nodes for which the sum of weights of the edges that cross from S to its complement $\bar{S} = V \backslash S$ is maximized.

Set $x_j = 1$, $j \in S$, and $x_j = -1$ for $j \in \bar{S}$. Then we can formulate our max cut problem as:

$$\begin{array}{rl} \textbf{maxcut} : \max & \frac{1}{4}\sum_{i,j=1}^{n} w_{ij}(1 - x_i x_j) \\ s.t. & x_j \in \{1, -1\},\ j = 1, \ldots, n. \end{array}$$

Every term in this sum is of the form $1 - x_i x_j$, which is equal to 0 if both x_i, x_j have the same sign, and equal to 2 if x_i, x_j have reverse signs. Having the same sign means that both nodes i, j are either in S or $\bar{S}$, so they should not contribute to the cut, and indeed $1 - x_i x_j = 0$ in this case. When x_i, x_j have reverse signs, then it means that one node i or j is in S and the other is in $\bar{S}$. However, since the sum is over i, j, we will encounter in the sum once the term $w_{ij}(1 - x_i x_j)$ and once the term $w_{ji}(1 - x_j x_i)$, accounting for $4w_{ij}$, which explains the factor $1/4$.



Then let

$$Y = xx^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} (x_1, \ldots, x_n)$$

so that $y_{ij} = x_i x_j$. Set $W = (w_{ij})$ to be the matrix containing all the weights. The coefficients x_i, x_j must be ± 1 by construction, we can capture this by asking $y_{jj} = 1$. Indeed $y_{jj} = x_j^2 = 1$ implies that $x_j = \pm 1$ for all j. Then we can reformulate the max cut problem as follows:

$$\begin{aligned} \textbf{maxcut} : \max \quad & \tfrac{1}{4} \textstyle\sum_{i,j=1}^{n} w_{ij} - W \bullet Y \\ s.t. \quad & y_{jj} = 1, \; j = 1, \dots, n \\ & Y = xx^T. \end{aligned}$$

We note that Y is a rank 1 positive semidefinite matrix. We "relax" this condition by removing the rank 1 restriction. This gives

$$\begin{aligned} \textbf{relaxmaxcut} : \max \quad & \tfrac{1}{4} \textstyle\sum_{i,j=1}^{n} w_{ij} - W \bullet Y \\ s.t. \quad & y_{jj} = 1, \; j = 1, \dots, n \\ & Y \succeq 0. \end{aligned}$$

Now we have that the optimal solution for the relax-max-cut problem is greater or equal to that for the max-cut problem, we write MAXCUT $\leq$ RELAX. However, it was proven by Goemans and Williamson (1995) that in fact 0.87856·RELAX $\leq$ MAXCUT $\leq$ RELAX, that is, the optimal value of the SDP relaxation is guaranteed to be no more than 12% higher than the optimal value for the NP-hard max cut problem.

6.3 An SDP Relaxation of the Independent Set Problem

Let $G = (V, E)$ be an undirected graph with $|V| = n$ vertices.

Problem 7. The *stable/independent set* problem consists of determining a subset S of nodes, no two of which are connected by an edge. The size $\alpha(G)$ of the largest stable set is called the *stability number of a graph.*

A natural integer programming formulation of $\alpha(G)$ is

$$\begin{aligned} \alpha(G) = \max \quad & \textstyle\sum_{i=1}^{n} x_i \\ s.t. \quad & x_i + x_j \leq 1, \; \{i, j\} \in E \\ & x_i \in \{0, 1\}, \; i = 1, \dots, n. \end{aligned}$$

Label every node by 0 or 1, the first condition tells that if there is an edge between i and j, then both labels cannot be 1, it is either 0 and 1 or 0 and 0, capturing the property that nodes with label 1 form a stable set.

We can give an LP relaxation of this problem by letting x_i range from 0 to 1:

$$\begin{aligned} LP = \max \quad & \textstyle\sum_{i=1}^{n} x_i \\ s.t. \quad & x_i + x_j \leq 1, \; \{i, j\} \in E \\ & 0 \leq x_i \leq 1, \; i = 1, \dots, n. \end{aligned}$$