Positive-Definite Programming

Lieven Vandenberghe
Electrical Engineering Department
K.U. Leuven
Leuven, Belgium

Stephen Boyd
Electrical Engineering Department
Stanford University
Stanford, California

1 Introduction

1.1 Positive-definite programming

We consider the problem of minimizing a linear function of a variable $x \in \mathbf{R}^m$ subject to a *linear matrix inequality:*

$$\begin{array}{ll} \text{minimize} & c^T x, \\ \text{subject to} & F(x) \geq 0, \end{array} \tag{1}$$

where

$$F(x) \triangleq F_0 + \sum_{i=1}^{m} x_i F_i. \tag{2}$$

The problem data are the vector $c \in \mathbf{R}^m$ and $m+1$ symmetric matrices $F_0, \ldots, F_m \in \mathbf{R}^{n \times n}$. The inequality sign in $F(x) \geq 0$ means that $F(x)$ is positive-semidefinite, i.e., $z^T F(x) z \geq 0$ for all $z \in \mathbf{R}^n$. This problem is called a *positive-definite program* (PDP), following Nesterov and Nemirovsky [43].

Problem (2) is a convex optimization problem since its objective and constraint are convex: if $F(x) \geq 0$ and $F(y) \geq 0$, then, for all $\lambda, 0 \leq \lambda \leq 1$,

$$F(\lambda x + (1-\lambda y)) = \lambda F(x) + (1-\lambda) F(y) \geq 0.$$

Although the PDP (1) may appear quite specialized, we will see that it includes many important optimization problems as special cases. For instance, consider the linear program

$$\begin{array}{ll} \text{minimize} & c^T x, \\ \text{subject to} & Ax \geq b, \end{array} \tag{3}$$

where the inequality denotes *componentwise* inequality. Since a vector $v \geq 0$ (componentwise) if and only if the matrix $diag(v)$ (i.e., the diagonal matrix with the components of v on its diagonal) is positive-semidefinite, we can express the LP (3) as a PDP with $F(x) = diag(Ax+b)$, i.e.,

$$F_0 = -diag(b), \quad F_i = diag(a_i), \quad i = 1, \ldots, m,$$

where $A = [a_1 \ldots a_m] \in \mathbf{R}^{n \times m}$.

Positive-definite programming can therefore 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. We can also view the PDP (1) as a semi-infinite linear program, since the matrix inequality $F(x) \geq 0$ is equivalent to an infinite set of linear constraints on x, i.e., $z^T F(x) z \geq 0$ for each $z \in \mathbf{R}^n$. It is therefore not surprising that the theory of positive-definite programming closely parallels linear programming theory, or that many algorithms for solving linear programs should have generalizations that handle PDPs. There are many important differences, however. For instance, the duality results are weaker for PDPs than for LPs. As another important difference, there is no simple or obvious analog of the simplex method for PDPs.

Before proceeding further we give a simple example of a *nonlinear* (convex) optimization problem that can be cast as a PDP, but not as a linear program. Consider the problem

$$\begin{array}{ll} \text{minimize} & \frac{(c^T x)^2}{d^T x} \\ \text{subject to} & Ax + b \geq 0, \end{array} \tag{4}$$

where we assume that $d^T x > 0$ whenever $Ax + b \geq 0$. We start with the standard trick of introducing an auxiliary variable t that serves as an upper bound on the objective:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & Ax + b \geq 0 \\ & \dfrac{(c^T x)^2}{d^T x} \leq t. \end{array} \tag{5}$$

In this formulation, the objective is a linear function of the variables x and t; the nonlinear (convex) objective in (4) shows up as a nonlinear (convex) constraint in (5). These constraints, in turn, can be expressed as a linear matrix inequality in the variables x and t:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} diag(Ax+b) & 0 & 0 \\ 0 & t & c^T x \\ 0 & c^T x & d^T x \end{bmatrix} \geq 0. \end{array} \tag{6}$$

(Here, again, $diag(v)$ represents the diagonal matrix with the elements of v on its diagonal.) Thus we have reformulated the nonlinear (convex) problem (4) as the PDP (6). More examples and applications will be given in the next section.

There are good reasons for studying positive-definite programming problems. First, positive-definiteness constraints arise directly in a number of important applications. Secondly, many convex optimization problems, e.g., linear programming and (convex) quadratically constrained quadratic programming, can be cast as PDPs. Positive-definite programming therefore offers a unified way to study the properties of and derive algorithms for a wide variety of convex optimization problems. Most importantly, however, *PDPs can be solved very efficiently, both in theory and in practice.*

Theoretical tractability follows from convexity, along with the observation that we can construct, in polynomial time, a cutting plane for the constraint set through any given infeasible point (see, e.g., [8, §2.3]). One can therefore apply the ellipsoid method of Yudin and Nemirovsky, and Shor (see [64, 56]) to solve problem (1) in polynomial time. In practice, however, the ellipsoid method is slow.

In this paper we concentrate on recently developed interior-point methods for positive-definite programming. Of course general-purpose nonlinear optimization methods (trust region methods, sequential quadratic optimization, ...) could be used, possibly after modification, to solve PDPs. Interior-point methods, however, enjoy several properties that make them especially interesting.

- It is now generally accepted that interior-point methods for linear programming are competitive with the simplex method and even faster for problems with more than $10,000$ variables or constraints (see, e.g., [32]). One can expect to see the same trend more generally, especially since the very efficient simplex method has no counterpart in positive-definite programming. In our experience with positive-definite programming for control applications, we have found interior-point methods to be very efficient.
- Interior-point methods have a polynomial worst-case complexity.
- Interior-point methods are ideally suited for structured problems. We will see that every iteration of an interior-point method involves the solution of a least-squares problem with the same structure as $F(x)$ in (1). These matrices are often highly structured but not necessarily sparse. The structure can be exploited by combining interior-point methods with iterative least-squares methods such as conjugate-gradients [21] or LSQR [49]. This is not possible in the simplex method, for instance, nor in many other classical methods.

1.2 Examples and Applications

In this section we list a few examples and applications. The list is not exhaustive, and the purpose is more to give an idea of the generality of the problem. More examples are described in [43] and [8].

Quadratically Constrained Quadratic Programming

A convex quadratic constraint $(Ax+b)^T(Ax+b) - c^Tx - d \le 0$ can be written as

$$\begin{bmatrix} I & Ax+b \\ (Ax+b)^T & c^Tx+d \end{bmatrix} \ge 0.$$

The left-hand side depends affinely on the vector x: it can be expressed as $F_0 + x_1F_1 + \cdots + x_mF_m \ge 0$, with

$$F_0 = \begin{bmatrix} I & b \\ b^T & d \end{bmatrix}, \quad F_i = \begin{bmatrix} 0 & a_i \\ a_i^T & c_i \end{bmatrix}, \; i = 1, \ldots, m,$$

where $A = [a_1 \ldots a_m]$. Therefore, a general *quadratically constrained quadratic program*

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \; i = 1, \ldots, L, \end{array}$$

where each f_i is a convex quadratic function $f_i(x) = (A_ix+b)^T(A_ix+b) - c_i^Tx - d_i$, can be written as

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} I & A_0x+b_0 \\ (A_0x+b_0)^T & c_0^Tx_0+d_0+t \end{bmatrix} \ge 0, \\ & \begin{bmatrix} I & A_ix+b_i \\ (A_ix+b_i)^T & c_i^Tx+d_i \end{bmatrix} \ge 0, \; i = 1, \ldots, L. \end{array}$$

This is a PDP in x and t, since one can think of the $L+1$ matrix inequalities as diagonal blocks of one block diagonal matrix inequality $F(x,t) \ge 0$.

Matrix Norm and Maximum Eigenvalue Minimization

Suppose $A(x)$ is a (possibly rectangular) matrix that depends affinely on x: $A(x) = A_0 + x_1A_1 + \cdots + x_mA_m$. The problem of minimizing the (spectral, or maximum singular value) norm $\|A(x)\|$ over x is a PDP:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} tI & A(x) \\ A(x)^T & tI \end{bmatrix} \ge 0. \end{array} \tag{7}$$

If $A(x)$ is a symmetric matrix, a related problem is to minimize the maximum eigenvalue of A:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & tI - A(x) \ge 0. \end{array}$$

Note that both $\|A(x)\|$ and the maximum eigenvalue $\lambda_{max}(A(x))$ are nondifferentiable functions of x.

Logarithmic Chebychev Approximation

Suppose we wish to solve $Ax \approx b$ approximately, where $A = [a_1 \cdots a_n]^T \in \mathbf{R}^{n\times m}$ and $b \in \mathbf{R}^n$. In Chebychev approximation we minimize the infinity norm of the residual, i.e., we solve

$$\text{minimize} \;\; \max_i |a_i^Tx - b_i|.$$

This can be cast as a linear program, with x and an auxiliary variable t as variables:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t \le a_i^T x - b_i \le t, \quad i = 1, \ldots, n. \end{array}$$

In some applications b_i has the dimension of a power or intensity, and is typically expressed on a logarithmic scale. In such cases the more natural optimization problem is

$$\text{minimize} \ \ \max_i | \log(a_i^T x) - \log(b_i)| \tag{8}$$

(assuming $b_i > 0$, and interpreting $\log(a_i^T x)$ as $-\infty$ when $a_i^T x \le 0$).

This *logarithmic Chebychev approximation* problem can be cast as a PDP. To see this, note that

$$|\log(a_i^T x) - \log(b_i)| = \log \max(a_i^T x/b_i, b_i/a_i^T x)$$

(assuming $a_i^T x > 0$). Problem (8) is therefore equivalent to

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & 1/t \le a_i^T x/b_i \le t, \quad i = 1, \ldots, n, \end{array}$$

or,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} t - a_i^T x/b_i & 0 & 0 \\ 0 & a_i^T x/b_i & 1 \\ 0 & 1 & t \end{bmatrix} \ge 0, \quad i = 1, \ldots, n, \end{array}$$

which is a PDP. This example illustrates two important points. It shows that positive-definite programming includes many optimization problems that do not look like (1) at first sight. And secondly, it shows that the problem is much more general than linear programming, despite the close analogy.

Control and System Theory

Positive-definite programming problems arise frequently in control and system theory. Boyd, El Ghaoui, Feron and Balakrishnan catalog many examples in [8]. We will describe one simple example here.

Consider the *differential inclusion*

$$\frac{dx}{dt} \in Co\{A_1, \ldots, A_L\} x(t), \tag{9}$$

where $x(t) \in \mathbf{R}^n$ and the matrices $A_1, \ldots, A_L$ are given, and $Co\{A_1, \ldots, A_L\}$ denotes the *convex hull* of $A_1, \ldots, A_L$. We seek an ellipsoidal invariant set, i.e., an ellipsoid $\mathcal{E}$ such that for any x that satisfies (9), $x(T) \in \mathcal{E}$ implies $x(t) \in \mathcal{E}$ for all $t \ge T$. The existence of such an ellipsoid implies, for example, that all solutions of the differential inclusion (9) are bounded.

The ellipsoid $\mathcal{E} = \{x \mid x^T P x \le 1\}$, where $P = P^T > 0$, is invariant if and only if the function $V(t) = x(t)^T P x(t)$ is nonincreasing for any solution x of (9). (In this case

we say that V is a quadratic Lyapunov function that proves stability of the differential inclusion (9).) Thus, $\mathcal{E}$ is invariant if and only if

$$\frac{d}{dt} V(x(t)) = x(t)^T \left(A(t)^T P + P A(t) \right) x(t) \le 0,$$

for any $x(t) \in \mathbf{R}^n$ and $A(t) \in Co\{A_1, \ldots, A_L\}$. This is equivalent to $A^T P + PA \le 0$ for all $A \in Co\{A_1, \ldots, A_L\}$, which in turn is equivalent to the condition

$$A_k^T P + P A_k \le 0, \quad k = 1, \ldots, L.$$

This is a linear matrix inequality constraint in the matrix P, considered as the variable.

To find an invariant ellipsoid for the differential inclusion (9) (or verify that none exists), we need to solve the feasibility problem

$$P > 0, \quad A_k^T P + P A_k \le 0, \quad k = 1, \ldots, L \tag{10}$$

for the (matrix) variable P. Several standard methods can be used to convert this feasibility problem into a PDP that has an obvious initial feasible point. For instance, we can solve the PDP with variables $P = P^T \in \mathbf{R}^{n \times n}$ and $t \in \mathbf{R}$,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & A_k^T P + P A_k \le 0, \quad k = 1, \ldots, L, \\ & P \ge -tI, \\ & P \le I. \end{array}$$

(The last constraint is added, without loss of generality, to normalize the otherwise homogeneous problem.) This PDP can be initialized with $P = 0$, $t = 1$ and then solved; the optimum value of t is negative if and only if (10) is feasible.

Structural Optimization

Ben-Tal and Bendsøe in [11] consider the following problem from structural optimization. A structure of L linear elastic bars connect a set of N nodes. The geometry (topology and lengths of the bars) and the material (Young's modulus) are fixed; the task is to size the bars, i.e., determine appropriate cross-sectional areas for the bars. In the simplest version of the problem we consider one fixed set of externally applied nodal forces f_i, $i = 1, \ldots, N$. (More complicated versions consider mulitple loading scenarios.) The vector of (small) node displacements resulting from the load forces f will be denoted d. The objective is the elastic stored energy $\frac{1}{2} f^T d$, which is a measure of the inverse of the stiffness of the structure. We also need to take into account constraints on the total volume (or equivalently, weight), and upper and lower bounds on the cross-sectional area of each bar.

The design variables are the cross-sectional areas x_i. The relation between f and d is linear: $A(x)d = f$, where

$$A(x) \triangleq \sum_{i=1}^{N} x_i A_i$$

is called the stiffness matrix. The matrices A_i are all symmetric positive-semidefinite and depend only on fixed parameters (Young's modulus, length of the bars, and geometry). The optimization problem then becomes (see [11])

$$\begin{array}{ll} \text{minimize} & f^T d \\ \text{subject to} & A(x)d = f, \\ & \sum_{i=1}^{L} l_i x_i \leq v, \\ & \underline{x}_i \leq x_i \leq \overline{x}_i \quad i = 1, \ldots, L, \end{array}$$

where d and x are the variables, v is maximum volume, l_i are the lengths of the bars, and $\underline{x}_i$, $\overline{x}_i$ the upper and lower bounds on the cross-sectional areas. For simplicity, we assume that $\underline{x}_i > 0$, and that $A(x)$ is positive-definite for all positive values of x_i. We can then eliminate d and write

$$\begin{array}{ll} \text{minimize} & f^T A(x)^{-1} f \\ \text{subject to} & \sum_{i=1}^{L} l_i x_i \leq v, \\ & \underline{x}_i \leq x_i \leq \overline{x}_i, \quad i = 1, \ldots, L, \end{array}$$

or,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} t & f^T \\ f & A(x) \end{bmatrix} \geq 0, \\ & \sum_{i=1}^{L} l_i x_i \leq v, \\ & \underline{x}_i \leq x_i \leq \overline{x}_i, \quad i = 1, \ldots, L, \end{array}$$

which is a PDP in x and t.

Pattern Separation by Ellipsoids

The simplest classifiers in pattern recognition use hyperplanes to separate two sets. A hyperplane $a^T x + b = 0$ separates two sets of points $\{x_i\}$ and $\{y_j\}$ if

$$\begin{array}{ll} a^T x_i + b < 0 & \text{for all } i, \\ a^T y_j + b > 0 & \text{for all } j. \end{array}$$

This is a set of linear inequalities in $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$, and a solution can be found by linear programming. If the two sets cannot be separated by a hyperplane, we can try to separate them by a quadratic surface. In other words we seek a quadratic function $f(x) = x^T P x + b^T x + c$ that satisfies

$$(x_i)^T P x_i + b^T x_i + c < 0 \quad \text{for all } i, \tag{11}$$

$$(y_j)^T P y_j + b^T y_j + c > 0 \quad \text{for all } j. \tag{12}$$

These inequalities are a set of linear inequalities in the variables $P = P^T \in \mathbf{R}^n$, $b \in \mathbf{R}^n$, and $c \in \mathbf{R}$, and again can be solved using linear programming.

We can put further restrictions on the quadratic surface separating the two sets. For instance, for cluster analysis we might try to find an ellipsoid that contains all the points x_i and none of the y_j (see [53]). This constraint imposes the condition $P > 0$ in addition to the linear inequalities (11) and (12) on the variables P, b, and c. Thus finding an ellipsoid that contains all the x_i variables but none of the y_i variables (or determining that no such ellipsoid exists) can be done by solving a linear matrix inequality feasibility problem.

We can optimize the shape and the size of the ellipsoid by adding an objective function and other constraints. For instance, the ratio of the largest to the smallest semi-axis length is the square root of the condition number of P. In order to make the ellipsoid as spherical as possible, one can introduce an additional variable γ, add the constraint

$$I \leq P \leq \gamma I, \tag{13}$$

and minimize γ over (11), (12) and (13). This is a PDP in the variables γ, P, b and c. This PDP will be feasible if and only if there is an ellipsoid that contains all the x_i and none of the y_i; its optimum value is one if and only there is a sphere that separates the two sets of points.

Geometrical Problems Involving Quadratic Forms

Many geometrical problems involving quadratic functions can be expressed as PDPs. We will give one simple example. Suppose we are given m ellipsoids $\mathcal{E}_1, \ldots, \mathcal{E}_m$ described as the sublevel sets of the quadratic functions

$$f_i(x) = x^T A_i x + 2b_i^T x + c_i, \quad i = 1, \ldots, m,$$

i.e., $\mathcal{E}_i = \{x | f_i(x) \leq 0\}$. The goal is to find the smallest sphere that contains all m of these ellipsoids (or equivalently, contains the convex hull of their union).

The condition that one ellipsoid contain another can be expressed in terms of a matrix inequality. Suppose that the ellipsoids $\mathcal{E} = \{x | f(x) \leq 0\}$ and $\tilde{\mathcal{E}} = \{x | \tilde{f}(x) \leq 0\}$, with

$$f(x) = x^T A x + 2b^T x + c, \quad \tilde{f}(x) = x^T \tilde{A} x + 2\tilde{b}^T x + \tilde{c},$$

have nonempty interior. Then it can be shown that $\mathcal{E}$ contains $\tilde{\mathcal{E}}$ if and only if there is a $\tau \geq 0$ such that

$$\begin{bmatrix} A & b \\ b^T & c \end{bmatrix} \leq \tau \begin{bmatrix} \tilde{A} & \tilde{b} \\ \tilde{b}^T & \tilde{c} \end{bmatrix}.$$

(The 'if' part is trivial; the 'only if' part is less trivial. See [8, 59]).

Returning to our problem, consider the sphere $\mathcal{S}$ represented by $f(x) = x^T x - 2x_c^T x + \gamma \leq 0$. $\mathcal{S}$ contains the ellipsoids $\mathcal{E}_1, \ldots, \mathcal{E}_m$ if and only if there are nonnegative $\tau_1, \ldots, \tau_m$ such that

$$\begin{bmatrix} I & -x_c \\ -x_c^T & \gamma \end{bmatrix} \leq \tau_i \begin{bmatrix} A_i & b_i \\ b_i^T & c_i \end{bmatrix}.$$

Note that these conditions can be considered one large linear matrix inequality in the variables x_c, γ, and $\tau_1, \ldots, \tau_m$.

Our goal is to minimize the radius of the sphere $\mathcal{S}$, which is $r = \sqrt{x_c^T x_c - \gamma}$. To do this we express the condition $r^2 \leq t$ as the matrix inequality

$$\begin{bmatrix} I & x_c \\ x_c^T & t+\gamma \end{bmatrix} \geq 0$$

and minimize the variable t.

Putting it all together we see that we can find the smallest sphere containing the ellipsoids $\mathcal{E}_1, \ldots, \mathcal{E}_m$ by solving the PDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \begin{bmatrix} I & -x_c \\ -x_c^T & \gamma \end{bmatrix} \leq \tau_i \begin{bmatrix} A_i & b_i \\ b_i^T & c_i \end{bmatrix}, \quad i=1,\ldots,m, \\ & \tau_i \geq 0, \quad i=1,\ldots,m, \\ & \begin{bmatrix} I & x_c \\ x_c^T & t+\gamma \end{bmatrix} \geq 0. \end{array}$$

The variables here are $x_c, \tau_1, \ldots, \tau_m, \gamma$, and t.

This example demonstrates once again the breadth of problems that can be reformulated as PDPs. It also demonstrates that the task of this reformulation can be nontrivial.

Other Fields

- PDPs occur in statistics, in minimum trace factor analysis (see Watson [61]), as the *educational testing problem* (see [17, 18]), and in optimum experiment design (see Pukelsheim [50]).
- PDPs have been used to compute upper or lower bounds for combinatorial optimization problems. Examples are Lovász's famous upper bound on the Shannon capacity of a graph [33], Shor's bounds for integer programming, and Alizadeh's work [4, 2, 3].
- The problem of minimizing the maximum eigenvalue of a matrix has been studied extensively by Overton; see [45] for a list of applications.

1.3 Historical Overview

A very early paper on the theoretical properties of PDPs is Bellman and Fan [9]. Other references discussing optimality conditions are Craven and Mond [13], Shapiro [55], Fletcher [18], and Allwright [5].

Many researchers have worked on the problem of minimizing the maximum eigenvalue of a symmetric matrix. See, for instance, Cullum, Donath and Wolfe [12], Goh and Teo [24], Panier [48], Allwright [6], Overton [44, 45], Overton and Womersley [47, 46], Ringertz [51], Fan and Nekooie [20], Fan [16], and Hiriart-Urruty and Ye [26].

The history of interior-point methods is relatively young. Interior-point methods for linear programming were introduced by Karmarkar in 1984 [28], although many of the underlying principles are older (see, e.g., Fiacco and McCormick [19], Lieu and Huard [31], and Dikin [15]). Karmarkar's algorithm, and the interior-point methods developed afterwards, combine a very low, polynomial, worst-case complexity with excellent behavior in practice.

Karmarkar's paper has had an enormous impact, and several variants of his method have been developed (see, e.g., the survey by Gonzaga [23]). Interior-point methods have also been extended and generalized to convex quadratic programming, and to certain linear complementarity problems (see Kojima, Megiddo, Noma and Yoshise [29]).

An important breakthrough was achieved by Nesterov and Nemirovsky in 1988 [38, 40, 39, 41, 41]. They showed that the interior-point methods for linear programming can be generalized to all convex optimization problems. The key element is the knowledge of a barrier function with certain properties (*self-concordance*). Unfortunately, although Nesterov and Nemirovsky prove that a self-concordant barrier function exists for every convex set, it is not always known how to compute it in practice.

PDPs are an important class of convex optimization problems for which self-concordant barrier functions are known, and, therefore, interior-point methods are applicable. At the same time, they offer a simple conceptual framework and make possible a self-contained treatment of interior-point methods for many convex optimization problems.

Independently of Nesterov and Nemirovsky, Alizadeh [4] has generalized interior-point methods from linear programming to positive-definite programming. Other recent articles are Jarre [27], Vandenberghe and Boyd [60], Rendl, Vanderbei and Wolkowicz [54], Yoshise [65], and Alizadeh, Haeberly and Overton [1]. An excellent reference on interior-point methods for general convex problems is Den Hertog [14].

1.4 Outline

This paper gives a survey of interior-point methods for positive-definite programming. We start with a section on duality theory. In Section 3 we introduce the barrier function for PDPs, and the concepts of central points and central path. The notion of central path is heavily used in Section 4, which discusses primal-dual methods.

This survey is not meant to be exhaustive and emphasizes primal-dual methods. The most important omissions are the projective methods of Karmarkar, and of Nesterov and Nemirovsky [43]. Our motivation for the restriction to primal-dual methods is twofold. Primal-dual methods are commonly held to be more efficient in practice, and, secondly, their behavior is often easier to analyze. Moreover, all interior-point methods are based on similar principles, and we hope that the material discussed here is sufficient as a tutorial introduction to the entire field.

2 Duality

2.1 The Dual PDP

The *dual* problem associated with the PDP (1) is

$$\begin{array}{ll} \text{maximize} & -TrF_0Z \\ \text{subject to} & TrF_iZ = c_i,\ i = 1, \dots, m, \\ & Z \geq 0. \end{array} \quad (14)$$

Here the variable is the symmetric matrix Z, which is subject to m equality constraints and the nonnegativity condition. We write TrX for the trace of a symmetric matrix, i.e., $TrX = X_{11} + \cdots + X_{nn}$. Note that the objective function in (14) is a linear function of Z.

The dual problem (14) is also a PDP, i.e., it can be put in the same form as the *primal* problem (1). Let us assume for simplicity that the matrices $F_1, \dots, F_m$ are linearly independent. Then we can express

$$\{ Z \mid Z = Z^T \in \mathbf{R}^{n \times n},\ TrF_iZ = c_i \}$$

in the form

$$\{ G(y) = G_0 + y_1G_1 + \cdots + y_pG_p \mid y \in \mathbf{R}^p, \}$$

where $p = \frac{n(n+1)}{2} - m$ and the G_i are appropriate matrices. We define $d \in \mathbf{R}^p$ by $d_i = TrF_0G_i$, so that $d^Ty = TrF_0(G(y) - G_0)$. Then the dual problem becomes (except for a constant term in the objective and a change of sign to transform maximization into minimization)

$$\begin{array}{ll} \text{minimize} & d^Ty \\ \text{subject to} & G(y) \geq 0, \end{array}$$

which is a PDP. It is possible to use notation that, unlike ours, emphasizes the complete symmetry between the primal and dual problems (see, e.g., Nesterov and Nemirovsky). Our notation was designed to make the primal problem as "explicit" as possible, with x denoting a "free" variable.

As an example of the dual PDP, let us apply the definition to the linear program (3), i.e., take $F_0 = -diag(b)$ and $F_i = diag(a_i)$. In this case, the diagonal structure makes it possible to simplify the dual problem. The objective function and the equality constraints only involve the diagonal elements of Z, and, obviously, replacing the off-diagonal of a positive-definite matrix by zeros does not alter its positive-definiteness either. Instead of optimizing over all symmetric $n \times n$ matrices Z, we can therefore limit ourselves to diagonal matrices $Z = diag(z)$. Problem (14) then reduces to

$$\begin{array}{ll} \text{maximize} & b^Tz \\ \text{subject to} & z \geq 0, \\ & a_i^Tz = c_i,\ \ i = 1, \dots, m, \end{array} \quad (15)$$

which is the familiar dual of the LP (3).

This example demonstrates an important point. In general, it is often the case that the dual problem can be simplified when the matrices F_i are structured. For example, if the matrix $F(x)$ is block diagonal, the dual variables Z can be assumed to have the same block diagonal structure.

Linear programming duality is very strong owing to the polyhedral character of the feasible set: The optimum values of (3) and (15) are always equal, except in the pathological case where both problems are infeasible. (We adopt the standard convention that the optimum value of (3) is $+\infty$ if the problem is infeasible, and the optimum value of (15) is $-\infty$ if the dual problem is infeasible.) Duality results for general PDPs are weaker, as we will see below.

Let us return to our discussion of the dual PDP. The key property of the dual PDP is that it yields bounds on the optimal value of the primal PDP, and vice versa. Suppose that Z is dual feasible, and x is primal feasible. Then we have:

$$-TrF_0Z \leq c^Tx. \quad (16)$$

The inequality follows from the simple calculation

$$c^Tx + TrZF_0 = \sum_{i=1}^{m} TrZF_ix_i + TrZF_0 = TrZF(x) \geq 0.$$

(We used the fact that $TrAB \geq 0$ when $A = A^T \geq 0$ and $B = B^T \geq 0$.)

Since (16) holds for *any* feasible x, we conclude that $p^* \geq -TrZF_0$, where p^* is the optimal value of the PDP (1),

$$p^* \triangleq inf \{ c^Tx \mid F(x) \geq 0 \}. \quad (17)$$

In other words: *Dual feasible matrices yield lower bounds for the primal problem.* We can interpret x as a *suboptimal point* which gives the upper bound $p^* \leq c^Tx$ and Z as a *certificate* that proves the lower bound $p^* \geq -TrF_0Z$.

If x is primal feasible and Z is dual feasible, we refer to the quantity $\eta \triangleq c^Tx + TrF_0Z$ as the *duality gap* associated with x and Z. The duality gap is the difference between the upper and lower bound; it is the width of the interval in which we have localized p^*. If we define d^* to be the optimal value in the dual problem,

$$d^* \triangleq sup \{ -TrF_0Z \mid Z = Z^T \geq 0,\ TrF_iZ = c_i,\ i = 1, \dots, m \}, \quad (18)$$

then we can restate the result (16) as $p^* \geq d^*$, i.e., the optimal value of the dual problem is less than or equal to the optimal value of the primal problem. Note that when the primal (or dual) problem is infeasible, the right-hand (left-hand) side becomes ∞ $(-\infty)$ so the inequality trivially holds. In fact equality usually obtains, as stated in the following theorem (see Nesterov and Nemirovsky [43], or Rockafellar [52]).

Theorem 1 *We have $p^* = d^*$ if any of the following conditions holds.*

1. *The primal problem (1) is strictly feasible.*
2. *The dual problem (14) is strictly feasible.*

3. *The primal solution set*

$$X_{opt} \triangleq \{x \mid F(x) \geq 0 \text{ and } c^T x = p^*\}$$

is nonempty and bounded.

4. *The dual solution set*

$$Z_{opt} \triangleq \{Z \mid Z \geq 0,\ TrF_i Z = c_i, \text{ and } -TrF_0 Z = d^*\}$$

is nonempty and bounded.

Example
Consider the PDP

$$\begin{array}{ll} \text{minimize} & x_1 \\ \text{subject to} & \begin{bmatrix} 0 & x_1 & 0 \\ x_1 & x_2 & 0 \\ 0 & 0 & x_1+1 \end{bmatrix} \geq 0. \end{array}$$

The feasible set is $\{(x_1, x_2) \mid x_1 = 0, x_2 \geq 0\}$, and therefore $p^* = 0$. The dual problem can be simplified as

$$\begin{array}{ll} \text{maximize} & -z_2 \\ \text{subject to} & \begin{bmatrix} z_1 & (1-z_2)/2 & 0 \\ (1-z_2)/2 & 0 & 0 \\ 0 & 0 & z_2 \end{bmatrix} \geq 0, \end{array}$$

and the feasible set is $\{(z_1, z_2) \mid z_1 \geq 0, z_2 = 1\}$. The dual problem therefore has optimal value $d^* = -1$. This PDP violates all four conditions mentioned in the theorem. Both problems are feasible, but not strictly feasible, and the optimal sets X_{opt} and Z_{opt} are both unbounded. Note also the contrast with linear programming, where it is impossible to have a finite nonzero duality gap at the optimum.

Example
We take the matrix norm minimization problem mentioned in Section 1.2:

$$\begin{array}{l} \text{minimize } \|A(x)\| \\ x \in \mathbf{R}^m \end{array} \tag{19}$$

where $A(x) = A_0 + x_1 A_1 + \cdots + x_m A_m$, and we remind the reader that $\|A(x)\|$ is the maximum singular value of $A(x)$.

The problem (19) is a basic problem in the theory of Banach spaces; its optimum value is the norm of (the image of) A_0 in the quotient space of all $p \times q$ matrices modulo the span of $A_1, \ldots, A_m$. In this theory we encounter the following dual of (19):

$$\begin{array}{ll} \text{maximize} & TrA_0^T Q \\ \text{subject to} & TrA_i^T Q = 0,\ \ i = 1, \ldots, m, \\ & \|Q\|_* \leq 1, \end{array} \tag{20}$$

where $\|Q\|_* = \sum \sigma_i(Q)$ is the *nuclear norm* of Q, which is the norm dual to the maximum singular value. It is also known that the optimal values of (19) and (20) are always equal.

Let us verify that this (Banach space) notion of duality coincides with PDP duality. The dual PDP of problem (7) is

$$\begin{array}{ll} \text{maximize} & -2TrA_0^T Z_{12} \\ \text{subject to} & TrA_i^T Z_{12} = 0,\ \ i = 1, \ldots, m, \\ & TrZ_{11} + TrZ_{22} = 1, \\ & \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{12}^T & Z_{22} \end{bmatrix} \geq 0. \end{array} \tag{21}$$

This can be simplified. The positive-definite constraint can be rewritten as

$$\begin{bmatrix} Z_{11} & 0 \\ 0 & Z_{22} \end{bmatrix} \geq \begin{bmatrix} 0 & -Z_{12} \\ -Z_{12}^T & 0 \end{bmatrix}. \tag{22}$$

The eigenvalues of the matrix on the right are the singular values of Z_{12}, each singular value appearing twice. It is well known that if A and B are two symmetric matrices, then $A \geq B$ implies $\lambda_k(A) \geq \lambda_k(B)$, assuming the eigenvalues of A and B are taken in the same order. As a consequence, inequality (22) implies

$$2\sum \sigma_i(Z_{12}) \leq TrZ_{11} + TrZ_{22} \leq 1.$$

Since the matrices Z_{11} and Z_{22} do not appear in any other constraint, nor in the objective, we see that problem (21) reduces to

$$\begin{array}{ll} \text{maximize} & -2TrA_0^T Z_{12} \\ \text{subject to} & TrA_i^T Z_{12} = 0,\ \ i = 1, \ldots, m, \\ & 2\sum \sigma_i(Z_{12}) \leq 1, \end{array}$$

which is the same as (20) with $Q = 2Z_{12}$.

Problem (19) is always strictly feasible; it suffices to choose $x = 0$ and $t > \|A_0\|$. Applying Theorem 1, we conclude that the optimal duality gap is always zero.

2.2 The Primal-Dual Formulation

Theorem 1 has important consequences for PDP algorithms. It gives conditions under which the primal-dual optimization problem

$$\begin{array}{ll} \text{minimize} & c^T x + TrF_0 Z \\ \text{subject to} & F(x) \geq 0, \\ & Z \geq 0, \\ & TrF_i Z = c_i,\ \ i = 1, \ldots, m, \end{array} \tag{23}$$

has optimum value zero. Here we minimize the duality gap $c^T x + TrF_0 Z$ over all primal and dual feasible points. The duality gap is a linear function of x and Z, and therefore problem (23) is a PDP in x and Z.

Primal-dual methods for PDPs solve (23), assuming the primal and dual problems are strictly feasible. They generate a sequence of feasible points $x^{(k)}$ and $Z^{(k)}$, and in each step use the dual information in $Z^{(k)}$ to find good updates for $x^{(k)}$ and vice-versa.

This means that at every stage of the algorithm, we have available suboptimal primal and dual solutions x, Z. The primal solution x proves an upper bound $c^T x > p^*$ on the optimal value; the dual solution proves a lower bound $p^* > -TrF_0 Z$. The iteration continues until the duality gap is less than a given tolerance ϵ.

3 The Barrier Function

In this section, we introduce a barrier function for linear matrix inequality constraints and discuss its properties. This leads us to the fundamental concept of *centrality*, and the definition of *central points* and *central path*. From now on we will assume that the matrices F_i are independent.

3.1 Definition

The function

$$\phi(x) \triangleq \begin{cases} \log\det F(x)^{-1} & \text{if } F(x) > 0 \\ +\infty & \text{otherwise} \end{cases} \tag{24}$$

is a *barrier function* for $X \triangleq \{x \mid F(x) \geq 0\}$, i.e., $\phi(x)$ is finite if and only if $F(x) > 0$, and becomes infinite as x approaches the boundary of X. There are many other barrier functions for X (for example, trace can be substituted for determinant in (24)), but this one enjoys many special properties (see [43]). In particular, when $F(x) > 0$, it is analytic and strictly convex.

In the case of a set of linear inequalities $Ax \geq b$, where $A = [a_1 \dots a_n]^T$, we have $F(x) = diag(Ax - b)$, and the definition reduces to the familiar logarithmic barrier function

$$\phi(x) = \begin{cases} -\sum_{i=1}^n \log(a_i^T x - b_i) & \text{if } Ax \geq b, \\ +\infty & \text{otherwise.} \end{cases}$$

We first give formulas for the gradient $g(x)$ and Hessian $H(x)$ of ϕ. Recall that $TrAB$ is the standard inner product of two symmetric matrices A and B; the corresponding norm is the Frobenius norm, $\|A\|_F = (TrA^2)^{1/2}$.

The gradient and the Hessian of ϕ can be readily derived from the following second order approximation of the function $-\log\det X$. If $X > 0$ is $n \times n$ and symmetric, then

$$\log\det(X+Y)^{-1} = \log\det X^{-1} - TrX^{-1}Y + \tfrac{1}{2}TrX^{-1}YX^{-1}Y + o\left(\|Y\|^2\right). \tag{25}$$

From equation (25), one can immediately derive a second order approximation for $\phi(x)$:

$$\begin{aligned} \phi(x+v) &\approx \phi(x) - TrF(x)^{-1}\left(\sum_{i=1}^m v_i F_i\right) + \\ &\quad \tfrac{1}{2}TrF(x)^{-1}\left(\sum_{i=1}^m v_i F_i\right)F(x)^{-1}\left(\sum_{j=1}^m v_j F_j\right) \\ &= \phi(x) - \sum_{i=1}^m v_i TrF(x)^{-1}F_i + \tfrac{1}{2}\sum_{i=1}^m\sum_{j=1}^m v_i v_j TrF(x)^{-1}F_iF(x)^{-1}F_j. \end{aligned}$$

We conclude that the gradient $g(x)$ and the Hessian $H(x)$ of $\phi(x)$ are given by

$$g_i(x) = -TrF(x)^{-1}F_i = -TrF(x)^{-1/2}F_iF(x)^{-1/2}, \tag{26}$$

and

$$\begin{aligned} H_{ij}(x) &= TrF(x)^{-1}F_iF(x)^{-1}F_j \\ &= Tr\left(F(x)^{-1/2}F_iF(x)^{-1/2}\right)\left(F(x)^{-1/2}F_jF(x)^{-1/2}\right), \end{aligned} \tag{27}$$

for $i, j = 1, \dots, m$.

From expression (27) we can verify that ϕ is strictly convex for strictly feasible x. For $x, y \in \mathbf{R}^m$ with $F(x) > 0$,

$$\begin{aligned} y^T H(x) y &= \sum_{i,j=1}^m y_i y_j Tr\left(F(x)^{-1/2}F_iF(x)^{-1/2}\right)\left(F(x)^{-1/2}F_jF(x)^{-1/2}\right) \\ &= Tr\left(F(x)^{-1/2}\left(\sum_{i=1}^m y_iF_i\right)F(x)^{-1/2}\right)^2 \\ &= \left\|F(x)^{-1/2}\left(\sum_{i=1}^m y_iF_i\right)F(x)^{-1/2}\right\|_F^2 \geq 0. \end{aligned} \tag{28}$$

We see that $y^T H(x) y = 0$ if and only if $\sum_{i=1}^m y_i F_i = 0$. By independence of $F_1, \dots, F_m$, we conclude that $H(x) > 0$, i.e., ϕ is strictly convex.

Finally, we note that the barrier function ϕ is bounded below if and only if the feasible set X is bounded.

3.2 Analytic Center

3.2.1 Definition

We suppose now that the linear matrix inequality $F(x) \geq 0$ is strictly feasible and that its feasible set is bounded. Since ϕ is strictly convex, it has a unique minimizer, which we denote

$$x^* \triangleq \underset{x}{\text{argmin}}\ \phi(x). \tag{29}$$

We will refer to x^* as the *analytic center* of the linear matrix inequality $F(x) > 0$. It is important to note that the analytic center is a property of a linear matrix inequality and not of its solution set X. The same set X can be represented by different matrix inequalities, which have different analytic centers.

From (26) we see that x^* is characterized by

$$Tr F(x^*)^{-1} F_i = 0, \quad i = 1, \ldots, m. \tag{30}$$

Thus, $F(x^*)^{-1}$ is orthogonal to the span of $F_1, \ldots, F_m$.

In the case of a set of linear inequalities, the definition coincides with Sonnevend's definition [57, 58], i.e.,

$$x^* = \underset{Ax \geq b}{\text{argmax}} \prod_{i=1}^{n} (a_i^T x - b_i).$$

3.2.2 Computing the Analytic Center

Newton's method, with appropriate step length selection, can be used to efficiently compute the analytic center. Starting with a strictly feasible point $x^{(0)}$, the algorithm follows the iteration:

$$x^{(k+1)} := x^{(k)} - \alpha^{(k)} H(x^{(k)})^{-1} g(x^{(k)}), \tag{31}$$

where $\alpha^{(k)}$ is the damping factor at the kth iteration, and $g(x)$ and $H(x)$ are the gradient and Hessian of the barrier function in x.

Nesterov and Nemirovsky [43] give a simple step length rule appropriate for the general class of self-concordant barrier functions mentioned earlier. The damping factor depends on a quantity called the *Newton decrement* of ϕ at x:

$$\delta(x) \triangleq \left\| H(x)^{-1/2} g(x) \right\|.$$

The name comes from the observation that $\delta(x)^2/2$ is the difference between $\phi(x)$ and the minimum value of the quadratic approximation of ϕ at x. Alternatively, $\delta(x)$ is the length of the Newton step $-H(x)^{-1}g(x)$ measured in the norm induced by the Hessian $H(x)$.

The damping factor is:

$$\alpha^{(k)} := \begin{cases} 1 & \text{if } \delta(x^{(k)}) \leq 1/4, \\ 1/(1+\delta(x^{(k)})) & \text{if } \delta(x^{(k)}) > 1/4. \end{cases} \tag{32}$$

Nesterov and Nemirovsky show that this step length always results in $F(x^{(k+1)}) > 0$. Moreover, for $\delta(x^{(k)}) < 1/4$, we have $\delta(x^{(k+1)}) \leq 2\delta(x^{(k)})^2$, i.e., the algorithm converges quadratically. They also give a complete convergence analysis. The main results can be summarized as follows.

- Until the region of quadratic convergence ($\delta(x) < 1/4$) is reached, the objective $\log\det F(x)^{-1}$ decreases at least by the absolute constant 0.3068 at each Newton step. (By absolute constant we mean it does not depend on n, the problem data, or the required accuracy of computing x^*.)

- Once the region of quadratic convergence is reached, at most a *constant* number c of Newton steps is required to compute x^* to a given accuracy. (The constant c does not depend on n or the problem data, but only on the required accuracy ϵ. Since the convergence is quadratic in this region, c grows as $\log\log 1/\epsilon$ if ϵ decreases.)

In other words, the number of Newton steps required to compute x^* given x can be bounded in terms of $-\log\det F(x) + \log\det F(x^*)$:

$$\#\text{Newton steps} \leq c + 3.26(-\log\det F(x) + \log\det F(x^*)), \tag{33}$$

where c depends only on the required accuracy of computing x^* and grows extremely slowly.

Therefore, the quantity $\psi(x) \triangleq -\log\det F(x) + \log\det F(x^*)$ has a very natural interpretation as the 'deviation from centrality' of a point x. In general, however, $\psi(x)$ can only be evaluated by computing the center x^*.

3.3 The Primal-Dual Central Path

We now return to the primal-dual formulation of Section 2.2.

For $\alpha > 0$ consider the set of strictly feasible pairs x, Z with duality gap α, i.e., $c^T x + Tr F_0 Z = \alpha$. The *analytic center* of this set is the minimizer of the barrier term $-\log\det F(x) - \log\det Z$. We denote the analytic center as $x^*(\alpha), Z^*(\alpha)$:

$$\begin{array}{rll} (x^*(\alpha), Z^*(\alpha)) = & \text{argmin} & -\log\det F(x) - \log\det Z \\ & \text{subject to} & F(x) \geq 0, \\ & & Z \geq 0, \\ & & Tr F_i Z = c_i, i = 1, \ldots, m, \\ & & c^T x + Tr F_0 Z = \alpha. \end{array} \tag{34}$$

Thus, among all feasible pairs x, Z with the duality gap α, the pair x^*, Z^* maximizes $det(F(x)Z)$. The pair (x^*, Z^*) converges to a primal and dual optimal pair as $\alpha \to 0$, and the curve given by (x^*, Z^*) for $\alpha > 0$ is called the *central path* for the problem (23).

The central pair (x^*, Z^*) has many important properties. For our purposes here we need:

Theorem 2 $F(x^*(\alpha))Z^*(\alpha) = (\alpha/n)I$. *Conversely, if x and Z are a feasible pair and $F(x)Z = (\alpha/n)I$ then $x = x^*(\alpha)$ and $Z = Z^*(\alpha)$.*

In other words, centrality is characterized by $F(x)$ and Z being inverses of each other, up to a constant.

Now consider a feasible pair (x, Z), and define $\alpha = Tr F(x) Z$. Then $(x^*(\alpha), Z^*(\alpha))$ is the central pair with the same duality gap as x, Z. Therefore

$$\log\det F(x)Z \geq \log\det F(x^*(\alpha))Z^*(\alpha) = n\log n - n\log Tr F(x)Z$$

with equality holding only when x, Z are central.

As in Section 3.2.2 we can say that the difference

$$\begin{aligned}\psi(x,Z) &\triangleq -\log\det F(x)Z + \log\det F(x^*(\alpha))Z^*(\alpha)\\ &= -\log\det F(x)Z + n\log TrF(x)Z - n\log n \qquad (35)\end{aligned}$$

is a measure of the deviation of (x, Z) from centrality: $\psi(x, Z)$ is, up to a constant, an upper bound on the computational effort required to "center" (x, Z) (meaning, compute the central pair with the same duality gap). From (35), we also note that $\psi(x, Z)$ can be evaluated as a function of x and Z, without knowing x^* and Z^*.

The function ψ is not convex or quasiconvex (except of course when restricted to $TrF(x)Z$ constant). We also note that ψ depends only on the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $F(x)Z$:

$$\psi(x,Z) = n\log\frac{\left(\sum_{i=1}^n \lambda_i\right)/n}{\left(\prod_{i=1}^n \lambda_i\right)^{1/n}}.$$

Thus, $\psi(x, Z)$ is n times the logarithm of the ratio of the arithmetic to the geometric mean of the eigenvalues of $F(x)Z$. (From which we see again that ψ is nonnegative, and zero only when $F(x)Z$ is a multiple of the identity.) We can also think of ψ as a smooth measure of condition number of the matrix $F(x)Z$ since

$$\log\kappa - 2\log 2 \leq \psi(x,Z) \leq (n-1)\log\kappa, \qquad (36)$$

where $\kappa = \lambda_{max}/\lambda_{min}$ is the condition number of $F(x)Z$.

4 Primal-Dual Potential Reduction Methods

The methods in this section take full advantage of duality theory by updating primal and dual feasible points in each iteration. The basic idea is the following. We choose a potential function $\varphi(x, Z)$ such that

- φ is smooth on the interior of the feasible set
- φ is unbounded below if x and Z approach the optimal solution
- φ increases unboundedly as x or Z approach all other points on the boundary of the feasible set

The algorithms minimize the potential function by some modification of Newton's method. If the potential function is properly chosen, this leads to algorithms with a polynomial worst-case complexity.

Theoretically there is a perfect symmetry between primal and dual problems. In practice, it is important to keep in mind that the dimension of the dual problem is usually much larger than the dimension of the primal problem. Computations involving the dual problem will therefore have to be arranged carefully.

4.1 Primal-Dual Potential Function

Define the primal-dual potential function

$$\begin{aligned}\varphi(x,Z) &\triangleq \nu\sqrt{n}\log\left(c^Tx + TrF_0Z\right) + \psi(x,Z)\\ &= q\log\left(c^Tx + TrF_0Z\right) - \log\det F(x) - \log\det Z - n\log n, \qquad (37)\end{aligned}$$

where $q = n+\nu\sqrt{n}$. The first term, $\nu\sqrt{n}\log\left(c^Tx + TrF_0Z\right)$, measures a decrease in objective function. The second term, $\psi(x, Z)$, is the deviation from centrality defined in the previous section. The parameter ν controls the relative weight of both terms. To minimize the worst-case complexity it has to be of $O(1)$, but in practice a larger value can be more efficient. Note that along the central path, $\varphi(x^*(\alpha), Z^*(\alpha)) = \nu\sqrt{n}\log\alpha$ which decreases to $-\infty$ as α converges to zero.

A fixed decrease in the first term corresponds to a fixed fractional reduction of the duality gap. A fixed decrease in the second term corresponds to a fixed amount of "centering" in the following sense: up to a constant, it is the reduction in the (bound on) computational effort required to "center" the current pair.

If ψ decreases by one, the new pair is 3.26 Newton steps closer to centrality than the original pair (or more precisely, the upper bound on the number of Newton steps required to center the new pair is 3.26 smaller than the upper bound for the original pair). If the other term, $\nu\sqrt{n}\log TrF(x)Z$, decreases by one, then the duality gap is reduced by the factor $exp(-1/\nu\sqrt{n}) \approx 1 - 1/(\nu\sqrt{n})$. In other words,

$$\nu\sqrt{n} \text{ fewer Newton steps to center } \sim 31\% \text{ duality gap reduction,}$$

where $\sim$ means that the left and right-hand sides result in an equal decrease in φ.

By minimizing the smooth function φ, we solve the primal and dual problems. Indeed since $\psi(x, Z) \geq 0$ for feasible x and Z, we have

$$TrF(x)Z \leq exp\frac{\varphi(x,Z)}{\nu\sqrt{n}}, \qquad (38)$$

which shows that small potential implies small duality gap.

The basic idea of the primal-dual algorithm is to generate iterations of primal and dual feasible matrices satisfying

$$\varphi(x^{(k+1)}, Z^{(k+1)}) \leq \varphi(x^{(k)}, Z^{(k)}) - \delta, \qquad (39)$$

for some absolute positive constant δ. By (38) we therefore have:

$$TrF(x^{(k)})Z^{(k)} \leq exp\frac{\varphi(x^{(0)},Z^{(0)}) - k\delta}{\nu\sqrt{n}} = c_0c_1^kTrF(x^{(0)})Z^{(0)}, \qquad (40)$$

where

$$c_0 = exp\frac{\psi(x^{(0)},Z^{(0)})}{\nu\sqrt{n}}, \quad c_1 = exp\frac{-\delta}{\nu\sqrt{n}}.$$

We can interpret the result (40) as follows: the duality gap converges to zero at least exponentially at a rate given by the constant c_1. The constant c_0 depends only on the centrality of the initial pair, and is one if the initial pair is central.

In other words, we have polynomial convergence:

Theorem 3 *Assume that (39) holds with some $\delta > 0$ that does not depend on n or ϵ, where $0 < \epsilon < 1$. Then for*

$$k \geq \frac{\nu\sqrt{n}\, log(1/\epsilon) + \psi(x^{(0)}, Z^{(0)})}{\delta},$$

we have $TrF(x^{(k)})Z^{(k)} < \epsilon TrF(x^{(0)})Z^{(0)}$.

Roughly speaking, we have convergence in $O(\sqrt{n})$ steps, provided the initial pair is sufficiently centered.

We conclude this section with a conceptual outline of the potential reduction algorithm.

Potential Reduction Algorithm

given strictly feasible x and Z.
repeat
1. Find primal and dual feasible directions δx and δZ.
2. Find $p, q \in \mathbf{R}$ that minimize $\varphi(x + p\delta x, Z + q\delta Z)$.
3. Update: $x := x + p\delta x$ and $Z := Z + q\delta Z$.

until duality gap $\leq \epsilon$.

The key task, then, is to show how to update $(x^{(k)}, Z^{(k)})$ into $(x^{(k+1)}, Z^{(k+1)})$ such that (39) holds.

4.2 Plane Search

In the next sections we will describe several possibilities for computing directions δx and δZ. Here we assume that these search directions are given, and we will show how the potential function $\varphi(x + p\delta x, Z + q\delta Z)$ can be minimized over the plane defined by these two directions. We will use the notation $F \stackrel{\Delta}{=} F(x)$ and $\delta F \stackrel{\Delta}{=} \sum_{i=1}^{m} F_i \delta x_i$.

The two-dimensional minimization can be done very efficiently if we first compute the eigenvalues $\mu_1, \ldots, \mu_n$ of $F^{-1/2}\delta F F^{-1/2}$ and the eigenvalues $\nu_1, \ldots, \nu_n$ of $Z^{-1/2}\delta Z Z^{-1/2}$. The potential function can then be written in terms of p and q as

$$\varphi(p,q) = (n + \nu\sqrt{n})\, log(c_1 + c_2 p + c_3 q) + f(p,q) - n\, log\, n,$$

where $c_1 = TrFZ$, $c_2 = TrZ\delta F$, and $c_3 = TrF\delta Z$, and f is the restriction of the barrier term to the plane, i.e.,

$$\begin{aligned} f(p,q) &= -\,log\,det(F + p\delta F) - log\,det(Z + q\delta Z) \\ &= -\sum_{i=1}^{n} log(1 + p\mu_i) - \sum_{i=1}^{n} log(1 + q\nu_i) - log\,det(FZ). \end{aligned}$$

Note that once we have the eigenvalues μ_i and ν_i, we can compute $\varphi(p,q)$ and its derivatives in $O(n)$ operations.

An efficient way to minimize φ is Newton's method with Nesterov and Nemirovsky's step length. Although $\varphi(p,q)$ is not convex (it is the sum of a concave and a convex term), it can be replaced by a convex upper bound at every iteration of the plane search by linearizing the concave term. Newton's method can then be safely applied to this convex approximation (see next section for more details).

There is no need to calculate the minimum of $\varphi(p,q)$ very accurately. One can take a fixed number of steps, for example, as Ben Tal and Nemirovsky in [10].

The main cost of this scheme is in the initial computation of the eigenvalues μ_i and ν_i. Once these are known, each step in the plane search can be carried out at a cost of $O(n)$ operations.

4.3 Potential Reduction Method 1

The first, and most obvious, method for computing search directions δx and δZ, is by Newton's method. We need a slight modification to Newton's method, however, because the potential function $\varphi(x, Z)$ is not convex itself, but the sum of a concave and a convex function. The modification can be interpreted in several ways.

- We use exactly Newton's method, but apply it to a different potential function $\tilde{\varphi}$. The modified function $\tilde{\varphi}$ is obtained from φ by linearizing the concave term $q\, log(c^T x + TrF_0 Z)$ around the current iterate $x^{(k)}, Z^{(k)}$. Therefore, $\tilde{\varphi}$ is a convex function. It has the property that

  $$\tilde{\varphi}(x, Z) \geq \varphi(x, Z),$$

  for all x and Z, and $\tilde{\varphi}(x^{(k)}, Z^{(k)}) = \varphi(x^{(k)}, Z^{(k)})$. In other words, if a particular choice of $x^{(k+1)}$, $Z^{(k+1)}$ reduces $\tilde{\varphi}$ by a certain amount, then the reduction in φ itself will be even bigger.

- It is a *quasi-Newton method* applied to φ. A quasi-Newton method uses the exact gradient of φ, but replaces the Hessian by an approximation. In this case the approximation is to drop the second derivatives of the concave term, and to consider only the barrier term when forming the Hessian.

Following the first interpretation, let us linearize the concave term:

$$\begin{aligned} \tilde{\varphi}(x, Z) \;&\stackrel{\Delta}{=}\; \varphi(x^{(k)}, Z^{(k)}) + \rho\left(c^T(x - x^{(k)}) + TrF_0(Z - Z^{(k)}\right) - \\ &\quad log\,det\,F(x) - log\,det\,Z - n\,log\,n, \end{aligned}$$

where $\rho \stackrel{\Delta}{=} q \big/ \left(c^t x^{(k)} + TrF_0 Z^{(k)}\right)$. Note that the function $\tilde{\varphi}$ is separable in x and Z:

$$\tilde{\varphi}(x, Z) = \tilde{\varphi}^p(x) + \tilde{\varphi}^d(Z) + \text{constant},$$

where

$$\tilde{\varphi}^p(x) \;\stackrel{\Delta}{=}\; \rho c^T x - log\,det\,F(x), \qquad (41)$$

$$\tilde{\varphi}^d(Z) \;\stackrel{\Delta}{=}\; \rho TrF_0 Z - log\,det\,Z. \qquad (42)$$

In the derivation of the Newton directions δx^p and δZ^d, we can therefore consider the primal and dual parts separately.

4.3.1 Primal Least-Squares Problem

The Newton direction for $\tilde{\varphi}^p$ is the direction that minimizes the best quadratic approximation to the function. Recall from Section 3 that the best quadratic approximation to the barrier function $-\log\det V$ in a given matrix X, is given by

$$-\log\det(X+V) = -TrX^{-1}V + \tfrac{1}{2}TrVX^{-1}VX^{-1}.$$

We therefore have (with $F \triangleq F(x)$):

$$\begin{aligned} \delta x^p &= \operatorname*{argmin}_{v \in \mathbf{R}^m} \ \rho c^T v - TrF^{-1}\left(\sum_{i=1}^m F_i v_i\right) + \\ &\quad \tfrac{1}{2}Tr\left(\sum_{i=1}^m F_i v_i\right)F^{-1}\left(\sum_{j=1}^m F_j v_j\right)F^{-1} \qquad (43) \\ &= \operatorname*{argmin}_{v \in \mathbf{R}^m} \ \rho c^T v - \sum_{i=1}^m v_i\left(TrF^{-1}F_i\right) + \\ &\quad \tfrac{1}{2}\sum_{i=1}^m\sum_{j=1}^m v_i v_j\left(TrF_iF^{-1}F_jF^{-1}\right). \qquad (44) \end{aligned}$$

Let $\tilde{F}_i = F^{-1/2}F_iF^{-1/2}$, for $i = 1,\ldots,m$, and C be an $n \times n$ matrix satisfying $TrF_iC = c_i$, $i = 1,\ldots,m$ (take $C = Z$, for instance). Then the expression for δx^p can be further simplified to

$$\begin{aligned} \delta x^p &= \operatorname*{argmin}_{v \in \mathbf{R}^m} \ \sum_{i=1}^m v_i TrF_i\left(\rho C - F^{-1}\right) + \qquad (45) \\ &\quad \tfrac{1}{2}\sum_{i=1}^m\sum_{j=1}^m v_i v_j Tr\left(F^{-1/2}F_iF^{-1/2}F^{-1/2}F_jF^{-1/2}\right) \\ &= \operatorname*{argmin}_{v \in \mathbf{R}^m} \ \sum_{i=1}^m v_i Tr\tilde{F}_i\left(\rho F^{1/2}CF^{1/2} - I\right) + \tfrac{1}{2}\sum_{i=1}^m\sum_{j=1}^m v_i v_j Tr\tilde{F}_i\tilde{F}_j \\ &= \operatorname*{argmin}_{v \in \mathbf{R}^m} \ \left\|\rho F^{1/2}CF^{1/2} - I + \sum_{i=1}^m \tilde{F}_i v_i\right\|_F. \qquad (46) \end{aligned}$$

This is a least squares problem with m variables and $n(n+1)/2$ equations.

Let us consider the LP (3) as an illustration. In this case, all matrices in (46) are diagonal, and $F = diag(Ax-b)$. The least-squares problem then reduces to minimizing the norm of the diagonal elements

$$\text{minimize} \ \left\|\rho X\tilde{c} - e + X^{-1}Av\right\|,$$

where $X = diag(Ax-b)$, e is a vector with all components one, and $\tilde{c}$ is any vector that solves $A^T\tilde{c} = c$.

4.3.2 Dual Least-Squares Problem

In a similar way, among all feasible directions, δZ^d will be the direction that minimizes the second order approximation of $\tilde{\varphi}^d(Z)$:

$$\delta Z^d = \operatorname*{argmin}_{\substack{V:\,TrF_iV=0\\ i=1,\ldots,m}} \ \rho TrF_0V - TrZ^{-1}V + \tfrac{1}{2}TrVZ^{-1}VZ^{-1}. \qquad (47)$$

This involves a least-squares problem with exactly the same dimension as the primal problem. The easiest way to see this, is to introduce Lagrange multipliers v_i, $i = 1,\ldots,m$ for the constraints on V, and to write the optimality conditions for (47):

$$\rho F_0 - Z^{-1} + Z^{-1}\delta Z^dZ^{-1} + \sum_{i=1}^m v_iF_i = 0,$$

$$TrF_i\delta Z^d = 0, \ \ i = 1,\ldots,m,$$

or, equivalently, if we write $\tilde{F}_i = Z^{1/2}F_iZ^{1/2}$, $i = 1,\ldots,m$:

$$\rho Z^{1/2}F_0Z^{1/2} - I + Z^{-1/2}\delta Z^dZ^{-1/2} + \sum_{i=1}^m v_i\tilde{F}_i = 0, \qquad (48)$$

$$Tr\tilde{F}_iZ^{-1/2}\delta Z^dZ^{-1/2} = 0, \ \ i = 1,\ldots,m.$$

Hence the multipliers v_i, $i = 1,\ldots,m$, are the solution of the least squares problem

$$\operatorname*{minimize}_{v \in R^m} \ \left\|\rho Z^{1/2}F_0Z^{1/2} - I + \sum_{i=1}^m \tilde{F}_iv_i\right\|_F, \qquad (49)$$

which, again, has $n(n+1)/2$ equations and m variables. From the least-squares solution v, one then obtains δZ^d by substitution in equation (48).

In the case of an LP where $F(x) = diag(Ax-b)$ and $Z = diag(z)$, problem (49) becomes

$$\text{minimize} \ \|-\rho Zb - e + ZAv\|.$$

4.3.3 Summary

The following outline summarizes the two preceding sections. The method is due to Vandenberghe and Boyd [60], and is a generalization of Gonzaga and Todd's method for linear programming [25].

Potential Reduction Algorithm 1

given strictly feasible x and Z.

repeat

1. Solve the least-squares problem (46) for δx^p.

2. Solve the least-squares problem (49), and compute δZ^d from (48).
3. Find $p, q \in \mathbf{R}$ that minimize $\varphi(x + p\delta x^p, Z + q\delta Z^d)$.
4. Update: $x := x + p\delta x^p$ and $Z := Z + q\delta Z^d$.

until duality gap $\leq \epsilon$.

We refer to [60] for a complexity analysis. The basic ideas are summarized in the following two theorems. Define

$$\lambda^p \triangleq \left\| \sum_{i=1}^m F^{-1/2} F_i F^{-1/2} \delta x_i^p \right\|_F, \quad \lambda^d \triangleq \left\| Z^{-1/2} \delta Z^d Z^{-1/2} \right\|_F.$$

The first theorem states that it is always possible to reduce the potential as long as λ^p and λ^d are not both small.

Theorem 4 *Assume that x and Z are strictly feasible, and that δx^p and δZ^d are the corresponding Newton directions. Let $p = 1/(1+\lambda^p)$ and $q = 1/(1+\lambda^q)$. Then*

$$\begin{aligned} \varphi(x + p\delta x^p, Z) &\leq \varphi(x, Z) - \lambda^p + log(1+\lambda^p), \\ \varphi(x, Z + q\delta Z^d) &\leq \varphi(x, Z) - \lambda^d + log(1+\lambda^d). \end{aligned}$$

The second theorem states that λ^p and λ^d never become small at the same time.

Theorem 5 $\max\{\lambda^p, \lambda^d\} \geq 0.35$.

As a consequence of these two theorems, we see that it is always possible to reduce the potential function by at least $log\, 0.35 - log(1 - 0.35) = 0.78$ per iteration. We have seen in Section 4.1 that this implies convergence of the algorithm in $O(\sqrt{n})$ iterations.

4.4 Potential Reduction Method 2

The above algorithm has the disadvantage of requiring the solution of two least-squares problems per iteration. In this section we will show that a complete primal-dual algorithm can be based on the primal least-squares problem only. In linear programming this primal-dual method is called Ye's method [62], but the extension to PDPs is due to Nesterov and Nemirovsky [43] and Alizadeh [2].

We first write down the optimality conditions that characterize δx^p as the minimizer of (44):

$$\rho c_i - Tr F^{-1} F_i + \sum_{j=1}^m \delta x_j^p Tr\left(F_i F^{-1} F_j F^{-1}\right) = 0,$$

for $i = 1, \ldots, m$, or

$$(1/\rho) Tr F_i \left(F^{-1} - \sum_{j=1}^m \delta x_j^p F^{-1} F_j F^{-1} \right) = c_i.$$

In other words, the matrix $(1/\rho)\left(F^{-1} - \sum_{j=1}^m F^{-1} F_j F^{-1} \delta x_j^p\right)$ satisfies the equality constraints needed for dual feasibility, and the matrix

$$\delta Z^p \triangleq (1/\rho) \left(F^{-1} - \sum_{j=1}^m \delta x_j^p F^{-1} F_j F^{-1} \right) - Z \tag{50}$$

is a dual feasible direction.

The second potential reduction algorithm takes this δZ^p as dual search direction, and performs a plane search in the plane defined by δx^p and δZ^p:

Potential Reduction Algorithm 2

given strictly feasible x and Z.

repeat

1. Solve the least-squares problem (46) for δx^p.
2. Find δZ^p from (50).
3. Find $p, q \in \mathbf{R}$ that minimize $\varphi(x + p\delta x^p, Z + q\delta Z^d)$.
4. Update: $x := x + p\delta x^p$ and $Z := Z + q\delta Z^d$.

until duality gap $\leq \epsilon$.

The effectiveness of this algorithm is a consequence of Theorem 4 and Theorem 6 below. Theorem 4 implies that

$$\varphi(x + \frac{1}{1+\lambda^p}\delta x^p, Z) \leq \varphi(x, Z) - \lambda^p + log(1 + \lambda_p).$$

Theorem 6 *If $\lambda^p < 1$, then*

$$\varphi(x, Z + \delta Z^p) \leq \varphi(x, Z) + \nu(\lambda^p - 1/2) - \lambda^p - log(1 - \lambda^p).$$

Hence, again, there is always at least one way to reduce the potential function substantially. If λ^p is large, then a primal update only would be enough; when λ^p becomes too small, the dual update takes over. The plane search guarantees a smooth and optimal transition between both extreme cases.

4.5 Combined Algorithms

By duality, the considerations of Section 4.4 can be repeated for the dual least-squares problem (49). Solving the dual least-squares problem, we not only find a dual Newton direction δZ^d but also a primal feasible direction δx^d, and a complete primal-dual algorithm can be based on the dual least-squares problem alone.

This means that four search directions are available at each strictly feasible pair x, Z:

primal least squares	δx^p	δZ^p
dual least squares	δx^d	δZ^d

In this table, three pairs of primal and dual directions are known to reduce the potential function:

- Method 1 uses δx^p, δZ^d.
- Method 2 uses δx^p, δZ^p.
- The dual version of method 2 method uses δx^d, δZ^d.

These three methods all have a worst-case complexity of $O(\sqrt{n}L\,log(1/\epsilon))$ iterations, and also in practice, it is not clear which one is faster (in terms of iterations). The first method has the disadvantage of requiring two least-squares problems per update, but one can think of combinations, where in each step either the primal or the dual least-squares problem (for instance, alternatingly) is solved.

4.6 Long-Step Path-Following Methods

A number of recent articles describe path-following linear programming methods with superlinear or quadratic asymptotic convergence (see Zhang and Tapia [66], Ye, Güler, Tapia and Zhang [63], Mizuno, Todd and Ye [35]). In this section we discuss a PDP variant of these methods. We use a conceptual version of an algorithm presented by Nesterov [37]. For a different approach, see Alizadeh, Haeberly and Overton [1].

The algorithm traces the central path using a predictor-corrector approach. The predictor phase makes the largest possible step without leaving a pre-determined neighborhood of the central path. Our discussion in Section 3.3 suggests using $\psi(x,Z)$ as a measure for deviation from the central path. The parameter ψ_{max} in the following algorithm is an upper bound on ψ and determines how close to the central path the iterates stay. We assume that the initial points are strictly feasible and sufficiently centered.

Long-Step Path-Following Algorithm

given strictly feasible x and Z with $\psi(x,Z) \le \psi_{max}$.
repeat
1. *Corrector step.*
   (Approximately) center x, Z, keeping the duality gap constant.
2. *Predictor step.*
   Compute directions δx^{prd}, δZ^{prd} (approximately) tangent to central path.
   Compute $p > 0$ such that $\psi(x + p\delta x^{prd}, Z + p\delta Z^{prd}) = \psi_{max}$.
   Take $x = x + p\delta x^{prd}$, $Z = Z + p\delta Z^{prd}$.

until duality gap $\le \epsilon$.

4.6.1 Predictor Step

Assume x and Z are central with duality gap α: $F(x)Z = (\alpha/n)I$. We first derive expressions for the tangent directions to the central path at x, Z. From the first order

expansion of the equality $F(x+\delta x)(Z+\delta Z) = ((\alpha - \delta\alpha)/n)I$, we find that the tangent directions δx^{prd} and δZ^{prd} satisfy

$$-I = F(x)\delta Z^{prd} + \left(\sum_{i=1}^{m} \delta x_i^{prd} F_i\right) Z, \tag{51}$$

or, using $F(x)Z = (\alpha/n)I$,

$$-I = F^{1/2}\delta Z^{prd}F^{1/2} + (\alpha/n)F^{-1/2}\left(\sum_{i=1}^{m} \delta x_i^{prd} F_i\right) F^{-1/2}, \tag{52}$$

where, as before, $F \triangleq F(x)$.

This equation, together with the dual feasibility conditions $TrF_i\delta Z^{prd} = 0$, is equivalent to the optimality conditions of the least-squares problem,

$$\delta x^{prd} = \underset{v \in \mathbf{R}^m}{\text{argmin}} \left\| (n/\alpha)I + \sum_{i=1}^{m} v_i F^{-1/2}F_iF^{-1/2} \right\|, \tag{53}$$

and the dual direction follows from $\delta Z^{prd} = -F^{-1} - (\alpha/n)F^{-1}\left(\sum_{i=1}^m \delta x_i^{prd} F_i\right)F^{-1}$.

Theorem 7 *If ψ_{max} is a positive constant, and $p > 0$ is computed from*

$$\psi(x + p\delta x^{prd}, Z + p\delta Z^{prd}) = \psi_{max},$$

then

$$TrF(x + p\delta x^{prd})(Z + p\delta Z^{prd}) \le (1 - \tau/\sqrt{n})TrF(x)Z,$$

where τ is a positive constant less than one.

In other words, every predictor step reduces the duality gap by $1 - \tau/\sqrt{n}$. The number of predictor steps needed to reduce the duality gap by a factor ϵ is at most $\frac{\sqrt{n}}{\tau}\,log(1/\epsilon)$. This establishes polynomial convergence. In practice, the steps typically become larger as the algorithm approaches the solution. This results in fast (superlinear) asymptotic convergence.

4.6.2 Corrector Step

In the corrector phase one can use Newton's method, possibly combined with a plane search, to find a primal-dual central pair with a given duality gap α. As explained in Section 3.3, the number of corrector steps in every outer iteration will be bounded by an absolute constant. Therefore, the corrector steps do not contribute to the order of the worst-case complexity.

5 Conclusions and Extensions

5.1 Solving Structured Problems

A natural question is: What is the computational effort required to solve a PDP using the methods described above? The answer depends on the amount of structure in the matrix inequality $F(x)$. The three methods have the same worst-case complexity: The number of iterations to solve a PDP to a given accuracy grows as $O(\sqrt{n})$. In practice, it has been observed by many researchers that the number grows more slowly, as $\log n$ or as $n^{1/4}$, and can often be assumed to be almost constant (see Nesterov and Nemirovsky [43], or Gonzaga and Todd [25] for comments on the average behavior). Typical numbers range from 10 to 50.

The overall cost is therefore determined by the amount of work per iteration. As we have seen, the dominant part here is the solution of a least-squares problem of the form

$$\text{minimize} \left\| D - \sum_{i=1}^{m} v_i \tilde{F}_i \right\|_F , \tag{54}$$

where $\tilde{F}_i = V^T F_i V$. The matrices V and D change from iteration to iteration. Problem (54) is a least-squares problem with m variables and $n(n+1)/2$ equations. Using direct methods it can be solved in $O(m^2n^2)$ operations.

Important savings are possible when the matrices F_i are structured. The easiest type is block-diagonal structure. Assume $F(x)$ consists of L diagonal blocks of size n_i, $i = 1, \dots, L$. Then the number of equations in (54) is $\sum_{i=1}^{L} n_i(n_i+1)/2$, which is often an order less than $n(n+1)/2$. For instance, in the LP case (diagonal matrix $F(x)$), the number of variables is n, and solving the least-squares problem requires only $O(m^2n)$ operations.

In more complicated situations, one can use iterative methods to solve (54). The LSQR algorithm of Paige and Saunders [49] appears to be very well suited. At a high level, and in exact arithmetic, it has the following properties. It solves (54) in $m+1$ iterations. Every iteration evaluates two linear mappings

$$(v_1, \dots, v_m) \mapsto \sum_{i=1}^{m} v_i \tilde{F}_i, \quad \text{and} \quad W \mapsto (Tr\tilde{F}_1 W, \dots, Tr\tilde{F}_m W) \tag{55}$$

on a vector v and a symmetric matrix $W = W^T$. When the matrices F_i are unstructured, these two operations take mn^2 operations. Hence, the cost of solving (54) using LSQR is $O(n^2m^2)$, and nothing is gained over direct methods.

In most cases, however, the two operations (55) are much cheaper than mn^2 because of the special structure of the matrices F_i. A well-known example is a sparse LP. Iterative methods in interior-point algorithms for sparse LPs are addressed by Mehrotra [34], Kim and Nazareth [30], and Gill, Murray, Ponceleon and Saunders [22].

Sparsity is not the only example, however. The equations are often dense, but still highly structured in the sense that the (55) are easy to evaluate. Reference [60] discusses iterative methods for exploiting structure in PDPs arising in control theory.

Iterative methods have one important drawback. Due to round-off errors, convergence can be slow unless preconditioning or some type of reorthogonalization is used. The implementation of iterative methods is therefore very problem-dependent.

5.2 Generalized Eigenvalue Problems

As a final note, we would like to mention that great progress has recently been made in interior-point methods for generalized linear-fractional problems

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & tB(x) - A(x) \geq 0, \\ & B(x) \geq 0, \end{array}$$

where $B(x)$ and $A(x)$ are affine in x. This problem is quasiconvex, and generalizes the PDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & tI - A(x) \geq 0. \end{array}$$

See Boyd and El Ghaoui [7], and Nesterov and Nemirovsky [42, 36] for details.

Acknowledgment

We want to thank Hervé Lebret who helped us with an early draft of this paper, and also coded and tested most of the algorithms described here. We also thank Michael Grant and Benoît Lecinq who have been involved in several software projects related to this work.

L. Vandenberghe is Postdoctoral Researcher of the Belgian National Fund for Scientific Research (NFWO). His research was supported in part by the Belgian program on interuniversity attraction poles (IUAP 17 and 50) initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The research of S. Boyd was supported in part by AFOSR (under F49620-92-J-0013), NSF (under ECS-9222391), and ARPA (under F49620-93-1-0085).

References

[1] F. Alizadeh, J. P. A. Haeberly, and M. L. Overton. "A new primal-dual interior-point method for semidefinite programming," *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, Snowbird, Utah*, June 1994.

[2] F. Alizadeh, *Combinatorial Optimization with Interior Point Methods and Semi-Definite Matrices*, Ph.D. thesis, Univ. of Minnesota, October 1991.

[3] F. Alizadeh, "Combinatorial optimization with semidefinite matrices," *Proceedings of Second Annual Integer Programming and Combinatorial Optimization conference*, Carnegie-Mellon University, 1992.

[4] F. Alizadeh, "Optimization over the positive-definite cone: interior point methods and combinatorial applications," *Advances in Optimization and Parallel Computing*, Panos Pardalos (editor), North-Holland, 1992.

[5] J. C. Allwright, "Positive semidefinite matrices: characterization via conical hulls and least-squares solution of a matrix equation," *SIAM Journal on Control and Optimization*, 26(3), 1988, 537-556.

[6] J. Allwright, "On maximizing the minimum eigenvalue of a linear combination of symmetric matrices," *SIAM J. on Matrix Analysis and Applications*, 10 (1989), 347-382.

[7] S. Boyd and L. El Ghaoui, "Method of centers for minimizing generalized eigenvalues," *Linear Algebra and Applications, special issue on Numerical Linear Algebra Methods in Control, Signals and Systems*, 188 July (1993), 63-111.

[8] 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.

[9] R. Bellman and K. Fan, "On systems of linear inequalities in Hermitian matrix variables," V. L. Klee (editor), *Convexity*, 7 of *Proceedings of Symposia in Pure Mathematics*, American Mathematical Society, 1963, 1-11.

[10] 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.

[11] A. Ben-Tal and M. P. Bendsøe, "A new method for optimal truss topology design," *SIAM Journal on Optimization*, 3 (1993), 322-358.

[12] J. Cullum, W. Donath, and P. Wolfe, "The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices," *Math. Programming Study*, 3 (1975), 35-55.

[13] B. Craven and B. Mond, "Linear programming with matrix variables," *Linear Algebra and Appl.*, **38** (1981), 73-80.

[14] D. den Hertog, *Interior Point Approach to Linear, Quadratic and Convex Programming*, Kluwer, 1993.

[15] I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Soviet Math. Dokl.*, 8(3) 1967, 674-675.

[16] M. K. H. Fan, "A quadratically convergent algorithm on minimizing the largest eigenvalue of a symmetric matrix," *Linear Algebra and Appl.*, (1993) 188-189:231-253.

[17] R. Fletcher, "A nonlinear programming problem in statistics (educational testing)," *SIAM Journal on Scientific and Statistical Computing*, 2(3) 1981, 257-267.

[18] R. Fletcher, "Semidefinite matrix constraints in optimization," *SIAM J. Control and Opt.*, 23 (1985), 493-513.

[19] A. Fiacco and G. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques*, Wiley, 1968. Reprinted in the *SIAM Classics in Applied Mathematics series* 1990.

[20] M. K. H. Fan and B. Nekooie, "On minimizing the largest eigenvalue of a symmetric matrix," *Proc. IEEE Conf. on Decision and Control*, December 1992, 134-139.

[21] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, Baltimore, second edition, 1989.

[22] P. E. Gill, W. Murray, D. B. Ponceleon, and M. A. Saunders, "Preconditioners for indefinite systems arising in optimization," *SIAM J. on Matrix Analysis and Applications*, 13 (1992), 292-311.

[23] C. C. Gonzaga, "Path-following methods for linear programming," *SIAM Review*, 34(2) June 1992, 167-224.

[24] C. Goh and D. Teo, "On minimax eigenvalue problems via constrained optimization," *Journal of Optimization Theory and Applications*, 57(1) 1988, 59-68.

[25] C. C. Gonzaga and M. J. Todd, "An $O(\sqrt{n}L)$-iteration large-step primal-dual affine algorithm for linear programming," *SIAM Journal on Optimization*, 2(3) August 1992, 349-359.

[26] J. B. Hiriart-Urruty and D. Ye, "Sensitivity analysis of all eigenvalues of a symmetric matrix," Technical report, Univ. Paul Sabatier, Toulouse, 1992.

[27] F. Jarre, "An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices," *SIAM J. Control and Opt.*, 31(5) September 1993, 1360-1377.

[28] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, 4(4) 1984, 373-395.

[29] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise, *A Univied Approach to Interior Point Algorithms for Linear Complementarity Problems*, Lecture Notes in Computer Science, Springer-Verlag, 1991.

[30] K. Kim and J. L. Nazareth, "Implementation of a primal null-space affine scaling method and its extensions," Technical report, Department of Pure and Applied Mathematics, Washington State University, January 1992.

[31] B. Lieu and P. Huard, "La méthode des centres dans un espace topologique," *Numerische Mathematik*, 8 (1966), 56-67.

[32] I. J. Lustig, R. E. Marsten, and D. F. Shanno, "Interior point methods for linear programming: Computational state of the art," *ORSA Journal on Computing*, 6(1) 1994.

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

[34] S. Mehrotra, "Implementations of affine scaling methods: approximate solutions of systems of linear equations using preconditioned conjugate gradient methods," *ORSA Journal on Computing*, 4(2) Spring 1992, 103-118.

[35] S. Mizuno, M. J. Todd, and Y. Ye, "On adaptive-step primal-dual interior-point algorithms for linear programming," *Mathematics of Operations Research*, 18 (1993), 964-981.

[36] A. Nemirovski, "Long-step method of analytic centers for fractional problems," Technical Report 3/94, Technion, Haifa, Israel, 1994.

[37] Y. Nesterov, "Primal-dual methods," Seminar at CORE, Université Catholique de Louvain, February 1994.

[38] Yu. Nesterov and A. Nemirovsky, "A general approach to polynomial-time algorithms design for convex programming," Technical report, Centr. Econ. & Math. Inst., USSR Acad. Sci., Moscow, USSR, 1988.

[39] Yu. Nesterov and A. Nemirovsky, *Optimization over positive semidefinite matrices: Mathematical background and user's manual.* USSR Acad. Sci. Centr. Econ. & Math. Inst., 32 Krasikova St., Moscow 117418 USSR, 1990.

[40] Yu. Nesterov and A. Nemirovsky, "Self-concordant functions and polynomial time methods in convex programming," Technical report, Centr. Econ. & Math. Inst., USSR Acad. Sci., Moscow, USSR, April 1990.

[41] Yu. Nesterov and A. Nemirovsky, "Conic formulation of a convex programming problem and duality," Technical report, Centr. Econ. & Math. Inst., USSR Academy of Sciences, Moscow USSR, 1991.

[42] Yu. Nesterov and A. Nemirovsky, "An interior point method for generalized linear-fractional programming," Submitted to *Math. Programming, Series B.*, 1991.

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

[44] M. Overton, "On minimizing the maximum eigenvalue of a symmetric matrix," *SIAM J. on Matrix Analysis and Applications*, 9(2) 1988, 256-268.

[45] M. Overton, "Large-scale optimization of eigenvalues," *SIAM J. Optimization*, (1992), 88-120.

[46] M. Overton and R. Womersley, "On the sum of the largest eigenvalues of a symmetric matrix," *SIAM J. on Matrix Analysis and Applications*, (1992), 41-45.

[47] M. Overton and R. Womersley, "Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices," *Mathematical Programming*, **62** (1993), 321-357.

[48] E. Panier, "On the need for special purpose algorithms for minimax eigenvalue problems," *Journal Opt. Theory Appl.*, 62(2) August 1989, 279-287.

[49] C. C. Paige and M. S. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software*, 8(1) March 1982, 43-71.

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

[51] U. T. Ringertz, "Optimal design of nonlinear shell structures," Technical Report FFA TN 1991-18, The Aeronautical Research Institute of Sweden, 1991.

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

[53] J. B. Rosen, "Pattern separation by convex programming," *Journal of Mathematical Analysis and Applications*, **10** (1965), 123-134.

[54] F. Rendl, R. Vanderbei, and H. Wolkowicz, "A primal-dual interior-point method for the max-min eigenvalue problem," Technical report, University of Waterloo, Dept. of Combinatorics and Optimization, Waterloo, Ontario, Canada, 1993.

[55] A. L. Shapiro, "Optimal block diagonal ℓ_2 scaling of matrices," *SIAM J. on Numerical Analysis*, 22(1) February 1985, 81-94.

[56] N. Z. Shor, "Cut-off method with space extension in convex programming problems," *Cybernetics*, 13(1) 1977, 94-96.

[57] G. Sonnevend, "An analytical center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," *Lecture Notes in Control and Information Sciences*, **84**, Springer-Verlag, 1986, 866-878.

[58] G. Sonnevend, "New algorithms in convex programming based on a notion of 'centre' (for systems of analytic inequalities) and on rational extrapolation," *International Series of Numerical Mathematics*, **84** (1988), 311-326.

[59] F. Uhlig, "A recurring theorem about pairs of quadratic forms and extensions: A survey," *Linear Algebra and Appl.*, **25** (1979), 219-237.

[60] L. Vandenberghe and S. Boyd, "Primal-dual potential reduction method for problems involving matrix inequalities," To be published in *Math. Programming*, 1993.

[61] G. A. Watson, "Algorithms for minimum trace factor analysis," *SIAM J. on Matrix Analysis and Applications*, **13**(4) 1992, 1039-1053.

[62] Y. Ye, "An $O(n^3 L)$ potential reduction algorithm for linear programming," *Mathematical Programming*, **50** (1991), 239-258.

[63] Y. Ye, O. Güler, R. A. Tapia, and Y. Zhang, "A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming," *Mathematical Programming*, **59** (1993), 151-162.

[64] D. B. Yudin and A. S. Nemirovski, "Informational complexity and efficient methods for solving complex extremal problems," *Matekon*, **13** (1977), 25-45.

[65] A. Yoshise, "An optimization method for convex programs–interior-point method and analytical center," *Systems, Control and Information*, Special issue on Numerical Approaches in Control Theory. In Japanese, **38**(3) March 1994, 155-160.

[66] Y. Zhang and R. A. Tapia, "A superlinearly convergent polynomial primal-dual interior-point algorithm for linear programming," *SIAM J. on Optimization*, **3** (1993), 118-133.