Lecture 11 — October 3

Lecturer: Laurent El Ghaoui Scribe: Nikhil Shetty

11.1 Outline

- SDP Duality
- SDP Example: Combinatorial Optimization

11.2 SDP Duality

The standard model for the Semi Definite Programming (SDP) is

$$\begin{aligned} &\min \ c^T x \\ &s.t. \ F(x) = F_0 + \sum_{i=1}^{m} x_i F_i \succeq 0, \end{aligned}$$

where each of the F_i's are symmetric matrices.

We define the Lagrangian:

$$\mathcal{L}(x, Z) = c^T x - \mathbf{Tr}(ZF(x)),$$

where the dual variable Z is a psd matrix.

The Lagrangian is constructed so that

$$\max_{Z \succeq 0} \mathcal{L}(x, Z) = \begin{cases} c^T x & \text{if } F(x) \succeq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Thus, we can express p^* as the solution to an unconstrained problem:

$$p^* = \min_x \max_{Z \succeq 0} \mathcal{L}(x, Z)$$

The dual problem is

$$d^* = \max_{Z \succeq 0} \min_x \mathcal{L}(x, Z).$$

The minimum over x is simple to obtain:

$$\min_x \mathcal{L}(x, Z) = \begin{cases} -\mathbf{Tr}\, F_0 Z & \text{if } \mathbf{Tr}\, F_i Z = c_i, \ \ i = 1, \ldots, m, \\ -\infty & \text{otherwise.} \end{cases}$$

We obtain

$$\begin{array}{lll} d^* = \max\limits_{Z} & -\mathbf{Tr}(F_0 Z) & \\ s.t. & Z \succeq 0 & \\ & \mathbf{Tr}(F_i Z) & = c_i, \qquad i = 1, 2, \ldots m. \end{array}$$

We always have $p^* \geq d^*$. If, in addition, the primal problem is strictly feasible, then equality holds (strong duality).

11.3 An Example: Combinatorial Optimization

Let $W = W^T \in \mathbf{R}^{n \times n}$. Consider the problem

$$p^* := \max_x \; x^T W x \; : \; x_i^2 = 1, \;\; i = 1, \ldots, n. \tag{11.1}$$

11.3.1 Application: Maximum Cut Problem

The above problem arises in many problems of combinatorial optimization. Consider for example the problem of maximum-cut of a graph. An undirected graph is given, with weight $\omega_{ij} \geq 0$ given to edge (i, j), with the convention that $\omega_{ij} = 0$ if no edge connects nodes i and j. The maximum-cut problem is to cut the graph in two (that is, separate the nodes into two classes) so that the total weight of the cut (the weight of any edge that links two nodes that are in different classes) is maximized. Denote by x a vector of boolean variables x_i that takes the value 1 if the node i is in class A and -1 if it is in class B. The weight of the cut corresponding to x is then

$$\frac{1}{4} \sum_{i,j} \omega_{ij} (1 - x_i x_j).$$

Maximizing the cut corresponds to a problem of the form (11.1), with W set to

$$W_{ij} = -\frac{1}{4} \omega_{ij}, \;\; i, j = 1, \ldots, n.$$

11.3.2 Inequality form

Without loss of generality, we can assume $W \succ 0$. Indeed, if this is not the case, we can always add αI to W to make it p.d.; the objective is simply changed by adding a constant $n\alpha$.

When $W \succ 0$, we can express the problem in the equivalent form:

$$p^* := \max_x \; x^T W x \; : \; x_i^2 \leq 1, \;\; i = 1, \ldots, n. \tag{11.2}$$

In this problem, the non-convex constraints are relaxed to convex ones, but the problem is still not convex, as the objective is convex but has to be maximized (if we had a min instead of the max in the above, problem (11.2) would be convex, and in fact, trivial).

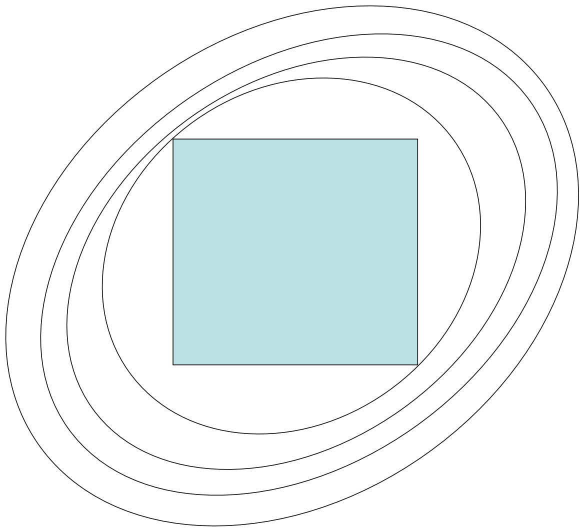

Figure 11.1: Geometric Intuition of Combinatorial Optimization

To see why the problem above is equivalent to our original problem, observe that the maximum of a convex function f over an arbitrary set $\mathcal{C}$ is the same as the maximum over its convex hull. Indeed, for every point $z \in \mathbf{Co}\mathcal{C}$, there exist points $x_k \in \mathcal{C}$ and a vector $\theta \geq 0$, $\sum_k \theta_k = 1$, such that $z = \sum_k \theta_k x_k$. By convexity of f, we have

$$f(z) = f(\sum_k \theta_k x_k) \leq \sum_k \theta_k f(x_k) \leq \max_{x \in \mathcal{C}} \ f(x),$$

hence,

$$\max_{z \in \mathbf{Co}\mathcal{C}} \ f(z) \leq \max_{x \in \mathcal{C}} \ f(x),$$

while the converse inequality holds trivially since $\mathcal{C} \subseteq \mathbf{Co}\mathcal{C}$.

Geometric intepretation

With the inequality form in place, due to our assumption $W \succ 0$, we can interpret our problem geometrically, as follows.

For $t \geq 0$, define the ellipsoid $\mathcal{E}_t = \{x : x^T W x \leq t\}$. It turns out that the problem (11.2) can be restated as

$$\min_{x,t} \ t \ : \ \mathcal{E}_t \supseteq \mathcal{B}, \tag{11.3}$$

where $\mathcal{B}$ is the unit ball for the l_∞-norm:

$$\mathcal{B} := \left\{ x \ : \ x_i^2 \leq 1, \ \ i = 1, \ldots, n \right\}.$$

The geometric interpretation is therefore that we are seeking to deflate ellipsoids $\mathcal{E}_t$ (of shape determined by W, and size determined by t) so that they contain the unit ball $\mathcal{B}$. The deflation process stops when the ellipsoids touch at least one of the vertices of $\mathcal{B}$. This interpretation is displayed in figure 11.1.

11.3.3 Lagrangian dual

The Lagrangian for problem (11.2) may be expressed as

$$\mathcal{L}(x,\lambda) = x^T W x + \sum_{i=1}^{m} \lambda_i (1 - x_i^2)$$

As usual, we have

$$p^* = \max_x \min_{\lambda \geq 0} \mathcal{L}(x,\lambda)$$

and

$$d^* = \min_{\lambda \geq 0} \max_x \mathcal{L}(x,\lambda)$$

The dual objective is given by by

$$g(\lambda) = \max_x \mathcal{L}(x,\lambda) = \max_x x^T (W - D(\lambda))x + \mathbf{Tr}(D(\lambda))$$

where $D(\lambda) = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

For a given symmetric matrix A, we have

$$\max_x x^T A x = \begin{cases} 0 & \text{if } A \preceq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Hence,

$$g(\lambda) = \begin{cases} \mathbf{Tr}\, D(\lambda) & \text{if } W \preceq D(\lambda) \\ +\infty & \text{otherwise.} \end{cases}$$

Therefore,

$$d^* = \min_D \mathbf{Tr}\, D \ : \ D \text{ diagonal}, \ D \succeq W. \tag{11.4}$$

The above problem is an SDP, which provides an upper bound for our original (non-convex) problem.

The Lagrangian relaxation above has a geometric interpretation in terms of the formulation (11.3). For a diagonal p.d. matrix D, we define the ellipsoid $\mathcal{E}_D = \{x : x^T D x \leq \mathbf{Tr}\, D\}$. This ellipsoid is oriented parallel to the axes, and in addition, by construction it contains the unit ball $\mathcal{B}$, since

$$x^T D x = \sum_i D_{ii} x_i^2 = \sum_i D_{ii} = \mathbf{Tr}\, D$$

This means that if, instead of solving the (hard) geometric problem (11.3), we instead seek to solve

$$\min_t t \ : \ \mathcal{E}_t \supseteq \mathcal{E}(D), \tag{11.5}$$

we will obtain an upper bound on the original problem.

The condition $\mathcal{E}_t \supseteq \mathcal{E}(D)$ is equivalent to

$$t \geq \max_x \{x^T W x \ : \ x^T D x \leq \mathbf{Tr}\, D\} = (\mathbf{Tr}\, D)\lambda_{\max}(D^{-1/2} W D^{-1/2}),$$

where $\lambda_{\max}$ denotes the largest eigenvalue. In turn, the above is the same as $tD \succeq (\mathbf{Tr}\, D)W$. By homogeneity with respect to D in problem (11.5), we can assume $t = \mathbf{Tr}\, D$, hence the latter problem becomes

$$\min_{t,D} \ t \ : \ D \text{ diagonal}, \ t = \mathbf{Tr}\, D, \ \ D \succeq W,$$

which is exactly the same as the dual (11.4).

Geometrically, the SDP relaxation is based on inserting and object simple to handle (an ellipsoid parallel to the axes) between the ball $\mathcal{B}$ and the ellipsoid $\mathcal{E}_t$.

11.3.4 The bidual

Taking the dual a second time will not, in general, result in the original problem, since the latter is not convex.

In this case, taking the dual of the dual gives us another problem, which is equivalent to the dual. To obtain the bidual, we express the dual in an unconstrained way, as usual:

$$d^* = \min_{D \text{ diagonal}} \ \max_{X \succeq 0} \ \mathcal{L}(D, X)$$

where $\mathcal{L}(D, X) := \mathbf{Tr}(D) + \mathbf{Tr}(X(W - D))$. The bidual is

$$p^{**} := \max_{X \succeq 0} \ \min_{D \text{ diagonal}} \ \mathcal{L}(D, X).$$

Note that the problem of computing d^* is convex, and satisfies Slater's condition (strict feasibility), hence strong duality holds:

$$p^{**} = d^*.$$

The (bi-)dual function is here

$$g(X) := \min_{D \text{ diagonal}} \ \mathcal{L}(D, X) = \begin{cases} \mathbf{Tr}\, WX & \text{if } X_{ii} = 1, \ i = 1, \ldots, n, \\ -\infty & \text{otherwise.} \end{cases}$$

The bidual has the expression:

$$p^{**} = d^* = \max_X \ \mathbf{Tr}\, WX \ : \ X \succeq 0, \ \ X_{ii} = 1, \ \ i = 1, \ldots, n. \tag{11.6}$$

We can see that analogy between the primal problem p^* and the dual d^* as obtained in (11.6) by rewriting the primal as a rank-constrained problem:

$$p^* = \max_X \ \mathbf{Tr}\, WX \ : \ X \succeq 0, \ \ X_{ii} = 1, \ \ i = 1, \ldots, n, \ \ \mathbf{Rank}\, X = 1. \tag{11.7}$$

This is obtained by setting $X = xx^T$. Relaxing (that is, ignoring) the rank constraint in the above leads directly to the upper bound (11.6).

It has been shown that the quality of the approximation obtained by the SDP is independent of problem size. Precisely,

$$\frac{2}{\pi} d^* \leq p^* \leq d^*.$$

In fact, one can find in polynomial time, an $x \in \{-1, +1\}^n$ such that $x^T W x \geq (2/\pi) d^*$.