

Lecture 13: SDP Duality

Lecturer: Laurent El Ghaoui

Reading assignment: Chapter 5 of BV.

13.1 Direct approach

13.1.1 Primal problem

Consider the SDP in standard form:

$$p^* := \max_X \langle C, X \rangle : \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \quad X \succeq 0, \quad (13.1)$$

where C, A_i are given symmetric matrices, $\langle A, B \rangle = \mathbf{Tr} AB$ denotes the scalar product between two symmetric matrices, and $b \in \mathbf{R}^m$ is given.

13.1.2 Dual problem

At first glance, the problem (13.1) is not amenable to the duality theory developed so far, since the constraint $X \succeq 0$ is not a scalar one.

Minimum eigenvalue representation. We develop a dual based on a representation of the problem via the minimum eigenvalue, as

$$p^* = \max_X \langle C, X \rangle : \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \quad \lambda_{\min}(X) \geq 0, \quad (13.2)$$

where we have used the minimum eigenvalue function of a symmetric matrix A , given by

$$\lambda_{\min}(A) := \min_Y \langle Y, A \rangle : Y \succeq 0, \quad \mathbf{Tr} Y = 1 \quad (13.3)$$

to represent the positive semi-definiteness condition $X \succeq 0$ in the SDP. The proof of the above representation of the minimum eigenvalue can be obtained by first showing that we can without loss of generality assume that A is diagonal, and noticing that we can then restrict Y to be diagonal as well. Note that the above representation proves that λ_{\min} is concave, so problem (13.2) is indeed convex as written.

Lagrangian and dual function. The Lagrangian for the *maximization* problem (13.2) is

$$\begin{aligned}\mathcal{L}(X, \lambda, \nu) &= \langle C, X \rangle + \sum_{i=1}^m \nu_i (b_i - \langle A_i, X \rangle) + \lambda \cdot \lambda_{\min}(X) \\ &= \nu^T b + \langle C - \sum_{i=1}^m \nu_i A_i, X \rangle + \lambda \cdot \lambda_{\min}(X),\end{aligned}$$

where $\nu \in \mathbf{R}^m$ and $\lambda \geq 0$ are the dual variables. The corresponding dual function

$$g(\lambda, \nu) := \max_X \mathcal{L}(X, \lambda, \nu).$$

involves the following subproblem, in which $Z = C - \sum_{i=1}^m \nu_i A_i$ and $\lambda \geq 0$ are given:

$$G(\lambda, Z) := \max_X \langle Z, X \rangle + \lambda \lambda_{\min}(X). \quad (13.4)$$

For fixed $\lambda \geq 0$, the function $G(\cdot, \lambda)$ is the conjugate of the convex function $-\lambda \lambda_{\min}$.

We have

$$\begin{aligned}G(\lambda, Z) &= \max_X \left(\langle Z, X \rangle + \lambda \min_{Y \succeq 0, \text{Tr} Y = 1} \langle Y, X \rangle \right) \\ &= \max_X \min_{Y \succeq 0, \text{Tr} Y = 1} \langle Z + \lambda Y, X \rangle \text{ [eqn. (13.3)]} \\ &= \max_X \min_{Y \succeq 0, \text{Tr} Y = \lambda} \langle Z + Y, X \rangle \text{ [replace } Y \text{ by } \lambda Y\text{]} \\ &\leq \min_{Y \succeq 0, \text{Tr} Y = \lambda} \max_X \langle Z + Y, X \rangle \text{ [minimax inequality]} \\ &= \min_{Y \succeq 0, \text{Tr} Y = \lambda} \begin{cases} 0 & \text{if } Z + Y = 0 \\ +\infty & \text{otherwise} \end{cases} \\ &= \underline{G}(\lambda, Z),\end{aligned}$$

where

$$\underline{G}(\lambda, Z) := \begin{cases} 0 & \text{if } \mathbf{Tr} Z = -\lambda \leq 0, \quad Z \preceq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

We now show that $G(\lambda, Z) = \underline{G}(\lambda, Z)$. To prove this, first note that $G(\lambda, Z) \geq 0$ (since $X = 0$ is feasible). This shows that $G(\lambda, Z)$ itself is 0 if (λ, Z) is in the domain of \underline{G} . Conversely, if $\mathbf{Tr} Z + \lambda \neq 0$, choosing $X = \epsilon t I$ with $\epsilon = \mathbf{sign}(\mathbf{Tr} Z + \lambda)$ and $t \rightarrow +\infty$ implies $G(\lambda, Z) = +\infty$. Likewise, if $\mathbf{Tr} Z + \lambda = 0$ and $\lambda < 0$, we choose $X = t I$ with $t \rightarrow +\infty$, with the same result. Finally, if $\mathbf{Tr} Z = -\lambda \leq 0$ but $\lambda_{\max}(Z) > 0$, choose $X = t u u^T$, with u a unit-norm eigenvector corresponding to the largest eigenvalue of Z , and $t \rightarrow +\infty$. Here, we have

$$\langle Z, X \rangle + \lambda \lambda_{\min}(X) = t(\lambda_{\max}(Z) + \lambda) \rightarrow +\infty,$$

where we have exploited the fact that $\lambda_{\max}(Z) + \lambda > \lambda \geq 0$.

Dual problem. Coming back to the Lagrangian, we need to apply our result to $Z = C - \sum_{i=1}^m \nu_i A_i$. The dual function is

$$g(\lambda, \nu) = \begin{cases} 0 & \text{if } Z := C - \sum_{i=1}^m \nu_i A_i \preceq 0, \quad \mathbf{Tr} Z = -\lambda \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

We obtain the dual problem

$$d^* = \min_{\lambda, \nu, Z} \nu^T b : Z = C - \sum_{i=1}^m \nu_i A_i \preceq 0, \quad \mathbf{Tr} Z = -\lambda \geq 0,$$

or, after elimination of λ , and noticing that $Z \preceq 0$ implies $\mathbf{Tr} Z \leq 0$:

$$d^* = \min_{\nu} \nu^T b : \sum_{i=1}^m \nu_i A_i \succeq C \quad (13.5)$$

The dual problem is also an SDP, in standard inequality form.

13.2 Conic approach

13.2.1 Conic Lagrangian

The same dual can be obtained with the “conic” Lagrangian

$$\mathcal{L}(X, \nu, Y) := \langle C, X \rangle + \sum_{i=1}^m \nu_i (b_i - \langle A_i, X \rangle) + \langle Y, X \rangle,$$

where now we associate a *matrix* dual variable Y to the constraint $X \succeq 0$.

Let us check that the Lagrangian above “works”, in the sense that we can represent the constrained maximization problem (13.1) as an unconstrained, maximin problem:

$$p^* = \max_X \min_{Y \succeq 0} \mathcal{L}(X, \nu, Y).$$

We need to check that, for an arbitrary matrix Z , we have

$$\min_{Y \succeq 0} \langle Y, X \rangle = \begin{cases} 0 & \text{if } X \succeq 0 \\ -\infty & \text{otherwise.} \end{cases} \quad (13.6)$$

This is an immediate consequence of the following:

$$\min_{Y \succeq 0} \langle Y, X \rangle = \min_{t \geq 0} \min_{Y \succeq 0, \mathbf{Tr} Y = t} \langle Y, X \rangle = \min_{t \geq 0} t \lambda_{\min}(X),$$

where we have exploited the representation of the minimum eigenvalue given in (13.3). The geometric interpretation is that the cone of positive-semidefinite matrices has a 90° angle at the origin.

13.2.2 Dual problem

The minimax inequality then implies

$$p^* \leq d^* := \min_{\nu, Y \succeq 0} \max_X \mathcal{L}(X, \nu, Y).$$

The corresponding dual function is

$$g(Y, \nu) = \max_X \mathcal{L}(X, \nu, Y) = \begin{cases} \nu^T b & \text{if } C - \sum_{i=1}^m \nu_i A_i + Y = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem then writes

$$d^* = \min_{\nu, Y \succeq 0} g(Y, \nu) = \min_{\nu, Y \succeq 0} \nu^T b : C - \sum_{i=1}^m \nu_i A_i = -Y \preceq 0.$$

After elimination of the variable Y , we find the same problem as before, namely (13.5).

13.3 Weak and strong duality

13.3.1 Weak duality

For the maximization problem (13.2), weak duality states that $p^* \leq d^*$. Note that the fact that weak duality inequality

$$\nu^T b \geq \langle C, X \rangle$$

holds for any primal-dual feasible pair (X, ν) , is a direct consequence of (13.6).

13.3.2 Strong duality

From Slater's theorem, strong duality will hold if the primal problem is strictly feasible, that is, if there exist $X \succ 0$ such that $\langle A_i, X \rangle = b_i$, $i = 1, \dots, m$.

Using the same approach as above, one can show that the dual of problem (13.5) is precisely the primal problem (13.2). Hence, if the dual problem is strictly feasible, then strong duality also holds. Recall that we say that a problem is *attained* if its optimal set is not empty. It turns out that if both problems are strictly feasible, then both problems are attained.

A strong duality theorem. The following theorem summarizes our results.

Theorem 1. Consider the SDP

$$p^* := \max_X \langle C, X \rangle : \langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \quad X \succeq 0$$

and its dual

$$d^* = \min_{\nu} \nu^T b : \sum_{i=1}^m \nu_i A_i \succeq C.$$

The following holds:

- Duality is symmetric, in the sense that the dual of the dual is the primal.
- Weak duality always holds: $p^* \leq d^*$, so that, for any primal-dual feasible pair (X, ν) , we have $\nu^T b \geq \langle C, X \rangle$.
- If the primal (resp. dual) problem is bounded above (resp. below), and strictly feasible, then $p^* = d^*$ and the dual (resp. primal) is attained.
- If both problems are strictly feasible, then $p^* = d^*$ and both problems are attained.

13.4 Examples

13.4.1 An SDP where strong duality fails

Contrarily to linear optimization problems, SDPs can fail to have a zero duality gap, even when they are feasible. Consider the example:

$$p^* = \min_x x_2 : \begin{pmatrix} x_2 + 1 & 0 & 0 \\ 0 & x_1 & x_2 \\ 0 & x_2 & 0 \end{pmatrix} \succeq 0.$$

Any primal feasible x satisfies $x_2 = 0$. Indeed, positive-semidefiniteness of the lower-right 2×2 block in the LMI of the above problem writes, using Schur complements, as $x_1 \leq 0$, $x_2^2 \leq 0$. Hence, we have $p^* = 0$. The dual is

$$d^* = \max_{Y \in \mathcal{S}^3} -Y_{11} : Y \succeq 0, Y_{22} = 0, 1 - Y_{11} - 2Y_{23} = 0.$$

Any dual feasible Y satisfies $Y_{23} = 0$ (since $Y_{22} = 0$), thus $Y_{11} = -1 = d^*$.

13.4.2 An eigenvalue problem

For a matrix $A \in \mathcal{S}_+^n$, we consider the SDP

$$p^* = \max_X \langle A, X \rangle : \mathbf{Tr} X = 1, X \succeq 0. \quad (13.7)$$

The associated Lagrangian, using the conic approach, is

$$\mathcal{L}(X, Y, \nu) = \langle A, X \rangle + \nu(1 - \mathbf{Tr} X) + \langle Y, X \rangle,$$

with the matrix dual variable $Y \succeq 0$, while $\nu \in \mathbf{R}$ is free.

The dual function is

$$g(Y, \nu) = \max_X \mathcal{L}(X, Y, \nu) = \begin{cases} \nu & \text{if } \nu I = Y + A \\ +\infty & \text{otherwise.} \end{cases}$$

We obtain the dual problem

$$p^* \leq d^* = \min_{\nu, Y} \nu : Y + A = \lambda I, Y \succeq 0.$$

Eliminating Y leads to

$$d^* = \min_{\nu} \{\nu : \nu I \succeq A\} = \lambda_{\max}(A).$$

Both the primal and dual problems are strictly feasible, so $p^* = d^*$, and both values are attained. This proves the representation (13.7) for the largest eigenvalue of A .

13.4.3 SDP relaxations for a non-convex quadratic problem

In lectures 7 and 11, we have seen two kinds of relaxation for the non-convex problem

$$p^* := \max_x x^T W x : x_i^2 \leq 1, i = 1, \dots, n,$$

where the symmetric matrix $W \in \mathcal{S}^n$ is given.

One relaxation is based on a standard relaxation of the constraints, and leads to (see lecture 11)

$$p^* \leq d^{\text{lag}} := \min_D \mathbf{Tr} D : D \succeq W, D \text{ diagonal.} \quad (13.8)$$

Another relaxation (lecture 7) involved expressing the problem as an SDP with rank constraints on the a matrix $X = xx^T$:

$$d^{\text{rank}} := \max_X \langle W, X \rangle : X \succeq 0, X_{ii} = 1, i = 1, \dots, m.$$

Let us examine the dual of the first relaxation (13.8). We note that the problem is strictly feasible, so strong duality holds. Using the conic approach, we have

$$\begin{aligned} d^{\text{lag}} &:= \min_D \max_{Y \succeq 0} \mathbf{Tr} D + \langle Y, W - D \rangle \\ &= \max_{Y \succeq 0} \min_D \mathbf{Tr} D + \langle Y, W - D \rangle \\ &= \max_Y \langle Y, W \rangle : Y \succeq 0, Y_{ii} = 1, i = 1, \dots, m \\ &= d^{\text{rank}}. \end{aligned}$$

This shows that both Lagrange and rank relaxations give the same value, and are dual of each other.

In general, for arbitrary non-convex quadratic problems, the rank relaxation can be shown to be always better than the Lagrange relaxation, as the former is the (conic) dual to the latter. If either is strictly feasible, then they have the same optimal value.