MIT 6.7230 - Algebraic techniques and semidefinite optimization February 9, 2024

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 $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$, and is called *positive definite* if $x^T Ax > 0$ 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 $\mathcal{S}^n_{++}$. As we shall prove soon, $\mathcal{S}^n_+$ is a proper cone (i.e., closed, convex, pointed, and solid). We will use the inequality signs "$\preceq$" and "$\succeq$" to denote the partial order induced by $\mathcal{S}^n_+$ (usually called the *Löwner* partial order).

1 Positive semidefinite (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 $x \in \mathbb{R}^n$, $x^T Ax \geq 0$.*
- *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 definite ($A \succ 0$).*
- *For all nonzero $x \in \mathbb{R}^n$, $x^T Ax > 0$.*
- *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, $A \succ 0 \Leftrightarrow T^T AT \succ 0$.
- Schur complement. The following conditions are equivalent:

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

We now prove the following result:

Theorem 3. *The set $\mathcal{S}_+^n$ of positive semidefinite matrices is a proper cone.*

Proof. Invariance under nonnegative scalings follows directly from the definition, so S_+^n is a cone. By the second statement in Proposition 1, S_+^n is the intersection of infinitely many closed halfspaces, and hence it is both closed and convex. To show pointedness, notice that if there is a symmetric matrix A that belongs to both S_+^n and $-S_+^n$, then $x^T A x$ must vanish for all $x \in \mathbb{R}^n$, thus A must be the zero matrix. Finally, the cone is solid since $I_n + X$ is positive definite for all X provided $\|X\|$ is small enough (e.g., by continuity of eigenvalues). $\square$

We state next some additional facts on the geometry of the cone $\mathcal{S}_+^n$ of positive semidefinite matrices.

- If S_n is equipped with the inner product $\langle X, Y\rangle := X \bullet Y = \mathrm{Tr}(XY)$, then $\mathcal{S}_+^n$ is a self-dual cone.
- The cone $\mathcal{S}_+^n$ is *not* polyhedral, and its extreme rays are the rank one matrices.

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:

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

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

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

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. We then 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 $C \bullet X$. Furthermore, in the case of feasibility problems (i.e., $C = 0$), the dual problem can be used to certify the nonexistence of solutions to the primal problem. This property will be crucial in our later 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 with 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 *adjoint 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:

$$\begin{array}{ll} \text{minimize} & \langle c, x \rangle_S \\ \text{subject to} & \left\{ \begin{array}{rcl} \mathcal{A}x & = & b \\ x & \in & \mathcal{K} \end{array} \right. \end{array} \qquad \begin{array}{ll} \text{maximize} & \langle y, b \rangle_T \\ \text{subject to} & c - \mathcal{A}^* y \in \mathcal{K}^*, \end{array}$$

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

We give here a simple interpretation of duality, in terms of the separating hyperplane theorem. For simplicity, we concentrate on the case of feasibility only, i.e., where we are interested in deciding the existence of a solution x to the equations

$$\mathcal{A}x = b, \qquad x \in \mathcal{K}, \tag{4}$$

where as before $\mathcal{K}$ is a proper cone in the vector space S.

Consider now the image $\mathcal{A}(\mathcal{K})$ of the cone under the linear mapping. Notice that feasibility of (4) is equivalent to the point b being contained on $\mathcal{A}(\mathcal{K})$. We have now two convex sets in T, namely $\mathcal{A}(\mathcal{K})$ and $\{b\}$, and we are interested in knowing whether they intersect or not. *If* these sets satisfy certain properties (for instance, closedness and compactness) then we could apply the separating hyperplane theorem, to produce a linear functional y that will be positive on one set, and negative on the other. In particular, nonnegativity on $\mathcal{A}(\mathcal{K})$ implies

$$\langle y, \mathcal{A}(x) \rangle \geq 0 \;\; \forall x \in \mathcal{K} \qquad \Leftrightarrow \qquad \langle \mathcal{A}^*(y), x \rangle \geq 0 \;\; \forall x \in \mathcal{K} \qquad \Leftrightarrow \qquad \mathcal{A}^*(y) \in \mathcal{K}^*.$$

Thus, under these conditions, if (4) is infeasible, there is a linear functional y satisfying

$$\langle y, b \rangle < 0, \qquad \mathcal{A}^* y \in \mathcal{K}^*.$$

This yields a *certificate* of the infeasibility of the conic system (4).

2.3 Strong duality in SDP

Despite the formal similarities, there are a number of differences between linear programming and general conic programming (and in particular, SDP). Among them, we notice that in SDP optimal solutions may not necessarily exist (even if the optimal value is finite), and there can be a nonzero duality gap.

Nevertheless, we have seen that weak duality always holds for conic programming problems. As opposed to the LP case, *strong* duality can fail in general SDP. 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 are usually satisfied in practice, SDP problems have strong duality, and thus zero duality gap.

Theorem 4. *Assume that both the primal and dual problems are* strictly *feasible. Then, both achieve their optimal solutions, and there is no 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 easy to prove) that $x(k)$ converges to zero for all initial conditions x_0 iff $|\lambda_i(A)| < 1$, for $i = 1, \ldots, n$.

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

Theorem 5. *Given an $n \times n$ real matrix A, the following conditions are equivalent:*

(i) All eigenvalues of A are inside the unit circle, i.e., $|\lambda_i(A)| < 1$ for $i = 1, \ldots, n$.

(ii) There exist a matrix $P \in \mathcal{S}^n$ such that

$$P \succ 0, \qquad A^T P A - P \prec 0.$$

Proof. $(ii) \Rightarrow (i)$: Let $Av = \lambda v$. Then,

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

and therefore $|\lambda| < 1$.
$(i) \Rightarrow (ii)$: Let $P := \sum_{k=0}^{\infty} (A^k)^T Q A^k$, where $Q \succ 0$. The sum converges by the eigenvalue assumption. Then,

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

□

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:

$$(A+BK)^T P (A+BK) - P \prec 0, \qquad P \succ 0$$
$$\Updownarrow$$
$$\begin{bmatrix} P & (A+BK)^T P \\ P(A+BK) & P \end{bmatrix} \succ 0$$

This condition is not simultaneously convex in (P, K) (since it is bilinear). 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 $K = YQ^{-1}$.

3.2 Theta function

Given a graph $G = (V, E)$, a *stable set* (or *independent 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 upper bounds on the *Shannon capacity of a graph* [Lov79], a problem of that appears in coding theory (when computing the zero-error capacity of a noisy channel [Sha56]). 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 :

$$\max\ J \bullet X \qquad \text{s.t.} \begin{cases} \mathrm{Tr}(X) = 1 & \\ X_{ij} = 0 & (i,j) \in E \\ X \succeq 0 & \end{cases} \tag{5}$$

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 it achieves an objective value equal to $|S|$. As a consequence, the inequality above directly follows.

3.3 Euclidean distance matrices

Assume we are given a list of pairwise distances between a finite number of points. Under what conditions can the points be embedded in some finite-dimensional space, and those distances be realized as the *Euclidean* metric between the embedded points? This problem appears in a large number of applications, including distance geometry, computational chemistry, and machine learning.

Concretely, assume we have a list of distances d_{ij}, for $(i,j) \in [1,n] \times [1,n]$. We would like to find points $x_i \in \mathbb{R}^k$ (for some value of k), such that $\|x_i - x_j\| = d_{ij}$ for all i, j. What are necessary and sufficient conditions for such an embedding to exist? In 1935, Schoenberg [Sch35] gave an exact characterization in terms of the semidefiniteness of the matrix of squared distances:

Theorem 6. *The distances d_{ij} can be embedded in a Euclidean space if and only if the $n \times n$ matrix*

$$D := \begin{bmatrix} 0 & d_{12}^2 & d_{13}^2 & \dots & d_{1n}^2 \\ d_{12}^2 & 0 & d_{23}^2 & \dots & d_{2n}^2 \\ d_{13}^2 & d_{23}^2 & 0 & \dots & d_{3n}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{1n}^2 & d_{2n}^2 & d_{3n}^2 & \dots & 0 \end{bmatrix}$$

is negative semidefinite on the subspace orthogonal to the vector $e := (1, 1, \ldots, 1)$.

Proof. We show only the necessity of the condition. Assume an embedding exists, i.e., there are points $x_i \in \mathbb{R}^k$ such that $d_{ij} = \|x_i - x_j\|$. Consider now the Gram matrix G of inner products

$$G := \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \ldots & \langle x_1, x_n \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \ldots & \langle x_2, x_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_n, x_1 \rangle & \langle x_n, x_2 \rangle & \ldots & \langle x_n, x_n \rangle \end{bmatrix} = [x_1, \ldots, x_n]^T [x_1, \ldots, x_n],$$

which is positive semidefinite by construction. Since $D_{ij} = \|x_i - x_j\|^2 = \langle x_i, x_i \rangle + \langle x_j, x_j \rangle - 2\langle x_i, x_j \rangle$, we have

$$D = \operatorname{diag}(G) \cdot e^T + e \cdot \operatorname{diag}(G)^T - 2G,$$

from where the result directly follows. □

Notice that the dimension of the embedding is given by the rank k of the Gram matrix G.

For more on this and related embeddings problems, good starting points are Schoenberg's original paper [Sch35], as well as the book [DL97].

4 Software

Remark 7. *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 Lehigh:* `http://sedumi.ie.lehigh.edu`
- *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://sdpa.indsys.chuo-u.ac.jp/sdpa/`
- *CSDP, originally by Brian Borchers, now a COIN-OR project:* `https://projects.coin-or.org/Csdp/`
- *MOSEK, a commercial high performance software for large-scale conic programming (including SDP):* `http://www.mosek.com`*.*

A very convenient way of using these (and other) SDP solvers under MATLAB is through the YALMIP parser/solver (Johan Löfberg, `http://users.isy.liu.se/johanl/yalmip/`*), or the disciplined convex programming software CVX (Michael Grant and Stephen Boyd,* `http://www.stanford.edu/~boyd/cvx`*. If you use Julia instead, you can use the modeling languages/environments JuMP or Convex.jl; see* `https://www.juliaopt.org`*.*

References

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

[DL97] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

[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.

[Sch35] I. J. Schoenberg. Remarks to Maurice Fréchet's article "Sur la définition axiomatique d'une classe d'espace distanciés vectoriellement applicable sur l'espace de Hilbert". *Ann. of Math. (2)*, 36(3):724–732, 1935.

[Sha56] C. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, September 1956.

[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.