DETERMINANT MAXIMIZATION WITH LINEAR MATRIX INEQUALITY CONSTRAINTS[†]

LIEVEN VANDENBERGHE[‡], STEPHEN BOYD[§], AND SHAO-PO WU[§]

Abstract. The problem of maximizing the determinant of a matrix subject to linear matrix inequalities arises in many fields, including computational geometry, statistics, system identification, experiment design, and information and communication theory. It can also be considered as a generalization of the semidefinite programming problem.

We give an overview of the applications of the determinant maximization problem, pointing out simple cases where specialized algorithms or analytical solutions are known. We then describe an interior-point method, with a simplified analysis of the worst-case complexity and numerical results that indicate that the method is very efficient, both in theory and in practice. Compared to existing specialized algorithms (where they are available), the interior-point method will generally be slower; the advantage is that it handles a much wider variety of problems.

1. Introduction.

We consider the optimization problem

$$\begin{array}{ll} \text{minimize} & c^T x + \log\det G(x)^{-1} \\ \text{subject to} & G(x) \succ 0 \\ & F(x) \succeq 0, \end{array} \tag{1.1}$$

where the optimization variable is the vector $x \in \mathbf{R}^m$. The functions $G : \mathbf{R}^m \to \mathbf{R}^{l\times l}$ and $F : \mathbf{R}^m \to \mathbf{R}^{n\times n}$ are affine:

$$\begin{aligned} G(x) &= G_0 + x_1 G_1 + \cdots + x_m G_m, \\ F(x) &= F_0 + x_1 F_1 + \cdots + x_m F_m, \end{aligned}$$

where $G_i = G_i^T$ and $F_i = F_i^T$. The inequality signs in (1.1) denote matrix inequalities, *i.e.*, $G(x) \succ 0$ means $z^T G(x) z > 0$ for all nonzero z and $F(x) \succeq 0$ means $z^T F(x) z \geq 0$ for all z. We call $G(x) \succ 0$ and $F(x) \succeq 0$ (strict and nonstrict, respectively) *linear matrix inequalities* (LMIs) in the variable x. We will refer to problem (1.1) as a max-det problem, since in many cases the term $c^T x$ is absent, so the problem reduces to maximizing the determinant of $G(x)$ subject to LMI constraints.

The max-det problem is a convex optimization problem, *i.e.*, the objective function $c^T x + \log\det G(x)^{-1}$ is convex (on $\{x \mid G(x) \succ 0\}$), and the constraint set is convex. Indeed, LMI constraints can represent many common convex constraints, including linear inequalities, convex quadratic inequalities, and matrix norm and eigenvalue constraints (see Alizadeh[1], Boyd, El Ghaoui, Feron and Balakrishnan[13], Lewis and Overton[47], Nesterov and Nemirovsky[51, §6.4], and Vandenberghe and Boyd[69]).

In this paper we describe an interior-point method that solves the max-det problem very efficiently, both in worst-case complexity theory and in practice. The method we describe shares many features of interior-point methods for linear and semidefinite programming. In particular, our computational experience (which is limited to problems of moderate size — several hundred variables, with matrices up to 100×100)

[†] Research supported in part by AFOSR (under F49620-95-1-0318), NSF (under ECS-9222391 and EEC-9420565), MURI (under F49620-95-1-0525), and a NATO Collaborative Research Grant (CRG-941269). Associated software is available at URL `http://www-isl.stanford.edu/people/boyd` and from anonymous FTP to `isl.stanford.edu` in `pub/boyd/maxdet`.

[‡] Electrical Engineering Department, University of California, Los Angeles CA 90095-1594. E-mail: `vandenbe@ee.ucla.edu`.

[§] Information Systems Laboratory, Electrical Engineering Department, Stanford University, Stanford CA 94305. E-mail: `boyd@isl.stanford.edu`, `clive@isl.stanford.edu`.

indicates that the method we describe solves the max-det problem (1.1) in a number of iterations that hardly varies with problem size, and typically ranges between 5 and 50; each iteration involves solving a system of linear equations.

Max-det problems arise in many fields, including computational geometry, statistics, and information and communication theory, so the duality theory and algorithms we develop have wide application. In some of these applications, and for very simple forms of the problem, the max-det problems can be solved by specialized algorithms or, in some cases, analytically. Our interior-point algorithm will generally be *slower* than the specialized algorithms (when the specialized algorithms can be used). The *advantage* of our approach is that it is much more general; it handles a much wider variety of problems. The analytical solutions or specialized algorithms, for example, cannot handle the addition of (convex) constraints; our algorithm for general max-det problems does.

In the remainder of §1, we describe some interesting special cases of the max-det problem, such as semidefinite programming and analytic centering. In §2 we describe examples and applications of max-det problems, pointing out analytical solutions where they are known, and interesting extensions that can be handled as general max-det problems. In §3 we describe a duality theory for max-det problems, pointing out connections to semidefinite programming duality. Our interior-point method for solving the max-det problem (1.1) is developed in §4–§9. We describe two variations: a simple 'short-step' method, for which we can prove polynomial worst-case complexity, and a 'long-step' or adaptive step predictor-corrector method which has the same worst-case complexity, but is much more efficient in practice. We finish with some numerical experiments. For the sake of brevity, we omit most proofs and some important numerical details, and refer the interested reader to the technical report [70]. A C implementation of the method described in this paper is also available [76].

Let us now describe some special cases of the max-det problem.

Semidefinite programming. When $G(x) = 1$, the max-det problem reduces to

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) \succeq 0, \end{array} \tag{1.2}$$

which is known as a *semidefinite program* (SDP). Semidefinite programming unifies a wide variety of convex optimization problems, *e.g.*, linear programming,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

which can be expressed as an SDP with $F(x) = \mathbf{diag}(b - Ax)$. For surveys of the theory and applications of semidefinite programming, see [1], [13], [46], [47], [51, §6.4], and [69].

Analytic centering. When $c = 0$ and $F(x) = 1$, the max-det problem (1.1) reduces to

$$\begin{array}{ll} \text{minimize} & \log\det G(x)^{-1} \\ \text{subject to} & G(x) \succ 0, \end{array} \tag{1.3}$$

which we call the *analytic centering* problem. We will assume that the feasible set $\{x \mid G(x) \succ 0\}$ is nonempty and bounded, which implies that the matrices G_i, $i = 1, \ldots, m$, are linearly independent, and that the objective $\phi(x) = \log\det G(x)^{-1}$

is strictly convex (see, *e.g.*, [69] or [12]). Since the objective function grows without bound as x approaches the boundary of the feasible set, there is a unique solution $x^\star$ of (1.3). We call $x^\star$ the *analytic center* of the LMI $G(x) \succ 0$. The analytic center of an LMI generalizes the analytic center of a set of linear inequalities, introduced by Sonnevend[64, 65].

Since the constraint cannot be active at the analytic center, $x^\star$ is characterized by the optimality condition $\nabla\phi(x^\star) = 0$:

$$(\nabla\phi(x^\star))_i = -\mathbf{Tr}G_iG(x^\star)^{-1} = 0, \quad i = 1, \ldots, m \tag{1.4}$$

(see for example Boyd and El Ghaoui[12]).

The analytic center of an LMI is important for several reasons. We will see in §5 that the analytic center can be computed very efficiently, so it can be used as an easily computed robust solution of the LMI. Analytic centering also plays an important role in interior-point methods for solving the more general max-det problem (1.1). Roughly speaking, the interior-point methods solve the general problem by solving a sequence of analytic centering problems.

Parametrization of LMI feasible set. Let us restore the term c^Tx:

$$\begin{array}{ll} \text{minimize} & c^Tx + \log\det G(x)^{-1} \\ \text{subject to} & G(x) \succ 0, \end{array} \tag{1.5}$$

retaining our assumption that the feasible set $\mathbf{X} = \{x \mid G(x) \succ 0\}$ is nonempty and bounded, so the matrices G_i are linearly independent and the objective function is strictly convex. Thus, problem (1.5) has a unique solution $x^\star(c)$, which satisfies the optimality conditions $c + \nabla\phi(x^\star(c)) = 0$, *i.e.*,

$$\mathbf{Tr}G_iG(x^\star(c))^{-1} = c_i, \quad i = 1, \ldots, m.$$

Thus for each $c \in \mathbf{R}^m$, we have a (readily computed) point $x^\star(c)$ in the set $\mathbf{X}$.

Conversely, given a point $x \in \mathbf{X}$, define $c \in \mathbf{R}^m$ by $c_i = \mathbf{Tr}G(x)^{-1}G_i$, $i = 1, \ldots, m$. Evidently we have $x = x^\star(c)$. In other words, there is a one-to-one correspondence between vectors $c \in \mathbf{R}^m$ and feasible vectors $x \in \mathbf{X}$: the mapping $c \mapsto x^\star(c)$ is a parametrization of the feasible set $\mathbf{X}$ of the strict LMI $G(x) \succ 0$, with parameter $c \in \mathbf{R}^m$. This parametrization of the set $\mathbf{X}$ is related to the *Legendre transform* of the convex function $\log\det G(x)^{-1}$, defined by

$$\mathcal{L}(y) = -\inf\{-y^Tx + \log\det G(x)^{-1} \mid G(x) \succ 0\}.$$

Maximal lower bounds in the positive definite cone. Here we consider a simple example of the max-det problem. Let $A_i = A_i^T$, $i = 1, \ldots, L$, be positive definite matrices in $\mathbf{R}^{p\times p}$. A matrix X is a lower bound of the matrices A_i if $X \preceq A_i$, $i = 1, \ldots, L$; it is a maximal lower bound if there is no lower bound Y with $Y \neq X$, $Y \succeq X$.

Since the function $\log\det X^{-1}$ is monotone decreasing with respect to the positive semidefinite cone, *i.e.*,

$$0 \prec X \preceq Y \Longrightarrow \log\det Y^{-1} \leq \log\det X^{-1},$$

we can compute a maximal lower bound A_{mlb} by solving

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1} \\ \text{subject to} & X \succ 0 \\ & X \preceq A_i, \quad i = 1, \ldots, L. \end{array} \tag{1.6}$$

This is a max-det problem with $p(p+1)/2$ variables (the elements of the matrix X), and L LMI constraints $A_i - X \succeq 0$, which we can also consider as diagonal blocks of one single block diagonal LMI

$$\mathbf{diag}(A_1 - X, A_2 - X, \ldots, A_L - X) \succeq 0.$$

Of course there are other maximal lower bounds; replacing $\log\det X^{-1}$ by any other monotone decreasing matrix function, *e.g.*, $-\mathbf{Tr}X$ or $\mathbf{Tr}X^{-1}$, will also yield (other) maximal lower bounds. The maximal lower bound $A_{\rm mlb}$ obtained by solving (1.6), however, has the property that it is invariant under congruence transformations, *i.e.*, if the matrices A_i are transformed to TA_iT^T, where $T \in \mathbf{R}^{p\times p}$ is nonsingular, then the maximal lower bound obtained from (1.6) is $TA_{\rm mlb}T^T$.

2. Examples and applications. In this section we catalog examples and applications. The reader interested only in duality theory and solution methods for the max-det problem can skip directly to §3.

2.1. Minimum volume ellipsoid containing given points. Perhaps the earliest and best known application of the max-det problem arises in the problem of determining the minimum volume ellipsoid that contains given points $x^1, \ldots, x^K$ in $\mathbf{R}^n$ (or, equivalently, their convex hull $\mathbf{Co}\{x^1, \ldots, x^K\}$). This problem has applications in cluster analysis (Rosen[58], Barnes[9]), and robust statistics (in ellipsoidal peeling methods for outlier detection; see Rousseeuw and Leroy[59, §7]).

We describe the ellipsoid as $\mathcal{E} = \{x \mid \|Ax + b\| \leq 1\}$, where $A = A^T \succ 0$, so the volume of $\mathcal{E}$ is proportional to $\det A^{-1}$. Hence the minimum volume ellipsoid that contains the points x^i can be computed by solving the convex problem

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & \|Ax^i + b\| \leq 1,\ i = 1, \ldots, K \\ & A = A^T \succ 0, \end{array} \tag{2.1}$$

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 just convex quadratic inequalities in the variables A and b, can be expressed as LMIs

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

These LMIs can in turn be expressed as one large block diagonal LMI, so (2.1) is a max-det problem in the variables A and b.

Nesterov and Nemirovsky[51, §6.5], and Khachiyan and Todd[43] describe interior-point algorithms for computing the maximum-volume ellipsoid in a polyhedron described by linear inequalities (as well as the minimum-volume ellipsoid covering a polytope described by its vertices).

Many other geometrical problems involving ellipsoidal approximations can be formulated as max-det problems. References [13, §3.7], [16] and [68] give several examples, including the maximum volume ellipsoid contained in the intersection or in the sum of given ellipsoids, and the minimum volume ellipsoid containing the sum of given ellipsoids. For other ellipsoidal approximation problems, suboptimal solutions can be computed via max-det problems.

Ellipsoidal approximations of convex sets are used in control theory and signal processing in *bounded-noise* or *set-membership* techniques. These techniques were

first introduced for state estimation (see, *e.g.*, Schweppe[62, 63], Witsenhausen[75], Bertsekas and Rhodes[11], Chernousko[16, 17]), and later applied to system identification (Fogel and Huang[35, 36], Norton[52, 53, §8.6], Walter and Piet-Lahanier[71], Cheung, Yurkovich and Passino[18]), and signal processing Deller[23]. (For a survey emphasizing signal processing applications, see Deller *et al.* [24]).

Other applications include the *method of inscribed ellipsoids* developed by Tarasov, Khachiyan, and Erlikh[66], and design centering (Sapatnekar[60]).

2.2. Matrix completion problems.

Positive definite matrix completion. In a positive definite matrix completion problem we are given a symmetric matrix $A_f \in \mathbf{R}^{n\times n}$, some entries of which are fixed; the remaining entries are to be chosen so that the resulting matrix is positive definite.

Let the positions of the free (unspecified) entries be given by the index pairs (i_k, j_k), (j_k, i_k), $k = 1, \ldots, m$. We can assume that the diagonal elements are fixed, *i.e.*, $i_k \neq j_k$ for all k. (If a diagonal element, say the (l,l)th, is free, we take it to be very large, which makes the lth row and column of A_f irrelevant.) The positive definite completion problem can be cast as an SDP feasibility problem:

$$\begin{array}{ll} \text{find} & x \in \mathbf{R}^m \\ \text{such that} & A(x) \stackrel{\Delta}{=} A_f + \displaystyle\sum_{k=1}^{m} x_k \left(E_{i_k j_k} + E_{j_k i_k}\right) \succ 0, \end{array}$$

where E_{ij} denotes the matrix with all elements zero except the (i,j) element, which is equal to one. Note that the set $\{x \mid A(x) \succ 0\}$ is bounded since the diagonal elements of $A(x)$ are fixed.

Maximum entropy completion. The analytic center of the LMI $A(x) \succ 0$ is sometimes called the *maximum entropy completion* of A_f. From the optimality conditions (1.4), we see that the maximum entropy completion $x^\star$ satisfies

$$2\mathbf{Tr} E_{i_k j_k} A(x^\star)^{-1} = 2\left(A(x^\star)^{-1}\right)_{i_k j_k} = 0,\ k = 1, \ldots, m,$$

i.e., the matrix $A(x^*)^{-1}$ has a zero entry in every location corresponding to an unspecified entry in the original matrix. This is a very useful property in many applications; see, for example, Dempster[27], or Dewilde and Ning[30].

Parametrization of all positive definite completions. As an extension of the maximum entropy completion problem, consider

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr} C A(x) + \log\det A(x)^{-1} \\ \text{subject to} & A(x) \succ 0, \end{array} \tag{2.2}$$

where $C = C^T$ is given. This problem is of the form (1.5); the optimality conditions are

$$A(x^*) \succ 0, \quad \left(A(x^*)^{-1}\right)_{i_k j_k} = C_{i_k j_k},\ \ k = 1, \ldots, m, \tag{2.3}$$

i.e., the inverse of the optimal completion matches the given matrix C in every free entry. Indeed, this gives a parametrization of all positive definite completions: a positive definite completion $A(x)$ is uniquely characterized by specifying the elements of its inverse in the free locations, *i.e.*, $\left(A(x)^{-1}\right)_{i_k j_k}$. Problem (2.2) has been studied by Bakonyi and Woerdeman[8].

Contractive completion. A related problem is the contractive completion problem: given a (possibly nonsymmetric) matrix A_f and m index pairs (i_k, j_k), $k = 1, \ldots, m$, find a matrix

$$A(x) = A_f + \sum_{k=1}^{m} x_k E_{i_k,j_k}.$$

with spectral norm (maximum singular value) less than one.

This can be cast as a semidefinite programming feasibility problem [69]: find x such that

$$\begin{bmatrix} I & A(x) \\ A(x)^T & I \end{bmatrix} \succ 0. \tag{2.4}$$

One can define a maximum entropy solution as the solution that maximizes the determinant of (2.4), *i.e.*, solves the max-det problem

$$\begin{array}{ll} \text{maximize} & \log\det(I - A(x)^T A(x)) \\ \text{subject to} & \begin{bmatrix} I & A(x) \\ A(x)^T & I \end{bmatrix} \succ 0. \end{array} \tag{2.5}$$

See Nævdal and Woerdeman[50], Helton and Woerdeman[38]. For a statistical interpretation of (2.5), see §2.3.

Specialized algorithms and references. Very efficient algorithms have been developed for certain specialized types of completion problems. A well known example is the maximum entropy completion of a positive definite banded Toeplitz matrix (Dym and Gohberg[31], Dewilde and Deprettere[29]). Davis, Kahan, and Weinberger[22] discuss an analytic solution for a contractive completion problem with a special (block matrix) form. The methods discussed in this paper solve the *general* problem efficiently, although they are slower than the specialized algorithms where they are applicable. Moreover they have the advantage that other convex constraints, *e.g.*, upper and lower bounds on certain entries, are readily incorporated.

Completion problems, and specialized algorithms for computing completions, have been discussed by many authors, see, *e.g.*, Dym and Gohberg[31], Grone, Johnson, Sá and Wolkowicz[37], Barrett, Johnson and Lundquist[10], Lundquist and Johnson[49], Dewilde and Deprettere[29], Dembo, Mallows, and Shepp[26]. Johnson gives a survey in [41]. An interior-point method for an approximate completion problem is discussed in Johnson, Kroschel, and Wolkowicz[42].

We refer to Boyd *et al.* [13, §3.5], and El Ghaoui[32], for further discussion and additional references.

2.3. Risk-averse linear estimation. Let $y = Ax + w$ with $w \sim \mathcal{N}(0, I)$ and $A \in \mathbf{R}^{q\times p}$. Here x is an unknown quantity that we wish to estimate, y is the measurement, and w is the measurement noise. We assume that $p \leq q$ and that A has full column rank.

A linear estimator $\widehat{x} = My$, with $M \in \mathbf{R}^{p\times q}$, is unbiased if $\mathbf{E}\widehat{x} = x$ where $\mathbf{E}$ means expected value, *i.e.*, the estimator is unbiased if $MA = I$. The minimum-variance unbiased estimator is the unbiased estimator that minimizes the error variance

$$\mathbf{E}\|My - x\|^2 = \mathbf{Tr}MM^T = \sum_{i=1}^{p} \sigma_i^2(M),$$

where $\sigma_i(M)$ is the ith largest singular value of M. It is given by $M = A^+$, where $A^+ = (A^TA)^{-1}A^T$ is the pseudo-inverse of A. In fact the minimum-variance estimator is optimal in a stronger sense: it not only minimizes $\sum_i \sigma_i^2(M)$, but each singular value $\sigma_i(M)$ separately:

$$MA = I \Longrightarrow \sigma_i(A^+) \leq \sigma_i(M), \quad i = 1, \ldots, p. \tag{2.6}$$

In some applications estimation errors larger than the mean value are more costly, or less desirable, than errors less than the mean value. To capture this idea of *risk aversion* we can consider the objective or cost function

$$2\gamma^2 \log \mathbf{E} \exp\left(\frac{1}{2\gamma^2}\|My - x\|^2\right) \tag{2.7}$$

where the parameter γ is called the *risk-sensitivity parameter*. This cost function was introduced by Whittle in the more sophisticated setting of stochastic optimal control; see [72, §19]. Note that as $\gamma \to \infty$, the risk-sensitive cost (2.7) converges to the cost $\mathbf{E}\|My - x\|^2$, and is always larger (by convexity of exp). We can gain further insight from the first terms of the series expansion in $1/\gamma^2$:

$$\begin{aligned} 2\gamma^2 \log \mathbf{E} \exp\left(\frac{1}{2\gamma^2}\|\widehat{x} - x\|^2\right) &\simeq \mathbf{E}\|\widehat{x} - x\|^2 + \frac{1}{4\gamma^2}\left(\mathbf{E}\|\widehat{x} - x\|^4 - \left(\mathbf{E}\|\widehat{x} - x\|^2\right)^2\right) \\ &= \mathbf{E}z + \frac{1}{4\gamma^2}\mathrm{var}\ z, \end{aligned}$$

where $z = \|\widehat{x} - x\|^2$ is the squared error. Thus for large γ, the risk-averse cost (2.7) augments the mean-square error with a term proportional to the variance of the squared error.

The unbiased, risk-averse optimal estimator can be found by solving

$$\begin{array}{ll} \text{minimize} & 2\gamma^2 \log \mathbf{E} \exp\left(\frac{1}{2\gamma^2}\|My - x\|^2\right) \\ \text{subject to} & MA = I, \end{array}$$

which can be expressed as a max-det problem. The objective function can be written as

$$\begin{aligned} & 2\gamma^2 \log \mathbf{E} \exp\left(\frac{1}{2\gamma^2}\|My - x\|^2\right) \\ &= 2\gamma^2 \log \mathbf{E} \exp\left(\frac{1}{2\gamma^2} w^T M^T M w\right) \\ &= \begin{cases} 2\gamma^2 \log \det(I - (1/\gamma^2)M^TM)^{-1/2} & \text{if } M^TM \prec \gamma^2 I \\ \infty & \text{otherwise} \end{cases} \\ &= \begin{cases} \gamma^2 \log \det \begin{bmatrix} I & \gamma^{-1}M^T \\ \gamma^{-1}M & I \end{bmatrix}^{-1} & \text{if } \begin{bmatrix} I & \gamma^{-1}M^T \\ \gamma^{-1}M & I \end{bmatrix} \succ 0 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

so the unbiased risk-averse optimal estimator solves the max-det problem

$$\begin{array}{ll} \text{minimize} & \gamma^2 \log\det \left[\begin{array}{cc} I & \gamma^{-1}M^T \\ \gamma^{-1}M & I \end{array} \right]^{-1} \\ \text{subject to} & \left[\begin{array}{cc} I & \gamma^{-1}M^T \\ \gamma^{-1}M & I \end{array} \right] \succ 0 \\ & MA = I. \end{array} \tag{2.8}$$

This is in fact an analytic centering problem, and has a simple analytic solution: the least squares estimator $M = A^+$. To see this we express the objective in terms of the singular values of M:

$$\gamma^2 \log\det \left[\begin{array}{cc} I & \gamma^{-1}M^T \\ \gamma^{-1}M & I \end{array} \right]^{-1} = \left\{ \begin{array}{ll} -\gamma^2 \displaystyle\sum_{i=1}^{p} \log(1 - \sigma_i^2(M)/\gamma^2)^{-1} & \text{if } \sigma_1(M) < \gamma \\ \infty & \text{otherwise.} \end{array} \right.$$

It follows from property (2.6) that the solution is $M = A^+$ if $\|A^+\| < \gamma$, and that the problem is infeasible otherwise. (Whittle refers to the infeasible case, in which the risk-averse cost is always infinite, as 'neurotic breakdown'.)

In the simple case discussed above, the optimal risk-averse and the minimum-variance estimators coincide (so there is certainly no advantage in a max-det problem formulation). When additional convex constraints on the matrix M are added, *e.g.*, a given sparsity pattern, or triangular or Toeplitz structure, the optimal risk-averse estimator can be found by including these constraints in the max-det problem (2.8) (and will not, in general, coincide with the constrained minimum-variance estimator).

2.4. Experiment design.

Optimal experiment design. As in the previous section, 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 $A^+(A^+)^T = (A^TA)^{-1}$. We suppose that the rows of the matrix $A = [a_1 \ \ldots \ a_q]^T$ can be chosen among M possible test vectors $v^{(i)} \in \mathbf{R}^p$, $i = 1, \ldots, M$:

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

The goal of experiment design is to choose the vectors a_i so that the error covariance $(A^TA)^{-1}$ 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 $A^TA = q\sum_{i=1}^{M} \lambda_i v^{(i)} {v^{(i)}}^T$, 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 $(A^TA)^{-1}$ have been proposed. For example, in E-optimal design, we minimize the norm of the error covariance, $\lambda_{\max}((A^TA)^{-1})$, which is equivalent to maximizing the smallest eigenvalue

of A^TA. This is readily cast as the SDP

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

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

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

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 $(A^TA)^{-1}$, which leads to the max-det problem

$$\begin{array}{ll} \text{minimize} & \log\det\left(\displaystyle\sum_{i=1}^M \lambda_i v^{(i)} {v^{(i)}}^T\right)^{-1} \\ \text{subject to} & \lambda_i \geq 0, \ i = 1, \ldots, M \\ & \displaystyle\sum_{i=1}^M \lambda_i = 1. \end{array} \tag{2.9}$$

In §3 we will derive an interesting geometrical interpretation of the D-optimal matrix A, and show that A^TA determines the minimum volume ellipsoid, centered at the origin, that contains $v^{(1)}, \ldots, v^{(M)}$.

Fedorov[33], Atkinson and Donev[7], Pukelsheim[55], and Cook and Fedorov[19] give surveys and additional references on optimal experiment design. Wilhelm[74, 73] discusses nondifferentiable optimization methods for experiment design. Jávorzky *et al.* [40] describe an application in frequency domain system identification, and compare the interior-point method discussed later in this paper with conventional algorithms. Lee *et al.* [45, 5] discuss a non-convex experiment design problem and a relaxation solved by an interior-point method.

Extensions of D-optimal experiment design. The formulation of D-optimal design as an max-det problem has the advantage that one can easily incorporate additional useful convex constraints. For example, one can add linear inequalities $c_i^T\lambda \leq \alpha_i$, which can reflect bounds on the total cost of, or time required to carry out, the experiments.

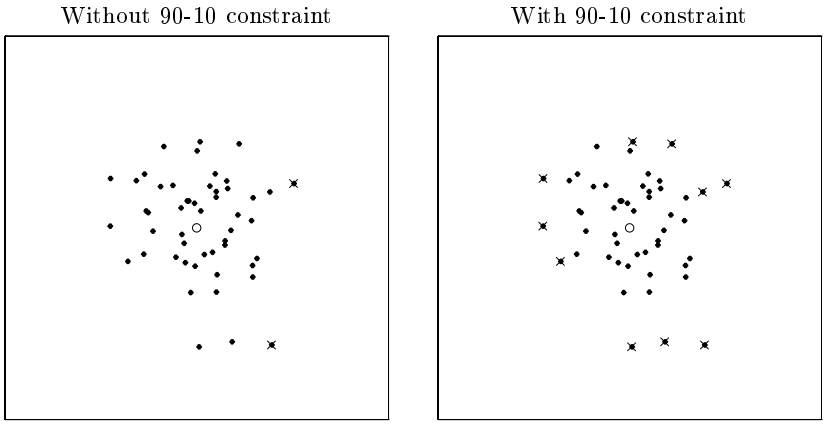

FIG. 2.1. *A D-optimal experiment design involving 50 test vectors in* $\mathbf{R}^2$*, with and without the 90-10 constraint. The circle is the origin; the dots are the test vectors that are not used in the experiment (*i.e.*, have a weight* $\lambda_i = 0$*); the crosses are the test vectors that are used (*i.e.*, have a weight* $\lambda_i > 0$*). Without the 90-10 constraint, the optimal design allocates all measurements to only two test vectors. With the constraint, the measurements are spread over ten vectors, with no more than 90% of the measurements allocated to any group of five vectors. See also Figure 2.2.*

We can also consider the case where each experiment yields several measurements, *i.e.*, the vectors a_i and $v^{(k)}$ become matrices. The max-det problem formulation (2.9) remains the same, except that the terms $v^{(k)}v^{(k)T}$ can now have rank larger than one. This extension is useful in conjunction with additional linear inequalities representing limits on cost or time: we can model discounts or time savings associated with performing groups of measurements simultaneously. Suppose, for example, that the cost of simultaneously making measurements $v^{(1)}$ and $v^{(2)}$ is less than the sum of the costs of making them separately. We can take $v^{(3)}$ to be the matrix

$$v^{(3)} = \left[\begin{array}{cc} v^{(1)} & v^{(2)} \end{array} \right]$$

and assign costs c_1, c_2, and c_3 associated with making the first measurement alone, the second measurement alone, and the two simultaneously, respectively.

Let us describe in more detail another useful additional constraint that can be imposed: that no more than a certain fraction of the total number of experiments, say 90%, is concentrated in less than a given fraction, say 10%, of the possible measurements. Thus we require

$$\sum_{i=1}^{\lfloor M/10 \rfloor} \lambda_{[i]} \leq 0.9, \tag{2.10}$$

where $\lambda_{[i]}$ denotes the ith largest component of λ. The effect on the experiment design will be to spread out the measurements over more points (at the cost of increasing the determinant of the error covariance). (See Figures 2.1 and 2.2.)

The constraint (2.10) is convex; it is satisfied if and only if there exists $x \in \mathbf{R}^M$ and t such that

$$\begin{array}{l} \lfloor M/10 \rfloor \, t + \sum_{i=1}^{M} x_i \leq 0.9 \\ t + x_i \geq \lambda_i, \quad i = 1, \ldots, M \\ x \geq 0 \end{array} \tag{2.11}$$

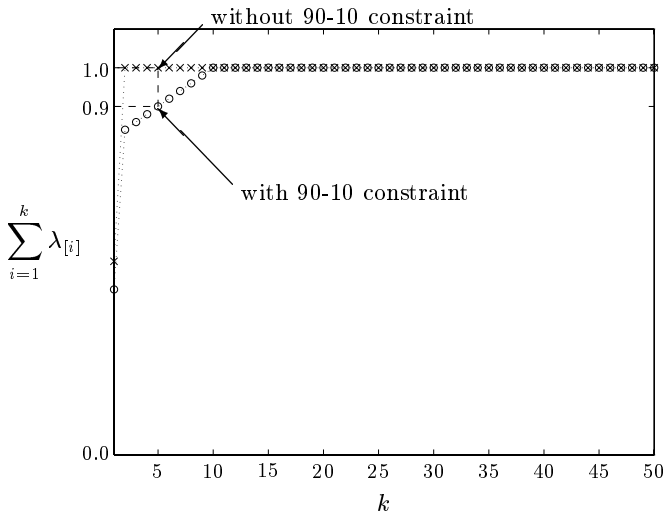

FIG. 2.2. *Experiment design of Figure 2.1. The curves show the sum of the largest k components of λ as a function of k, without the 90-10 constraint ('$\times$'), and with the constraint ('$\circ$'). The constraint specifies that the sum of the largest five components should be less than 0.9,* i.e., *the curve should avoid the area inside the dashed rectangle.*

(see [14, p.318]). One can therefore compute the D-optimal design subject to the 90-10 constraint (2.10) by adding the linear inequalities (2.11) to the constraints in (2.9) and solving the resulting max-det problem in the variables λ, x, t.

2.5. Maximum likelihood estimation of structured covariance matrices. The next example is the maximum likelihood (ML) estimation of structured covariance matrices of a normal distribution. This problem has a long history; see *e.g.*, Anderson[2, 3].

Let $y^{(1)}, \ldots, y^{(M)}$ be M samples from a normal distribution $\mathcal{N}(0, \Sigma)$. The ML estimate for Σ is the positive definite matrix that maximizes the log-likelihood function $\log \prod_{i=1}^{M} p(y^{(i)})$, where

$$p(x) = ((2\pi)^p \det \Sigma)^{-1/2} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right).$$

In other words, Σ can be found by solving

$$\begin{array}{ll} \text{maximize} & \log\det \Sigma^{-1} - \frac{1}{M} \displaystyle\sum_{i=1}^{N} {y^{(i)}}^T \Sigma^{-1} y^{(i)} \\ \text{subject to} & \Sigma \succ 0. \end{array} \tag{2.12}$$

This can be expressed as a max-det problem in the *inverse* $R = \Sigma^{-1}$:

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr} SR + \log\det R^{-1} \\ \text{subject to} & R \succ 0, \end{array} \tag{2.13}$$

where $S = \frac{1}{M} \sum_{i=1}^{n} y^{(i)} {y^{(i)}}^T$. Problem (2.13) has the straightforward analytical solution $R = S^{-1}$ (provided S is nonsingular).

It is often useful to impose additional structure on the covariance matrix Σ or its inverse R (Anderson[2, 3], Burg, Luenberger, Wenger[15], Scharf[61, §6.13],

Dembo[25]). In some special cases (*e.g.*, Σ is circulant) analytical solutions are known; in other cases where the constraints can be expressed as LMIs in R, the ML estimate can be obtained from a max-det problem. To give a simple illustration, bounds on the variances Σ_{ii} can be expressed as LMIs in R:

$$\Sigma_{ii} = e_i^T R^{-1} e_i \leq \alpha \Longleftrightarrow \left[\begin{array}{cc} R & e_i \\ e_i^T & \alpha \end{array} \right] \succeq 0.$$

The formulation as a max-det problem is also useful when the matrix S is singular (for example, because the number of samples is too small) and, as a consequence, the max-det problem (2.13) is unbounded below. In this case we can impose constraints (*i.e.*, prior information) on Σ, for example lower and upper bounds on the diagonal elements of R.

2.6. Gaussian channel capacity.

The Gaussian channel and the water-filling algorithm. The entropy of a normal distribution $\mathcal{N}(\mu, \Sigma)$ is, up to a constant, equal to $\frac{1}{2} \log\det \Sigma$ (see Cover and Thomas[21, Chapter 9]). It is therefore not surprising that max-det problems arise naturally in information theory and communications. One example is the computation of channel capacity.

Consider a simple Gaussian communication channel: $y = x + v$, where y, x, and v are random vectors in $\mathbf{R}^n$; $x \sim \mathcal{N}(0, X)$ is the input; y is the output, and $v \sim \mathcal{N}(0, R)$ is additive noise, independent of x. This model can represent n parallel channels, or one single channel at n different time instants or n different frequencies.

We assume the noise covariance R is known and given; the input covariance X is the variable to be determined, subject to constraints (such as power limits) that we will describe below. Our goal is to maximize the *mutual information* between input and output, given by

$$\frac{1}{2} \left(\log\det(X + R) - \log\det R \right) = \frac{1}{2} \log\det(I + R^{-1/2} X R^{-1/2})$$

(see [21]). The *channel capacity* is defined as the maximum mutual information over all input covariances X that satisfy the constraints. (Thus, the channel capacity depends on R and the constraints.)

The simplest and most common constraint is a limit on the average total power in the input, *i.e.*,

$$\mathbf{E} x^T x / n = \mathbf{Tr} X / n \leq P. \tag{2.14}$$

The information capacity subject to this average power constraint is the optimal value of

$$\begin{array}{ll} \text{maximize} & \frac{1}{2} \log\det(I + R^{-1/2} X R^{-1/2}) \\ \text{subject to} & \mathbf{Tr} X \leq nP \\ & X \succeq 0 \end{array} \tag{2.15}$$

(see [21, §10]). This is a max-det problem in the variable $X = X^T$.

There is a straightforward solution to (2.15), known in information theory as the *water-filling* algorithm (see [21, §10], [20]). Let $R = V \Lambda V^T$ be the eigenvalue

decomposition of R. By introducing a new variable $\widetilde{X} = V^T X V$, we can rewrite the problem as

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\log\det(I + \Lambda^{-1/2}\widetilde{X}\Lambda^{-1/2}) \\ \text{subject to} & \mathbf{Tr}\widetilde{X} \leq nP \\ & \widetilde{X} \succeq 0. \end{array}$$

Since the off-diagonal elements of $\widetilde{X}$ do not appear in the constraints, but decrease the objective, the optimal $\widetilde{X}$ is diagonal. Using Lagrange multipliers one can show that the solution is $\widetilde{X}_{ii} = \max(\nu - \lambda_i, 0)$, $i = 1, \ldots, n$, where the Lagrange multiplier ν is to be determined from $\sum \widetilde{X}_{ii} = nP$. The term 'water-filling' refers to a visual description of this procedure (see [21, §10], [20]).

Average power constraints on each channel. Problem (2.15) can be extended and modified in many ways. For example, we can replace the average *total* power constraint by an average power constraint on the individual channels, *i.e.*, we can replace (2.14) by $\mathbf{E}x_k^2 = X_{kk} \leq P$, $k = 1, \ldots, n$. The capacity subject to this constraint can be determined by solving the max-det problem

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\log\det\left(I + R^{-1/2}XR^{-1/2}\right) \\ \text{subject to} & X \succeq 0 \\ & X_{kk} \leq P, \quad k = 1, \ldots, n. \end{array}$$

The water-filling algorithm does not apply here, but the capacity is readily computed by solving this max-det problem in X. Moreover, we can easily add other constraints, such as power limits on subsets of individual channels, or an upper bound on the correlation coefficient between two components of x:

$$\frac{|X_{ij}|}{\sqrt{X_{ii}X_{jj}}} \leq \rho_{\max} \iff \left[\begin{array}{cc} \sqrt{\rho_{\max}}X_{ii} & X_{ij} \\ X_{ij} & \sqrt{\rho_{\max}}X_{jj} \end{array}\right] \succeq 0.$$

Gaussian channel capacity with feedback. Suppose that the n components of x, y, and v are consecutive values in a time series. The question whether knowledge of the past values v_k helps in increasing the capacity of the channel is of great interest in information theory [21, §10.6]). In the Gaussian channel with feedback one uses, instead of x, the vector $\tilde{x} = Bv + x$ as input to the channel, where B is a strictly lower triangular matrix. The output of the channel is $y = \tilde{x} + v = x + (B + I)v$. We assume there is an average total power constraint: $\mathbf{E}\tilde{x}^T\tilde{x}/n \leq P$.

The mutual information between $\tilde{x}$ and y is

$$\frac{1}{2}\left(\log\det((B+I)R(B+I)^T + X) - \log\det R\right),$$

so we maximize the mutual information by solving

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\left(\log\det((B+I)R(B+I)^T + X) - \log\det R\right) \\ \text{subject to} & \mathbf{Tr}(BRB^T + X) \leq nP \\ & X \succeq 0 \\ & B \text{ strictly lower triangular} \end{array}$$

over the matrix variables B and X. To cast this problem as a max-det problem, we introduce a new variable $Y = (B + I)R(B + I)^T + X$ (*i.e.*, the covariance of y), and

obtain

$$\begin{array}{ll} \text{maximize} & \log\det Y \\ \text{subject to} & \mathbf{Tr}(Y - RB^T - BR - R) \leq nP \\ & Y - (B+I)R(B+I)^T \succeq 0 \\ & B \text{ strictly lower triangular.} \end{array} \tag{2.16}$$

The second constraint can be expressed as an LMI in B and Y,

$$\left[\begin{array}{cc} Y & B+I \\ (B+I)^T & R^{-1} \end{array}\right] \succeq 0,$$

so (2.16) is a max-det problem in B and Y.

Capacity of channel with cross-talk. Suppose the n channels are independent, *i.e.*, all covariances are diagonal, and that the noise covariance depends on X: $R_{ii} = r_i + a_i X_{ii}$, with $a_i > 0$. This has been used as a model of near-end cross-talk (see [6]). The capacity (with the total average power constraint) is the optimal value of

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\sum_{i=1}^{n} \log\left(1 + \dfrac{X_{ii}}{r_i + a_i X_{ii}}\right) \\ \text{subject to} & X_{ii} \geq 0, \quad i = 1, \ldots, n \\ & \sum_{i=1}^{n} X_{ii} \leq nP, \end{array}$$

which can be cast as a max-det problem

$$\begin{array}{ll} \text{maximize} & \frac{1}{2}\sum_{i=1}^{n} \log(1+t_i) \\ \text{subject to} & X_{ii} \geq 0,\ t_i \geq 0, \quad i = 1, \ldots, n, \\ & \left[\begin{array}{cc} 1 - a_i t_i & \sqrt{r}_i \\ \sqrt{r}_i & a_i X_{ii} + r_i \end{array}\right] \succeq 0, \quad i = 1, \ldots, n, \\ & \sum_{i=1}^{n} X_{ii} \leq nP. \end{array}$$

The LMI is equivalent to $t_i \leq X_{ii}/(r_i + a_i X_{ii})$. This problem can be solved using standard methods; the advantage of a max-det problem formulation is that we can add other (LMI) constraints on X, *e.g.*, individual power limits. As another interesting possibility, we could impose constraints that distribute the power across the channels more uniformly, *e.g.*, a 90-10 type constraint (see §2.4).

3. The dual problem. We associate with (1.1) the *dual* problem

$$\begin{array}{ll} \text{maximize} & \log\det W - \mathbf{Tr} G_0 W - \mathbf{Tr} F_0 Z + l \\ \text{subject to} & \mathbf{Tr} G_i W + \mathbf{Tr} F_i Z = c_i, \quad i = 1, \ldots, m, \\ & W = W^T \succ 0, \quad Z = Z^T \succeq 0. \end{array} \tag{3.1}$$

The variables are $W \in \mathbf{R}^{l \times l}$ and $Z \in \mathbf{R}^{n \times n}$. Problem (3.1) is also a max-det problem, and can be converted into a problem of the form (1.1) by elimination of the equality constraints.

We say W and Z are dual feasible if they satisfy the constraints in (3.1), and strictly dual feasible if in addition $Z \succ 0$. We also refer to the max-det problem (1.1) as the primal problem and say x is primal feasible if $F(x) \succeq 0$ and $G(x) \succ 0$, and strictly primal feasible if $F(x) \succ 0$ and $G(x) \succ 0$.

Let p^* and d^* be the optimal values of problem (1.1) and (3.1), respectively (with the convention that $p^* = +\infty$ if the primal problem is infeasible, and $d^* = -\infty$ if the dual problem is infeasible).

The optimization problem (3.1) is the Lagrange dual of problem (1.1), rewritten as

$$\begin{array}{ll} \text{minimize} & c^T x + \log\det X^{-1} \\ \text{subject to} & X = G(x) \\ & F(x) \succeq 0, \quad X \succ 0. \end{array}$$

(We introduce a new variable $X = X^T \in \mathbf{R}^{l\times l}$ and add an equality constraint.) To derive the dual problem, we associate a Lagrange mulitiplier $Z = Z^T \succeq 0$ with the LMI $F(x) \succeq 0$, and a multiplier $W = W^T$ with the equality constraint $X = G(x)$. The optimal value can then be expressed as

$$p^* = \inf_{x,X} \sup_{Z\succeq 0, W} \left(c^T x + \log\det X^{-1} - \mathbf{Tr} ZF(x) + \mathbf{Tr} W(X - G(x))\right).$$

Changing the order of the supremum and the infimum, and solving the inner unconstrained minimization over x and X analytically, yields a lower bound on p^*:

$$\begin{aligned} p^* &\geq \sup_{Z\succeq 0, W} \inf_{x,X} \left(c^T x + \log\det X^{-1} - \mathbf{Tr} ZF(x) + \mathbf{Tr} W(X - G(x))\right) \\ &= \sup_{Z\succeq 0,\, W\succ 0,\, c_i = \mathbf{Tr} ZF_i + \mathbf{Tr} WG_i} (\log\det W - \mathbf{Tr} ZF_0 - \mathbf{Tr} WG_0 + l) \\ &= d^*. \end{aligned}$$

The inequality $p^* \geq d^*$ holds with equality if a constraint qualification holds, as stated in the following theorem.

THEOREM 3.1. *$p^* \geq d^*$. If (1.1) is strictly feasible, the dual optimum is achieved; if (3.1) is strictly feasible, the primal optimum is achieved. In both cases, $p^* = d^*$.*

The theorem follows from standard results in convex optimization (Luenberger[48, Chapter 8], Rockafellar[57, §29,30], Hiriart-Urruty and Lemaréchal[39, Chapter XII]), so we will not prove it here. See also Lewis[46] for a more general discussion of convex analysis of functions of symmetric matrices.

The difference between the primal and dual objective, *i.e.*, the expression

$$\begin{aligned} & c^T x + \log\det G(x)^{-1} + \log\det W^{-1} + \mathbf{Tr} G_0 W + \mathbf{Tr} F_0 Z - l \\ &= \sum_{i=1}^m x_i \mathbf{Tr} G_i W + \mathbf{Tr} G_0 W + \sum_{i=1}^m x_i \mathbf{Tr} F_i Z + \mathbf{Tr} F_0 Z - \log\det G(x) W - l \\ &= \mathbf{Tr} G(x) W - \log\det G(x) W - l + \mathbf{Tr} F(x) Z, \end{aligned} \tag{3.2}$$

is called the *duality gap* associated with x, W and Z. Theorem 3.1 states that the duality gap is always nonnegative, and zero only if x, W and Z are optimal.

Note that zero duality gap (3.2) implies $G(x)W = I$ and $F(x)Z = 0$. This gives the optimality condition for the max-det problem (1.1): a primal feasible x is optimal if there exists a $Z \succeq 0$, such that $F(x)Z = 0$ and

$$\mathbf{Tr} G_i G(x)^{-1} + \mathbf{Tr} F_i Z = c_i, \quad i = 1, \ldots, m.$$

This optimality condition is always sufficient; it is also necessary if the primal problem is strictly feasible.

In the remainder of the paper we will assume that the max-det problem is strictly primal and dual feasible. By Theorem 3.1, this assumption implies that the primal problem is bounded below and the dual problem is bounded above, with equality at the optimum, and that the primal and dual optimal sets are nonempty.

Example: semidefinite programming dual. As an illustration, we derive from (3.1) the dual problem for the SDP (1.2). Substituting $G_0 = 1$, $G_i = 0$, $l = 1$, in (3.1) yields

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

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

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

which is the dual SDP (in the notation used in [69]).

Example: D-optimal experiment design. As a second example we derive the dual of the experiment design problem (2.9). After a few simplifications we obtain

$$\begin{array}{ll} \text{maximize} & \log\det W + p - z \\ \text{subject to} & W = W^T \succ 0 \\ & {v^{(i)}}^T W v^{(i)} \leq z, \; i = 1, \ldots, M, \end{array} \tag{3.3}$$

where the variables are the matrix W and the scalar variable z. Problem (3.3) can be further simplified. The constraints are homogeneous in W and z, so for each dual feasible W, z we have a ray of dual feasible solutions tW, tz, $t > 0$. It turns out that we can analytically optimize over t: replacing W by tW and z by tz changes the objective to $\log\det W + p\log t + p - tz$, which is maximized for $t = p/z$. After this simplification, and with a new variable $\widetilde{W} = (p/z)W$, problem (3.3) becomes

$$\begin{array}{ll} \text{maximize} & \log\det \widetilde{W} \\ \text{subject to} & \widetilde{W} \succ 0 \\ & {v^{(i)}}^T \widetilde{W} v^{(i)} \leq p, \; i = 1, \ldots, M. \end{array} \tag{3.4}$$

Problem (3.4) has an interesting geometrical meaning: the constraints state that $\widetilde{W}$ determines an ellipsoid $\{x \mid x^T \widetilde{W} x \leq p\}$, centered at the origin, that contains the points $v^{(i)}$, $i = 1, \ldots, M$; the objective is to maximize $\det \widetilde{W}$, *i.e.*, to minimize the volume of the ellipsoid.

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

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

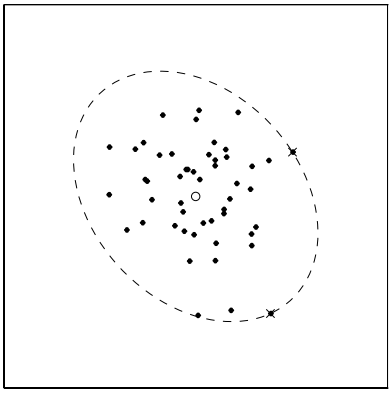

FIG. 3.1. *In the dual of the D-optimal experiment design problem we compute the minimum-volume ellipsoid, centered at the origin, that contains the test vectors. The test vectors with a nonzero weight lie on the boundary of the optimal ellipsoid. Same data and notation as in Figure 2.1.*

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

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

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

$$\begin{aligned} &\sum_{j=1}^M \lambda_j \left(p - v^{(j)\,T} \left(\sum_{i=1}^M \lambda_i v^{(i)} v^{(i)\,T} \right)^{-1} v^{(j)} \right) \\ &= p - \mathbf{Tr} \left(\sum_{j=1}^M \lambda_j v^{(j)} v^{(j)\,T} \right) \left(\sum_{i=1}^M \lambda_i v^{(i)} v^{(i)\,T} \right)^{-1} \\ &= 0. \end{aligned}$$

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

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

Geometrically, λ_j is nonzero only if $v^{(j)}$ lies on the boundary of the minimum volume ellipsoid. This makes more precise the intuitive idea that an optimal experiment only uses 'extreme' test vectors. Figure 3.1 shows the optimal ellipsoid for the experiment design example of Figure 2.1.

The duality between D-optimal experiment designs and minimum-volume ellipsoids also extends to non-finite compacts sets (Titterington[67], Pronzato and Walter[54]). The D-optimal experiment design problem on a compact set $C \subset \mathbf{R}^p$ is

$$\text{maximize } \log\det \mathbf{E} v v^T \tag{3.5}$$

over all probability measures on C. This is a convex but semi-infinite optimization problem, with dual ([67])

$$\begin{array}{ll} \text{maximize} & \log\det \widetilde{W} \\ \text{subject to} & \widetilde{W} \succ 0 \\ & v^T \widetilde{W} v \leq p, \quad v \in C. \end{array} \tag{3.6}$$

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

General methods for solving the semi-infinite optimization problems (3.5) and (3.6) fall outside the scope of this paper. In particular cases, however, these problems can be solved as max-det problems. One interesting example arises when C is the union of a finite number of ellipsoids. In this case, the dual (3.6) can be cast as a max-det problem (see [70]) and hence efficiently solved; by duality, we can recover from the dual solution the probability distribution that solves (3.5).

4. The central path. In this section we describe the *central path* of the max-det problem (1.1), and give some of its properties. The central path plays a key role in interior point methods for the max-det problem.

The primal central path. For strictly feasible x and $t \geq 1$, we define

$$\varphi_p(t,x) \stackrel{\Delta}{=} t\left(c^T x + \log\det G(x)^{-1}\right) + \log\det F(x)^{-1}. \tag{4.1}$$

This function is the sum of two convex functions: the first term is a positive multiple of the objective function in (1.1); the second term, $\log\det F(x)^{-1}$, is a barrier function for the set $\{x \mid F(x) \succ 0\}$. For future use, we note that the gradient and Hessian of $\varphi_p(x,t)$ are given by the expressions

$$\left(\nabla\varphi_p(t,x)\right)_i = t\left(c_i - \mathbf{Tr}G(x)^{-1}G_i\right) - \mathbf{Tr}F(x)^{-1}F_i, \tag{4.2}$$

$$\left(\nabla^2\varphi_p(t,x)\right)_{ij} = t\mathbf{Tr}G(x)^{-1}G_iG(x)^{-1}G_j + \mathbf{Tr}F(x)^{-1}F_iF(x)^{-1}F_j, \tag{4.3}$$

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

It can be shown that $\varphi_p(t,x)$ is a *strictly* convex function of x if the m matrices $\mathbf{diag}(G_i,F_i)$, $i = 1,\ldots,m$, are linearly independent, and that it is bounded below (since we assume the problem is strictly dual feasible). We define $x^\star(t)$ as the unique minimizer of $\varphi_p(t,x)$:

$$x^\star(t) = \operatorname{argmin}\left\{\varphi_p(t,x) \mid G(x) \succ 0, \quad F(x) \succ 0\right\}.$$

The curve $x^\star(t)$, parametrized by $t \geq 1$, is called the *central path*.

The dual central path. Points $x^\star(t)$ on the central path are characterized by the optimality conditions $\nabla\varphi_p(t,x^\star(t)) = 0$, *i.e.*, using the expression (4.2),

$$\mathbf{Tr}G(x^\star(t))^{-1}G_i + \frac{1}{t}\mathbf{Tr}F(x^\star(t))^{-1}F_i = c_i, \quad i = 1,\ldots,m.$$

From this we see that the matrices

$$W^\star(t) = G(x^\star(t))^{-1}, \quad Z^\star(t) = \frac{1}{t}F(x^\star(t))^{-1} \tag{4.4}$$

are strictly dual feasible. The duality gap associated with $x^\star(t)$, $W^\star(t)$ and $Z^\star(t)$ is, from expression (3.2),

$$\mathbf{Tr}F(x^\star(t))Z^\star(t) + \mathbf{Tr}G(x^\star(t))W^\star(t) - \log\det G(x^\star(t))W^\star(t) - l = \frac{n}{t},$$

which shows that $x^\star(t)$ converges to the solution of the max-det problem as $t \to \infty$.

It can be shown that the pair $(W^\star(t), Z^\star(t))$ actually lies on the *dual* central path, defined as

$$(W^\star(t), Z^\star(t)) = \operatorname{argmin}\left\{ \varphi_d(t,W,Z) \;\middle|\; \begin{array}{l} W = W^T \succeq 0, \quad Z = Z^T \succ 0, \\ \mathbf{Tr}G_iW + \mathbf{Tr}F_iZ = c_i, \quad i = 1,\ldots,m \end{array} \right\}$$

where

$$\varphi_d(t,W,Z) \triangleq t\left(\log\det W^{-1} + \mathbf{Tr}G_0W + \mathbf{Tr}F_0Z - l\right) + \log\det Z^{-1}.$$

The close connections between primal and dual central path are summarized in the following theorem.

THEOREM 4.1. *If x is strictly primal feasible, and W, Z are strictly dual feasible, then*

$$\varphi_p(t,x) + \varphi_d(t,W,Z) \geq n(1+\log t) \tag{4.5}$$

with equality if and only if $x = x^\star(t)$, $W = W^\star(t)$, $Z = Z^\star(t)$.

Proof. If $A = A^T \in \mathbf{R}^{p\times p}$ and $A \succ 0$, then $-\log\det A \geq -\mathbf{Tr}A + p$ (by convexity of $-\log\det A$ on the cone of positive semidefinite matrices). Applying this inequality, we find

$$\begin{aligned}
\varphi_p(t,x) + \varphi_d(t,W,Z) &= t\left(\mathbf{Tr}G(x)W + \mathbf{Tr}F(x)Z - \log\det G(x)W - l\right) - \log\det F(x)Z \\
&= t\left(-\log\det W^{1/2}G(x)W^{1/2} + \mathbf{Tr}W^{1/2}G(x)W^{1/2}\right) \\
&\quad - \log\det tZ^{1/2}F(x)Z^{1/2} + \mathbf{Tr}tZ^{1/2}F(x)Z^{1/2} + n\log t - tl \\
&\geq tl + n + n\log t - tl = n(1+\log t).
\end{aligned}$$

The equality for $x = x^\star(t)$, $W = W^\star(t)$, $Z = Z^\star(t)$ can be verified by substitution.

Tangent to the central path. We conclude this section by describing how the tangent direction to the central path can be computed. Let $\phi_1(x) = -\log\det G(x)$ and $\phi_2(x) = -\log\det F(x)$. A point $x^\star(t)$ on the central path is characterized by

$$t\left(c + \nabla\phi_1(x^\star(t))\right) + \nabla\phi_2(x^\star(t)) = 0.$$

The tangent direction $\frac{\partial x^\star(t)}{\partial t}$ can be found by differentiating with respect to t:

$$c + \nabla\phi_1(x^\star(t)) + \left(t\nabla^2\phi_1(x^\star(t)) + \nabla^2\phi_2(x^\star(t))\right)\frac{\partial x^\star(t)}{\partial t} = 0,$$

so that

$$\frac{\partial x^\star(t)}{\partial t} = -\left(t\nabla^2\phi_1(x^\star(t)) + \nabla^2\phi_2(x^\star(t))\right)^{-1}\left(c + \nabla\phi_1(x^\star(t))\right). \tag{4.6}$$

By differentiating (4.4), we obtain the tangent to the dual central path,

$$\frac{\partial W^\star(t)}{\partial t} = -G(x^\star(t))^{-1}\left(\sum_{i=1}^m \frac{\partial x^\star{}_i(t)}{\partial t}G_i\right)G(x^\star(t))^{-1}, \tag{4.7}$$

$$\frac{\partial Z^\star(t)}{\partial t} = -\frac{1}{t^2}F(x^\star(t))^{-1} - \frac{1}{t}F(x^\star(t))^{-1}\left(\sum_{i=1}^m \frac{\partial x^\star{}_i(t)}{\partial t}F_i\right)F(x^\star(t))^{-1}. \tag{4.8}$$

5. Newton's method. In this section we consider the problem of minimizing $\varphi_p(t,x)$ for fixed t, *i.e.*, computing $x^\star(t)$, given a strictly feasible initial point:

$$\begin{array}{ll} \text{minimize} & \varphi_p(t,x) \\ \text{subject to} & G(x) \succ 0 \\ & F(x) \succ 0. \end{array} \tag{5.1}$$

This includes, as a special case, the analytic centering problem ($t = 1$ and $F(x) = 1$). Our main motivation for studying (5.1) will become clear in next section, when we discuss an interior-point method based on minimizing $\varphi_p(t,x)$ for a sequence of values t.

Newton's method with line search can be used to solve problem (5.1) efficiently.

Newton method for minimizing $\varphi_p(t,x)$

given strictly feasible x, tolerance δ $(0 < \delta \leq 0.5)$

repeat

1. Compute the Newton direction $\delta x^N = -\left(\nabla^2\varphi_p(t,x)\right)^{-1}\nabla\varphi_p(t,x)$
2. Compute $\lambda = (\delta x^N \nabla^2\varphi_p(t,x)\delta x^N)^{1/2}$
3. **if** $(\lambda > 0.5)$, compute $\widehat{h} = \operatorname{argmin}\varphi_p(t, x + h\delta x^N)$
   else $\widehat{h} = 1$
4. Update: $x := x + \widehat{h}\delta x^N$

until $\lambda \leq \delta$

The quantity

$$\lambda = (\delta x^N \nabla^2\varphi_p(t,x)\delta x^N)^{1/2} \tag{5.2}$$

is called the *Newton decrement* at x. The cost of Step 3 (the *line search*) is very small, usually negligible compared with the cost of computing the Newton direction; see §8 for details.

It is well known that the asymptotic convergence of Newton's method is quadratic. Nesterov and Nemirovsky in [51, §2.2] give a complete analysis of the *global* speed of convergence. The main result of their convergence analysis applied to problem (5.1) is the following theorem.

THEOREM 5.1. *The algorithm terminates in fewer than*

$$11(\varphi_p(t,x^{(0)}) - \varphi_p(t,x^\star(t))) + \log_2\log_2(1/\delta) \tag{5.3}$$

iterations, and when it terminates, $\varphi_p(t,x) - \varphi_p(t,x^\star(t)) \leq \delta$.

A self-contained proof is given in Ref.[70].

Note that the right-hand side of (5.3) does not depend on the problem size (*i.e.*, m, n, or l) at all, and only depends on the problem data through the difference between the value of the function $\varphi_p(t,\cdot)$ at the initial point $x^{(0)}$ and at the central point $x^\star(t)$.

The term $\log_2\log_2(1/\delta)$, which is characteristic of quadratic convergence, grows *extremely slowly* with required accuracy δ. For all practical purposes it can be considered a constant, say, five (which guarantees an accuracy of $\delta = 2.33\cdot 10^{-10}$). Not quite precisely, then, the theorem says we can compute $x^\star(t)$ in at most $11(\varphi_p(t,x^{(0)}) - \varphi_p(t,x^\star(t))) + 5$ Newton steps. The precise statement is that within this number of

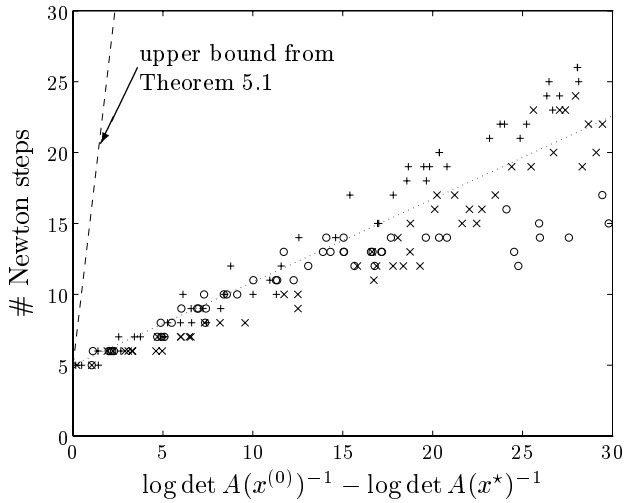

FIG. 5.1. *Number of Newton iterations to minimize* $\log\det A(x)^{-1}$ *versus* $\log\det A(x^{(0)})^{-1} - \log\det A(x^\star)^{-1}$ *(with* $\delta = 2.33\cdot 10^{-10}$, i.e., $\log_2\log_2(1/\delta) = 5$*). Random matrix completion problems of three sizes ('+':* $m = 20$*;* $l = 20$*, '×':* $m = 100$*,* $l = 20$*, 'o':* $m = 20$*,* $l = 100$*). The dotted line is a least-squares fit of the data and is given by* $5 + 0.59(\log\det A(x^{(0)})^{-1} - \log\det A(x^\star)^{-1})$*. The dashed line is the upper bound of Theorem 5.1 (*$5 + 11(\log\det A(x^{(0)})^{-1} - \log\det A(x^\star)^{-1})$*).*

iterations we can compute an extremely good approximation of $x^\star(t)$. In the sequel, we will speak of 'computing the central point $x^\star(t)$' when we really mean computing an extremely good approximation. We can justify this on several grounds. It is possible to adapt our exposition to account for the extremely small approximation error incurred by terminating the Newton process after $11(\varphi_p(t, x^{(0)}) - \varphi_p(t, x^\star(t))) + 5$ steps. Indeed, the errors involved are certainly on the same scale as computer arithmetic (roundoff) errors, so if a complexity analysis is to be carried out with such precision, it should also account for roundoff error.

Theorem 5.1 holds for an 'implementable' version of the algorithm as well, in which an appropriate *approximate* line search is used instead of the exact line search.

Numerical experiment. The bound provided by Theorem 5.1 on the number of Newton steps required to compute $x^\star(t)$, starting from $x^{(0)}$, will play an important role in our path-following method. It is therefore useful to examine how the bound compares to the actual number of Newton steps required in practice to compute $x^\star(t)$.

Figure 5.1 shows the results of a numerical experiment that compares the actual convergence of Newton's method with the bound (5.3). The test problem is a matrix completion problem

$$\begin{array}{ll} \text{minimize} & \log\det A(x)^{-1} \\ \text{subject to} & A(x) = A_f + \displaystyle\sum_{k=1}^{m} x_k \left(E_{i_k j_k} + E_{j_k i_k}\right) \succ 0, \end{array}$$

which is a particular case of (5.1) with $c = 0$, $G(x) = A(x)$, $F(x) = 1$, and $\varphi_p(t, x) = \log\det A(x)^{-1}$. We considered problems of three different sizes: $m = 20$, $l = 20$ (indicated by '+'); $m = 100$, $l = 20$ (indicated by '×'); $m = 20$, $l = 100$ (indicated by 'o'). Each point on the figure corresponds to a different problem instance, generated as follows.

- The matrices A_f were constructed as $A_f = UU^T$ with the elements of U drawn from a normal distribution $\mathcal{N}(0,1)$. This guarantees that $x = 0$ is strictly feasible. The m index pairs (i_k, j_k), $i_k \neq j_k$, were chosen randomly with a uniform distribution over the off-diagonal index pairs. For each of the three problem sizes, 50 instances were generated.
- For each problem instance, we first computed $x^\star$ using $x = 0$ as starting point. We then selected a value γ (uniformly in the interval $(0, 30)$), generated a random $\widehat{x} \in \mathbf{R}^m$ (with distribution $\mathcal{N}(0, I)$), and then computed $x^{(0)} = x^\star + t(\widehat{x} - x^\star)$ such that

$$\log\det(A(x^{(0)}))^{-1} - \log\det(A(x^*))^{-1} = \gamma.$$

  This point $x^{(0)}$ was used as starting point for the Newton algorithm.

Our experience with other problems shows that the results for this family of random problems are quite typical.

From the results we can draw two important conclusions.

- The quantity $\log\det(A(x^{(0)}))^{-1} - \log\det(A(x^\star))^{-1}$ not only provides an upper bound on the number of Newton iterations via Theorem 5.1; it is also a very good predictor of the number of iterations in practice. The dimensions m and l on the other hand have much less influence (except of course, through $\log\det(A(x^{(0)}))^{-1} - \log\det(A(x^\star))^{-1}$).
- The average number of Newton iterations seems to grow as

$$\alpha + \beta\left(\log\det(A(x^{(0)}))^{-1} - \log\det(A(x^\star))^{-1}\right),$$

  with $\alpha \simeq 5$, $\beta \simeq 0.6$. This is significantly smaller than the upper bound of Theorem 5.1 ($\alpha = 5$, $\beta = 11$).

In summary, we conclude that the difference $\varphi_p(t, x^{(0)}) - \varphi_p(t, x^\star(t))$ is a good measure, in theory and in practice, of the effort required to compute $x^\star(t)$ using Newton's method, starting at $x^{(0)}$.

A computable upper bound on the number of Newton steps. Note that $\varphi_p(t, x^\star(t))$ is not known explicitly as a function of t. To evaluate the bound (5.3) one has to compute $x^\star(t)$, *i.e.*, carry out the Newton algorithm. (Which, at the very least, would seem to defeat the purpose of trying to estimate or bound the number of Newton steps required to compute $x^\star(t)$.) Therefore the bound of Theorem 5.1 is not (directly) useful in practice. From Theorem 4.1, however, it follows that every dual feasible point W,Z provides a lower bound for $\varphi_p(t, x^\star(t))$:

$$\varphi_p(t, x^\star(t)) \geq -\varphi_d(t, W, Z) + n(1 + \log t).$$

and that the bound is exact if $W = W^\star(t)$ and $Z = Z^\star(t)$.

We can therefore replace the bound (5.3) by a weaker, but more easily computed bound, provided we have a dual feasible pair W, Z:

$$\begin{aligned} &11\left(\varphi_p(t, x^{(0)}) - \varphi_p(t, x^\star(t))\right) + \log_2\log_2(1/\delta) \\ &\leq 11\psi_{\rm ub}(t, x^{(0)}, W, Z) + \log_2\log_2(1/\delta), \end{aligned} \tag{5.4}$$

where

$$\psi_{\rm ub}(t, x, W, Z) = \varphi_p(t, x) + \varphi_d(t, W, Z) - n(1 + \log t). \tag{5.5}$$

This is the bound we will use in practice (and in our complexity analysis): it gives a readily computed bound on the number of Newton steps required to compute $x^\star(t)$, starting from $x^{(0)}$, given any dual feasible W, Z.

6. Path-following algorithms. Path-following methods for convex optimization have a long history. In their 1968 book [34], Fiacco and McCormick work out many general properties, *e.g.*, convergence to an optimal point, connections with duality, etc. No attempt was made to give a worst-case convergence analysis, until Renegar[56] proved polynomial convergence of a path-following algorithm for linear programming. Nesterov and Nemirovsky[51, §3] studied the convergence for nonlinear convex problems and provided proofs of polynomial worst-case complexity. See [51, pp.379–386] and Den Hertog[28] for a historical overview.

We will present two variants of a path-following method for the max-det problem. The short-step version of §6.2 is basically the path-following method of [34, 51], with a simplified, self-contained complexity analysis (see also Anstreicher and Fampa[4] for a very related analysis of an interior-point method for semidefinite programming). In the long-step verion of §6.3 we combine the method with predictor steps to accelerate convergence. This, too, is a well known technique, originally proposed by Fiacco and McCormick; our addition is a new step selection rule.

6.1. General idea. One iteration proceeds as follows. The algorithm starts at a point $x^\star(t)$ on the central path. As we have seen above, the duality gap associated with $x^\star(t)$ is n/t. We then select a new value $t^+ > t$, and choose a strictly feasible starting point $\widehat{x}$ (which may or may not be equal to $x^\star(t)$). The point $\widehat{x}$ serves as an approximation of $x^\star(t^+)$ and is called the *predictor* of $x^\star(t^+)$. Starting at the predictor $\widehat{x}$, the algorithm computes $x^\star(t^+)$ using Newton's method. This reduces the duality gap by a factor t^+/t. The step from $x^\star(t)$ to $x^\star(t^+)$ is called an *outer iteration.*

The choice of t^+ and $\widehat{x}$ involves a tradeoff. A large value of t^+/t means fast duality gap reduction, and hence fewer outer iterations. On the other hand it makes it more difficult to find a good predictor $\widehat{x}$, and hence more Newton iterations may be needed to compute $x^\star(t^+)$.

In the method discussed below, we impose a bound on the maximum number of Newton iterations per outer iteration, by requiring that the predictor $\widehat{x}$ and the new value of t^+ satisfy

$$\varphi_p(t^+, \widehat{x}) - \varphi_p(t^+, x^\star(t^+)) \leq \gamma. \tag{6.1}$$

This implies that no more than $5 + 11\gamma$ Newton iterations are required to compute $x^\star(t^+)$ starting at $\widehat{x}$. Of course, the exact value of the left hand side is not known, unless we carry out the Newton minimization, but as we have seen above, we can replace the condition by

$$\psi_{\rm ub}(t^+, \widehat{x}, \widehat{W}, \widehat{Z}) = \gamma, \tag{6.2}$$

where $\widehat{W}$ and $\widehat{Z}$ are conveniently chosen dual feasible points.

The parameters in the algorithm are $\gamma > 0$ and the desired accuracy ϵ.

Path-following algorithm

given $\gamma > 0$, $t \geq 1$, $x := x^\star(t)$
repeat
1. Select t^+, $\widehat{x}$, $\widehat{W}$, $\widehat{Z}$ such that $t^+ > t$ and $\psi_{\rm ub}(t^+, \widehat{x}, \widehat{W}, \widehat{Z}) = \gamma$
2. Compute $x^\star(t^+)$ starting at $\widehat{x}$, using the Newton algorithm of §5
3. $t := t^+$, $x := x^\star(t^+)$

until $n/t \leq \epsilon$

Step 1 in this outline is not completely specified. In the next sections we will discuss in detail different choices. We will show that one can always find $\widehat{x}$, $\widehat{W}$, $\widehat{Z}$ and t^+ that satisfy

$$\frac{t^+}{t} \geq 1 + \sqrt{\frac{2\gamma}{n}}. \tag{6.3}$$

This fact allows us to estimate the total complexity of the method, *i.e.*, to derive a bound on the total number of Newton iterations required to reduce the duality gap to ϵ. The algorithm starts on the central path, at $x^\star(t^{(0)})$, with initial duality gap $\epsilon^{(0)} = n/t^{(0)}$. Each iteration reduces the duality gap by t^+/t. Therefore the total number of outer iterations required to reduce the initial gap of $\epsilon^{(0)}$ to a final value below ϵ is at most

$$\left\lceil \frac{\log(\epsilon^{(0)}/\epsilon)}{\log(1+\sqrt{2\gamma/n})} \right\rceil \leq \left\lceil \sqrt{n}\frac{\log(\epsilon^{(0)}/\epsilon)}{\log(1+\sqrt{2\gamma})} \right\rceil.$$

(The inequality follows from the concavity of $\log(1+x)$.) The total number of Newton steps can therefore be bounded as

$$\begin{aligned} \text{Total \#Newton iterations} &\leq \lceil 5 + 11\gamma \rceil \left\lceil \sqrt{n}\frac{\log(\epsilon^{(0)}/\epsilon)}{\log(1+\sqrt{2\gamma})} \right\rceil \\ &= O\left(\sqrt{n}\log(\epsilon^{(0)}/\epsilon)\right). \end{aligned} \tag{6.4}$$

This upper bound increases slowly with the problem dimensions: it grows as $\sqrt{n}$, and is independent of l and m. We will see later that the performance in practice is even better.

Note that we assume that the minimization in Step 2 of the algorithm is exact. The justification of this assumption lies in the very fast local convergence of Newton's method: we have seen in §5 that it takes only a few iterations to improve a solution with Newton decrement $\lambda \leq 0.5$ to one with a very high accuracy.

Nevertheless, in a practical implementation (as well as in a rigorous theoretical analysis), one has to take into account the fact that $x^\star(t)$ can only be computed approximately. For example, the stopping criterion $n/t \leq \epsilon$ is based on the duality gap associated with exactly central points $x^\star(t)$, $W^\star(t)$, and $Z^\star(t)$, and is therefore not quite accurate if $x^\star(t)$ is only known approximately. We give a suitably modified criterion in Ref.[70], where we show that dual feasible points are easily computed during the centering step (Step 2) once the Newton decrement is less than one. Using the associated duality gap yields a completely rigorous stopping criterion. We will briefly point out some other modifications, as we develop different variants of the algorithm in the next sections; full details are described in [70]. With these modifications, the algorithm works well even when $x^\star(t)$ is computed approximately. (We often use a value $\delta = 10^{-3}$ in the Newton algorithm.)

It is also possible to extend the simple worst-case complexity analysis to take into account incomplete centering, but we will not attempt such an analysis here. For the fixed reduction algorithm (described immediately below), such a complete analysis can be found in Nesterov and Nemirovsky [51, §3.2].

6.2. Fixed-reduction algorithm. The simplest variant uses $\widehat{x} = x^\star(t)$, $\widehat{W} = W^\star(t)$, and $\widehat{Z} = Z^\star(t)$ in Step 1 of the algorithm. Substitution in condition (6.2)

gives

$$\begin{aligned}
&\psi_{\rm ub}(t^+, \widehat{x}, \widehat{W}, \widehat{Z}) \\
&= t^+(\mathbf{Tr}G(x^\star(t))W^\star(t) + \mathbf{Tr}F(x^\star(t))Z^\star(t) - \log\det G(x^\star(t))W^\star(t) - l) \\
&\quad - \log\det F(x^\star(t))Z^\star(t) - n(1+\log t^+) \\
&= n(t^+/t - 1 - \log(t^+/t)) = \gamma, \qquad (6.5)
\end{aligned}$$

which is a simple nonlinear equation in one variable, with a unique solution $t^+ > t$. We call this variant of the algorithm the *fixed-reduction* algorithm because it uses the same value of t^+/t — and hence achieves a fixed duality gap reduction factor — in each outer iteration. The outline of the fixed-reduction algorithm is as follows.

Fixed-reduction algorithm

given $\gamma > 0$, $t \geq 1$, $x := x^\star(t)$
Find α such that $n(\alpha - 1 - \log\alpha) = \gamma$
repeat
1. $t^+ := \alpha t$
2. Compute $x^\star(t^+)$ starting at x, using the Newton algorithm of §5
3. $t := t^+$, $x := x^\star(t^+)$

until $n/t \leq \epsilon$

We can be brief in the convergence analysis of the method. Each outer iteration reduces the duality gap by a factor α, so the number of outer iterations is *exactly*

$$\left\lceil \frac{\log(\epsilon^{(0)}/\epsilon)}{\log\alpha} \right\rceil.$$

The inequality (6.3), which was used in the complexity analysis of the previous section, follows from the fact that for $y \geq 1$

$$n(y - 1 - \log y) \leq \frac{n}{2}(y-1)^2,$$

and hence $\alpha \geq 1 + \sqrt{2\gamma/n}$.

This convergence analysis also reveals the limitation of the fixed reduction method: the number of outer iterations is never better than the number predicted by the theoretical analysis. The upper bound on the total number of Newton iterations (6.4) is also a good estimate in practice, provided we replace the constant $5 + 11\gamma$ with an empirically determined estimate such as $3 + 0.7\gamma$ (see Figure 5.1). The purpose of the next section is to develop a method with the same worst-case complexity as the fixed-reduction algorithm, but a much better performance in practice.

6.3. Primal-dual long-step algorithm. It is possible to use much larger values of t^+/t, and hence achieve larger gap reduction per outer iteration, by using a better choice for $\widehat{x}$, $\widehat{W}$, and $\widehat{Z}$ in Step 1 of the path-following algorithm.

A natural choice for $\widehat{x}$ is to take a point along the tangent to the central path, *i.e.*,

$$\widehat{x} = x^\star(t) + p\frac{\partial x^\star(t)}{\partial t},$$

for some $p > 0$, where the tangent direction is given by (4.6). Substitution in (6.2) gives a nonlinear equation from which t^+ and p can be determined. Taking the idea one step further, one can allow $\widehat{W}$ and $\widehat{Z}$ to vary along the tangent to the dual central path, *i.e.*, take

$$\widehat{W} = W^\star(t) + q\frac{\partial W^\star(t)}{\partial t}, \quad \widehat{Z} = Z^\star(t) + q\frac{\partial Z^\star(t)}{\partial t}$$

for some $q > 0$, with the tangent directions given by (4.7) and (4.8). Equation (6.2) then has three unknowns: t^+, the primal step length p, and the dual step length q. The fixed-reduction update of previous section uses the solution $t^+ = \alpha t$, $p = q = 0$; an efficient method for finding a solution with larger t^+ is described below.

The outline of the long-step algorithm is as follows.

Primal-dual long-step algorithm

given $\gamma > 0$, $t \geq 1$, $x := x^\star(t)$, $W := W^\star(t)$, $Z := Z^\star(t)$
Find α such that $n(\alpha - 1 - \log\alpha) = \gamma$
repeat
1. *Compute tangent to central path.* $\delta x := \frac{\partial x^\star(t)}{\partial t}$, $\delta W := \frac{\partial W^\star(t)}{\partial t}$, $\delta Z := \frac{\partial Z^\star(t)}{\partial t}$
2. *Parameter selection and predictor step.*
   2a. $t^+ := \alpha t$
   repeat {
      2b. $\widehat{p}, \widehat{q} = \mathrm{argmin}_{p,q}\ \psi_{ub}(t^+, x + p\delta x, W + q\delta W, Z + q\delta Z)$
      2c. Compute t^+ from $\psi_{\mathrm{ub}}(t^+, x + \widehat{p}\delta x, W + \widehat{q}\delta W, Z + \widehat{q}\delta Z) = \gamma$
   }
   2d. $\widehat{x} = x + \widehat{p}\delta x$
3. *Centering step.* Compute $x^\star(t^+)$ starting at $\widehat{x}$, using the Newton algorithm of §5
4. *Update.* $t := t^+$, $x := x^\star(t^+)$, $W := W^\star(t^+)$, $Z := Z^\star(t^+)$

until $n/t \leq \epsilon$

Again we assume exact centering in Step 3. In practice, approximate minimization works, provided one includes a small correction to the formulas of the tangent directions; see Ref.[70].

Step 2 computes a solution to (6.2), using a technique illustrated in Figure 6.1. The figure shows four iterations of the inner loop of Step 2 (for an instance of the problem family described in §9). With a slight abuse of notation, we write $\psi_{\mathrm{ub}}(t^+, p, q)$ instead of

$$\psi_{\mathrm{ub}}(t^+, x^\star(t) + p\delta x, W^\star(t) + q\delta W, Z^\star(t) + q\delta Z). \tag{6.6}$$

We start at the value $t^{(0)} = t$, at the left end of the horizontal axis. The first curve (marked $\psi_{\mathrm{ub}}(t^+, 0, 0)$) shows (6.6) as a function of t^+, with $p = q = 0$, which simplifies to

$$\psi_{\mathrm{ub}}(t^+, x^\star(t), W^\star(t), Z^\star(t)) = n(t^+/t - 1 - \log(t^+/t))$$

(see §6.2). This function is equal to zero for $t^+ = t$, and equal to γ for the short-step update $t^+ = \alpha t$. We then do the first iteration of the inner loop of Step 2. Keeping t^+ fixed at its value $t^{(1)}$, we minimize the function (6.6) over p and q (Step 2b). This produces new values $\widehat{p} = p^{(1)}$ and $\widehat{q} = q^{(1)}$ with a value of $\psi_{\mathrm{ub}} < \gamma$. This allows us to

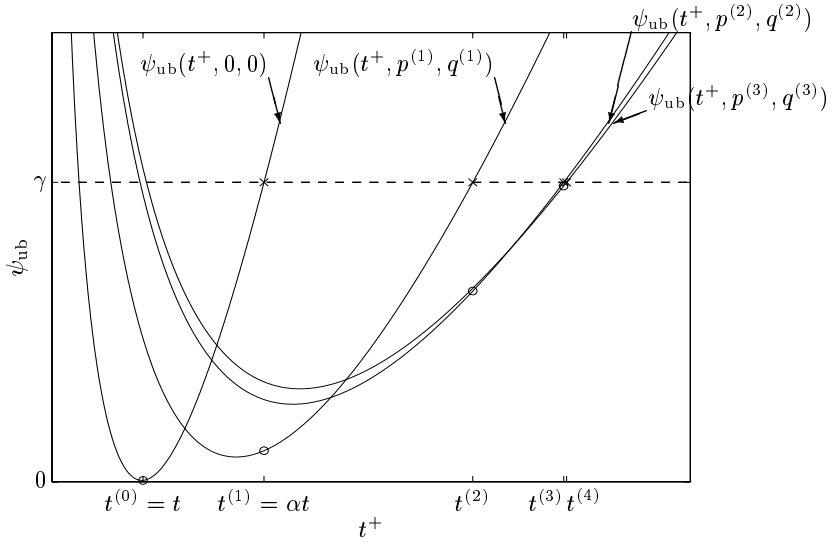

FIG. 6.1. *Parameter selection and predictor step in long-step algorithm alternates between minimizing $\psi_{\rm ub}(t^+, p, q)$ over primal step length p and dual step length q, and then increasing t^+ until $\psi_{\rm ub}(t^+, p, q) = \gamma$.*

increase t^+ again (Step 2c). The second curve in the figure (labeled $\psi_{\rm ub}(t^+, p^{(1)}, q^{(1)})$) shows the function (6.6) as a function of t^+ with fixed values $p = p^{(1)}$, $q = q^{(1)}$. The intersection with $\psi_{\rm ub} = \gamma$ gives the next value $t^+ = t^{(2)}$.

These two steps (2b, 2c) are repeated either for a fixed number of iterations or until t^+ converges (which in the example of Figure 6.1 happens after four or five iterations). Note that in each step 2c, we increase t^+, so that in particular, the final value of t^+ will be at least as large as its initial (short-step) value, $t^+ = \alpha t$. Thus, the complexity analysis for the short-step method still applies.

In practice, the inner loop (2b, 2c) often yields a value of t^+ considerably larger than the short-step value αt, while maintaining the same upper bound on the number of Newton step required to compute the next iterate $x^\star(t^+)$. In the example shown in the figure, the final value of t^+ is about a factor of 2.5 larger than the short-step value; in general, a factor of 10 is not uncommon.

Using some preprocessing we will describe in §8, the cost of the inner loop (2b, 2c) is very small, in most cases negligible compared with the cost of computing the tangent vectors.

Finally, note that the dual variables Z and W are not used in the fixed-reduction algorithm. In the primal-dual long-step algorithm they are used only in the predictor step to allow a larger step size p.

7. Preliminary phases. The algorithm starts at a central point $x^\star(t)$, for some $t \geq 1$. In this section we discuss how to select the initial t, and how to compute such a point.

Feasibility. If no strictly primal feasible point is known, one has to precede the algorithm with a first phase to solve the (SDP) feasibility problem: Find x that satisfies $G(x) > 0$, $F(x) > 0$. More details can be found in [69].

Choice of initial t. We now consider the situation where a strictly primal feasible point $x^{(0)}$ is known, but $x^{(0)}$ is not on the central path. In that case one has to select an appropriate initial value of t and compute a central point by Newton's

method starting at $x^{(0)}$. In theory (and often in practice) the simple choice $t = 1$ works.

It is not hard, however, to imagine cases where the choice $t = 1$ would be inefficient in practice. Suppose, for example, that the initial $x^{(0)}$ is very near $x^\star(100)$, so a reasonable initial value of t is 100 (but we don't know this). If we set $t = 1$, we expend many Newton iterations 'going backwards' up the central path towards the point $x^\star(1)$. Several outer iterations, and many Newton steps later, we find ourselves back near where we started, around $x^\star(100)$.

If strictly dual feasible points $W^{(0)}$, $Z^{(0)}$ are known, then we start with a known duality gap α associated with $x^{(0)}$, $W^{(0)}$ and $Z^{(0)}$. A very reasonable initial choice for t is then $t = \max\{1, n/\alpha\}$, since when $t = n/\alpha$, the centering stage computes central points with the same duality gap as the initial primal and dual solutions. In particular, the preliminary centering stage does not increase the duality gap (as it would in the scenario sketched above).

We can also interpret and motivate the initial value $t = n/\alpha$ in terms of the function $\psi_{\rm ub}(t, x^{(0)}, W^{(0)}, Z^{(0)})$, which provides an upper bound on the number of Newton steps required to compute $x^\star(t)$ starting at $x^{(0)}$. From the definition (5.5) we have

$$\psi_{\rm ub}(t, x^{(0)}, W^{(0)}, Z^{(0)}) = t\alpha + \log\det F(x^{(0)})^{-1} + \log\det {Z^{(0)}}^{-1} - n(1 + \log t),$$

which shows that the value $t = n/\alpha$ minimizes $\psi_{\rm ub}(t, x^{(0)}, W^{(0)}, Z^{(0)})$. Thus, the value $t = n/\alpha$ is the value which minimizes the upper bound on the number of Newton steps required in the preliminary centering stage.

A heuristic preliminary stage. When no initial dual feasible Z, W (and hence duality gap) are known, choosing an appropriate initial value of t can be difficult. We have had practical success with a variation on Newton's method that adapts the value of t at each step based on the (square of) the Newton decrement $\lambda(x, t)$,

$$\lambda(x,t)^2 = \nabla\varphi_p(t,x)^T \left(\nabla^2\varphi_p(t,x)\right)^{-1} \nabla\varphi_p(t,x),$$

which serves as a measure of proximity to the central path. It is a convex function of t, and is readily minimized in t for fixed x. Our heuristic preliminary phase is:

Preliminary centering phase

given strictly feasible x
$t := 1$
repeat {
1. $t := \max\{1, \operatorname{argmin} \lambda(x,t)\}$
2. $\delta x = -\left(\nabla^2\varphi_p(t,x)\right)^{-1} \nabla\varphi_p(t,x)$
3. $\widehat{h} = \operatorname{argmin}\, \varphi_p(t, x + h\delta x^N)$

} **until** $\lambda \leq \delta$

Thus, we adjust t each iteration to make the Newton decrement for the current x as small as possible (subject to the condition that t remains greater than 1).

8. Efficient line and plane searches. In this section we describe some simple preprocessing that allows us to implement the line search in the Newton method of §5 and the plane search of §6.3 very efficiently.

Line search in Newton's method. We first consider the line search in Newton's method of §5. Let λ_k, $k = 1, \ldots, l$, be the generalized eigenvalues of the pair $\sum_{i=1}^m \delta x_i^N G_i$, $G(x)$, and λ_k, $k = l+1, \ldots, l+n$, be the generalized eigenvalues of the pair $\sum_{i=1}^m \delta x_i^N F_i$, $F(x)$, where δx^N is the Newton direction at x. We can write $\varphi_p(t, x + h\delta x^N)$ in terms of these eigenvalues as

$$f(h) = \varphi_p(t, x + h\delta x^N) = \varphi_p(t, x) + hc^T \delta x^N + t \sum_{k=1}^{l} \log \frac{1}{1 + h\lambda_k} + \sum_{k=l+1}^{l+n} \log \frac{1}{1 + h\lambda_k}.$$

Evaluating the first and second derivatives $f'(h)$, $f''(h)$ of this (convex) function of $h \in \mathbf{R}$ requires only $O(n + l)$ operations (once the generalized eigenvalues λ_i have been computed). In most cases, the cost of the preprocessing, *i.e.*, computing the generalized eigenvalues λ_i, exceeds the cost of minimizing over h, but is small compared with the cost of computing the Newton direction. The function $\varphi_p(t, x + h\delta x^N)$ can therefore be efficiently minimized using standard line search techniques.

Plane search in long-step path-following method. A similar idea applies to the plane search of §6.3. In Step 2c of the primal-dual long-step algorithm we minimize the function $\psi_{\rm ub}(t, x + p\delta x, W + q\delta W, Z + q\delta Z)$ over p and q, where δx, δW, δZ are tangent directions to the central path. We can again reduce the function to a convenient form

$$\begin{aligned} &\psi_{\rm ub}(t, x + p\delta x, W + q\delta W, Z + q\delta Z) \\ &= \psi_{\rm ub}(t, x, W, Z) + p\beta_1 + q\beta_2 + t \sum_{k=1}^{l} \log \frac{1}{1 + p\lambda_k} + \sum_{k=l+1}^{l+n} \log \frac{1}{1 + p\lambda_k} \\ &\quad + t \sum_{k=1}^{l} \log \frac{1}{1 + q\mu_k} + \sum_{k=l+1}^{l+n} \log \frac{1}{1 + q\mu_k}, \end{aligned} \tag{8.1}$$

where λ_k, $k = 1, \ldots, l$, are the generalized eigenvalues of the pair $\sum_{i=1}^m \delta x_i G_i$, $G(x)$ and λ_k, $k = l+1, \ldots, l+n$, are the generalized eigenvalues of the pair $\sum_{i=1}^m \delta x_i F_i$, $F(x)$; μ_k, $k = 1, \ldots, l$, are the generalized eigenvalues of the pair δW, W, and μ_k, $k = l+1, \ldots, l+n$, are the generalized eigenvalues of the pair δZ, Z. The coefficients β_1 and β_2 are

$$\beta_1 = c^T \delta x, \quad \beta_2 = \mathbf{Tr} G_0 \delta W + \mathbf{Tr} F_0 \delta Z.$$

The first and second derivatives of the function (8.1) with respect to p and q can again be computed at a low cost of $O(l + n)$, and therefore the minimum of $\psi_{\rm ub}$ over the plane can be determined very cheaply, once the generalized eigenvalues have been computed.

In summary, the cost of line or plane search is basically the cost of preprocessing (computing certain generalized eigenvalues), which is usually negligible compared to the rest of algorithm (*e.g.*, determining a Newton or tangent direction).

One implication of efficient line and plane searches is that the total number of Newton steps serves as a good measure of the overall computing effort.

9. Numerical examples.

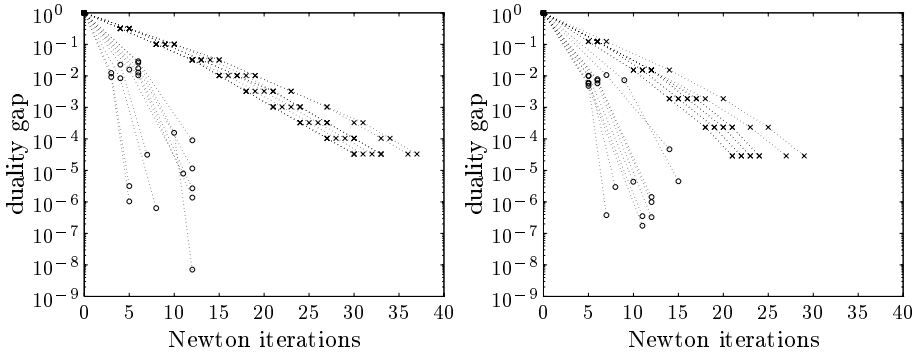

FIG. 9.1. *Duality gap versus number of Newton steps for randomly generated max-det problems of dimension $l = 10$, $n = 10$, $m = 10$. Left: $\gamma = 10$. Right: $\gamma = 50$. The crosses are the results for the fixed-reduction method; the circles are the results for the long-step method. Every cross/circle represents the gap at the end of an outer iteration.*

Typical convergence. The first experiment (Figure 9.1) compares the convergence of the fixed-reduction method and the long-step method. The lefthand plot shows the convergence of both methods for $\gamma = 10$; the righthand plot shows the convergence for $\gamma = 50$. Duality gap is shown vertically on a logarithmic scale ranging from 10^0 at the top to 10^{-9} at the bottom; the horizontal axis is the total number of Newton steps. Each outer iteration is shown as a symbol on the plot ('$\circ$' for the long-step and '$\times$' for the short-step method). Thus, the horizontal distance between two consecutive symbols shows directly the number of Newton steps required for that particular outer iteration; the vertical distance shows directly the duality gap reduction factor t^+/t.

Problem instances were generated as follows: $G_0 \in \mathbf{R}^{l\times l}$, $F_0 \in \mathbf{R}^{n\times n}$ were chosen random positive definite (constructed as U^TU with the elements of U drawn from a normal distribution $\mathcal{N}(0,1)$); the matrices G_i, F_i, $i = 1,\ldots,m$, were random symmetric matrices, with elements drawn from $\mathcal{N}(0,1)$; $c_i = \mathbf{Tr}G_i + \mathbf{Tr}F_i$, $i = 1,\ldots,m$. This procedure ensures that the problem is primal and dual feasible ($x = 0$ is primal feasible; $Z = I$, $W = I$ is dual feasible), and hence bounded below. We start on the central path, with initial duality gap one.

We can make the following observations.

- The convergence is very similar over all problem instances. The number of iterations required to reduce the duality gap by a factor 1000 ranges between 5 and 50. As expected, the long-step method performs much better than the fixed-reduction method, and typically converges in less than 15 iterations.
- The fixed-reduction method converges almost linearly. The duality gap reduction t^+/t per outer iteration can be computed from equation (6.5): $t^+/t = 3.14$ for $\gamma = 10$, and $t^+/t = 8.0$ for $\gamma = 50$. The number of Newton iterations per outer iteration is less than five in almost all cases, which is much less than the upper bound $5 + 11\gamma$. (Remember that this bound is a combination of two conservative estimates: Theorem 5.1 is conservative; see Figure 5.1. In addition we have replaced (6.1) with the weaker condition (6.2).)
- The long-step method takes a few more Newton iterations per centering step, but achieves a much larger duality gap reduction. Moreover the convergence accelerates near the optimum.
- Increasing γ has a large effect on the fixed-reduction method, but only little

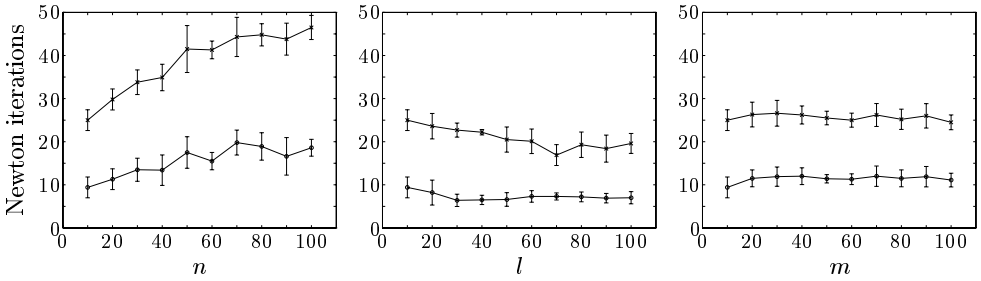

FIG. 9.2. *Newton iterations versus problem size for family of random problems. Fixed reduction method (top curve) and long-step method (lower curve).* $\gamma = 10$. Left. $l = 10$, $n = 10$–100, $m = 10$. Middle. $l = 10$–100, $n = 10$, $m = 10$. Right. $l = 10$, $n = 10$, $m = 10$–100. *The curves give the average over 10 problem instances. The error bars indicate the standard deviation.*

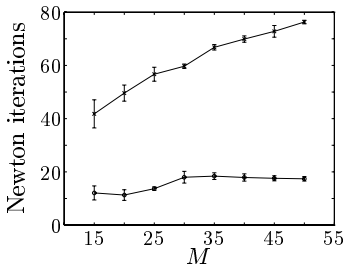

FIG. 9.3. *Newton iterations versus problem size for family of experiment design problems of §2.4 including 90-10 rule. Fixed reduction method (top curve) and long-step method (lower curve).* $\gamma = 10$, $p = 10$, $M = 15$–50. *The curves show the average over 10 problem instances. The error bars indicate the standard deviation.*

effect on the long-step method.

Complexity versus problem size. Figure 9.2 shows the influence of the problem dimension on the convergence. For each triplet (m, n, l) we generated 10 problem instances as above. We plot the number of Newton iterations to reduce the duality gap by a factor 1000, starting with duality gap 1. The plot shows the average number of Newton steps and the standard deviation. The top curve shows the results for the fixed-reduction method, the lower curve is for the long-step method.

- The number of Newton iterations in the short-step method depends on n as $O(\sqrt{n})$. This is easily explained from the convergence analysis of §6.2: We have seen that the number of outer iterations grows as $\sqrt{n}$, in theory and in practice, and hence the practical behavior of the fixed-reduction method is very close to the worst case behavior.
- We see that the number of iterations for the long-step method lies between 5 and 20, and is very weakly dependent on problem size.

Figure 9.3 shows similar results for a family of experiment design problems (2.9) in $\mathbf{R}^{10}$, including a 90-10 constraint (2.10). The points $v^{(i)}$, $i = 1, \ldots, M$, were generated from a normal distribution $\mathcal{N}(0, I)$ on $\mathbf{R}^p$. Note that the dimensions of the corresponding max-det problem are $m = 2M$, $n = 3M + 1$, $l = p$. Figure 9.3 confirms the conclusions of the previous experiment: It shows that the complexity of the fixed-reduction method grows as $\sqrt{n}$, while the complexity of the long-step method is almost independent of problem size.

10. Conclusion. The max-det problem (1.1) is a (quite specific) convex extension of the semidefinite programming problem, and hence includes a wide variety of

convex optimization problems as special cases. Perhaps more importantly, max-det problems arise naturally in many areas, including computational geometry, linear algebra, experiment design, linear estimation, and information and communication theory. We have described several of these applications.

Some of the applications have been studied extensively in the literature, and in some cases analytic solutions or efficient specialized algorithms have been developed. We have presented an interior-point algorithm that solves *general* max-det problems efficiently. The method can be applied to solve max-det problems for which no specialized algorithm is known; in cases where such a method exists, it opens the possibility of adding useful LMI constraints, which is an important advantage in practice.

We have proved a worst-case complexity of $O(\sqrt{n})$ Newton iterations. Numerical experiments indicate that the behavior is much better in practice: the method typically requires a number of iterations that lies between 5 and 50, almost independently of problem dimension. The total computational effort is therefore determined by the amount of work per iteration, *i.e.*, the computation of the Newton directions, and therefore depends heavily on the problem structure. When no structure is exploited, the Newton directions can be computed from the least-squares formulas in [70], which requires $O((n^2 + l^2)m^2)$ operations, but important savings are possible whenever we specialize the general method of this paper to a specific problem class.

Acknowledgments. We have many people to thank for useful comments, suggestions, and pointers to references and applications (including some applications that did not make it into the final version). In particular, we thank Paul Algoet, Kurt Anstreicher, Eric Beran, Patrick Dewilde, Valerii Fedorov, Gene Golub, Bijit Halder, Bill Helton, Istvan Kollar, Lennart Ljung, Tomas MacKelvey, Erik Ordentlich, Art Owen, Randy Tobias, Rick Wesel, and Henry Wolkowicz. We also thank the two reviewers and Michael Overton for very useful comments.

REFERENCES

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

[2] T. W. Anderson, *Statistical inference for covariance matrices with linear structure*, in Multivariate Analysis II. Proceedings of the Second International Symposium on Multivariate Analysis held at Wright State University, Dayton, Ohio, June 17-21, 1968, P. R. Krishnaiah, ed., Academic Press, New York, 1969, pp. 55–66.

[3] ———, *Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices*, in Essays in Probability and Statistics, R. C. Bose et al., eds., University of North Carolina Press, 1970, pp. 1–24.

[4] K. M. Anstreicher and M. Fampa, *A long-step path following algorithm for semidefinite programming problems*, tech. report, Department of Management Sciences, University of Iowa, Jan. 1996.

[5] K. M. Anstreicher, M. Fampa, J. Lee, and J. Williams, *Continuous relaxations for constrained maximum-entropy sampling*, in Integer Programming and Combinatorial Optimization, W. H. Cunningham, S. T. McCormick, and M. Queyranne, eds., no. 1084 in Lecture Notes in Computer Science, Springer Verlag, 1996, pp. 234–248.

[6] J. T. Aslanis and J. M. Cioffi, *Achievable information rates on digital subscriber loops: limiting information rates with crosstalk noise*, IEEE Transactions on Communications, 40 (1992), pp. 361–372.

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

[8] M. Bakonyi and H. J. Woerdeman, *Maximum entropy elements in the intersection of an affine space and the cone of positive definite matrices*, SIAM J. on Matrix Analysis and Applications, 16 (1995), pp. 360–376.

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

[10] W. W. Barrett, C. R. Johnson, and M. Lundquist, *Determinantal formulation for matrix completions associated with chordal graphs*, Linear Algebra and Appl., 121 (1989), pp. 265–289.

[11] D. Bertsekas and I. Rhodes, *Recursive state estimation for a set-membership description of uncertainty*, IEEE Trans. Aut. Control, AC-16 (1971), pp. 117–128.

[12] 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 (1993), pp. 63–111.

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

[14] S. Boyd and L. Vandenberghe, *Introduction to convex optimization with engineering applications*. Lecture Notes, Information Systems Laboratory. Stanford University, 1995. `http://www-isl.stanford.edu/people/boyd/392/ee392x.html`.

[15] J. P. Burg, D. G. Luenberger, and D. L. Wenger, *Estimation of structured covariance matrices*, Proceedings of the IEEE, 30 (1982), pp. 963–974.

[16] F. L. Chernousko, *Guaranteed estimates of undetermined quantities by means of ellipsoids*, Soviet Math. Doklady, 21 (1980), pp. 396–399.

[17] ———, *State Estimation for Dynamic Systems*, CRC Press, Boca Raton, Florida, 1994.

[18] M. Cheung, S. Yurkovich, and K. M. Passino, *An optimal volume ellipsoid algorithm for parameter estimation*, IEEE Trans. Aut. Control, AC-38 (1993), pp. 1292–1296.

[19] D. Cook and V. Fedorov, *Constrained optimization of experimental design*, Statistics, 26 (1995), pp. 129–178.

[20] T. M. Cover and S. Pombra, *Gaussian feedback capacity*, IEEE Transactions on Information Theory, 35 (1989), pp. 37–43.

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[22] C. Davis, W. M. Kahan, and H. F. Weinberger, *Norm-preserving dilations and their applications to optimal error bounds*, SIAM J. Numerical Anal., 19 (1982), pp. 445–469.

[23] J. Deller, *Set membership identification in digital signal processing*, IEEE Trans. Acoust., Speech, Signal Processing, 6 (1989), pp. 4–20.

[24] J. R. Deller, M. Nayeri, and S. F. Odeh, *Least-square identification with error bounds for real-time signal processing and control*, Proceedings of the IEEE, 81 (1993), pp. 815–849.

[25] A. Dembo, *The relation between maximum likelihood estimation of structured covariance matrices and periodograms*, IEEE Transactions on Acoustics, Speech and Signal Processing, 34 (1986), pp. 1661–1662.

[26] A. Dembo, C. L. Mallows, and L. A. Shepp, *Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation*, IEEE Transactions on Information Theory, 35 (1989), pp. 1206–1212.

[27] A. P. Dempster, *Covariance selection*, Biometrics, 28 (1972), pp. 157–175.

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

[29] P. Dewilde and E. F. A. Deprettere, *The generalized Schur algorithm: approximations and hierarchy*, in Topics in Operator Theory and Interpolation, I. Gohberg, ed., Birkhaüser Verlag, 1988, pp. 97–116.

[30] P. Dewilde and Z. Q. Ning, *Models for Large Integrated Circuits*, Kluwer, 1990.

[31] H. Dym and I. Gohberg, *Extensions of band matrices with band inverses*, Linear Algebra and Appl., 36 (1981), pp. 1–24.

[32] L. El Ghaoui, *Robustness analysis of semidefinite programs and applications to matrix completion problems*, in Proceedings of MTNS-96, 1996.

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

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

[35] E. Fogel, *System identification via membership set constraints with energy constrained noise*, IEEE Trans. Aut. Control, AC-24 (1979), pp. 752–757.

[36] E. Fogel and Y. Huang, *On the value of information in system identification- bounded noise case*, Automatica, 18 (1982), pp. 229–238.

[37] R. Grone, C. R. Johnson, E. M. Sá, and H. Wolkowicz, *Positive definite completions of partial Hermitian matrices*, Linear Algebra and Appl., 58 (1984), pp. 109–124.

[38] J. W. Helton and H. J. Woerdeman, *Symmetric Hankel operators: Minimal norm extensions and eigenstructures*, Linear Algebra and Appl., 185 (1993), pp. 1–19.

[39] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, vol. 305 of Grundlehren der mathematischen Wissenschaften, Springer-Verlag, New York, 1993.

[40] G. B. Jávorzky, I. Kollár, L. Vandenberghe, S. Boyd, and S.-P. Wu, *Optimal excitation signal design for frequency domain system identification using semidefinite programming*, in Proceedings of the 8th IMEKO TC4 Symposium on Recent Advances in Electrical Measurements, Budapest, Hungary, 1996.

[41] C. R. Johnson, *Positive definite completions: a guide to selected literature*, in Signal Processing. Part I: Signal Processing Theory, L. Auslander, T. Kailath, and S. Mitter, eds., The IMA volumes in mathematics and its applications, Springer, 1990, pp. 169–188.

[42] C. R. Johnson, B. Kroschel, and H. Wolkowicz, *An interior-point method for approximate positive semidefinite completions*, Tech. Report CORR Report 95-11, University of Waterloo, 1995. to appear in *Computational Optimization and Applications*.

[43] L. G. Khachiyan and M. J. Todd, *On the complexity of approximating the maximal inscribed ellipsoid for a polytope*, Mathematical Programming, 61 (1993), pp. 137–159.

[44] C. W. Ko, J. Lee, and K. Wayne, *A spectral bound for D-optimality*, tech. report, Department of Mathematics. University of Kentucky, 1996.

[45] J. Lee, *Constrained maximum entropy sampling*, tech. report, Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506-0027, 1995. to appear in *Operations Research*.

[46] A. S. Lewis, *Convex analysis on the Hermitian matrices*, SIAM J. on Optimization, 6 (1996), pp. 164–177.

[47] A. S. Lewis and M. L. Overton, *Eigenvalue optimization*, Acta Numerica, (1996), pp. 149–190.

[48] D. G. Luenberger, *Optimization By Vector Space Methods*, John Wiley & Sons, New York, N. Y., 1969.

[49] M. E. Lundquist and C. R. Johnson, *Linearly constrained positive definite completions*, Linear Algebra and Appl., 150 (1991), pp. 195–207.

[50] G. Naevdal and H. J. Woerdeman, *Partial matrix contractions and intersections of matrix balls*, Linear Algebra and Appl., 175 (1992), pp. 225–238.

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

[52] J. P. Norton, *An Introduction to Identification*, Academic Press, 1986.

[53] ———, *Identification and application of bounded-parameter models*, Automatica, 23 (1987), pp. 497–507.

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

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

[56] J. Renegar, *A polynomial-time algorithm, based on Newton's method, for linear programming*, Mathematical Programming, 40 (1988), pp. 59–93. North-Holland Press.

[57] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, second ed., 1970.

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

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

[60] S. S. Sapatnekar, *A convex programming approach to problems in VLSI design*, PhD thesis, Coordinated Science Laboratory, University of Illinois at Urbana-Campaign, Aug. 1992.

[61] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, 1991.

[62] F. C. Schweppe, *Recursive state estimation: Unknown but bounded errors and system inputs*, IEEE Trans. Aut. Control, AC-13 (1968), pp. 22–28.

[63] ———, *Uncertain dynamic systems*, Prentice-Hall, Englewood Cliffs, N.J., 1973.

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

[65] ———, *Applications of analytic centers*, in Numerical Linear Algebra, Digital Signal Processing, and Parallel Algorithms, G. Golub and P. V. Dooren, eds., vol. F70 of NATO ASI Series, 1991, pp. 617–632.

[66] S. Tarasov, L. Khachiyan, and I. Erlikh, *The method of inscribed ellipsoids*, Soviet Math. Dokl., 37 (1988), pp. 226–230. 1988 American Mathematical Society.

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

[68] L. Vandenberghe and S. Boyd, *Connections between semi-infinite and semidefinite programming*, in Proceedings of the International Workshop on Semi-Infinite Programming, R. Reemtsen and J.-J. Rueckmann, eds., 1996. To appear.

[69] ———, *Semidefinite programming*, SIAM Review, 38 (1996), pp. 49–95.

[70] L. Vandenberghe, S. Boyd, and S.-P. Wu, *Determinant maximization with linear matrix inequality constraints*, tech. report, Information Systems Laboratory. Stanford University, Feb. 1996.

[71] E. Walter and H. Piet-Lahanier, *Estimation of parameter bounds from bounded-error data: a survey*, Mathematics and Computers in Simulation, 32 (1990), pp. 449–468.

[72] P. Whittle, *Optimization over Time. Dynamic Programming and Stochastic Control*, John Wiley & Sons, 1982.

[73] A. Wilhelm, *Computing optimal designs by bundle trust methods*, Tech. Report 314, Institüt für Mathematik. Universität Augsburg, 1994.

[74] ———, *Algorithms for differentiable and nondifferentiable experimental design problems*, Proceedings SoftStat, (1995).

[75] H. S. Witsenhausen, *Sets of possible states of linear systems given perturbed observations*, IEEE Trans. Aut. Control, 13 (1968), pp. 556–558.

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