Applications of Semidefinite Programming

Lieven Vandenberghe[1] Stephen Boyd[2]

August 25, 1998

[1]Electrical Engineering Department, University of California, Los Angeles (vandenbe@ee.ucla.edu).
[2]Information Systems Laboratory, Electrical Engineering Department, Stanford University (boyd@isl.stanford.edu).

Abstract

A wide variety of nonlinear convex optimization problems can be cast as problems involving linear matrix inequalities (LMIs), and hence efficiently solved using recently developed interior-point methods. In this paper, we will consider two classes of optimization problems with LMI constraints:

- The *semidefinite programming problem*, *i.e.*, the problem of minimizing a linear function subject to a linear matrix inequality. Semidefinite programming is an important numerical tool for analysis and synthesis in systems and control theory. It has also been recognized in combinatorial optimization as a valuable technique for obtaining bounds on the solution of NP-hard problems.

- The problem of maximizing the determinant of a positive definite matrix subject to linear matrix inequalities. This problem has applications in computational geometry, experiment design, information and communication theory, and other fields.

We review some of these applications, including some interesting applications that are less well known and arise in statistics, optimal experiment design and VLSI.

1 Optimization problems involving LMI constraints

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

$$F(x) = F_0 + x_1F_1 + \cdots + x_mF_m \succeq 0, \tag{1}$$

where the matrices $F_i = F_i^T \in \mathbf{R}^{n\times n}$ are given, and the inequality $F(x) \succeq 0$ means $F(x)$ is positive semidefinite. The LMI (1) is a convex constraint in the variable $x \in \mathbf{R}^m$. Conversely, a wide variety of nonlinear convex constraints can be expressed as LMIs (see the recent surveys by Alizadeh [Ali95], Boyd, El Ghaoui, Feron and Balakrihnan [BEFB94], Lewis and Overton [LO96], Nesterov and Nemirovsky [NN94] and Vandenberghe and Boyd [VB96]).

The purpose of the paper is to illustrate the use of linear matrix inequalities with a number of applications from different areas. The examples fall in two categories. The first problem 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:

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & F(x) = F_0 + x_1F_1 + \cdots + x_mF_m \succeq 0. \end{array} \tag{2}$$

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 (2) looks very specialized, it is much more general than a 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.

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:

$$\begin{array}{ll} \text{maximize} & \det G(x) \\ \text{subject to} & G(x) = G_0 + x_1G_1 + \cdots + x_mG_m \succ 0 \\ & F(x) = F_0 + x_1F_1 + \cdots + x_mF_m \succeq 0. \end{array}$$

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 in applications in computational geometry, control, information theory, and statistics.

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

$$\begin{array}{ll} \text{minimize} & c^Tx + \log\det G(x)^{-1} \\ \text{subject to} & G(x) \succ 0 \\ & F(x) \succeq 0. \end{array} \tag{3}$$

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

2 Ellipsoidal approximation

Our first class of examples are ellipsoidal approximation problems. We can distinguish two basic problems. 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 first problem, it is convenient to parametrize the ellipsoid as the pre-image of a unit ball under an affine transformation, *i.e.*,

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

It can be assumed without loss of generality that $A = A^T \succ 0$, 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

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & A = A^T \succ 0 \\ & \|Av + b\| \le 1, \quad \forall v \in C, \end{array} \tag{4}$$

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

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

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

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & B = B^T \succ 0 \\ & By + d \in C \;\; \forall y, \|y\| \le 1 \end{array} \tag{5}$$

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 we shrink the minimum volume outer ellipsoid of a convex set $C \subset \mathbf{R}^n$ by a factor n about its center, we obtain 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 ambient dimension, 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]), and robust statistics (in ellipsoidal peeling methods for outlier detection; see Rousseeuw and Leroy [RL87, §7]).

Applying (4), we can write this problem as

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & \|Ax^i + b\| \leq 1, \ i = 1, \ldots, K \\ & A = A^T \succ 0, \end{array} \tag{6}$$

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

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

so (6) is a maxdet-problem in the variables A and b.

Maximum volume ellipsoid in polytope

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

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

To apply (5) we first work out the last constraint:

$$\begin{array}{rcl} By + d \in C \text{ for all } \|y\| \leq 1 & \Longleftrightarrow & a_i^T(By + d) \leq b_i \text{ for all } \|y\| \leq 1 \\ & \Longleftrightarrow & \displaystyle\max_{\|y\|\leq 1} a_i^T By + a_i^T d \leq b_i \\ & \Longleftrightarrow & \|Ba_i\| + a_i^T d \leq b_i. \end{array}$$

This is a convex constraints in B and d, and equivalent to the LMI

$$\left[\begin{array}{cc} (b_i - a_i^T d)I & Ba_i \\ a_i^T B & b_i - a_i^T d \end{array}\right] \succeq 0.$$

We can therefore formulate (5) as a maxdet-problem in the variables B and d:

$$\begin{array}{ll} \text{minimize} & \log\det B^{-1} \\ \text{subject to} & B \succ 0 \\ & \left[\begin{array}{cc} (b_i - a_i^T d)I & Ba_i \\ (Ba_i)^T & b_i - a_i^T d \end{array}\right] \succeq 0, \quad i = 1, \ldots, L. \end{array}$$

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:

$$\begin{array}{ll} \text{minimize} & \log\det A_0^{-1} \\ \text{subject to} & A_0 = A_0^T \succ 0 \\ & \tau_1 \geq 0, \ldots, \tau_K \geq 0 \\ & \begin{bmatrix} A_0 & b_0 & 0 \\ b_0^T & -1 & b_0^T \\ 0 & b_0 & -A_0 \end{bmatrix} - \tau_i \begin{bmatrix} A_i & b_i & 0 \\ b_i^T & c_i & 0 \\ 0 & 0 & 0 \end{bmatrix} \preceq 0, \quad i = 1, \ldots, K. \end{array} \tag{7}$$

(c_0 is given by $c_0 = b_0^T A_0^{-1} b_0 - 1$.) See [BEFB94, p.43] for details.

3 Wire and transistor sizing

We consider linear resistor-capacitor (RC) circuits described by the differential equation

$$C\frac{dv}{dt} = -G(v(t) - u(t)), \tag{8}$$

where $v(t) \in \mathbf{R}^n$ is the vector of node voltages, $u(t) \in \mathbf{R}^n$ is the vector of independent voltage sources, $C \in \mathbf{R}^{n\times n}$ is the capacitance matrix, and $G \in \mathbf{R}^{n\times n}$ is the conductance matrix. We assume that C and G are symmetric and positive definite (*i.e.*, that the capacitive and resistive subcircuits are reciprocal and strictly passive). We are interested in problems in which C and G depend on some design parameters $x \in \mathbf{R}^m$. Specifically we assume that the matrices C and G are *affine* functions of x, *i.e.*,

$$C(x) = C_0 + x_1 C_1 + \cdots + x_m C_m, \quad G(x) = G_0 + x_1 G_1 + \cdots + x_m G_m, \tag{9}$$

where C_i and G_i are symmetric matrices.

Linear RC models of the form (8) are often used as approximate models for transistors and interconnect wires in VLSI circuits. When the design parameters are the physical widths of conductors or transistors, the conductance and capacitance matrices are affine in these parameters, *i.e.*, they have the form (9). An important example is wire sizing, where x_i

denotes the width of a segment of some conductor or interconnect line. A simple lumped model of the segment consists of a π section: a series conductance, with a capacitance to ground on each end. Here the conductance is linear in the width x_i, and the capacitances are linear or affine. We can also model each segment by many such π sections, and still have the general form (8), (9). Another important example is an MOS transistor circuit where x_i denotes the width of a transistor. When the transistor is 'on' it is modeled as a conductance that is proportional to x_i, and a source-to-ground capacitance and drain-to-ground capacitance that are linear or affine in x_i. See [VBE96] for details.

Signal propagation delay

We are interested in how fast a change in the input u propagates to the different nodes of the circuit, and in how this propagation delay varies as a function of the variables x.

We assume that for $t < 0$, the circuit is in static steady-state with $u(t) = v(t) = v_-$. For $t \geq 0$, the source switches to the constant value $u(t) = v_+$. As a result we have, for $t \geq 0$,

$$v(t) = v_+ + e^{-C^{-1}Gt}(v_- - v_+) \tag{10}$$

which converges, as $t \to \infty$, to v_+ (since our assumption $C \succ 0$, $G \succ 0$ implies stability). The difference between the node voltage and its ultimate value $\tilde{v}(t) = e^{-C^{-1}Gt}(v_- - v_+)$, and we are interested in how large t must be before this is small. To simplify notation, we will relabel $\tilde{v}$ as v, and from here on study the rate at which

$$v(t) = e^{-C^{-1}Gt}v(0) \tag{11}$$

becomes small. Note that this v satisfies the autonomous equation $Cdv/dt = -Gv$.

Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of the circuit, *i.e.*, the eigenvalues of $-C^{-1}G$, or equivalently, the roots of the characteristic polynomial $\det(sC + G)$. They are real and negative since the matrix is similar to the symmetric, negative definite matrix $-C^{-1/2}GC^{-1/2}$. We assume the eigenvalues are sorted in decreasing order, *i.e.*, $0 > \lambda_1 \geq \cdots \geq \lambda_n$. The largest eigenvalue, λ_1, is called the *dominant eigenvalue* or *dominant pole* of the RC circuit. The *(critical) dominant time constant* is defined as

$$T^{\mathrm{dom}} = -1/\lambda_1. \tag{12}$$

Note that the dominant time constant T^{dom} is a very complicated function of G and C, *i.e.*, the negative inverse of the largest zero of the polynomial $\det(sC + G)$. However it can be expressed in terms of an LMI, since

$$T^{\mathrm{dom}} = \min\{\, T \mid TG - C \succeq 0 \,\}.$$

In particular,

$$T^{\mathrm{dom}}(x) \leq T_{\max} \Longleftrightarrow T_{\max}G(x) - C(x) \succeq 0. \tag{13}$$

We can conclude that T^{dom} is a *quasiconvex* function of x, *i.e.*, its sublevel sets

$$\left\{ x \;\middle|\; T^{\mathrm{dom}}(x) \leq T_{\max} \right\}$$

are convex sets for all values of $T_{\max}$.

As a consequence, various optimization problems involving dominant time constant, area, and power can be cast as convex optimization problems. Therefore we can compute *exact, optimal* tradeoff curves between these quantities. We discuss this in more detail now. For more extensive discussion and examples, we refer to [VBE96, VBE97], where this approach to delay optimization is explored in more detail, and compared with conventional techniques based on geometric programming.

Minimum area subject to bound on delay

Suppose the area of the circuit is a linear (or affine) function of the variables x_i. This occurs when the variables represent the widths of transistors or conductors (with lengths fixed as l_i), in which case the circuit area has the form

$$a_0 + x_1 l_1 + \cdots + x_m l_m$$

where a_0 is the area of the fixed part of the circuit.

We can minimize the area subject to a bound on the dominant time constant $T^{\mathrm{dom}} \le T_{\max}$, and subject to upper and lower bounds on the widths by solving the SDP

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} l_i x_i \\ \text{subject to} & T_{\max} G(x) - C(x) \succeq 0 \\ & x_{\min} \le x_i \le x_{\max}, \quad i = 1, \ldots, m. \end{array} \tag{14}$$

By solving this SDP for a sequence of values of $T_{\max}$, we can compute the exact optimal tradeoff between area and dominant time constant. The optimal solutions of (14) are on the tradeoff curve, *i.e.*, they are *Pareto optimal* for area and dominant time constant.

Minimum power dissipation subject to bound on delay

The total energy dissipated in the resistors during a transition from initial voltage $\overline{v}$ to final voltage 0 (or between 0 and $\overline{v}$) is the energy stored in the capacitors, *i.e.*, $(1/2)\overline{v}^T C \overline{v}$. Therefore for a fixed clock rate and fixed probability of transition, the average power dissipated in proportional to

$$\overline{v}^T C(x) \overline{v} = \sum_{i=1}^{m} x_k \left(\overline{v}^T C_i \overline{v} \right),$$

which is a linear function of the design parameters x.

Therefore we can minimize power dissipation subject to a constraint on the dominant time constant by solving the SDP

$$\begin{array}{ll} \text{minimize} & \overline{v}^T C(x) \overline{v} \\ \text{subject to} & T_{\max} G(x) - C(x) \succeq 0 \\ & x_{\min} \le x_i \le x_{\max}, \quad i = 1, \ldots, m. \end{array}$$

We can also add an upper bound on area, which is a linear inequality. By solving this SDP for a sequence of values of $T_{\max}$, we can compute the optimal tradeoff between power dissipation and dominant time constant. By adding a constraint that the area cannot exceed $A_{\max}$, and solving the SDP for a sequence of values of $T_{\max}$ and $A_{\max}$, we can compute the exact optimal tradeoff surface between power dissipation, area, and dominant time constant.

Minimum delay subject to area and power constraints

We can also directly minimize the delay subject to limits on area and power dissipation, by solving the (quasiconvex) optimization problem

$$\begin{array}{ll} \text{minimize} & T \\ \text{subject to} & TG(x) - C(x) \succeq 0 \\ & x_{\min} \leq x_i \leq x_{\max}, \quad i = 1, \dots, m \\ & f_i^T x \leq g_i, \quad i = 1, 2 \end{array}$$

with variables x and T, where the linear inequalities limit area and power dissipation.

4 Experiment design

As a third group of examples, we consider some 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. The error covariance of the minimum-variance estimator is equal to $A^\dagger (A^\dagger)^T = (A^T A)^{-1}$. We suppose that the rows of the matrix $A = [a_1 \ \dots \ a_q]^T$ can be chosen among M possible test vectors $v^{(i)} \in \mathbf{R}^p$, $i = 1, \dots, M$:

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

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 justified 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 E-optimal design, we minimize the norm of the error covariance, $\lambda_{\max}((A^T A)^{-1})$, which is equivalent to maximizing the smallest eigenvalue of $A^T A$. This is readily cast as the SDP

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & \sum_{i=1}^M \lambda_i v^{(i)} v^{(i)T} \succeq tI \\ & \sum_{i=1}^M \lambda_i = 1 \\ & \lambda_i \geq 0, \quad i = 1, \dots, M \end{array}$$

in the variables $\lambda_1, \ldots, \lambda_M$, and t. Another criterion is A-optimality, in which we minimize $\mathbf{Tr}(A^T A)^{-1}$. This can be cast as an SDP:

$$
\begin{array}{ll}
\text{minimize} & \sum_{i=1}^{p} t_i \\
\text{subject to} & \begin{bmatrix} \sum_{i=1}^{M} \lambda_i v^{(i)} v^{(i)T} & e_i \\ e_i^T & t_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, p, \\
& \lambda_i \geq 0, \quad i = 1, \ldots, M, \\
& \sum_{i=1}^{M} \lambda_i = 1,
\end{array}
$$

where e_i is the ith unit vector in $\mathbf{R}^p$, and the variables are λ_i, $i = 1, \ldots, M$, and t_i, $i = 1, \ldots, p$.

In D-optimal design, we minimize the determinant of the error covariance $(A^T A)^{-1}$, which leads to the maxdet-problem

$$
\begin{array}{ll}
\text{minimize} & \log\det \left(\sum_{i=1}^{M} \lambda_i v^{(i)} v^{(i)T} \right)^{-1} \\
\text{subject to} & \lambda_i \geq 0, \; i = 1, \ldots, M \\
& \sum_{i=1}^{M} \lambda_i = 1.
\end{array} \tag{15}
$$

Fedorov [Fed71], Atkinson and Donev [AD92], and Pukelsheim [Puk93] give surveys and additional references on optimal experiment design.

The formulation of optimal design as maxdet-problems or SDPs has the advantage that we can easily incorporate additional useful convex constraints (see [VBW98]).

5 Problems involving 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

$$
H(x_0, \ldots, x_{2n}) = \begin{bmatrix}
x_0 & x_1 & x_2 & \cdots & x_{n-1} & x_n \\
x_1 & x_2 & x_3 & \cdots & x_n & x_{n+1} \\
x_2 & x_3 & x_4 & \cdots & x_{n+1} & x_{n+2} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
x_{n-1} & x_n & x_{n+1} & \cdots & x_{2n-2} & x_{2n-1} \\
x_n & x_{n+1} & x_{n+2} & \cdots & x_{2n-1} & x_{2n}
\end{bmatrix} \succeq 0. \tag{16}
$$

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 (16) is an LMI in the variables x_k, *i.e.*, the condition that x_0, ..., 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 SDP

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

with variables x_1, ..., 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 we 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 [Ber95] and Schwerer [Sch96]).

Upper bound on the variance via semidefinite programming

As another example, we 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

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

with variables y, x_1, …, x_{2n}. The 2×2-LMI is equivalent to $y \leq x_2 - x_1^2$. More generally, we 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

$$\begin{array}{ll} \text{maximize} & \log\det H(1, x_1, \ldots, x_{2n}) \\ \text{subject to} & \underline{\mu}_k \leq x_k \leq \overline{\mu}_k, k = 1, \ldots, 2n \\ & H(1, x_1, \ldots, x_{2n}) \succ 0. \end{array} \tag{17}$$

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 (17) is equivalent to

$$\begin{array}{ll} \text{maximize} & \log\det \mathbf{E}\, f(t) f(t)^T \\ \text{subject to} & \underline{\mu} \leq \mathbf{E}\, f(t) \leq \overline{\mu} \end{array} \tag{18}$$

over all probability distributions on $\mathbf{R}$, where $f(t) = [1\ t\ t^2\ \ldots\ t^n]^T$. 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 Structural optimization

We consider a truss structure with m bars connecting a set of n nodes. External forces are applied at each node, which cause a (small) displacement in the node positions. $f \in \mathbf{R}^n$ will denote the vector of (components of) external forces, and $d \in \mathbf{R}^n$ will denote the vector of corresponding node displacements. (By 'corresponding' we mean if f_i is, say, the z-coordinate of the external force applied at node k, then d_i is the z-coordinate of the displacement of node k.) The vector f is called a *loading* or *load*.

We assume damping can be ignored and that the structure is linearly elastic, *i.e.*, the relation between node displacement and external forces can be described by the differential equation

$$M\ddot{d} + Kd = f. \tag{19}$$

The matrix $M = M^T \succ 0$ is called the *mass matrix* of the truss; the matrix $K = K^T \succ 0$ is called the *stiffness matrix*.

We assume that the geometry (unloaded bar lengths and node positions) of the truss is fixed; we are to design the cross-sectional areas of the bars. These cross-sectional areas will be the design variables x_i, $i = 1, \ldots, m$. The stiffness matrix K and the mass matrix M are linear functions of x:

$$K(x) = x_1 K_1 + \cdots + x_m K_m, \quad M(x) = x_1 M_1 + \cdots + x_m M_m,$$

where $K_i = K_i^T \succeq 0$, and $M_i = M_i^T \succeq 0$ depend on the truss geometry. We assume these matrices are given For simplicity we also assume that $K(x) \succ 0$ for all nonnegative x.

The total weight W_{tot} of the truss also depends on the bar cross-sectional areas:

$$W_{\mathrm{tot}}(x) = w_1 x_1 + \cdots + w_m x_m,$$

where $w_i > 0$ are known, given constants (density of the material times the length of bar i). Roughly speaking, the truss becomes stiffer, but also heavier, when we increase x_i; there is a tradeoff between stiffness and weight.

Static measures of stiffness

We first consider the equilibrium state, *i.e.*, static conditions. In equilibrium, we have $\ddot{d} = 0$ and d is determined by the set of linear equations

$$Kd = f.$$

Our goal is to design the stiffest truss, subject to bounds on the bar cross-sectional areas and total truss weight:

$$l \le x_i \le u, \ \ i = 1, \ldots, m, \qquad W_{\mathrm{tot}}(x) \le W, \tag{20}$$

where W is a given limit on truss weight.

There are several ways to form a scalar measure of how stiff a truss is, for a given load f. Perhaps the simplest is the norm (squared) of the resulting displacement vector d:

$$\mathcal{D}(x, f) \triangleq d^T d = f^T K(x)^{-2} f.$$

Another measure is the elastic stored energy,

$$\mathcal{E}(x, f) \triangleq \frac{1}{2} f^T K(x)^{-1} f.$$

Maximizing stiffness corresponds to minimizing $\mathcal{D}(x,f)$ or $\mathcal{E}(x,f)$. These measures are similar (they are both small when K is 'large'), but not exactly the same. In particular, we will see that using the elastic stored energy $\mathcal{E}$ leads to SDP problems, while $\mathcal{D}$ in general does not.

We can consider several different scenarios that reflect our knowledge about the possible loadings f that can occur. The simplest is that f is a single, fixed, known loading. The design that minimizes the stored energy is the solution of

$$\begin{array}{ll} \text{minimize} & f^T K(x)^{-1} f \\ \text{subject to} & W_{\text{tot}}(x) \le W \\ & l \le x_i \le u. \end{array} \tag{21}$$

This problem is convex and can be cast as the following SDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} K(x) & f \\ f^T & t \end{bmatrix} \succeq 0 \\ & W_{\text{tot}}(x) \le W \\ & l \le x_i \le u \end{array}$$

in the variables x, t. It should be noted, however, that the SDP is not the most efficient way to solve (21). More efficient methods that directly solve problems of the form (21) have been proposed by Nemirovsky and Ben Tal [BTN92, BTN95a].

In more sophisticated formulations, the loading f might vary over a given set of possible loads, and we are interested in optimizing the stiffness for the worst case scenario (see also [BTN95b]). Suppose for example that f is unknown but bounded, *i.e.*, it can take arbitrary values in a ball $B = \{f \mid \|f\| \le \alpha\}$. The maximum stiffness as f varies over B can be written as

$$\max_{f \in B} \mathcal{E}(x,f) = \frac{\alpha^2}{2} \lambda_{\max}(K^{-1}(x)).$$

Therefore we can find the worst-case design by solving

$$\begin{array}{ll} \text{minimize} & \lambda_{\max} K(x)^{-1} \\ \text{subject to} & W_{\text{tot}}(x) \le W \\ & l \le x_i \le u, \end{array}$$

or, equivalently, by maximizing the smallest eigenvalue of K. This can be cast as an SDP

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & K(x) \succeq tI \\ & W_{\text{tot}}(x) \preceq W \\ & l \le x_i \le u. \end{array}$$

Finally, we can consider situations where the load is random with known statistics. For example, suppose f is a random variable with known mean and covariance:

$$\mathbf{E}\, f = \hat{f}, \quad \mathbf{E}(f - \hat{f})(f - \hat{f})^T = \Sigma.$$

We are interested in minimizing the expected stored energy, *i.e.*,

$$\begin{array}{ll} \text{minimize} & \mathbf{E}\, f^T K^{-1} f \\ \text{subject to} & W_{\text{tot}}(x) \le W \\ & l \le x_i \le u. \end{array}$$

This problem can be cast as an SDP, by first writing the objective function as

$$\mathbf{E}\, f^T K^{-1} f = \widehat{f}^T K(x)^{-1} \widehat{f} + \mathbf{Tr} \Sigma K(x)^{-1},$$

and introducing two new variables t and $X = X^T$:

$$\begin{array}{ll} \text{minimize} & t + \mathbf{Tr} \Sigma X \\ \text{subject to} & \begin{bmatrix} K(x) & \widehat{f} \\ \widehat{f} & t \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} K(x) & I \\ I & X \end{bmatrix} \succeq 0 \\ & W_{\text{tot}}(x) \le W \\ & l \le x_i \le u. \end{array}$$

Dynamic measure of stiffness

We now return to the dynamical model (19). The *modal frequencies* of the structure described by (19) are defined as the values $\omega \ge 0$ that satisfy

$$\det(K - M\omega^2) = 0,$$

i.e., the squareroots of the (generalized) eigenvalues of the pair (M, K). The fundamental frequency is the smallest modal frequency, *i.e.*,

$$\omega_1 = \lambda_{\min}^{1/2}(M, K).$$

It is the lowest frequency at which the structure can oscillate. The *fundamental period* of the structure is $2\pi/\omega_1$.

To avoid oscillations we would like to make the fundamental frequency higher than the frequencies of the excitations, *i.e.*, we would like to impose a lower bound $\omega_1 \ge \Omega$. (Equivalently, we impose a maximum fundamental period.) This constraint can be expressed as an LMI, since

$$\omega_1 \ge \Omega \Longleftrightarrow M(x)\Omega^2 - K(x) \preceq 0.$$

Various interesting design problems with bounds on the fundamental frequency can therefore be expressed as SDPs. As an example, we can minimize the weight subject to the lower bound on the fundamental frequency constraint by solving the SDP

$$\begin{array}{ll} \text{minimize} & W_{\text{tot}}(x) \\ \text{subject to} & M(x)\Omega^2 - K(x) \preceq 0 \\ & l_i \le x_i \le u_i. \end{array}$$

7 Quasi-Newton updates

In quasi-Newton methods for unconstrained minimization of a convex function f, the Newton step $-\nabla^2 f(x)^{-1}\nabla f(x)$ is replaced by $-H^{-1}\nabla f(x)$, where $H = H^T \succ 0$ is an approximation of the Hessian matrix, based on prior information and previous gradient evaluations. In each iteration, as the algorithm moves from x to the next point x^+, a new approximation H^+ is determined, based on the current H, and on the difference between the gradients at x^+ and x. A good updating rule for H should satisfy several properties: H^+ should be close to H, it should be easy to compute (or, more precisely, the search direction $-{H^+}^{-1}\nabla f(x^+)$ should be easy to compute), and it should incorporate the new information obtained by evaluating the gradient $\nabla f(x^+)$. This last property is usually enforced by imposing the secant condition

$$H^+(x^+ - x) = \nabla f(x^+) - \nabla f(x). \tag{22}$$

Byrd and Nocedal [BN89] have proposed to measure the difference between H and H^+ by using the Kullback-Leibler divergence (or relative entropy), given by

$$\frac{1}{2}\left(\mathbf{Tr} H^{-1/2}H^+H^{-1/2} - \log\det H^{-1/2}H^+H^{-1/2} - n\right)$$

(see also Dennis and Wolkowicz [DW93]). The Kullback-Leibler divergence is nonnegative for all positive definite H and H^+, and zero only if $H^+ = H$. The update H^+ that satisfies the secant condition and minimizes the Kullback-Leibler divergence is the solution of the following optimization problem:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr} H^{-1/2}H^+H^{-1/2} - \log\det H^{-1/2}H^+H^{-1/2} - n \\ \text{subject to} & H^+ \succ 0 \\ & H^+(x^+ - x) = \nabla f(x^+) - \nabla f(x). \end{array} \tag{23}$$

Fletcher [Fle91] has shown that the solution is given by

$$H^+ = H - \frac{Hss^TH}{s^THs} + \frac{gg^T}{s^Tg}, \tag{24}$$

assuming that $s^Tg > 0$, where $s = x^+ - x$ and $g = \nabla f(x^+) - \nabla f(x)$. Formula (24) is well known in unconstrained optimization as the BFGS (Broyden, Fletcher, Goldfarb, Shanno) quasi-Newton update.

Fletcher's observation opens the possibility of adding more complicated LMI constraints to (23), and solving the resulting problem numerically. For example, we can impose a certain sparsity pattern on H^+, or we can relax the secant condition as

$$\|H^+(x^+ - x) - \nabla f(x^+) + \nabla f(x)\| \le \epsilon,$$

where ϵ is a given tolerance.

Updating H^+ by numerically solving a convex optimization problem will obviously involve far more computation than the BFGS update. Thus, this formulation for quasi-Newton updates is only interesting when gradient evaluations are very expensive.

8 Conclusion

The applications we described are only a few of the many applications of SDP that have been recently investigated, especially in the areas of combinatorial optimization and control theory. This recent research has been motivated by the development of interior-point methods for SDP. Most interior-point methods for linear programming have been generalized to SDPs. As in linear programming, these methods have polynomial worst-case complexity, and perform very well in practice. Several implementations have become available in the last few years. These include a commercial SDP solver [GNLC95], and several public domain general-purpose SDP software packages: SDPPACK [AHNO97], CSDP [Bor97], SDPMATH [BPS96], LMITOOL [END95], SDPA [FK95], SDPT3 [TTT96], SP [VB94], SDPSOL [WB96]. An important difference with LP is that most of these codes do not exploit problem structure (*e.g.*, sparsity), and if they do, only to a limited extent. Larger SDPs can be solved successfully with interior-point methods but require specialized techniques that take advantage of problem structure.

Acknowledgments

We thank Henry Wolkowicz for pointing out refs. [BN89] and [Fle91].

References

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

[AHNO97] F. Alizadeh, J. P. Haeberly, M. V. Nayakkankuppam, and M. L. Overton. SDPPACK *User's Guide, Version 0.8 Beta.* NYU, June 1997.

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

[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.

[BN89] R. H. Byrd and J. Nocedal. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM J. on Numerical Analysis*, pages 727–739, 1989.

[Bor97] B. Borchers. CSDP*, a C library for semidefinite programming.* New Mexico Tech, March 1997.

[BPS96] N. Brixius, F. A. Potra, and R. Sheng. Solving semidefinite programs with Mathematica. Technical Report 97/1996, Department of Mathematics, University of Iowa, 1996.

[BTN92] A. Ben-Tal and A. Nemirovskii. Interior point polynomial time method for truss topology design. Technical Report 3/92, Faculty of Industrial Engineering and Management, Technion, Israel Institute of Technology, June 1992.

[BTN95a] A. Ben-Tal and A. Nemirovski. Optimal design of engineering structures. *Optima*, pages 4–9, October 1995.

[BTN95b] A. Ben-Tal and A. Nemirovski. Robust truss topology design via semidefinite programming. Technical Report 4/95, Optimization Laboratory. Faculty of Industrial Engineering. Technion, August 1995.

[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.

[DW93] J. E. Dennis and H. Wolkowicz. Sizing and least change secant methods. Technical Report CORR 90-02, Department of Combinatorics and Optimization. University of Waterloo, 1993.

[END95] L. El Ghaoui, R. Nikoukha, and F. Delebecque. LMITOOL*: a front-end for LMI optimization, user's guide*, 1995. Available via anonymous ftp to `ftp.ensta.fr` under `/pub/elghaoui/lmitool`.

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

[FK95] K. Fujisawa and M. Kojima. SDPA (semidefinite programming algorithm) user's manual. Technical Report B-308, Department of Mathematical and Computing Sciences. Tokyo Institute of Technology, 1995.

[Fle91] R. Fletcher. A new variational result for quasi-Newton formulae. *SIAM J. on Optimization*, pages 18–21, 1991.

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

[GNLC95] P. Gahinet, A. Nemirovski, A. J. Laub, and M. Chilali. *LMI Control Toolbox*. The MathWorks, Inc., 1995.

[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.

[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.

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

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

[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.

[TTT96] K. C. Toh, M. J. Todd, and R. H. Tütüncü. SDPT3*: a* MATLAB *software package for semidefinite programming*, 1996. Available at `http://www.math.nus.sg/~mattohkc/index.html`.

[VB94] L. Vandenberghe and S. Boyd. SP*: Software for Semidefinite Programming. User's Guide, Beta Version.* Stanford University, October 1994. Available at `http://www-isl.stanford.edu/people/boyd`.

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

[VBE96] L. Vandenberghe, S. Boyd, and A. El Gamal. Optimizing dominant time constant in RC circuits. Technical report, Information Systems Laboratory, Stanford University, November 1996.

[VBE97] L. Vandenberghe, S. Boyd, and A. El Gamal. Optimal wire and transistor sizing for circuits with non-tree topology. In *Proceedings of the 1997 IEEE/ACM International Conference on Computer Aided Design*, pages 252–259, 1997.

[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.