Lecture 13 — October 11

Lecturer: Caramanis & Sanghavi *Scribe: Tong Zhao, Yingzhe Li*

13.1 Topics Covered Last Time

- Duality for Convex Quadratic
- Robust Optimization

In the last lecture, we covered duality for convex quadratic problems. We introduced Slater's condition, which essentially says that if the functions defining the objective and the constraints are convex, and if there is a point that is strictly feasible for all non-polyhedral constraints, then strong duality holds. We also discussed complementary slackness, and showed how we can use this, to solve the so-called water-filling problem. Then, we introduced the notion of Robust Optimization. We showed that the nominal problem, when robustified, seems to have infinitely many constraints. We showed how this can often be reformulated as a (finite dimensional) convex optimization problem, using duality. We saw several examples of Robust Optimization.

In this lecture, we study conic duality and convex conic programming, which is a powerful generalization of linear programming.

13.2 Conic Duality

13.2.1 From Linear to Cone Programming

In the standard linear programming (LP) problem

$$\begin{aligned} &\min c^{'}x \\ &\text{s.t. } Ax - b \geq 0 \text{ (A: } m \times n). \end{aligned}$$

The constraint inequality $Ax - b \geq 0$ in LP is an inequality between vector, and it holds when $Ax - b \in R^n_+$. Thus in LP, for any vector $x \in R^n$, we have:

$$x \geq 0 \Leftrightarrow x \in R^n_+.$$

While in the conic duality problems, we want to define the "vector inequality" based on a convex cone K, instead of R^n_+. Then, the vector inequality $\succeq$ is completely identified by the set **K** of $\succeq -nonnegative$ vectors, which leads to the following definition:

Definition 1. *(Vector Ineuqality) For any vector in Euclidean space **E**, the vector inequality $\succeq$ is completely identified by the set **K** of $\succeq -nonnegative$ vectors, where:*

$$\boldsymbol{K} = \{a \in \boldsymbol{E} : a \succeq 0\}.$$

And for any vectors a, b,

$$\begin{aligned} a \succeq b &\Leftrightarrow a - b \succeq 0 \\ &\Leftrightarrow a - b \in \boldsymbol{K} \end{aligned} \tag{13.1}$$

In Definition 1, the set **K** should be a convex cone, which needs to satisfy the following conditions:

1. **K** is non empty and closed under addition:
$$a, a^{'} \in \mathbf{K} \Rightarrow a + a^{'} \in \mathbf{K};$$

2. **K** is a conic set:
$$a \in \mathbf{K}, \lambda \geq 0 \Rightarrow \lambda a \in \mathbf{K};$$

3. **K** is pointed:
$$a \in \mathbf{K} \text{ and } -a \in \mathbf{K} \Rightarrow a = 0.$$

From now on, for any vector x and convex set **K**, we will use $\succeq_{\mathbf{K}}$ equivalently as:

$$x \succeq_{\mathbf{K}} 0 \Leftrightarrow x \in \mathbf{K}.$$

Next, we will provide some examples of the convex cones that can be applied to the coni duality problems:

- (Second Order Cone) The second order cone **K** is defined as for any vector $x \in \mathrm{R}^{n+1}$, $x \in \mathbf{K}$ if and only if:
$$x^{n+1} \geq (x_1^2 + x_2^2 + ... + x_{n^2})^{\frac{1}{2}}.$$
Specifically, for the 3-D second order cone, or ice-cream cone, is shown in the following figure:

- (Semi-Definite Cone) The semi-definite cone is defined as:
$$\begin{aligned} S_+^n &= \{X \in S^n, \lambda_i(X) \geq 0\} \\ &= \{X \in S^n, z^T X z \geq 0, \forall z \in R^n\}, \end{aligned}$$
where $\lambda_i(X)$ denotes the i-th eigenvalue of X.

  It is easy to see using the second definition above, that the semi-definte cone is a convex set. It is interesting to consider what **separation** means in this space. That is, recall we have

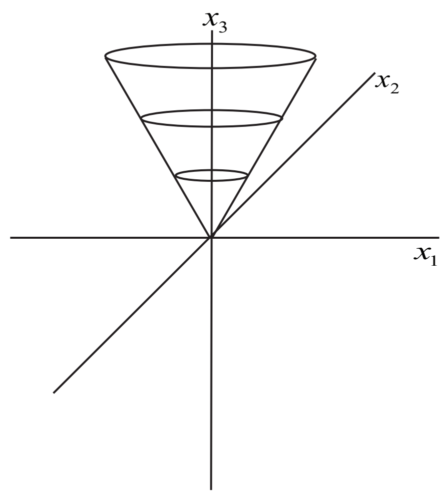

Figure 13.1. 3-D ice cream cone

shown in an earlier lecture, that if C is any closed convex set, and x a point not in C, then there is a hyperplane H such that $C \subseteq H^+$ and $x \in \mathrm{int}H^-$. We apply this to the semidefinite cone:

Example. Suppose we have a symmetric matrix $M \notin S^n_+$. Then there exists a symmetric vector $N \subset S^n$ and $\beta \in R$ which define the hyperplane $H = \{X : \langle N, X\rangle = \beta\}$ such that:

$$S^n_+ \subseteq H^+, M \in \mathrm{int}H^-.$$

Now, generically, a hyperplane is defined by its normal, and its offset, N and β, respectively. Since $S^n_+ \subseteq H^+$, and since S^n_+ is a cone (i.e., we can scale it by any nonnegative scalar) we can immediately see that we must have $\beta = 0$. Therefore we need only find N. Now, M is not semidefinite. Therefore, by definition, it must have an eigenvector with strictly negative eigenvalue. Let x denote this eigenvector, and $\lambda < 0$ the associated eigenvector. Then it follows that:

$$\begin{aligned}\langle M, xx^T\rangle &= \sum M_{ij}(xx^T)_{ij} = \sum M_{ij}x_ix_j \\ &= x^T Mx = x^T\lambda x \\ &= \lambda\|x\|^2 < 0.\end{aligned}$$

Meanwhile, $\langle A, xx^\top\rangle \geq 0$ for all $A \in S^n_+$, by definition. Therefore, $N = xx^\top$ is the normal we need to define our hyperplane.

13.2.2 Cone Programming

Having defined some of the motivation and a few examples of cones, we now define conic programming. It can be considered as a straightforward generalization of linear programming. Indeed, recall the linear programming problem:

$$\begin{aligned} &\text{min } c^{'}x \\ &\text{s.t. } Ax = b, \\ &\quad x \in R^n_+. \end{aligned}$$

We can define the cone programming problem as:

Definition 2. ***(Cone Programming):***

$$\begin{aligned} &\textit{min } \langle c, X \rangle \\ &\textit{s.t. } \langle A_i, X \rangle = b_i, \\ &\quad X \succeq_k 0. \end{aligned}$$

Recall from previous section, $X \succeq_k 0 \Leftrightarrow X \in K$, and $X \succeq_k Y \Leftrightarrow X - Y \succeq_k 0$. The semi-definite cone is a very popular example of a cone that is often used. We will see several uses of it in this and subsequent lectures.

Next, we analyze the dual functions of cone programming. First let's recall the LP problem:

$$\begin{aligned} &\text{min } c^{'}x \\ &\text{s.t. } b - Ax \geq 0. \end{aligned} \tag{13.2}$$

The dual function of (13.2) is defined as:

$$\begin{aligned} q(\lambda) &= \inf_x L(x, \lambda) \\ &= \inf_x c^{'}x + \lambda^T(Ax - b), \end{aligned} \tag{13.3}$$

where $\lambda \geq 0$.

Similar to LP, consider the following cone programming problem:

$$\begin{aligned} &\text{(P): min } \langle c, x \rangle \\ &\qquad \text{s.t. } B - \mathcal{A}x \succeq_K 0. \end{aligned} \tag{13.4}$$

The dual function of (13.4) is:

$$q(\Lambda) = \inf_x \langle c, x \rangle + \langle \Lambda, \mathcal{A}X - B \rangle. \tag{13.5}$$

Recall that in linear programming, we restricted λ to be nonnegative, i.e., in the positive outhunt $\mathbb{R}^n_+$. The reason for that restriction was that in the formulation of the Lagrangian, we had:

$$\min_x \max_\lambda : c^\top x + \lambda(Ax - b).$$

If x is chosen so that the constraint $Ax - b \leq 0$ is violated, then a nonnegative λ could drive the cost to $+\infty$. Depending on the sign of the constraints, the restriction on λ is chosen analogously. To this end, define the **dual** of a cone as follows:

$$K^* = \{Z : \langle Z, X \rangle \geq 0, \forall x \in K\}$$

The relation between K_* and K is illustrated in the following figure. The case of $K = R^n_+$, and

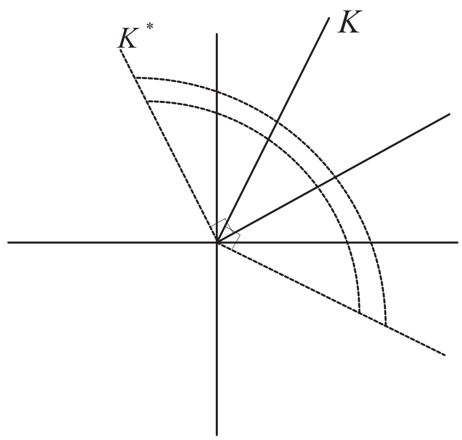

Figure 13.2. Iluustration of dual cone

also $K = S^n_+$, are special, as these are what is known as *self-dual*. That is, it can be derived (try it!) that:

$$\begin{aligned} K^* &= \{Z : \langle Z, X \rangle \geq 0, \forall x \geq 0\} \\ &= R^n_+ \end{aligned}$$

and similarly for $K = S^n_+$.

The dual cone is the key object we want. We now turn to the derivation of the conic dual.

13.2.3 Dual Problem of Cone Programming

Next, we will derive the dual problem of (13.4). Similar to LP, we can derive the weak duality of cone programming through Lagrange multipliers. Let's write the primal, as above:

$$\begin{aligned} \min : \quad & \langle c, x \rangle \\ \text{s.t.} : \quad & B - \mathcal{A}x \succeq_K 0. \end{aligned}$$

Now, using the dual, we can write this exactly, as

$$\min_x \max_{\Lambda \in K^*} : \; \langle c, x \rangle + \langle \Lambda, \mathcal{A}x - B \rangle.$$

Note that by the definition of K^*, if x is chosen in a way that violates the constraint, then the inner max can take the value to $+\infty$. Therefore, this $\min\max$ is equivalent to the original formulation. Flipping the min and the max gives us weak duality. Putting all this together, we get:

$$\begin{aligned}
&\min\ \langle c, x\rangle \\
&\text{s.t. } B - \mathcal{A}x \succeq_K 0 \\
&\quad = \min_x : \max_{\Lambda\in K^*} \langle c, x\rangle + \langle \Lambda, \mathcal{A}x - B\rangle \\
&\quad \geq \max_{\Lambda\in K^*} : \min_x \langle c, x\rangle + \langle \Lambda, \mathcal{A}x - B\rangle \\
&\quad = \max_{\Lambda\in -K^*} : \min_x \langle c - \mathcal{A}^*\Lambda, x\rangle + \langle \Lambda, B\rangle
\end{aligned} \tag{13.6}$$

where A^* is the adjoint operator of A. For real matrices, the adjoint operator is merely the transpose, but the adjoint is defined much more generally, for any two (linearly) paired spaces. See any text book on linear operators or functional analysis for many more details on this. In short, the adjoint of a linear operator A is defined by the relationship:

$$\langle Ax, y\rangle = \langle x, A^*y\rangle, \quad \forall x, y.$$

In (13.6), it can be seen that $q(\Lambda) = \min_x \langle c - \mathcal{A}^*\Lambda, x\rangle + \langle \Lambda, B\rangle$, which is only a function of Λ. The domain for $q(\Lambda)$ should be:

$$\begin{aligned}
\text{dom q: } &\Lambda \in -K^*, \\
&\mathcal{A}^*\Lambda = C.
\end{aligned} \tag{13.7}$$

This follows from the fact that x is unconstrained, thus if $\mathcal{A}^*\Lambda \neq C$, can choose x to make $q(\Lambda) \to -\infty$, under which $\Lambda \notin$ dom(q). Thus, the dual problem of (13.4) is:

$$\begin{aligned}
\text{(D): max: } &\langle \Lambda, B\rangle \\
\text{s.t.: } &\Lambda \in \text{dom q}.
\end{aligned} \tag{13.8}$$

Next, we will derive the dual of standard form conic prolem (2) in the following corollary.

Corollary 13.1. *For the standard form conic problem:*

$$\begin{aligned}
\text{(P): min } &\langle c, x\rangle \\
\text{s.t. } &\mathcal{A}x = B, \\
&x \succeq_K 0.
\end{aligned} \tag{13.9}$$

the dual of (13.9) is:

$$\begin{aligned}
\text{(D): max: } &\langle \Lambda, B\rangle \\
\text{s.t.: } &c - \langle \mathcal{A}^*, \Lambda\rangle \in K^*.
\end{aligned} \tag{13.10}$$

Proof: For $\forall \lambda \in K^*$ and Λ, the Lagrangian of (13.9) can be written as:

$$L(x, \lambda, \Lambda) = \langle c, x \rangle + \langle \Lambda, B - \mathcal{A}x \rangle - \lambda^T x.$$

λ is added so that

$$\max_{\lambda \in K^*} -\lambda^T x = \begin{cases} 0, & \text{if } x \in K, \\ +\infty, & \text{otherwise.} \end{cases}$$

Thus, we will have:

$$\begin{aligned} \text{(P) } &= \min_x \max_{\Lambda, \lambda \in K^*} L(x, \lambda, \Lambda) \\ &= \min_x \max_{\Lambda, \lambda \in K^*} \langle c, x \rangle + \langle \Lambda, B - \mathcal{A}x \rangle - \lambda^T x \\ &\geq \max_{\Lambda, \lambda \in K^*} \min_x \langle c - \mathcal{A}^* \Lambda - \lambda, x \rangle + \langle \Lambda, B \rangle \end{aligned}$$

while

$$\min_x \langle c - \mathcal{A}^* \Lambda - \lambda, x \rangle + \langle \Lambda, B \rangle = \begin{cases} \langle \Lambda, B \rangle & \text{if } c - \mathcal{A}^* \Lambda - \lambda = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Recall that $\lambda \in K^*$, thus we can derive the dual for standard form conic shown as (13.10). □

From 13.6, we have seen the weak duality holds for cone programming. Next, we will give the following theorem regarding when strong duality holds for cone programming:

Theorem 13.2. *(Strong Duality) For cone programming, if the primal problem is feasible with finite solution, and there exists strictly feasible point, then we can guarantee that (P)=(D).*

In Theorem (13.2), for the constraint that $g(x) \preccurlyeq_K 0$, a strictly feasible point refers the point x which satisfies that $g(x) \prec_K 0$, or equivalently $-g(x) \in \text{int } K$. The detailed proof of Theorem (13.2) can be found in Theorem 1.4.2 of Aharon Ben-Tal and Arkadi Nemirovski's book.

13.3 Semidefinite Programming

Definition 3. *(Semidefinite Programming)In the standard semidefinite programming (SDP) problem, we have $K = S_+^n$, i.e., it is a self-dual Cone*

$$\begin{aligned} \textit{(P): min } & \sum C_i x_i \\ \textit{st } & \sum A_i x_i - B \succeq 0 \end{aligned}$$

Lemma 13.3. *The dual of the SDP is*

$$\begin{aligned} \textit{(D): max } & \langle \Lambda, B \rangle \\ \textit{st } & \langle \Lambda, A_i \rangle = C_i, \Lambda \succeq 0 \end{aligned}$$

Proof: *Check the Dual with Primal*

$$\begin{aligned} Primal &= \min_x : \max_{\Lambda \in K^*} : \sum(\langle C_i, x_i\rangle + \langle \Lambda, A_i x_i - B_i\rangle) \\ &\geq \max_{\Lambda \in -K^*} : \min_x : \sum(\langle C_i, x_i\rangle + \langle -\Lambda, A_i x_i - B_i\rangle) \\ &= \max_{\Lambda \in -K^*} : \min_x : \sum(\langle C_i - A_i^* \Lambda, x_i\rangle + \langle x_i, B_i\rangle) \end{aligned}$$

where A^* *is adjoint A for minimization of x function, and we denote*

$$q(\lambda) = \min_x : \sum(\langle C_i - A_i^* \Lambda, x_i\rangle + \langle x_i, B_i\rangle)$$

so that we obtain,

$$\begin{aligned} &(D): \max \langle \Lambda, B\rangle \\ &\quad st\ = C_i, \Lambda\langle \Lambda, A_i\rangle \succeq 0 \end{aligned}$$

□

13.3.1 Application of semidefinite programming

Problem 1.
Given a set $C \subseteq R^n$*, find minimum volume of ellipsoid* ε*, so that* $C \subseteq \varepsilon$

$$\varepsilon = \{x : x^T Q x + Q^T + C \leq 0\} = \{x : \|Ax + b\|_2 \leq 1\} \tag{13.11}$$

Fact 1. *Volume* $\varepsilon \propto \det(A^{-1})$

Proof: *Check with the fact as below:*
Without loss in generality, we can assume that:
an arbitrarily oriented ellipsoid, centered at v, is defined by the equation.

$$(x - v)^T B^{-1} (x - v) = 1$$

where B is positive semidefinite with $B = AA^T$ *and x,v are vectors.*
Now, we can write the equation above with A in fact case:

$$(x - v)^T (AA^T)^{-1} (x - v) = 1$$

For the quadratic part of the above equation, we have

$$\varepsilon \propto \det(A^{-1})$$

□

Fact 2. $\log(\det(A^{-1}))$ *is a convex function of A*

So that the problem is converted to

$$\begin{aligned} \textit{min: } & \det(A^{-1}) \\ \textit{st: } & \|Ax+b\| \leq 1, \forall x \in C. \end{aligned}$$

Problem 2. *Considering the problem:*

$$\begin{aligned} \max_{x\in C} : & \|Ax+b\| \\ \textit{st: } & \|Ax+b\| \leq 1, \forall x \in C. \end{aligned} \tag{13.12}$$

This is not convex optimization problem, and not tractable for the solution, i.e., NP hard. Similarly, for the problem considering below:

$$\begin{aligned} & \min_{x\in C} : x^T A_0 x + 2b_0^T x + C_0 \\ & st : x^T A_1 x + 2b_1^T x + C_1 \leq 0. \end{aligned} \tag{13.13}$$

where A_0, A_1 are not positive semidefinite and this is also a non-convex optimization

Problem 3. *Consider the problem below:*

$$q(\lambda) = \inf_x f(x) + \lambda g(x) = \min[linear functions] \tag{13.14}$$

Obviously, $q(\lambda)$ is always a concave function. Now consider the dual:

$$\begin{aligned} \textit{(D): max: } & q(\lambda) \\ \textit{st: } & \lambda \geq 0, \\ & \lambda \in \textit{dom}(q). \end{aligned}$$

We obtain that the dual of a concave function is a convex function. This is because the max function of a negative concave function is a convex function.

$$q(\lambda) = \inf_x x^T(A_0 + \lambda A_1)x + 2(b_0 + \lambda b_1)_T x + (c_0 + \lambda c_1)$$

where

$$\begin{aligned} \textit{dom(q)} \ : & A_0 + \lambda A_1 \geq 0; \\ & b_0 + \lambda b_1 \in R(A_0 + \lambda A_1) \end{aligned}$$

In this way, we obtain that

$$x^* = -(A_0 + \lambda A_1)^+(b_0 + \lambda b_1), \tag{13.15}$$

and $(A_0 + \lambda A_1)^+$ is known as "Moore-Penrose Pseudo Inverse". So that the dual is now become

$$\begin{aligned} \textit{(D): max: } & (C_0 + \lambda C_1) - (b_0 + \lambda b_1)^T (A_0 + \lambda A_1)^+ (b_0 + \lambda b_1) \\ \textit{st: } & \lambda \geq 0 \\ & b_0 + \lambda b_1 \in R(A_0 + \lambda A_1) \\ & A_0 + \lambda A_1 \succeq 0 \end{aligned} \tag{13.16}$$

where R denoted as the Range of $A_0 + \lambda A_1$

In the following section, we will concentrate on Schur Complement, which helps us solve the dual function in SDP.

13.4 Schur Complement

Definition 4. *Lemma on Schur Complement. Let X, defined as:*

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

If A is invertible and $\det(A) \neq 0$*, then Schur Complement of A in X is*

$$S = C - B^T A^{-1} B.$$

Theorem 13.4. $X \succ 0$ *iff* $A \succ 0$ *and* $S \succ 0$*; If* $A \succ 0$*, then* $X \succeq 0$ *iff* $S \succ 0$.

Proof: *The positive semidefiniteness of X is equivalent to the fact that: For* $\forall x \in R^{\mathrm{k}}, y \in R^{\mathrm{l}}$,

$$0 \leq \begin{pmatrix} x^T & y^T \end{pmatrix} \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^T A x + 2x^T B y + y^T C y$$

or, which is same, to the fact that: $\forall y \in R^k$,

$$\inf_{x \in R^{\mathrm{k}}} [x^T A x + 2x^T B y + y^T C y] \geq 0.$$

Since A is positive definite by assumption, the infimum in x *can be computed explicitly for every fixed y: the optimal x is* $-A^{-1}By$*, and the optimal value is:*

$$y^T C y - y^T B^T A^{-1} C y = y^T [C - B^T A^{-1} B] y.$$

The positive definiteness/semidefiniteness of A is equivalent to the fact the latter expression is, respectively, positive/nonnegative for every $y \neq 0$*, i.e., to the positive definite-ness/semidefiniteness of the Schur Complement of A in X.* □

Property 1.

$$\det(X) = \det(A) \times \det(S) \tag{13.17}$$

Proof: *Check A,B,C in* 2×2 *matrix, thus:*

$$\det(X) = AC - BB^T,$$

$$\det(S) = \det(C) - \det(B)^2 / \det(A).$$

since A is invertible. Then multiplies $\det(A)$ *at LHS and RHS,*

$$\det(S)\det(A) = \det(C)\det(A) - \det(B)^2 = \det(X).$$

Thus, the statement holds. □

Property 2. *Minimization: if A is positive semidefinite, i.e.,* $A \succ 0$, *then:*

$$\min_x(\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}) = \min_x(x^T A x + y^T B^T x + x^T B y + y^T C y)$$

Obviously, it has optimal value:

$$x^* = -A^{-1}By,$$

with the minimum value of function that is:

$$\textit{min} = y^T S y.$$

13.4.1 SDP with Schur Complement

In this section, we will use Schur complement to analyze the non-convex optimization problem:

$$\begin{aligned} &\min_{x \in C} : x^T A_0 x + 2b_0^T x + C_0 \\ &st : x^T A_1 x + 2b_1^T x + C_1 \le 0. \end{aligned} \tag{13.18}$$

According to Schur Complement and recall the dual of (13.18) given in (13.16), we now introduce γ and denotes: $S = (C_0 + \lambda C_1 - \gamma) - (b_0 + \lambda b_1)^T (A_0 + \lambda A_1)(b_0 + \lambda b_1)$, then we have:

$$\begin{aligned} \text{(D): } \max: &\ \gamma \\ \text{st: } &S \succeq 0, \\ &\lambda \ge 0, \\ &B \in R(A), \\ &A \succeq 0. \end{aligned} \tag{13.19}$$

where A denotes $A_0 + \lambda A_1$, and B denotes $b_0 + \lambda b_1$ for short.

Now we can rewrite the dual as below with Schur Complement:

$$\begin{aligned} \text{(D): } \max: &\ \gamma \\ \text{st: } &\lambda \ge 0, \\ &\begin{bmatrix} A_0 + \lambda A_1 & b_0 + \lambda b_1, \\ b_0 + \lambda b_1^T & C_0 + \lambda C_1 - \gamma \end{bmatrix} \succeq 0. \end{aligned} \tag{13.20}$$

Now recall that:

$$\begin{aligned}\text{(P): } \min: &\ C^\top x\\ \text{st: } &\ Ax - b \succeq_K 0,\\ &\ x \geq 0.\end{aligned}$$

The dual is:

$$\begin{aligned}\text{(D): } \max: &\ \mu^\top b\\ \text{st: } &\ C - \mu^\top A \succeq_K 0,\\ &\ \mu \succeq_{k^*} 0.\end{aligned}$$

Now, let us rewrite the dual above, in order to get it closer to the form we have here, so that we can then easily take the dual. We have:

$$\begin{aligned}\text{(D): -min: } &\ (0,-1)^\top(\lambda,\gamma)\\ \text{st: } &\ \begin{bmatrix} A_1 & b_1 \\ b_1^\top & C_1 \end{bmatrix}\lambda + \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}\gamma - \begin{bmatrix} -A_0 & -b_0 \\ -b_0^\top & -C_0 \end{bmatrix} \succeq 0.\end{aligned} \tag{13.21}$$

Now, this is quite close to the generic form we gave above. and thus we can take the *dual of the dual* problem (13.21). With a bit of work, we see that this is:

$$\begin{aligned}\min: &\ \langle \begin{pmatrix} A_0 & b_0 \\ b_0^T & C_0 \end{pmatrix}, \begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \rangle\\ \text{st: } &\ \langle \begin{pmatrix} A_1 & b_1 \\ b_1^T & C_1 \end{pmatrix}, \begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \rangle \leq 0,\\ &\ X \succeq xx^T.\end{aligned}$$

Multiplying out, we can see that this is equivalent to the optimization problem:

$$\begin{aligned}\min: &\ \langle A_0, X\rangle + 2b_0^T x + C_0\\ \text{st: } &\ \langle A_1, X\rangle + 2b_1^T x + C_1 \leq 0,\\ &\ X \succeq xx^T.\end{aligned} \tag{13.22}$$

Note that, $\langle A_0, xx^T\rangle = x^T A_0 x$ and the problem above is convex optimization. Moreover, note that if we replace the convex constraint $X \succeq xx^T$ by the non-convex constraint, $X = xx^T$, then we precisely recover the primal problem. Thus, the dual of the dual, is a *convex relaxation* of the (non-convex) primal problem. Essentially, the convexification here involves dropping the rank 1 constraint. This is a common procedure, that has powerful implications for obtaining convex relaxations of combinatorial problems. In the next lecture, we discuss the application of this relaxation, and its quality, to the so-called MAXCUT problem.

In general, the relaxation of the primal and the primal do not have the same objective. That is, while weak duality (of course) holds, strong duality fails, primarily due to the non-convexity of the primal. Perhaps surprisingly, this is not true for the special case of minimizing an arbitrary quadratic subject to an arbitrary quadratic constraint. That is, we have the following theorem to show the relation between the primal problem (13.18) and the dual of dual problem (13.22):

Theorem 13.5. *(S-lemma, S-procedure): Primal (13.18)= Dual of dual (13.22).*

This is related to a hidden convexity phenomenon that says the following: Suppose A_0 and A_1 are two $n \times n$ symmetric matrices, possibly not positive semidefinite. Then the set

$$W \triangleq \{(z_1, z_2) = (x^\top A_0 x, x^\top A_1 x),\ x \in \mathbb{R}^n\} \subseteq \mathbb{R}^2,$$

is convex.

Exercise. Show that this is true, in the special (and easier) case when A_0 and A_1 are both positive semidefinite.