1

CONNECTIONS BETWEEN SEMI-INFINITE AND SEMIDEFINITE PROGRAMMING

Lieven Vandenberghe[1] and Stephen boyd[2]

[1] *Electrical Engineering Department, University of California, Los Angeles*

[2] *Electrical Engineering Department, Stanford University*

To appear in: R. Reemtsen and J.-J. Rueckmann, Eds, *Semi-infinite Programming*, Kluwer Academic Publishers.

1 INTRODUCTION

We consider convex optimization problems with *linear matrix inequality* (LMI) constraints, *i.e.*, constraints of the form

$$F(x) = F_0 + x_1 F_1 + \cdots + x_m F_m \geq 0, \tag{1.1}$$

where the matrices $F_i = F_i^T \in \mathbf{R}^{n\times n}$ are given, and the inequality $F(x) \geq 0$ means $F(x)$ is positive semidefinite. The LMI (1.1) is a convex constraint in the variable $x \in \mathbf{R}^m$. Conversely, many nonlinear convex constraints can be expressed as LMIs (see the recent surveys by Alizadeh [Ali95], Boyd, El Ghaoui, Feron and Balakrishnan [BEFB94], Lewis and Overton [LO96], Nesterov and Nemirovsky [NN94] and Vandenberghe and Boyd [VB96]).

The purpose of the paper is to explore some connections between optimization with LMI constraints and semi-infinite programming. We immediately note that the LMI (1.1) is equivalent to an infinite set of linear inequalities: $F(x) \geq 0$ if and only if

$$v^T F(x) v = v^T F_0 v + \sum_{i=1}^{m} x_i \left(v^T F_i v\right) \geq 0$$

for all v in the compact set $\{v \in \mathbf{R}^n \mid \|v\| = 1\}$. It is therefore clear that convex optimization problems with LMI constraints can be studied as special cases of semi-infinite programming. Perhaps more interestingly, we will see that some important semi-infinite optimization problems can be formulated in terms of linear matrix inequalities. Such a reduction, if possible, has important practical consequences: It means that those SIPs can be solved efficiently with recent interior-point methods for LMI problems. The emphasis of the paper will be on illustrating this point with examples from systems and control, signal processing, computational geometry, and statistics.

The examples in this paper will fall in two categories. The first is known as the *semidefinite programming problem* or SDP. In an SDP we minimize a linear function of a variable $x \in \mathbf{R}^m$ subject to an LMI:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) = F_0 + x_1 F_1 + \cdots + x_m F_m \geq 0. \end{array} \tag{1.2}$$

Semidefinite programming can be regarded as an extension of linear programming where the componentwise inequalities between vectors are replaced by matrix inequalities, or, equivalently, the first orthant is replaced by the cone of positive semidefinite matrices. Although the SDP (1.2) looks very specialized, it is much more general than a (finite-dimensional) linear program, and it has many applications in engineering and combinatorial optimization [Ali95, BEFB94, LO96, NN94, VB96]. Most interior-point methods for linear programming have been generalized to semidefinite programs. As in linear programming, these methods have polynomial worst-case complexity, and perform very well in practice.

We can express the SDP as a semi-infinite linear program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & v^T F(x) v \geq 0 \text{ for all } v. \end{array}$$

Lasserre [Las95] and Pataki [Pat95] have exploited this fact to formulate Simplex-like algorithms for SDP. The observation is also interesting for theoretical purposes since it allows us to apply, for example, duality results from SIP to SDP.

The second problem that we will encounter is the problem of maximizing the determinant of a matrix subject to LMI constraints. We call this the *determinant maximization* or maxdet-problem.

$$\begin{array}{ll} \text{maximize} & \det G(x) \\ \text{subject to} & G(x) = G_0 + x_1 G_1 + \cdots + x_m G_m > 0 \\ & F(x) = F_0 + x_1 F_1 + \cdots + x_m F_m \geq 0. \end{array}$$

The matrices $G_i = G_i^T \in \mathbf{R}^{l \times l}$ are given matrices. The problem is equivalent to minimizing the convex function $\log \det G(x)^{-1}$ subject to the LMI constraints.

The max-det objective arises naturally in applications in computational geometry, control, information theory, and statistics.

A unified form that includes both the SDP and the determinant maximization problem is

$$\begin{array}{ll} \text{minimize} & c^T x + \log\det G(x)^{-1} \\ \text{subject to} & G(x) > 0 \\ & F(x) \geq 0. \end{array} \tag{1.3}$$

This problem was studied in detail in Vandenberghe, Boyd and Wu [VBW98].

The basic facts about these two optimization problems, and of the unified form (1.3), can be summarized as follows.

- Both problems are convex.
- There is an extensive and useful duality theory for the problems.
- Very efficient interior-point methods for the problems have been developed recently [NN94].
- The problems look very specialized, but include a wide variety of convex optimization problems, with many applications in engineering.

2 DUALITY

In [VBW98] it was shown that we can associate with with (1.3) the *dual* problem

$$\begin{array}{ll} \text{maximize} & \log\det W - \mathbf{Tr}\ G_0 W - \mathbf{Tr}\ F_0 Z + l \\ \text{subject to} & \mathbf{Tr}\ G_i W + \mathbf{Tr}\ F_i Z = c_i, \quad i = 1, ..., m, \\ & W = W^T > 0, \quad Z = Z^T \geq 0. \end{array} \tag{2.1}$$

The variables are $W \in \mathbf{R}^{l \times l}$ and $Z \in \mathbf{R}^{n \times n}$. We say W and Z are dual feasible if they satisfy the constraints in (2.1), and strictly dual feasible if in addition $Z > 0$. We also refer to (1.3) as the primal problem and say x is primal feasible if $F(x) \geq 0$ and $G(x) > 0$, and strictly primal feasible if $F(x) > 0$ and $G(x) > 0$.

Let $p^\star$ and $d^\star$ be the optimal values of problem (1.3) and (2.1), respectively (with the convention that $p^\star = +\infty$ if the primal problem is infeasible, and $d^\star = -\infty$ if the dual problem is infeasible). The following theorem follows from standard results in convex analysis (Rockafellar [Roc70], see also [VBW98]).

Theorem 2.1 *$p^\star \geq d^\star$. If (1.3) is strictly feasible, the dual optimum is achieved; if (2.1) is strictly feasible, the primal optimum is achieved. In both cases, $p^\star = d^\star$.*

As an illustration, we derive the dual problem for the SDP (1.2). Substituting $G_0 = 1$, $G_i = 0$, $n = 1$, in (2.1) yields

$$\begin{array}{ll} \text{maximize} & \log W - W - \mathbf{Tr}\, F_0 Z + 1 \\ \text{subject to} & \mathbf{Tr}\, F_i Z = c_i, \quad i = 1, \ldots, m, \\ & W > 0, \quad Z \geq 0. \end{array}$$

The optimal value of W is one, so the dual problem reduces to

$$\begin{array}{ll} \text{maximize} & -\mathbf{Tr}\, F_0 Z \\ \text{subject to} & \mathbf{Tr}\, F_i Z = c_i, \quad i = 1, \ldots, m, \\ & Z \geq 0, \end{array} \tag{2.2}$$

which is the dual SDP (in the notation used in [VB96]). Applying the duality result of Theorem 2.1 we see that the the optimal values of (1.2) and (2.2) are equal if at least one of the problems is strictly feasible.

Examples of primal and dual problems with nonzero optimal duality gap are well known in the semi-infinite programming literature, and also arise in SDP (see [VB96] for an example).

3 ELLIPSOIDAL APPROXIMATION

Our first class of examples are ellipsoidal approximation problems. We can distinguish two basic forms. The first is the problem of finding the *minimum-volume* ellipsoid around a given set C. The second problem is the problem of finding the *maximum-volume* ellipsoid contained in a given convex set C. Both can be formulated as convex semi-infinite programming problems.

To solve the the first problem, it is convenient to parametrize the ellipsoid as the pre-image of a unit ball under an affine transformation, *i.e.*,

$$\mathcal{E} = \{ v \mid \|Av + b\| \leq 1 \}.$$

It can be assumed without loss of generality that $A = A^T > 0$, in which case the volume of $\mathcal{E}$ is proportional to $\det A^{-1}$. The problem of computing the minimum-volume ellipsoid containing C can be written as

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & A = A^T > 0 \\ & \|Av + b\| \leq 1, \quad \forall v \in C, \end{array} \tag{3.1}$$

where the variables are A and b. For general C, this is a semi-infinite programming problem. Note that both the objective function and the constraints are convex in A and b.

For the second problem, where we maximize the volume of ellipsoids enclosed in a convex set C, it is more convenient to represent the ellipsoid as the *image* of the unit ball under an affine transformation, *i.e.*, as

$$\mathcal{E} = \{By + d \mid \|y\| \leq 1\}.$$

Again it can be assumed that $B = B^T > 0$. The volume is proportional to $\det B$, so we can find the maximum volume ellipsoid inside C by solving the convex optimization problem

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & B = B^T > 0 \\ & By + d \in C \quad \forall y, \|y\| \leq 1, \end{array} \tag{3.2}$$

in the variables B and d. For general convex C, this is again a convex semi-infinite optimization problem.

The ellipsoid of least volume containing a set is often called the Löwner ellipsoid (after Danzer, Grünbaum, and Klee [DGK63, p.139]), or the Löwner-John ellipsoid (Grötschel, Lovász and Schrijver [GLS88, p.69]). John in [Joh85] has shown that if one shrinks the minimum volume outer ellipsoid of a convex set $C \subset \mathbf{R}^n$ by a factor n about its center, one obtains an ellipsoid contained in C. Thus the Löwner-John ellipsoid serves as an ellipsoidal approximation of a convex set, with bounds that depend only on the dimension of the ambient space, and not in any other way on the set C.

Minimum volume ellipsoid containing given points.

The best known example is the problem of determining the minimum volume ellipsoid that contains given points $x^1, \ldots, x^K$ in $\mathbf{R}^n$, *i.e.*,

$$C = \{x^1, \ldots, x^K\},$$

(or, equivalently, the convex hull $\mathbf{Co}\,\{x^1, \ldots, x^K\}$). This problem has applications in cluster analysis (Rosen [Ros65], Barnes [Bar82]), robust statistics (in

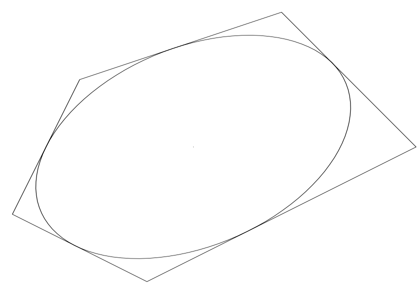

Figure 1 Maximum volume ellipsoid contained in a polyhedron.

ellipsoidal peeling methods for outlier detection (Rousseeuw and Leroy [RL87, §7]), and robotics (Rimon and Boyd [RB96]).

Applying (3.1), we can write this problem as

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & \|Ax^i + b\| \le 1, \; i = 1, \ldots, K \\ & A = A^T > 0, \end{array} \tag{3.3}$$

where the variables are $A = A^T \in \mathbf{R}^{n\times n}$ and $b \in \mathbf{R}^n$. The norm constraints $\|Ax^i + b\| \le 1$, which are just convex quadratic inequalities in the variables A and b, can be expressed as LMIs

$$\left[\begin{array}{cc} I & Ax^i + b \\ (Ax^i + b)^T & 1 \end{array}\right] \ge 0.$$

so (3.3) is a maxdet-problem in the variables A and b.

Maximum volume ellipsoid in polytope.

Assume the set C is a polytope described by a set of linear inequalities:

$$C = \{x \mid a_i^T x \le b_i, \; i = 1, \ldots, L\}$$

(see Figure 1). To apply (3.2) we first work out the last constraint:

$$By + d \in C \text{ if } \|y\| = 1 \iff a_i^T(By + d) \le b_i \text{ if } \|y\| \le 1 \tag{3.4}$$

$$\iff \max_{\|y\|\le 1} a_i^T By + a_i^T d \le b_i \tag{3.5}$$

$$\iff \|Ba_i\| + a_i^T d \le b_i, \quad i = 1, \ldots, L. \tag{3.6}$$

This is a set of L convex constraints in B and d, and equivalent to the L LMIs

$$\begin{bmatrix} (b_i - a_i^T d)I & Ba_i \\ a_i^T B & b_i - a_i^T d \end{bmatrix} \geq 0, \quad i = 1, \ldots, L.$$

We can therefore formulate (3.2) as a maxdet-problem in the variables B and d:

$$\begin{array}{ll} \text{minimize} & \log\det B^{-1} \\ \text{subject to} & B > 0 \\ & \begin{bmatrix} (b_i - a_i^T d)I & Ba_i \\ (Ba_i)^T & b_i - a_i^T d \end{bmatrix} \geq 0, \quad i = 1, \ldots, L. \end{array}$$

Minimum volume ellipsoid containing ellipsoids.

These techniques extend to several interesting cases where C is not finite or polyhedral, but is defined as a combination (the sum, union, or intersection) of ellipsoids. In particular, it is possible to compute the optimal inner approximation of the intersection or the sum of ellipsoids, and the optimal outer approximation of the union or sum of ellipsoids, by solving a maxdet problem. We refer to [BEFB94] and Chernousko [Che94] for details.

As an example, consider the problem of finding the minimum volume ellipsoid $\mathcal{E}_0$ containing K given ellipsoids $\mathcal{E}_1, \ldots, \mathcal{E}_K$. For this problem we describe the ellipsoids as sublevel sets of convex quadratic functions:

$$\mathcal{E}_i = \{x \mid x^T A_i x + 2b_i^T x + c_i \leq 0\}, \quad i = 0, \ldots, K.$$

The solution can be found by solving the following maxdet-problem in the variables $A_0 = A_0^T$, b_0, and K scalar variables τ_i:

$$\begin{array}{ll} \text{minimize} & \log\det A_0^{-1} \\ \text{subject to} & A_0 = A_0^T > 0 \\ & \tau_1 \geq 0, \ldots, \tau_K \geq 0 \\ & \begin{bmatrix} A_0 & b_0 & 0 \\ b_0^T & -1 & b_0^T \\ 0 & b_0 & -A_0 \end{bmatrix} - \tau_i \begin{bmatrix} A_i & b_i & 0 \\ b_i^T & c_i & 0 \\ 0 & 0 & 0 \end{bmatrix} \leq 0, \quad i = 1, \ldots, K. \end{array}$$

(c_0 is given by $c_0 = b_0^T A_0^{-1} b_0 - 1$.) See [BEFB94, p.43] for details. Figure 2 shows an instance of the problem.

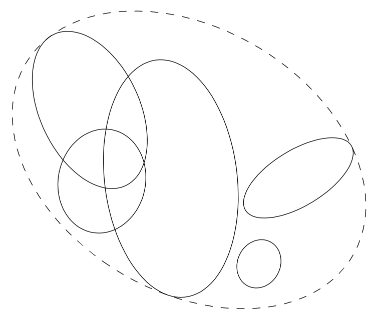

Figure 2 Minimum volume ellipsoid containing five given ellipsoids. Finding such an ellipsoid can be cast as a maxdet-problem, hence efficiently solved.

4 EXPERIMENT DESIGN

As a second group of examples, we consider problems in optimal experiment design. We consider the problem of estimating a vector x from a measurement $y = Ax + w$, where $w \sim \mathcal{N}(0, I)$ is measurement noise. We assume A has full column rank. The minimum-variance estimator is $\widehat{x} = A^+ y$, where A^+ is the pseudo-inverse of A, *i.e.*, $A^+ = (A^T A)^{-1} A^T$. The error covariance of the minimum-variance estimator is equal to $A^+ (A^+)^T = (A^T A)^{-1}$. We suppose that the rows of the matrix $A = [a_1 \ \ldots \ a_q]^T$ can be chosen among M possible test vectors $v^{(i)} \in \mathbf{R}^p$, $i = 1, \ldots, M$:

$$a_i \in \{v^{(1)}, \ldots, v^{(M)}\}, \quad i = 1, \ldots, q.$$

The goal of experiment design is to choose the vectors a_i so that the error covariance $(A^T A)^{-1}$ is 'small'. We can interpret each component of y as the result of an experiment or measurement that can be chosen from a fixed menu of possible experiments; our job is to find a set of measurements that (together) are maximally informative.

We can write $A^T A = q \sum_{i=1}^{M} \lambda_i v^{(i)} {v^{(i)}}^T$, where λ_i is the fraction of rows a_k equal to the vector $v^{(i)}$. We ignore the fact that the numbers λ_i are integer multiples of $1/q$, and instead treat them as continuous variables, which is jus-

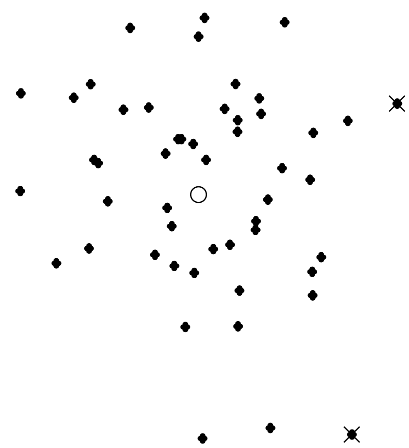

Figure 3 A D-optimal experiment design involving 50 test vectors in $\mathbf{R}^2$. The circle is the origin; the dots are the test vectors that are not used in the experiment (*i.e.*, have a weight $\lambda_i = 0$); the crosses are the test vectors that are used (*i.e.*, have a weight $\lambda_i > 0$). The D-optimal design allocates all measurements to only two test vectors.

tified in practice when q is large. (Alternatively, we can imagine that we are designing a random experiment: each experiment a_i has the form $v^{(k)}$ with probability λ_k.)

Many different criteria for measuring the size of the matrix $(A^T A)^{-1}$ have been proposed. For example, in D-optimal design, we minimize the determinant of the error covariance $(A^T A)^{-1}$, which leads to the maxdet-problem

$$\begin{array}{ll} \text{minimize} & \log\det\left(\sum_{i=1}^{M} \lambda_i v^{(i)} {v^{(i)}}^T\right)^{-1} \\ \text{subject to} & \lambda_i \geq 0, \; i = 1, \ldots, M \\ & \sum_{i=1}^{M} \lambda_i = 1. \end{array} \tag{4.1}$$

An example is shown in Figure 3.

Fedorov [Fed71], Atkinson and Donev [AD92], and Pukelsheim [Puk93] give surveys and additional references on optimal experiment design. The formulation of D-optimal design as a maxdet-problem has the advantage that one can easily incorporate additional useful convex constraints. See [VBW98] for examples.

There is an interesting relation between optimal experiment design and ellipsoidal approximation. We first derive the dual of the experiment design problem (4.1), applying (2.1). After a few simplifications we obtain

$$\begin{array}{ll} \text{maximize} & \log\det W + p - z \\ \text{subject to} & W = W^T > 0 \\ & {v^{(i)}}^T W v^{(i)} \leq z, \; i = 1, \ldots, M, \end{array} \tag{4.2}$$

where the variables are the matrix W and the scalar variable z. Problem (4.2) can be further simplified. The constraints are homogeneous in W and z, so for each dual feasible W, z we have a ray of dual feasible solutions tW, tz, $t > 0$. It turns out that we can analytically optimize over t: replacing W by tW and z by tz changes the objective to $\log\det W + p\log t + p - tz$, which is maximized for $t = p/z$. After this simplification, and with a new variable $\tilde{W} = (p/z)W$, problem (4.2) becomes

$$\begin{array}{ll} \text{maximize} & \log\det \tilde{W} \\ \text{subject to} & \tilde{W} > 0 \\ & {v^{(i)}}^T \tilde{W} v^{(i)} \leq p, \; i = 1, \ldots, M. \end{array} \tag{4.3}$$

Problem (4.3) has an interesting geometrical meaning: the constraints state that $\tilde{W}$ determines an ellipsoid $\{x \mid x^T \tilde{W} x \leq p\}$, centered at the origin, that contains the points $v^{(i)}$, $i = 1, \ldots, M$; the objective is to maximize $\det \tilde{W}$, *i.e.*, to minimize the volume of the ellipsoid.

There is an interesting connection between the optimal primal variables λ_i and the points $v^{(i)}$ that lie on the boundary of the optimal ellipsoid $\mathcal{E}$. First note that the duality gap associated with a primal feasible λ and a dual feasible $\tilde{W}$ is equal to

$$\log\det \left(\sum_{i=1}^{M} \lambda_i v^{(i)} {v^{(i)}}^T \right)^{-1} - \log\det \tilde{W},$$

and is zero (hence, λ is optimal) if and only if $\tilde{W} = \left(\sum_{i=1}^{M} \lambda_i v^{(i)} {v^{(i)}}^T \right)^{-1}$. Hence, λ is optimal if

$$\mathcal{E} = \left\{ x \in \mathbf{R}^p \;\middle|\; x^T \left(\sum_{i=1}^{M} \lambda_i v^{(i)} {v^{(i)}}^T \right)^{-1} x \leq p \right\}$$

is the minimum-volume ellipsoid, centered at the origin, that contains the points $v^{(j)}$, $j = 1, \ldots, M$. We also have (in fact, for any feasible λ)

$$\sum_{j=1}^{M} \lambda_j \left(p - v^{(j)T} \left(\sum_{i=1}^{M} \lambda_i v^{(i)} v^{(i)T} \right)^{-1} v^{(j)} \right)$$

$$= p - \mathbf{Tr} \left(\sum_{j=1}^{M} \lambda_j v^{(j)} v^{(j)T} \right) \left(\sum_{i=1}^{M} \lambda_i v^{(i)} v^{(i)T} \right)^{-1} = 0. \qquad (4.4)$$

If λ is optimal, then each term in the sum on the left hand side is positive (since $\mathcal{E}$ contains all vectors $v^{(j)}$), and therefore the sum can only be zero if each term is zero:

$$\lambda_j > 0 \Longrightarrow v^{(j)T} \left(\sum_{i=1}^{M} \lambda_i v^{(i)} v^{(i)T} \right)^{-1} v^{(j)} = p,$$

Geometrically, λ_j is nonzero only if $v^{(j)}$ lies on the boundary of the minimum volume ellipsoid. This makes more precise the intuitive idea that an optimal experiment only uses 'extreme' test vectors. Figure 4 shows the optimal ellipsoid for the experiment design example of Figure 3.

The duality between D-optimal experiment designs and minimum-volume ellipsoids also extends to non-finite compacts sets (Titterington [Tit75], Pronzato and Walter [PW94]). The D-optimal experiment design problem on a compact set $C \subset \mathbf{R}^{\mathbf{p}}$ is

$$\text{maximize} \;\; \log\det \mathbf{E} vv^T \qquad (4.5)$$

over all probability measures on C. This is a convex but semi-infinite optimization problem, with dual ([Tit75])

$$\begin{array}{ll} \text{maximize} & \log\det \tilde{W} \\ \text{subject to} & \tilde{W} > 0 \\ & v^T \tilde{W} v \leq p, \quad v \in C. \end{array} \qquad (4.6)$$

Again, we see that the dual is the problem of computing the minimum volume ellipsoid, centered at the origin, and covering the set C.

General methods for solving the semi-infinite optimization problems (4.5) and (4.6) fall outside the scope of this paper. In particular cases, however, these problems can be solved as maxdet-problems. One interesting example arises when C is the union of a finite number of ellipsoids. In this case, the dual (4.6) can be cast as a maxdet-problem (see §3) and hence efficiently solved; by duality, we can recover from the dual solution the probability distribution that solves (4.5).

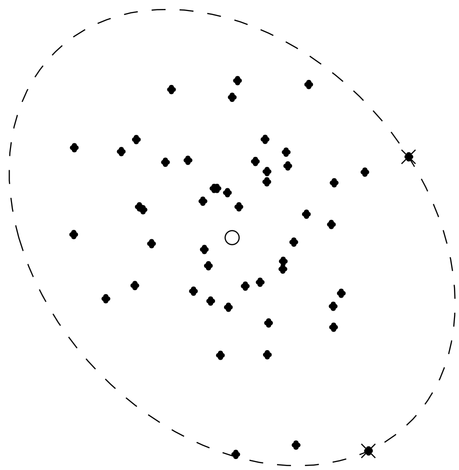

Figure 4 In the dual of the D-optimal experiment design problem we compute the minimum-volume ellipsoid, centered at the origin, that contains the test vectors. The test vectors with a nonzero weight lie on the boundary of the optimal ellipsoid. Same data and notation as in Figure 3.

5 PROBLEMS INVOLVING POWER MOMENTS

Bounds on expected values via semidefinite programming.

Let t be a random real variable. The expected values $\mathbf{E}t^k$ are called the (power) *moments* of the distribution of t. The following classical result gives a characterization of a moment sequence: There exists a probability distribution on $\mathbf{R}$ such that $x_k = \mathbf{E}t^k$, $k = 0, \ldots, 2n$, if and only if $x_0 = 1$ and

$$H(x_0, \ldots, x_{2n}) = \begin{bmatrix} x_0 & x_1 & x_2 & \ldots & x_{n-1} & x_n \\ x_1 & x_2 & x_3 & \ldots & x_n & x_{n+1} \\ x_2 & x_3 & x_4 & \ldots & x_{n+1} & x_{n+2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ x_{n-1} & x_n & x_{n+1} & \ldots & x_{2n-2} & x_{2n-1} \\ x_n & x_{n+1} & x_{n+2} & \ldots & x_{2n-1} & x_{2n} \end{bmatrix} \geq 0. \qquad (5.1)$$

It is easy to see that the condition is necessary: let $x_i = \mathbf{E}t^i$, $i = 0, \ldots, 2n$ be the moments of some distribution, and let $y = [y_0\; y_1\; \cdots\; y_n]^T \in \mathbf{R}^{n+1}$. Then we have

$$y^T H(x_0, \ldots, x_{2n}) y = \sum_{i,j=0}^{n} y_i y_j \mathbf{E}t^{i+j} = \mathbf{E}\left(y_0 + y_1 t^1 + \cdots + y_n t^n\right)^2 \geq 0.$$

Sufficiency is less obvious. The proof is classical (and based on convexity arguments); see *e.g.*, Krein and Nudelman [KN77, p.182] or Karlin and Studden [KS66, p.189–199]. There are similar conditions for distributions on finite or semi-infinite intervals.

Note that condition (5.1) is an LMI in the variables x_k, *i.e.*, the condition that $x_0, \ldots, x_{2n}$ be the moments of some distribution on $\mathbf{R}$ can be expressed as an LMI in x. Using this fact, we can cast some interesting moment problems as SDPs and maxdet-problems.

Suppose t is a random variable on $\mathbf{R}$. We do not know its distribution, but we do know some bounds on the moments, *i.e.*,

$$\underline{\mu}_k \leq \mathbf{E}t^k \leq \overline{\mu}_k$$

(which includes, as a special case, knowing exact values of some of the moments). Let $p(t) = c_0 + c_1 t + \cdots + c_{2n} t^{2n}$ be a given polynomial in t. The expected value of $p(t)$ is linear in the moments $\mathbf{E}t^i$:

$$\mathbf{E}p(t) = \sum_{i=0}^{2n} c_i \mathbf{E}t^i = \sum_{i=0}^{2n} c_i x_i.$$

We can compute upper and lower bounds for $\mathbf{E}p(t)$,

$$\begin{array}{ll} \text{minimize (maximize)} & \mathbf{E}p(t) \\ \text{subject to} & \underline{\mu}_k \leq \mathbf{E}t^k \leq \overline{\mu}_k, \quad k = 1, \ldots, 2n, \end{array}$$

over all probability distributions that satisfy the given moment bounds, by solving the SDPs

$$\begin{array}{ll} \text{minimize (maximize)} & c_1 x_1 + \cdots + c_{2n} x_{2n} \\ \text{subject to} & \underline{\mu}_k \leq x_k \leq \overline{\mu}_k, \quad k = 1, \ldots, 2n \\ & H(1, x_1, \ldots, x_{2n}) \geq 0 \end{array}$$

over the variables $x_1, \ldots, x_{2n}$. This gives bounds on $\mathbf{E}p(t)$, over all probability distributions that satisfy the known moment constraints. The bounds are sharp in the sense that there are distributions, whose moments satisfy the given moment bounds, for which $\mathbf{E}p(t)$ takes on the upper and lower bounds found by these SDPs.

A related problem was considered by Dahlquist, Eisenstat and Golub [DEG72], who analytically compute bounds on $\mathbf{E}t^{-1}$ and $\mathbf{E}t^{-2}$, given the moments $\mathbf{E}t^i$, $i = 1, \ldots, n$. (Here t is a random variable in a finite interval.) Using semidefinite programming one can solve more general problems where upper and lower bounds on $\mathbf{E}t^i$, $i = 1 \ldots, n$, (or the expected value of some polynomials) are known.

Another application arises in the optimal control of queuing networks (see Bertsimas *et al.* [BPT94, Ber95] and Schwerer [Sch96]).

Upper bound on the variance via semidefinite programming.

As another example, one can maximize the variance of t, over all probability distributions that satisfy the moment constraints (to obtain a sharp upper bound on the variance of t):

$$\begin{array}{ll} \text{maximize} & \mathbf{E}t^2 - (\mathbf{E}t)^2 \\ \text{subject to} & \underline{\mu}_k \leq \mathbf{E}t^k \leq \overline{\mu}_k, \quad k = 1, \ldots, 2n, \end{array}$$

which is equivalent to the SDP

$$\begin{array}{ll} \text{maximize} & y \\ \text{subject to} & \begin{bmatrix} x_2 - y & x_1 \\ x_1 & 1 \end{bmatrix} \geq 0 \\ & \underline{\mu}_k \leq x_k \leq \overline{\mu}_k, \quad k = 1, \ldots, 2n \\ & H(1, x_1, \ldots, x_{2n}) \geq 0 \end{array}$$

with variables y, x_1, $\ldots$, x_{2n}. The 2×2-LMI is equivalent to $y \leq x_2 - x_1^2$. More generally, one can compute an upper bound on the variance of a given polynomial $\mathbf{E}p(t)^2 - (\mathbf{E}p(t))^2$. Thus we can compute an upper bound on the variance of a polynomial $p(t)$, given some bounds on the moments.

A robust estimate of the moments.

Another interesting problem is the maxdet-problem

$$\begin{array}{ll} \text{maximize} & \log\det H(1, x_1, \ldots, x_{2n}) \\ \text{subject to} & \underline{\mu}_k \leq x_k \leq \overline{\mu}_k, k = 1, \ldots, 2n \\ & H(1, x_1, \ldots, x_{2n}) > 0. \end{array} \tag{5.2}$$

The solution can serve as a 'robust' solution to the feasibility problem of finding a probability distribution that satisfies given bounds on the moments. While the SDPs provide lower and upper bounds on $\mathbf{E}p(t)$, the maxdet-problem should provide a reasonable guess of $\mathbf{E}p(t)$.

Note that the maxdet-problem (5.2) is equivalent to

$$\begin{array}{ll} \text{maximize} & \log\det \mathbf{E} f(t) f(t)^T \\ \text{subject to} & \underline{\mu} \leq \mathbf{E} f(t) \leq \overline{\mu} \end{array} \tag{5.3}$$

over all probability distributions on $\mathbf{R}$, where $f(t) = \begin{bmatrix} 1 \; t \; t^2 \; \ldots \; t^n \end{bmatrix}^T$. We can interpret this as the problem of designing a random experiment to estimate the coefficients of a polynomial $p(t) = c_0 + c_1 t + \cdots + c_n t^n$.

6 POSITIVE-REAL LEMMA

Linear system theory provides numerous examples of semi-infinite constraints that can be cast as LMIs (see [BEFB94] for an extensive survey). One of the fundamental theorems, the positive-real lemma, can be interpreted in this light.

The positive-real lemma [AV73] gives a condition that guarantees that a rational function $H : \mathbf{C} \to \mathbf{R}^{m \times m}$, defined as

$$H(s) = C(sI - A)^{-1} B + D$$

where $A \in \mathbf{R}^{n \times n}$ (and of minimial dimension), $C \in \mathbf{R}^{m \times n}$, $B \in \mathbf{R}^{n \times m}$, $D \in \mathbf{R}^{m \times m}$, satisfies certain inequalities in the complex plane. The theorem states that

$$H(s) + H(s)^* \geq 0 \text{ for all } \Re s > 0 \tag{6.1}$$

if and only if there exists a $P = P^T$ such that

$$P > 0, \quad \begin{bmatrix} A^T P + PA & PB - C^T \\ B^T P - C & -D - D^T \end{bmatrix} \leq 0. \tag{6.2}$$

In other words, the infinite set of inequalities (6.1) is equivalent to the finite matrix inequality (6.2) with the auxiliary variable P.

Assume, for example, that A and B are given, and that the matrices C and D depend affinely on certain parameters $\theta \in \mathbf{R}^p$. Then (6.1) is an infinite set of LMIs in θ, while (6.2) is a finite LMI in θ and P.

Other examples in systems and control theory include the bounded-real lemma, and the Nevanlinna-Pick problem [BEFB94]. An application of the positive-real lemma in filter design is described in [WBV96, WBV97].

7 CONCLUSION

We have discussed examples of semi-infinite optimization problems that can be reduced to semidefinite programming or determinant maximization problems. It is clear that a reduction of SIPs to SDPs or maxdet-problems is not always possible. It is important, however, to recognize when such a reduction is possible, since it implies that the problems can be solved efficienlty using interior-point methods.

Acknowledgment.

We thank Shao-Po Wu for his help with the numerical examples in the paper, which were generated using the codes SDPSOL [WB96] and MAXDET [WVB96].

REFERENCES

[AD92] A. C. Atkinson and A. N. Donev. *Optimum experiment designs.* Oxford Statistical Science Series. Oxford University Press, 1992.

[Ali95] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, February 1995.

[AV73] B. Anderson and S. Vongpanitlerd. *Network analysis and synthesis: a modern systems theory approach.* Prentice-Hall, 1973.

[Bar82] E. R. Barnes. An algorithm for separating patterns by ellipsoids. *IBM Journal of Research and Development*, 26:759–764, 1982.

[BEFB94] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.

[Ber95] D. Bertsimas. The achievable region method in the optimal control of queuing systems; formulations, bounds and policies. *Queueing Systems*, 21:337–389, 1995.

[BPT94] D. Bertsimas, I. C. Paschalidis, and J. N. Tsitsiklis. Optimization of multiclass queueing networks:polyhedral and nonlinear characterizations of achievable performance. *Ann. Appl. Prob.*, 4(1):43–75, 1994.

[Che94] F. L. Chernousko. *State Estimation for Dynamic Systems*. CRC Press, Boca Raton, Florida, 1994.

[DEG72] G. Dahlquist, S. C. Eisenstat, and G. H. Golub. Bounds for the error of linear systems of equations using the theory of moments. *Journal of Mathematical Analysis and Applications*, 37:151–166, 1972.

[DGK63] L. Danzer, B. Grunbaum, and V. Klee. Helly's theorem and its relatives. In V. L. Klee, editor, *Convexity*, volume 7 of *Proceedings of Symposia in Pure Mathematics*, pages 101–180. American Mathematical Society, 1963.

[Fed71] V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, 1971.

[GLS88] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, 1988.

[Joh85] F. John. Extremum problems with inequalities as subsidiary conditions. In J. Moser, editor, *Fritz John, Collected Papers*, pages 543–560. Birkhauser, Boston, Massachussetts, 1985.

[KN77] M. G. Krein and A. A. Nudelman. *The Markov Moment Problem and Extremal Problems*, volume 50 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, Rhode Island, 1977.

[KS66] S. Karlin and W. J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley-Interscience, 1966.

[Las95] J. B. Lasserre. Linear programming with positive semi-definite matrices. Technical Report LAAS-94099, Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS, 1995.

[LO96] A. S. Lewis and M. L. Overton. Eigenvalue optimization. *Acta Numerica*, pages 149–190, 1996.

[NN94] Yu. Nesterov and A. Nemirovsky. *Interior-point polynomial methods in convex programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[Pat95] G. Pataki. Cone-LP's and semi-definite programs: facial structure, basic solutions, and the simplex method. Technical report, GSIA, Carnegie-Mellon University, 1995.

[Puk93] F. Pukelsheim. *Optimal Design of Experiments*. Wiley, 1993.

[PW94] L. Pronzato and E. Walter. Minimum-volume ellipsoids containing compact sets: Application to parameter bounding. *Automatica*, 30(11):1731–1739, 1994.

[RB96] E. Rimon and S. Boyd. Obstacle collision detection using best ellipsoid fit. *Journal of Intelligent and Robotic Systems*, pages 1–22, December 1996.

[RL87] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.

[Roc70] R. T. Rockafellar. *Convex Analysis*. Princeton Univ. Press, Princeton, second edition, 1970.

[Ros65] J. B. Rosen. Pattern separation by convex programming. *Journal of Mathematical Analysis and Applications*, 10:123–134, 1965.

[Sch96] E. Schwerer. *A Linear Programming Approach to the Steady-State Analysis of Markov Processes*. PhD thesis, Graduate School of Business, Stanford University, 1996. Draft.

[Tit75] D. M. Titterington. Optimal design: some geometric aspects of D-optimality. *Biometrika*, 62:313–320, 1975.

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

[VBW98] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM J. on Matrix Analysis and Applications*, April 1998. To appear.

[WB96] S.-P. Wu and S. Boyd. SDPSOL: *A Parser/Solver for Semidefinite Programming and Determinant Maximization Problems with Matrix Structure. User's Guide, Version Beta.* Stanford University, June 1996.

[WBV96] S.-P. Wu, S. Boyd, and L. Vandenberghe. FIR filter design via semidefinite programming and spectral factorization. In *Proc. IEEE Conf. on Decision and Control*, pages 271–276, 1996.

[WBV97] S.-P. Wu, S. Boyd, and L. Vandenberghe. Magnitude filter design via spectral factorization and convex optimization. *Applied and Computational Control, Signals and Circuits*, 1997. To appear.

[WVB96] S.-P. Wu, L. Vandenberghe, and S. Boyd. MAXDET*: Software for Determinant Maximization Problems. User's Guide, Alpha Version.* Stanford University, April 1996.