Lecture 2

Lecturer: Pablo A. Parrilo   *Scribe: Pablo A. Parrilo*

Notation: The set of real symmetric $n \times n$ matrices is denoted $\mathcal{S}^n$. A matrix $A \in \mathcal{S}^n$ is called *positive semidefinite* if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$, and is called *positive definite* if $x^T A x > 0$ for all nonzero $x \in \mathbb{R}^n$. The set of positive semidefinite matrices is denoted $\mathcal{S}^n_+$ and the set of positive definite matrices is denoted by $\mathcal{S}^n_{++}$. The cone $\mathcal{S}^n_+$ is a proper cone (i.e., closed, convex, pointed, and solid).

1   PSD matrices

There are several equivalent conditions for a matrix to be positive (semi)definite. We present below some of the most useful ones:

Proposition 1 *The following statements are equivalent:*

- *The matrix $A \in \mathcal{S}^n$ is positive semidefinite ($A \succeq 0$).*

- *For all $x \in \mathbb{R}^n$, $x^T A x \geq 0$.*

- *All eigenvalues of A are nonnegative.*

- *All $2^n - 1$ principal minors of A are nonnegative.*

- *There exists a factorization $A = B^T B$.*

For the definite case, we have a similar characterization:

Proposition 2 *The following statements are equivalent:*

- *The matrix $A \in \mathcal{S}^n$ is positive semidefinite ($A \succ 0$).*

- *For all nonzero $x \in \mathbb{R}^n$, $x^T A x > 0$.*

- *All eigenvalues of A are strictly positive.*

- *All n leading principal minors of A are positive.*

- *There exists a factorization $A = B^T B$, with B square and nonsingular.*

Here are some useful additional facts:

- If T is nonsingular, $A \succ 0 \Leftrightarrow T^T A T \succ 0$.

- Schur complement. The following conditions are equivalent:

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succ 0 \quad \Leftrightarrow \quad \left\{ \begin{array}{r} A \succ 0 \\ C - B^T A^{-1} B \succ 0 \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{r} C \succ 0 \\ A - B C^{-1} B^T \succ 0 \end{array} \right.$$

2 Semidefinite programming

Semidefinite programming (SDP) is a specific kind of convex optimization problem (e.g., [VB96, Tod01, BV04]), with very appealing numerical properties. An SDP problem corresponds to the optimization of a linear function subject to matrix inequality constraints.

An SDP problem in standard primal form is written as:

$$
\begin{aligned}
\text{minimize} \quad & C \bullet X \\
\text{subject to} \quad & A_i \bullet X = b_i, \quad i = 1, \ldots, m \\
& X \succeq 0,
\end{aligned}
\tag{1}
$$

where $C, A_i \in \mathcal{S}^n$, and $X \bullet Y := \mathrm{Tr}(XY)$. The matrix $X \in \mathcal{S}^n$ is the variable over which the maximization is performed. The inequality in the second line means that the matrix X must be positive semidefinite, i.e., all its eigenvalues should be greater than or equal to zero. The set of feasible solutions, i.e., the set of matrices X that satisfy the constraints, is always a convex set. In the particular case in which $C = 0$, the problem reduces to whether or not the inequality can be satisfied for some matrix X. In this case, the SDP is referred to as a *feasibility problem*. The convexity of SDP has made it possible to develop sophisticated and reliable analytical and numerical methods to solve them.

A very important feature of SDP problems, from both the theoretical and applied viewpoints, is the associated *duality theory*. For every SDP of the form (1) (usually called the *primal problem*), there is another associated SDP, called the *dual problem*, that can be stated as

$$
\begin{aligned}
\text{maximize} \quad & b^T \mathbf{y} \\
\text{subject to} \quad & \sum_{i=1}^{m} A_i y_i \preceq C,
\end{aligned}
\tag{2}
$$

where $b = (b_1, \ldots, b_m)$, and the vector $\mathbf{y} = (y_1, \ldots, y_m)$ contains the dual decision variables.

The key relationship between the primal and the dual problem is the fact that feasible solutions of one can be used to bound the values of the other problem. Indeed, let X and $\mathbf{y}$ be any two feasible solutions of the primal and dual problems respectively. Then we have the following inequality:

$$
C \bullet X - b^T \mathbf{y} = (C - \sum_{i=1}^{m} A_i y_i) \bullet X \geq 0,
\tag{3}
$$

where the last inequality follows from the fact that the two terms are positive semidefinite matrices. From (1) and (2) we can see that the left hand side of (3) is just the difference between the objective functions of the primal and dual problems. The inequality in (3) tells us that the value of the primal objective function evaluated at any feasible matrix X is always greater than or equal to the value of the dual objective function at any feasible vector $\mathbf{y}$. This property is known as *weak duality*. Thus, we can use any feasible X to compute an upper bound for the optimum of $b^T \mathbf{y}$, and we can also use any feasible $\mathbf{y}$ to compute a lower bound for the optimum of $\mathrm{Tr}(C \cdot X)$. Furthermore, in the case of feasibility problems (i.e., $C = 0$), the dual problem can be used to certify the nonexistence of solutions of the primal. This property will be crucial in our developments.

2.1 Conic duality

A general formulation, discussed briefly during the previous lecture, that unifies LP and SDP (as well as some other classes of optimization problems) is *conic programming*. We will be more careful than usual here (risking being a bit pedantic) in the definition of the respective spaces and mappings. It does not make much of a difference if we are working on $\mathbb{R}^n$ (since we can identify a space and its dual), but it is "good hygiene" to keep these distinctions in mind, and also useful when dealing on more complicated spaces.

We will start with two real vector spaces, S and T, and a linear mapping $\mathcal{A}: S \to T$. Every real vector space has an associated dual space, which is the vector space of real-valued linear functionals. We will denote these dual spaces by S^* and T^*, respectively, and the pairing between an element of a vector space and one of the dual as $\langle \cdot, \cdot \rangle$ (i.e., $f(x) = \langle f, x \rangle$). The *dual mapping* of $\mathcal{A}$ is the unique linear map $\mathcal{A}^*: T^* \to S^*$ defined through the property

$$\langle \mathcal{A}^* y, x \rangle_S = \langle y, \mathcal{A}x \rangle_T \qquad \forall x \in S, y \in T^*.$$

Notice here that the brackets on the left-hand side of the equation represent the pairing in S, and those on the right-hand side correspond to the pairing in T. We can then define the primal-dual pair of (conic) optimization problems:

$$\min \langle c, x \rangle_S \quad \text{s.t. } \begin{cases} \mathcal{A}x & = & b \\ x & \in & \mathcal{K} \end{cases} \qquad \max \langle y, b \rangle_T \quad \text{s.t.} \quad c - \mathcal{A}^* y \in \mathcal{K}^*,$$

where $b \in T$, $c \in S^*$, $\mathcal{K} \subset S$ is a proper cone, and $\mathcal{K}^* \subset S^*$ is the corresponding dual cone. Notice that exactly the same proof presented earlier works here to show weak duality:

$$
\begin{aligned}
\langle c, x \rangle_S - \langle y, b \rangle_T &= \langle c, x \rangle_S - \langle y, \mathcal{A}x \rangle_T \\
&= \langle c, x \rangle_S - \langle \mathcal{A}^* y, x \rangle_S \\
&= \langle c - \mathcal{A}^* y, x \rangle_S \\
&\geq 0.
\end{aligned}
$$

In the usual cases (e.g., LP and SDP), the vector spaces are finite dimensional, and thus isomorphic to their duals. The specific correspondence between these is given through whatever inner product we use.

Among the classes of problems that can be interpreted as particular cases of the general conic formulation we have *linear programs*, *second-order cone programs* (SOCP), and SDP, when we take the cone $\mathcal{K}$ to be the nonnegative orthant $\mathbb{R}^n_+$, the second order cone in n variables, or the PSD cone $\mathcal{S}^n_+$. We have then the following natural inclusion relationship among the different optimization classes.

$$\text{LP} \quad \subseteq \quad \text{SOCP} \quad \subseteq \quad \text{SDP}.$$

2.2 Geometric interpretation: separating hyperplanes

To be written.

2.3 Strong duality in SDP

We have seen that weak duality always holds for conic LP. As opposed to the LP case, strong duality can fail in SDP (and thus, in general conic programming). A nice example is given in [VB96, p. 65], where both the primal and dual problems are feasible, but their optimal values are different (i.e., there is a nonzero finite duality gap).

Nevertheless, under relatively mild constraint qualifications (Slater's condition, equivalent to the existence of strictly feasible primal and dual solutions) that usually satisfied in practice, SDP problems have strong duality, and thus zero duality gap.

There are several geometric interpretations of what causes the failure of strong duality for general SDP problems. A good one is based on the fact that the image of a proper cone under a linear transformation is not necessarily a proper cone. This fact seems quite surprising (or even wrong!) the first time one encounters it, but after a little while it starts being quite reasonable. Can you think of an example where this happens? What property will fail?

It should be mentioned that it is possible to formulate a more complicated SDP dual program (called the "Extended Lagrange-Slater Dual" in [Ram97]) for which strong duality always holds. For details, as well as a comparison with the more general "minimal cone" approach, we refer the reader to [Ram97, RTW97].

3 Applications

There have been *many* applications of SDP in a variety of areas of applied mathematics and engineering. We present here just a few, to give a flavor of what is possible. Many more will follow.

3.1 Lyapunov stability and control

Consider a linear difference equation (i.e., a discrete-time linear system) given by

$$x(k+1) = Ax(k), \qquad x(0) = x_0.$$

It is well-known (and relatively simple to prove) that $x(k)$ converges to zero for all initial conditiosn x_0 iff $|\lambda_i(A)| < 1, i = 1, \ldots n$.

There is a simple characterization of this spectral radius condition in terms of a quadratic *Lyapunov function* $V(x(k)) = x(k)^T P x(k)$:

$$|\lambda_i(A)| < 1 \quad \forall i \iff \exists P \succ 0 \quad A^T P A - P \prec 0$$

Proof

- ($\Longleftarrow$) Let $Av = \lambda v$. Then,

$$0 > v^T(A^T P A - P)v = (|\lambda|^2 - 1)\underbrace{v^T P v}_{>0},$$

and therefore $|\lambda| < 1$

- ($\Longrightarrow$) Let $P = \sum_{i=0}^{\infty}(A^i)^T Q A^i$, where $Q \succ 0$. The sum converges by the eigenvalue assumption. Then,

$$A^T P A - P = \sum_{i=1}^{\infty}(A^i)^T Q A^i - \sum_{i=0}^{\infty}(A^i)^T Q A^i = -Q \prec 0$$

∎

Consider now the case where A is not stable, but we can use linear state feedback, i.e., $A(K) = A + BK$, where K is a fixed matrix. We want to find a matrix K such that $A + BK$ is stable, i.e., all its eigenvalues have absolute value smaller than one.

Use Schur complements to rewrite the condition:

$$(A + BK)^T P(A + BK) - P \prec 0, \qquad P \succ 0$$

$$\Updownarrow$$

$$\begin{bmatrix} P & (A+BK)^T P \\ P(A+BK) & P \end{bmatrix} \succ 0$$

Condition is nonlinear in (P, K). However, we can do a congruence transformation with $Q := P^{-1}$, and obtain:

$$\begin{bmatrix} Q & Q(A+BK)^T \\ (A+BK)Q & Q \end{bmatrix} \succ 0$$

Now, defining a new variable $Y := KQ$ we have

$$\begin{bmatrix} Q & QA^T + Y^T B^T \\ AQ + BY & Q \end{bmatrix} \succ 0.$$

This problem is now linear in (Q, Y). In fact, it is an SDP problem. After solving it, we can recover the controller K via $K = Q^{-1}Y$.

3.2 Theta function

Given a graph $G = (V, E)$, a *stable set* is a subset of V with the property that the induced subgraph has no edges. In other words, none of the selected vertices are adjacent to each other.

The *stability number* of a graph, usually denoted by $\alpha(G)$, is the cardinality of the largest stable set. Computing the stability number of a graph is NP-hard. There are many interesting applications of the stable set problem. In particular, they can be used to provide bounds on the *Shannon capacity of a graph* [Lov79], a problem of importance in coding. In fact, this was one of the first appearances of what today is known as SDP.

The Lovász theta function is denoted by $\vartheta(G)$, and is defined as the solution of the SDP :

$$\max J \bullet X \qquad \text{s.t.} \quad \begin{cases} \text{Tr}(X) = 1 \\ X_{ij} = 0 \qquad (i, j) \in E \\ X \succeq 0 \end{cases} \tag{4}$$

where J is the matrix with all entries equal to one. The theta function is an upper bound on the stability number, i.e.,

$$\alpha(G) \leq \vartheta(G).$$

The inequality is easy to prove. Consider the indicator vector $\xi(S)$ of any stable set S, and define the matrix $X := \frac{1}{|S|} \xi \xi^T$. Is is easy to see that this X is a feasible solution of the SDP, and thus the inequality follows.

4 Software

Remark   There are many good software codes for semidefinite programming. Among the most well-known, we mention the following ones:

- SeDuMi, originally by Jos Sturm, now being maintained by the optimization group at McMaster: `http://sedumi.mcmaster.ca/`

- SDPT3, by Kim-Chuan Toh, Reha Tütüncü, and Mike Todd. `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`

- SDPA, by the research group of Masakazu Kojima, `http://grid.r.dendai.ac.jp/sdpa/`

- CSDP, by Brian Borchers, `http://infohost.nmt.edu/~borchers/csdp.html`

A very convenient way of using these (and other) SDP solvers under MATLAB is through the YALMIP parser/solver (Johan Löfberg, `http://control.ee.ethz.ch/~joloef/yalmip.php`).

References

[BV04]   S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[Lov79]   L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory,* 25(1):1–7, 1979.

[Ram97]   M. V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Math. Programming,* 77(2, Ser. B):129–162, 1997.

[RTW97]   M. V. Ramana, L. Tunçel, and H. Wolkowicz. Strong duality for semidefinite programming. *SIAM J. Optim.,* 7(3):641–662, 1997.

[Tod01]   M. Todd. Semidefinite optimization. *Acta Numerica*, 10:515–560, 2001.

[VB96]   L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.