MATHEMATICS OF OPERATIONS RESEARCH
Vol. 28, No. 4, November 2003, pp. 625–648
Printed in U.S.A.

ON THE COMPLEXITY OF COMPUTING ESTIMATES OF CONDITION MEASURES OF A CONIC LINEAR SYSTEM

ROBERT M. FREUND AND JORGE R. VERA

Condition numbers based on the "distance to ill-posedness" $\rho(d)$ have been shown to play a crucial role in the theoretical complexity of solving convex optimization models. In this paper, we present two algorithms and corresponding complexity analysis for computing estimates of $\rho(d)$ for a finite-dimensional convex feasibility problem $P(d)$ in standard primal form: find x that satisfies $Ax = b$, $x \in C_X$, where $d = (A, b)$ is the data for the problem $P(d)$. Under one choice of norms for the m- and n-dimensional spaces, the problem of estimating $\rho(d)$ is hard (co-NP complete even when $C_X = \Re^n_+$). However, when the norms are suitably chosen, the problem becomes much easier: We can estimate $\rho(d)$ to within a constant factor of its true value with complexity bounds that are linear in $\ln(C(d))$ (where $C(d)$ is the condition number of the data d for $P(d)$), plus other quantities that arise naturally in consideration of the problem $P(d)$. The first algorithm is an interior-point algorithm, and the second algorithm is a variant of the ellipsoid algorithm. The main conclusion of this work is that when the norms are suitably chosen, computing an estimate of the condition measures of $P(d)$ is essentially not much harder than computing a solution of $P(d)$ itself.

1. Introduction. This paper is concerned with the problem of computing estimates of condition measures of a conic linear system in primal standard form, namely

$$(1) \qquad P(d): \text{ find } x \text{ that solves } Ax = b, \; x \in C_X,$$

where $C_X \subset X$ is a closed convex cone in the (finite) n-dimensional normed linear vector space X (with norm $\|x\|$ for $x \in X$), $b \in Y$ where Y is a (finite) m-dimensional normed linear vector space (with norm $\|y\|$ for $y \in Y$), and $A \in L(X, Y)$ where $L(X, Y)$ denotes the set of all linear operators A: $X \longrightarrow Y$. The reader will recognize immediately that various formats for feasibility of linear programming (LP), semidefinite programming (SDP), and second-order cone programming (SOCP) are special cases of (1), either directly or by the introduction of slack variables, etc.

The problem $P(d)$ is a very general format for studying the feasible region of a convex optimization problem, and has been the focus of analysis using interior-point methods; see Nesterov and Nemirovskii (1994) and Renegar (1995b, 1996), as well as volume-reducing cutting-plane methods (Freund and Vera 2000a).

The concept of the "distance to ill-posedness" $\rho(d)$ and a closely related condition measure $C(d)$ for problems such as $P(d)$ was introduced by Renegar (1994) and in Vera (1996) in a more specific setting, but then generalized more fully in Renegar (1995a, b). Further properties of the distance to ill-posedness were developed in Freund and Vera (2000b), including implications for the geometry of the feasible region of $P(d)$.

In this paper, we are interested in the more specific problem of actually computing estimates of $\rho(d)$ and its relatives. This problem is relevant, not only from a theoretical point of view, but also potentially from a practical point of view. However, the efficiency

Received November 20, 1999; revised July 27, 2001, and September 27, 2002.
MSC2000 subject classification. Primary: 90C, 90C05, 90C60.
OR/MS subject classification. Primary: Programming/linear.
Key words. Complexity of linear programming, semidefinite programming, interior-point methods, conditioning, complexity theory, error analysis.

0364-765X/03/2804/0625
1526-5471 electronic ISSN,

of computing estimates of these condition measures necessarily depends on the choice of norms on X and Y. For example, consider the case where $X = \Re^n$ and $C_X = \Re^n_+$, which is LP feasibility. When $\|x\|$ is the L_1 norm and $\|y\|$ is the L_2 norm, it follows from Freund and Orlin (1985) that estimating $\rho(d)$ to within a fixed constant factor is co-NP complete. (This is discussed more completely in §4.)

Nevertheless, the potential of efficiently estimating $\rho(d)$ for other choices of norms has been already pointed out in Freund and Vera (2000b) as well as in Peña (1997). In the latter work, the author presents a method for estimating $\rho(d)$ using the developments in Renegar (1995b). The proposed estimate is guaranteed to be within a factor of $\sqrt{m}$ of $\rho(d)$, when the norms on X and Y are L_2 norms (or more generally, inner-product norms). While there is no formal analysis of the complexity of the method, it nevertheless shows excellent potential for use in practice so long as m is not unreasonably large; see the discussion in §6.

In this work, we start with the characterizations of $\rho(d)$ for Problem (1) developed in Freund and Vera (2000b), where it is shown that $\rho(d)$ can be characterized as the optimal value of certain optimization problems. As was noted in Freund and Vera (2000b), under a suitable choice of norm on Y (namely the L_1 norm in $\Re^m$), the characterization of $\rho(d)$ reduces to the solution of $2m$ convex optimization problems, and so might be amenable to efficient solution. In fact, in the case of linear programming feasibility ($X = \Re^n$ and $C_X = \Re^n_+$) and when the norm on X is the L_1 norm in $\Re^n$, the $2m$ optimization problems are each a linear program, and so the problem of computing $\rho(d)$ for an LP feasibility problem can be solved exactly via LP itself. This suggests that under a suitable choice of norms, that condition measures might in general be computable in an "efficient" way, and so leads to the following questions: (i) What is the computational complexity, in some appropriate model, of actually computing these condition measures to within some constant factor of their true values? (ii) Is it efficient to actually compute estimates of these condition measures "in practice"? In this paper, we address the first question; we show that when the norms are suitably chosen, an estimate of $\rho(d)$ within a given constant factor can be computed in not much more computation time than is needed to decide the consistency of $P(d)$. The second question will hopefully be addressed in future work, although the recent work by Peña (1997) indicates that the practical computation of estimates of condition measures is in fact possible as part of an algorithm for solving $P(d)$, without introducing excessive additional computation time. Our overarching goal in this respect is the eventual implementation of condition number estimation within the context of traditional optimization algorithms, with the least possible computational overhead.

The structure of the paper is as follows. Section 2 contains definitions of some notation used in the text, and some technical material which is needed in the development of the results. Section 3 contains a brief summary of the concept of the distance to ill-posedness and the condition number of a data instance, as used in this study, and provides an overview of the main results of the paper. Section 3 also reviews a variety of useful implications of these condition measures, including perturbation bounds for linear optimization, complexity bounds for interior-point methods for convex optimization, numerical precision requirements for these algorithms, and other results as well. (The purpose of the review is to motivate the reader in the sense that the use of the condition measures proposed in this line of research is of potential practical relevance.) Section 4 presents some of the characterizations of the distance to ill-posedness that will be used in our algorithms. Section 5 contains our main results, namely two algorithms (one based on interior-point methods, and one based on the ellipsoid algorithm) for estimating the distance to ill-posedness, together with complexity analysis of these algorithms. Section 6 contains remarks concerning extensions to arbitrary norms, and a discussion of more practical issues in estimating condition measures.

2. Notation. We work in the setup of finite-dimensional normed linear vector spaces. Both X and Y are normed linear spaces of finite dimension n and m, respectively, endowed with norms $\|x\|$ for $x \in X$ and $\|y\|$ for $y \in Y$. For $\bar{x} \in X$, let $B(\bar{x}, r)$ denote the ball centered at $\bar{x}$ with radius r, i.e., $B(\bar{x}, r) = \{x \in X \mid \|x - \bar{x}\| \le r\}$, and define $B(\bar{y}, r)$ analogously for $\bar{y} \in Y$. We denote the set of real numbers by R and the set of nonnegative real numbers by R_+.

We denote by $d = (A, b)$ the data for the problem, and for $d = (A, b) \in L(X, Y) \times Y$, we define the product norm on the cartesian product $L(X, Y) \times Y$ as

$$\|d\| = \|(A, b)\| = \max\{\|A\|, \|b\|\}, \tag{2}$$

where $\|b\|$ is the norm specified for Y and $\|A\|$ is the operator norm, namely

$$\|A\| = \max\{\|Ax\|\colon \|x\| \le 1\}. \tag{3}$$

For $\bar{d} = (\bar{A}, \bar{b})$ we define the ball $B(\bar{d}, r) = \{d = (A, b) \in L(X, Y) \times Y\colon \|d - \bar{d}\| \le r\}$. We associate with X and Y the dual spaces X^* and Y^* of linear functionals defined on X and Y, respectively, and whose induced (dual) norms are denoted by $\|u\|_*$ for $u \in X^*$ and $\|w\|_*$ for $w \in Y^*$. Let $c \in X^*$. In order to maintain consistency with standard linear algebra notation in mathematical programming, we will consider c to be a column vector in the space X^* and will denote the linear function $c(x)$ by $c^T x$. Similarly, for $A \in L(X, Y)$ and $f \in Y^*$, we denote $A(x)$ by Ax and $f(y)$ by $f^T y$. We denote the adjoint of A by A^T. If $X = L_p(\Re^n)$, the norm is given by

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p},$$

for $p \ge 1$. The norm dual to $\|x\|_p$ is $\|z\|_* = \|z\|_q$ where q satisfies $1/p + 1/q = 1$, with appropriate limits as $p \to 1$ and $p \to +\infty$.

Because X and Y are normed linear vector spaces of finite dimension, all norms on each space are equivalent, and one can choose a particular norm for X and a particular norm for Y if so desired. In the majority of our analysis we will assume that

$$X = L_2(\Re^n) \quad \text{and} \quad Y = L_1(\Re^m). \tag{4}$$

This choice of norms implies that $X^* = L_2(\Re^n)$ and $Y^* = L_\infty(\Re^m)$, and the resulting matrix norm on $L(X, Y)$ is given by $\|A\| = \max\{\|Ax\|_1\colon \|x\|_2 \le 1\}$, and satisfies $\|A\|_1 \le \|A\| \le \sqrt{n}\|A\|_1$, where $\|A\|_1 = \max\{\|A_{\cdot 1}\|_1, \ldots, \|A_{\cdot n}\|_1\}$.

Because all norms in finite-dimensional spaces are equivalent, there is not much loss of generality in assuming (4). If other norms are more appropriate for specific problem instances and settings, one can always convert to the norms assumed in (4) by using appropriate norm equivalence constants. However, these constants will affect key aspects of our main results; see the discussion in §6.

If C is a convex cone in X, C^* will denote the dual convex cone defined by $C^* = \{z \in X^* \mid z^T x \ge 0 \text{ for any } x \in C\}$. A cone C is *regular* if C is a closed convex cone, has a nonempty interior, and is pointed (i.e., contains no line). If C is a closed convex cone, then C is regular if and only if C^* is regular.

Let C be a regular cone in X. A critical component of our analysis is the "min-width" of a regular cone C defined as follows:

Definition 2.1. Let C be a regular cone. Let

$$\tau = \max\left\{\frac{t}{\|x\|}\colon B(x, t) \in C\right\}.$$

Note that τ has a natural interpretation as the least relative width of C. In Freund and Vera (2000b), we defined the "coefficient of linearity" of a regular cone as:

$$\beta = \sup_{u \in X^*,\, \|u\|_* = 1} \; \inf_{x \in C,\, \|x\| = 1} u^T x. \tag{5}$$

It follows from the duality theory of cones and norms that if β^* is the coefficient of linearity of the cone C^*, then $\tau = \beta^*$; see Proposition 2.1 of Freund and Vera (2000a). Likewise, if β is the coefficient of linearity of the cone C, then $\beta = \tau^*$, where τ^* is the min-width of the cone C^*. In this paper we quote results from Freund and Vera (2000a, b), where some results are expressed using β and β^*. They will be quoted here with the equivalent expression in terms of τ^* and τ.

Let $(\bar{x}, \bar{t})$ be such that $B(\bar{x}, \bar{t}) \subset C$ and $\tau = \bar{t}/\|\bar{x}\|$. We can normalize the point $\bar{x}$ so that $\|\bar{x}\| = 1$. This point is "central" in the cone C with respect to the norm $\|\cdot\|$. In Freund and Vera (2000b), we defined the "norm approximation vector" of the cone C as the point $\bar{u}$ where the supremum in (5) is attained. It is easy to see that the point $\bar{x}$ is the norm approximation vector of the dual cone C^*. Also, from (5), it follows that $\bar{u}$ has the property that $\tau^*\|x\| = \beta\|x\| \le \bar{u}^T x \le \|x\|$, for all $x \in C$. In what follows, we assume that whenever the cone C is given, the min-width and the norm approximation vector for both C and C^* are given as well.

It is illustrative to see the width construction of two oft-used families of cones, the nonnegative orthant $\mathbf{R}^k_+$ and the positive semidefinite cone $S^{k\times k}_+$. For the nonnegative orthant $C = \{x \in \mathbf{R}^k \colon x \ge 0\}$ with the Euclidean norm $\|x\| = \sqrt{x^T x}$, it is straightforward to show that $\tau = 1/\sqrt{k}$ and $\bar{x} = (1/\sqrt{k})(1, \ldots, 1)^T$ is the norm approximation vector. For the positive semidefinite cone $C = \{X \in \mathbf{R}^{k\times k} \colon X \succeq 0\}$ with the Frobenius norm $\|X\| = \sqrt{\text{trace}(X^T X)}$, it is easy to show that the width is $\tau = 1/\sqrt{k}$ and that $\bar{X} = (1/\sqrt{k})I$ is the norm approximation vector.

3. The concept of ill-posedness in optimization, condition measures, and the main results of the paper. We now present a brief description of concepts of condition measures for $P(d)$ related to a model of data perturbation, as originally developed by Renegar (1994, 1995a, b). We also highlight the use of these concepts in the context of sensitivity bounds for linear optimization and their connections with the complexity of algorithms. At the end of this section, we present a summary of the main results of this present work.

Recall that $d = (A, b)$ is the "data" for the problem $P(d)$; that is, we regard the cone C_X as fixed and given, and the data for the problem to be the linear operator A together with the vector b. We denote the set of solutions of $P(d)$ as X_d to emphasize the dependence on the data d, i.e., $X_d = \{x \in X \mid Ax = b, x \in C_X\}$. We define

$$\mathcal{F} = \{(A, b) \in L(X, Y) \times Y \mid \text{ there exists } x \text{ satisfying } Ax = b, x \in C_X\}. \tag{6}$$

Then $\mathcal{F}$ corresponds to those data instances (A, b) for which $P(d)$ is consistent; i.e., $P(d)$ has a solution. We denote the complement of $\mathcal{F}$ by $\mathcal{F}^c$. Then $\mathcal{F}^c$ consists precisely of those data instances $d = (A, b)$ for which $P(d)$ is inconsistent. The boundary of $\mathcal{F}$ and of $\mathcal{F}^c$ is the set,

$$\mathcal{B} = \partial\mathcal{F} = \partial\mathcal{F}^c = \text{cl}(\mathcal{F}) \cap \text{cl}(\mathcal{F}^c), \tag{7}$$

where ∂S denotes the boundary of a set S and $\text{cl}(S)$ is the closure of a set S. Note that if $d = (A, b) \in \mathcal{B}$, then $P(d)$ is ill-posed in the sense that arbitrary small changes in the data $d = (A, b)$ will yield consistent instances of $P(d)$ as well as inconsistent instances of $P(d)$. For any $d = (A, b) \in L(X, Y) \times Y$, we define

$$\begin{aligned} \rho(d) &= \inf_{\Delta d}\big\{\|\Delta d\| : d + \Delta d \in \mathcal{B}\big\} \\ &= \inf_{\Delta A,\, \Delta b}\big\{\|(\Delta A, \Delta b)\| : (A + \Delta A, b + \Delta b) \in \text{cl}(\mathcal{F}) \cap \text{cl}(\mathcal{F}^c)\big\}. \end{aligned} \tag{8}$$

Then $\rho(d)$ is the "distance to ill-posedness" of the data d; i.e., $\rho(d)$ is the distance from d to the set $\mathcal{B}$ of ill-posedness instances. In addition to the work of Renegar (1994, 1995a, b) cited earlier, further analysis of the distance to ill-posedness has been explored in Filipowski

(1997, 1999), Nunez and Freund (1998), as well as in Vera (1992, 1996) and Freund and Vera (2000a, b). Observe that by means of a theorem of alternative, if problem $P(d)$ is infeasible, its corresponding "alternative" problem is feasible and can be put into similar conic structure, and so the format presumed herein can handle both consistency and inconsistency questions about problems of the form $P(d)$; see Renegar (1995b), and Freund and Vera (2000b).

The condition number for the data instance $d = (A, b)$ is defined to be $C(d) = \|d\|/\rho(d)$, which is a scale-invariant reciprocal of the distance to ill-posedness. This condition measure is connected to many different properties related to the complexity and stability of the problem $P(d)$. If d corresponds to a consistent instance, the condition number $C(d)$ is connected to the size of solutions of $P(d)$ (Renegar 1994), as well as the size and location of inscribed balls in the feasible region of $P(d)$ (Freund and Vera 2000b). Furthermore, $C(d)$ is connected to relative error bounds of $P(d)$ (Renegar 1994): If x' satisfies $Ax' = b + \Delta b$, $x' \in C_X$ for some Δb, then there exists x feasible for $P(d)$ whose relative distance from x' is not too large, namely

$$\frac{\|x - x'\|}{\max\{1, \|x'\|\}} \le C(d)\frac{\|\Delta b\|}{\|d\|}.$$

The above notions can be extended directly to the setup of convex optimization. Suppose we want to solve the optimization problem,

$$\mathrm{OP}(d)\colon\ \max\{c^T x\colon Ax = b, x \in C_X\}.$$

Then this problem is well posed if the feasibility problem itself is well posed and if the level sets of the objective function in $\mathrm{OP}(d)$ are themselves bounded, this latter property implying the feasibility of the dual problem $\mathrm{OD}(d)$, which is

$$\mathrm{OD}(d)\colon\ \min\{b^T y\colon A^T y - c \in C_X^*, y \in Y^*\}.$$

We can then define a primal distance to ill-posedness, $\rho_P(d)$, which corresponds to the distance to ill-posedness for problem $\mathrm{OP}(d)$, for the primal data $d_P = (A, b)$, and a dual distance to ill-posedness, $\rho_D(d)$, which corresponds to the distance to ill-posedness for problem $\mathrm{OD}(d)$, for the data for the dual feasible region $d_D = (A^T, c)$. With these additional notions, we can define $\rho(d) = \min\{\rho_P(d), \rho_D(d)\}$ as the distance to ill-posedness for the instance $d = (A, b, c)$, and $C(d) = \|d\|/\rho(d)$ as the condition number. Renegar (1995a) has shown that if d and d' are two instances of the optimization problem with $z(d)$ and $z(d')$ the corresponding optimal values, then

$$|z(d) - z(d')| \le C(d)^2 \|d - d'\|,$$

as long as $\|d - d'\| \le \rho(d)/2$. It is important to notice that this perturbation bound is valid even for relatively "large" perturbations of the data, in contrast with other, more traditional, results concerning the sensitivity analysis of optimization problems which are based on local measures near the optimal solution (see, for instance, Mangasarian 1987).

Condition measures have also been used in connection with the complexity of certain algorithms. In Renegar (1995b), an interior-point algorithm is developed that will decide consistency or inconsistency of $P(d)$, and when consistent will compute a feasible solution of $P(d)$, where the upper bound on the number of iterations depends linearly on $\ln(C(d))$. In Vera (1998), the effect of conditioning on the numerical precision requirements of an algorithm for approximating a solution to a linear program is considered. It is shown there that when the interior-point algorithm is executed with some of the numerically significant operations running in finite precision arithmetic, the working precision needed, measured in terms of the number of digits, is proportional to $\ln(C(d))$. In Freund and Vera (2000a), we analyze the complexity of the ellipsoid algorithm applied to solving the optimization problem $\mathrm{OP}(d)$. It is shown that the number of main iterations needed is also proportional to $\ln(C(d))$. In Epelman and Freund (2000), the complexity of an elementary algorithm for resolving a conic linear system is studied, and this complexity depends polynomially on $C(d)$.

These results serve to illustrate that there might be some practical relevance in computing the above-mentioned condition measures. By analogy to numerical analysis tools for solving systems of equations, we envision the possibility of algorithms for linear programming, for example, that not only compute the optimal solution of the problem, but that also compute, without substantial additional effort, an estimate of the condition number of the problem instance. The complexity results that we obtain show that this is within the realm of possibility.

The main results. Assume from this point onward that the cone C_X is a regular cone and that the instance d of (1) is consistent.

For $\alpha > 1$, our goal is to compute an α-estimate of $\rho(d)$, which is defined to be a number $\hat{\rho}$ such that

$$\frac{\hat{\rho}}{\alpha} \le \rho(d) \le \hat{\rho}.$$

In the following sections, we will show explicit ways of computing a 2-estimate of $\rho(d)$, and as a consequence, an estimate of $C(d)$, under the choice of norms given in (4). We will describe two algorithms. One algorithm presumes the knowledge of a self-concordant barrier function for the cone C_X^*, and uses an interior-point (barrier) method. The other algorithm presumes knowledge of separation oracles for the cones C_X and C_X^*, and is based on the ellipsoid method. The consideration of these two algorithms makes use of two of the main theoretical developments in recent times in the area of convex optimization: path-following interior-point methods and volume-reducing cutting plane methods. The performance results for the two algorithms applied to a consistent instance of (1) with norms chosen via (4) are as follows:

- The first algorithm, EST-INT, has as input the data d, a starting point u^0 in the interior of C_X^*, and an upper bound $\bar{\delta}$ on $\|d\|$. The algorithm uses an interior-point method based on a self-concordant barrier function for C_X^*, whose parameter is ϑ_*. The algorithm will compute a 2-estimate of $\rho(d)$ in

$$O\left(m\sqrt{\vartheta_*}\ln\left(\vartheta_* + m + \frac{\|u^0\|}{\operatorname{dist}(u^0, \partial C_X^*)} + \frac{\bar{\delta}}{\|d\|} + C(d)\right)\right)$$

iterations of Newton steps.

- The second algorithm, EST-ELL, has as input the data d and an upper bound $\bar{\delta}$ on $\|d\|$. The algorithm will compute a 2-estimate for $\rho(d)$ in

$$O\left(m(m+n)^2\ln\left(m + \frac{1}{\tau} + \frac{1}{\tau^*} + \frac{\bar{\delta}}{\|d\|} + C(d)\right)\right)$$

iterations of the ellipsoid algorithm. Here τ and τ^* are the min-widths of the cones C_X and C_X^*, respectively.

In both cases of algorithms, notice the linear dependency with respect to $\ln(C(d))$ in the complexity bounds. Given that previous work by Renegar (1995b) for interior-point methods shows that the complexity of finding a solution of $P(d)$ depends linearly on $\ln(C(d))$, and Freund and Vera (2000a) show a similar conclusion for the ellipsoid method, we are led to the conclusion that estimating the condition measure is not much harder than solving the problem $P(d)$ itself, at least with these algorithms. Section 5 contains further elaboration of this theme.

The computation of an estimate within a factor of two might seem poor, but it is more than enough if we consider that the effect of the condition number on several properties of the problem enters in the form of $\ln(C(d))$. Furthermore, in §6 we discuss how these results can be modified to account for an arbitrary factor $\alpha > 1$.

4. Characterization of $\rho(d)$ via convex optimization problems. Several characterizations of the distance to ill-posedness $\rho(d)$ for the feasibility problem $P(d)$ given in (1) are presented in Freund and Vera (2000b), based on Renegar (1995b). We will concentrate on two of these characterizations for the case when d defines a feasible instance of (1).

Consider the following two problems:

$$\begin{aligned} P_r(d)\colon\ r(d) = \underset{v\in Y,\, \|v\|\le 1}{\text{minimum}}\ \underset{r,x,\theta}{\text{maximum}}\ & \theta \\ \text{s.t.}\ & br - Ax - v\theta = 0, \\ & |r| + \|x\| \le 1, \\ & r \ge 0 \qquad x \in C_X, \end{aligned} \tag{9}$$

and

$$\begin{aligned} P_j(d)\colon\ j(d) = \underset{y,q,g}{\text{minimum}}\ & \max\left\{\|A^T y - q\|_*, |b^T y + g|\right\} \\ \text{s.t.}\ & y \in Y^*,\ \|y\|_* = 1,\ q \in C_X^*,\ g \ge 0. \end{aligned} \tag{10}$$

In Renegar (1995b), it is shown that $\rho(d) = r(d)$ and in Freund and Vera (2000b), it is shown that $\rho(d) = j(d)$, and that problems $P_r(d)$ and $P_j(d)$ are duals, with strong duality holding. We summarize these results as:

THEOREM 4.1 (RENEGAR 1995b, FREUND AND VERA 2000b). *If $d \in \mathcal{F}$, then $\rho(d) = r(d) = j(d)$.*

Problem $P_r(d)$ measures, in a sense, how much the right-hand side of the homogenized version of $P(d)$, namely $br - Ax = 0$, $x \in C_X$, $r \ge 0$, can be perturbed and still maintain consistency. In a "dual" way, problem $P_j(d)$ measures how close the data is to satisfying a theorem of alternative that is a certificate of infeasibility.

We first show that under a particular choice of norms, problem $P_r(d)$ is a hard problem. Suppose that $X = \Re^n$, $Y = \Re^m$, and $C_X = \Re^n_+$, and that $\|x\| := \|x\|_1$ for $x \in X$ and $\|y\| := \|y\|_2$ for $y \in Y$. In this case, $P_r(d)$ is the problem of finding the largest inscribed Euclidean ball in $\Re^m$ centered at the origin and contained in the convex hull of the points $A_{\cdot 1}, A_{\cdot 2}, \ldots, A_{\cdot n}, -b$. However, the problem of simply testing if a Euclidean ball is contained in the convex hull of a given set of points is co-NP complete; see Freund and Orlin (1985). Therefore, computing an estimate of $\rho(d)$ to within any constant factor is co-NP complete even for the $C_X = \Re^n_+$, under this particular choice of norms.

Now let us return to the general case where C_X is any regular cone, and suppose instead that the norms on X and Y are chosen via (4). Then problem $P_r(d)$ can be interpreted as finding the largest L_1 ball in $\Re^m$ centered at the origin and contained in the set $\mathcal{H}_d := \{br - Ax\colon\ r \ge 0, x \in C_X, r + \|x\|_2 \le 1\}$. Then since the unit L_1 ball in $\Re^m$ is the convex hull of its $2m$ extreme points $e_1, \ldots, e_m, -e_1, \ldots, -e_m$ (here the vector e_i denotes the ith unit vector), we can solve this problem by computing the largest scaling θ of $e_1, \ldots, e_m, -e_1, \ldots, -e_m$ for which $\theta e_1, \ldots, \theta e_m, -\theta e_1, \ldots, -\theta e_m$ are all in $\mathcal{H}_d$. This in turn is solvable by separately solving the $2m$ convex problems:

$$\begin{aligned} S_{\pm i}(d)\colon\ s_{\pm i}(d) = \max_{r,x,\theta}\ & \theta \\ \text{s.t.}\ & br - Ax \pm e_i\theta = 0 \\ & x \in C_X, \qquad r \ge 0 \\ & r + \|x\|_2 \le 1, \end{aligned}$$

whose dual problems are

$$
\begin{aligned}
F_{\pm i}(d)\colon\ f_{\pm i}(d) = \min_{y,q,g}\ & \max\{\|A^T y - q\|_2, |b^T y + g|\} \\
\text{s.t.}\ & q \in C_X^*, \qquad g \geq 0, \qquad \pm y_i = 1,\ \|y\|_\infty \leq 1,
\end{aligned}
$$

where the notation $\pm i$ denotes the occurrence of the index i with the constraint $+y_i = 1$ and $-y_i = 1$ in $F_{\pm i}(d)$, respectively, and with the vector $+e_i$ and $-e_i$ in the equations of the $S_{\pm i}(d)$ problem, respectively. Observe that both of these families of problems are convex problems.

Notice that the selection of the L_1 norm for the Y space is important as it makes the problem tractable: the L_1 unit ball has only $2m$ extreme points.

Problem $F_{\pm i}(d)$ can be further simplified by relaxing the constraint $\|y\|_\infty \leq 1$. Let

$$
\begin{aligned}
\bar{F}_{\pm i}(d)\colon\ \bar{f}_{\pm i}(d) = \min_{y,q}\ & \max\{\|A^T y - q\|_2, b^T y\} \\
\text{s.t.}\ & q \in C_X^*, \qquad \pm y_i = 1.
\end{aligned}
$$

Note that all of these problems are convex problems and they all lead to $\rho(d)$, as is shown in the following result:

PROPOSITION 4.1. *Suppose that d is a feasible instance of* (1) *and that the norms on X and Y are chosen via* (4). *Then* $\rho(d) = f(d) = \min_{\pm i} f_{\pm i}(d) = \min_{\pm i} \bar{f}_{\pm i}(d) = s(d) = \min_{\pm i} s_{\pm i}(d)$.

PROOF. The equalities $f(d) = \min_{\pm i} f_{\pm i}(d) = s(d) = \min_{\pm i} s_{\pm i}(d)$ are obvious. We show now that $\min_{\pm i} f_{\pm i}(d) = \min_{\pm i} \bar{f}_{\pm i}(d)$. Suppose that $\min_{\pm i} f_{\pm i}(d)$ is actually attained at the index $+i_0$, so that $f(d) = f_{+i_0}(d)$. Let $(\bar{y}, \bar{q}, \bar{g})$ be such that $f_{+i_0}(d) = \max\{\|A^T\bar{y} - \bar{q}\|, |b^T\bar{y} + \bar{g}|\}$, $\bar{q} \in C_X^*$, $\|\bar{y}\|_\infty \leq 1$, $+y_{i_0} = 1$, $\bar{g} \geq 0$. Then, it is obvious that $(\bar{y}, \bar{q})$ is feasible for $\bar{f}_{+i_0}(d)$. Also, $b^T\bar{y} \leq b^T\bar{y} + \bar{g} \leq |b^T\bar{y} + \bar{g}|$ and so, $\max\{\|A^T\bar{y} - \bar{q}\|, b^T\bar{y}\} \leq f_{+i_0}(d)$. Therefore,

$$
\begin{aligned}
\min_{\pm i} \bar{f}_{\pm i}(d) &\leq \bar{f}_{+i_0}(d) \\
&\leq \max\{\|A^T\bar{y} - \bar{q}\|, b^T\bar{y}\} \\
&\leq f_{+i_0}(d) = \min_{\pm i} f_{\pm i}(d).
\end{aligned}
$$

Therefore $\min_{\pm i} \bar{f}_{\pm i}(d) \leq \min_{\pm i} f_{\pm i}(d)$. Next, let i_1 be such that $\min_{\pm i} \bar{f}_{\pm i}(d) = \bar{f}_{+i_1}(d)$, and let $(\bar{y}, \bar{q})$ be an optimal solution of $\bar{F}_{+i_1}(d)$, that is, $\bar{f}_{+i_1}(d) = \max\{\|A^T\bar{y} - \bar{q}\|, b^T\bar{y}\}$. Let

$$
\bar{g} = \begin{cases} -b^T\bar{y} & \text{if } b^T\bar{y} < 0, \\ 0 & \text{if } b^T\bar{y} \geq 0. \end{cases}
$$

Then $\max\{\|A^T\bar{y} - \bar{q}\|, |b^T\bar{y} + \bar{g}|\} = \max\{\|A^T\bar{y} - \bar{q}\|, b^T\bar{y}\}$. Now $\bar{y}_{+i_1} = 1$, and so $\|\bar{y}\|_\infty \geq 1$. Let $(\tilde{y}, \tilde{q}, \tilde{g}) = (\bar{y}, \bar{q}, \bar{g})/\|\bar{y}\|_\infty$. Then $\|\tilde{y}\|_\infty = 1$ and $\tilde{y}_j = \pm 1$ for some j. Therefore, $(\tilde{y}, \tilde{q}, \tilde{g})$ is feasible for $F_{+j}(d)$. (Without loss of generality, assume that $\tilde{y}_j = 1$.) Then

$$
\begin{aligned}
f_{+j}(d) &\leq \max\{\|A^T\tilde{y} - \tilde{q}\|, |b^T\tilde{y} + \tilde{g}|\} \\
&= \frac{1}{\|\bar{y}\|_\infty} \max\{\|A^T\bar{y} - \bar{q}\|, |b^T\bar{y} + \bar{g}|\} \\
&\leq \max\{\|A^T\bar{y} - \bar{q}\|, |b^T\bar{y} + \bar{g}|\} \\
&= \max\{\|A^T\bar{y} - \bar{q}\|, b^T\bar{y}\} = \bar{f}_{+i_1}(d).
\end{aligned}
$$

We conclude that $\min_{\pm i} f_{\pm i}(d) \leq f_{+j}(d) \leq \bar{f}_{+i_1}(d) = \min_{\pm i} \bar{f}_{\pm i}(d)$, completing the proof. □

5. On the complexity of computing a 2-estimate of $\rho(d)$. In this section, we discuss the complexity of computing a 2-estimate of the distance to ill-posedness $\rho(d)$ for a given data instance d of $P(d)$ under the choice of norms given in (4). This estimate of $\rho(d)$ can then be used to estimate the condition number $C(d)$, provided that $\|d\|$ can be estimated as well. We now discuss this last point briefly.

We assume throughout this section that an upper bound $\bar{\delta}$ on $\|d\|$ is known and given. Given the choice of norms specified in (4), one way to conveniently obtain an upper bound on $\|d\|$ is to simply compute

$$\bar{\delta} := \sqrt{n}\max\{\|A_{\cdot 1}\|_1, \ldots, \|A_{\cdot n}\|_1, \|b\|_1\}, \tag{11}$$

if $d = (A, b)$ is given as a real matrix and a real vector, respectively. (Here, $\|A_{\cdot j}\|_1$ denotes the L_1 norm of the jth column of A.) In this case, it is elementary to show that $(1/\sqrt{n})\bar{\delta} \le \|d\| \le \bar{\delta}$. With these estimates of $\rho(d)$ and of $\|d\|$, the final estimate for $C(d)$ would be $\widehat{C} = \bar{\delta}/\hat{\rho}$. If we use the above estimate for $\|d\|$, it is easy to prove that $\widehat{C}/\sqrt{n} \le C(d) \le 2\widehat{C}$.

The choice of algorithm and the complexity analysis of the algorithm for computing an estimate of $\rho(d)$ will depend on how the cone C_X is described. If C_X is described as the closure of the domain of a self-concordant barrier function, then we will compute an estimate of $\rho(d)$ using a suitably constructed interior-point algorithm. Our algorithm and its analysis in this case is presented in §5.1. If, on the other hand, C_X is described via a separation oracle, then we will compute an estimate of $\rho(d)$ using the ellipsoid algorithm, the analysis of which is presented in §5.2.

5.1. Estimation of $\rho(d)$ using a self-concordant barrier function, via an interior-point algorithm. In this section, we develop an interior-point algorithm called algorithm INT-EST, to compute a 2-estimate of $\rho(d)$. The algorithm works by using an interior-point algorithm to approximately solve the $2m$ convex optimization problems $\bar{F}_{\pm i}(d)$ to obtain an upper bound on $\rho(d)$; see Proposition 4.1. Our approach is based on the barrier method for solving a convex optimization problem using a self-concordant barrier function, as articulated in Renegar (1995b), based on the theory of self-concordant functions of Nesterov and Nemirovskii (1994). The barrier method is designed to approximately solve a problem of the form

$$P\colon\ \hat{z} = \min\{c^T x\colon\ x \in S\},$$

where $S \subset \Re^n$ is a compact convex set, and $c \in \Re^n$. The method requires the existence of a self-concordant barrier function $\phi(x)$ for the relative interior of the set S (see Renegar 1995b and Nesterov and Nemirovskii 1994 for details) and proceeds by approximately solving a sequence of problems of the form

$$P_\mu\colon\ \min\{c^T x + \mu\phi(x)\colon\ x \in \operatorname{relint} S\},$$

for a decreasing sequence of values of the barrier parameter μ. Here $\operatorname{relint} S$ denotes the relative interior of the set S. We base our complexity analysis on the general convergence results for the barrier method presented in Renegar (1995b), which are similar to (but are more accessible for our purposes than) related results found in Nesterov and Nemirovskii (1994). The barrier method starts at a given point $x^0 \in \operatorname{relint} S$. The method performs two stages. In Stage I, the method starts from x^0 and computes iterates based on Newton's method, ending when it has computed a point $\hat{x}$ that is an approximate solution of $P_{\hat{\mu}}$ for some penalty parameter $\hat{\mu}$ that is generated internally in Stage I. In Stage II, the barrier method computes a sequence of approximate solutions x^k of P_{μ_k}, again using Newton's method, for a decreasing sequence of penalty parameters μ_k converging to zero. One of the key properties of the iterates in Stage II is that they satisfy:

$$c^T x^k - 2\mu_k\vartheta \le \hat{z} \le c^T x^k, \tag{12}$$

and so the barrier method provides lower and upper bounds on $\hat{z}$ at each iteration of Stage II. Here ϑ is the complexity parameter associated with $\phi(\cdot)$. We mention that the constant "2" above can be replaced by any other suitable absolute constant $\alpha > 1$, depending on the specific implementation of the algorithm.

One description of the complexity of the barrier method is as follows:

- Stage I requires

$$O\left(\sqrt{\vartheta}\ln\left(\vartheta + \frac{1}{\mathrm{sym}(x^0)}\right)\right)$$

iterations, and Stage II requires

$$O\left(\sqrt{\vartheta}\ln\left(\vartheta + \frac{R}{\epsilon}\right)\right)$$

iterations in order to compute an ϵ-optimal solution of P, which is a feasible solution x of P for which $c^T x \le \hat{z} + \epsilon$.

In these expressions, R is the range of the objective function $c^T x$ over the set S, that is, $R = \max\{c^T x\colon x \in S\} - \min\{c^T x\colon x \in S\}$, and $\mathrm{sym}(x)$ is a measure of the "symmetry" of the point x with respect to the set S, and is defined as $\mathrm{sym}(x) = \max\{t\colon y \in S \Rightarrow x - t(y - x) \in S\}$. This term in the complexity of the barrier method arises since the closer the starting point is to the boundary, the larger is the value of the barrier function at this point, and so more effort is generally required to proceed from such a point.

We now return to our problem. Because the analysis of the complexity of the barrier method relies heavily on the feasible region S being a bounded set, rather than applying the barrier method directly to solve problem $\bar{F}_{\pm i}(d)$ (whose feasible region is unbounded), we will instead work with the following modification of problem $\bar{F}_{\pm i}(d)$ whose feasible region is bounded:

$$\tilde{F}_{\pm i}(d)\colon\ \tilde{f}_{\pm i}(d) = \min_{y,q,\gamma}\ \gamma$$

$$\text{s.t.}\ \|A^T y - q\|_2 \le \gamma, \tag{13}$$

$$b^T y \le \gamma, \tag{14}$$

$$\sqrt{y^T y} \le 2\sqrt{m}, \tag{15}$$

$$\gamma \le 7\bar{\delta}, \tag{16}$$

$$q \in C_X^*, \tag{17}$$

$$\pm y_i = 1. \tag{18}$$

The introduction of the quadratic bound on y is convenient for the barrier function we will use in our interior-point algorithm. However, any general bound on the norm of y will also work with a corresponding effect in the complexity estimates. Also, the specific numbers used in the right-hand side of this problem are defined technically to allow easy estimates of the symmetry of an initial point.

We now show that $\tilde{f}_{\pm i}(d)$ can still be used to compute $\rho(d)$:

PROPOSITION 5.1.

$$\min_{\pm i} \tilde{f}_{\pm i}(d) = \min_{\pm i} \bar{f}_{\pm i}(d) = \rho(d).$$

PROOF. Without loss of generality, let i_0 be an index such that $\tilde{f}_{+i_0}(d) = \min_{\pm i} \tilde{f}_{\pm i}(d)$, and let $(\bar{y}, \bar{q}, \bar{\gamma})$ be a point where the optimum is attained. Then, $\|A^T\bar{y} - \bar{q}\|_2 \le \bar{\gamma}$, $b^T\bar{y} \le \bar{\gamma}$, $\|\bar{y}\|_2 \le 2\sqrt{m}$, $\bar{\gamma} \le 7\bar{\delta}$, $+\bar{y}_{i_0} = 1$. Then, $(\bar{y}, \bar{q})$ is obviously feasible for $\bar{F}_{+i_0}(d)$. Notice that $\bar{\gamma} = \max\{\|A^T\bar{y} - \bar{q}\|_2, b^T\bar{y}\}$, which implies that $\bar{f}_{+i_0}(d) \le \bar{\gamma}$, from which it follows that $\min_{\pm i} \bar{f}_{\pm i}(d) \le \min_{\pm i} \tilde{f}_{\pm i}(d)$.

Next, also without loss of generality, let i_1 be an index such that $\bar{f}_{+i_1}(d) = \min_{\pm i} \bar{f}_{\pm i}(d) = \rho(d)$ and let $(\bar{y}, \bar{q})$ be an optimal solution of $\bar{F}_{+i_1}(d)$, and let $\bar{\gamma} = \bar{f}_{+i_1}(d)$. Note that $\|\bar{y}\|_\infty \geq 1$, and let

$$(\tilde{y}, \tilde{q}, \tilde{\gamma}) = \left(\frac{\bar{y}}{\|\bar{y}\|_\infty}, \frac{\bar{q}}{\|\bar{y}\|_\infty}, \frac{\bar{\gamma}}{\|\bar{y}\|_\infty} \right).$$

Therefore, $\|\tilde{y}\|_\infty = 1$ and assume without loss of generality that $\tilde{y}_j = 1$ for some index j.

We next argue that $(\tilde{y}, \tilde{q}, \tilde{\gamma})$ is feasible for $\tilde{F}_{+j}(d)$. To see this, note first that $\tilde{y}_j = 1$, $\sqrt{\tilde{y}^T\tilde{y}} \leq \sqrt{m}\|\tilde{y}\|_\infty = \sqrt{m} \leq 2\sqrt{m}$. Also, since $(\bar{y}, \bar{q})$ is optimal for $\bar{F}_{+i_1}(d)$, then $\bar{\gamma} = \max\{\|A^T\bar{y} - \bar{q}\|_2, b^T\bar{y}\} \leq \max\{\|A^T\bar{y}\|_2, b^T\bar{y}\} \leq \|d\|\|\bar{y}\|_\infty$, and so $\tilde{\gamma} = \bar{\gamma}/\|\bar{y}\|_\infty \leq \|d\| \leq \bar{\delta} \leq 7\bar{\delta}$. Therefore, $(\tilde{y}, \tilde{q}, \tilde{\gamma})$ is feasible for $\tilde{F}_{+j}(d)$. It follows that $\tilde{f}_{+j}(d) \leq \max\{\|A^T\tilde{y} - \tilde{q}\|_2, b^T\tilde{y}\} \leq \max\{\|A^T\bar{y} - \bar{q}\|_2, b^T\bar{y}\} = \bar{\gamma} = \bar{f}_{+i_1}(d)$. We conclude that $\min_{\pm i} \tilde{f}_{\pm i}(d) \leq \tilde{f}_{+j}(d) \leq \bar{f}_{+i_1}(d) \leq \min_{\pm i} \bar{f}_{\pm i}(d)$, completing the proof. □

We now specify the barrier function of the feasible region of problem $\tilde{F}_{\pm i}(d)$, and we analyze its complexity parameter. Let $B^*(\cdot)$ denote the self-concordant barrier function of the cone C_X^*, and let ϑ_* denote the complexity parameter for $B^*(\cdot)$. The barrier function for $\tilde{F}_{\pm i}(d)$ is constructed by simply adding the appropriate barrier functions for each of the constraints of $\tilde{F}_{\pm i}(d)$. Define:

$$\phi(y, q, \gamma) := B^*(q) - \ln(7\bar{\delta} - \gamma) - \ln(4m - y^Ty) - \ln(\gamma - b^Ty) - \ln\left(\gamma^2 - \|A^Ty - q\|_2^2\right).$$

The complexity parameter of each of the first three logarithm terms is 1, and the complexity parameter of the last logarithm term is 2. Therefore, from the barrier calculus of self-concordant functions, the complexity parameter for $\phi(y, q, \gamma)$ is at most $\vartheta := \vartheta_* + 5 = O(\vartheta_*)$. We next specify the starting point that will be used by the barrier method to approximately solve $\tilde{F}_{\pm i}(d)$. Let u^0 be a point in the interior of C_X^* and define

$$\bar{w}_{\pm i} = (\bar{y}, \bar{q}, \bar{\gamma}) = \left(\pm e_i, \frac{2\bar{\delta}}{\|u^0\|_2} u^0, 4\bar{\delta} \right).$$

Let us also define

$$\eta = \operatorname{dist}\left(\frac{1}{\|u^0\|_2} u^0, \partial C_X^* \right) = \frac{\operatorname{dist}(u^0, \partial C_X^*)}{\|u^0\|_2} \tag{19}$$

as the ratio of the distance from u^0 to the boundary of the cone C_X^* to the norm of u^0.

Let $\mathcal{D}_{\pm i}$ denote the feasible region of $\tilde{F}_{\pm i}(d)$. We will show later in this subsection that $\bar{w}_{\pm i}$ is, in fact, in the relative interior of the feasible region $\mathcal{D}_{\pm i}$.

We now formally state the algorithm EST-INT for computing a 2-estimate of $\rho(d)$ using an interior-point algorithm.

Algorithm EST-INT$(A, b, \bar{\delta}, u^0)$**.**

- **For** $\pm i = 1, \ldots, m$ **do**

Step 1. Apply Stage I of the barrier method to problem $\tilde{F}_{\pm i}(d)$, using the starting point $\bar{w}_{\pm i} = (\bar{y}, \bar{q}, \bar{\gamma}) = (\pm e_i, (2\bar{\delta}/\|u^0\|_2)u^0, 4\bar{\delta})$.

Step 2. Apply Stage II of the barrier algorithm to problem $\tilde{F}_{\pm i}(d)$, generating the sequence $\{w_{\pm i}^j = (y_{\pm i}^j, q_{\pm i}^j, \gamma_{\pm i}^j)\}_j$. **Stop at iteration** j **if**

$$4\mu_j\vartheta \leq \gamma_{\pm i}^j. \tag{20}$$

Let $\hat{w} = (\hat{y}_{\pm i}, \hat{q}_{\pm i}, \hat{\gamma}_{\pm i})$ denote the final iterate.

- Let $\hat{\rho} = \min_{\pm i}\{\hat{\gamma}_{\pm i}\}$.

The next theorem establishes the validity of algorithm EST-INT (in part (i)) and provides a complexity bound for the algorithm (in part (ii)).

THEOREM 5.1. *Suppose that d is a feasible instance of* (1) *and that the norms on X and Y are chosen via* (4). *Then*

(i) *The value $\hat{\rho}$ produced by algorithm EST-INT will satisfy*:

$$\frac{\hat{\rho}}{2} \le \rho(d) \le \hat{\rho}.$$

(ii) *Algorithm EST-INT will terminate in*

$$O\left(m\sqrt{\vartheta_*}\,\ln\left(\vartheta_* + m + \frac{\|u^0\|_2}{\mathrm{dist}_2(u^0, \partial C_X^*)} + \frac{\bar{\delta}}{\|d\|} + C(d)\right)\right)$$

iterations of Newton steps.

Notice that Theorem 5.1 states that the complexity of computing a 2-estimate of $\rho(d)$ is linear in $\ln(C(d))$. It has been shown in Renegar (1995b) that computing a feasible solution of $P(d)$ using the barrier method requires

$$O\left(\sqrt{\vartheta}\,\ln\left(\vartheta + \frac{\|x^0\|_2}{\mathrm{dist}(x^0, \partial C_X)} + C(d)\right)\right)$$

iterations, where the "ϑ" in this expression is the complexity parameter for a self-concordant barrier function for the cone C_X, and where x^0 is a starting point for the barrier method that satisfies $x^0 \in \operatorname{int} C_X$. Notice that both complexity bounds have the same sort of dependence on the complexity parameter for the respective barriers, and in fact from the theory of self-concordance we know that we can substitute ϑ_* for ϑ in the above expressions.

However, notice that the complexity bound for computing an estimate of $\rho(d)$ involves extra terms involving m and $\ln m$. We now partially explain where these two terms come from, and why we do not think that these terms can be eliminated through a different or more careful analysis. Recall from the discussion in §4 that problem $P_r(d)$ can be interpreted as finding the largest L_1 ball in $\Re^m$ centered at the origin and contained in the set $\mathscr{H}_d := \{br - Ax\colon\ r \ge 0, x \in C_X, r + \|x\|_2 \le 1\}$, and so the computation of $\rho(d)$ is accomplished by checking how large the $2m$ extreme points of the unit L_1 ball can be scaled and still lie in $\mathscr{H}_d$. We do not think that the "m" part of the extra operation count in Theorem 5.1 can be eliminated, as it arises precisely from the necessity of checking the $2m$ extreme points of the L_1 ball. The "$\ln m$" part of the extra operation count arises in the estimate of the symmetry of the starting point $\bar{w}$ in the feasible region of $\tilde{F}_{\pm i}(d)$, which in turn arises from the constraint (15). The use of the term "$\sqrt{m}$" in the right-hand side of (15) arises from converting between the L_∞ norm and the L_2 norm for Y. This conversion would be unnecessary if we replaced Constraint (15) with the constraint "$\|y\|_\infty \le 2$," but then the complexity parameter ϑ would increase by the factor m.

Observe also that the complexity bound in Theorem 5.1 is affected by the quality of the starting point u^0 chosen in the interior of the cone C_X^*. This is important as, in fact, for some specific cones, we know particular points for which the quantity $\eta^{-1} = \|u^0\|_2/\mathrm{dist}(u^0, \partial C_X^*)$ is nicely bounded from below. Using the definition of the "width" and norm approximation vector of a cone, from §2, it is straightforward to show that if $X = \Re^n$ with Euclidean norm $\|x\| = \|x\|_2 = \sqrt{x^T x}$, and $C_X = C_X^* = \Re^n_+ = \{x \in \Re^n \mid x \ge 0\}$, then by setting $u^0 = e = (1, \ldots, 1)^T$ we obtain a value of $\eta^{-1} = \sqrt{n}$. In the case of the positive semidefinite cone of real $k \times k$ symmetric matrices with Frobenius norm $\|x\| := \sqrt{\mathrm{trace}(x^T x)}$, it is easy to show by setting $u^0 = I$ that $\eta^{-1} = \sqrt{k}$.

Notice that the assertions of Theorem 5.1 are valid in the case when d is an ill-posed feasible instance, i.e., when $d \in \mathscr{F}$ but $\rho(d) = 0$. In this case, the optimal value of one of the problems $\tilde{F}_{\pm i}(d)$ will be equal to zero, and while the sequence of iterates generated by the algorithm will converge to the optimal value of zero, the stopping criteria might never be satisfied and so the algorithm might not terminate. However, even in this case, the complexity bound is vacuously valid since $C(d) = \infty$, although it will not be possible to provide a guaranteed estimate of the distance to ill-posedness in this case. This is, of course, the very primary effect of being ill-posed.

We now proceed to prove Theorem 5.1. We start with the following proposition, which shows that the proposed starting point of the barrier method is in the relative interior of $\mathscr{D}_{\pm i}$.

PROPOSITION 5.2. $\bar{w}_{\pm i} \in \operatorname{relint} \mathscr{D}_{\pm i}$.

PROOF. Notice that $\pm\bar{y}_i = 1$ and $\bar{q} \in \operatorname{int} C_X^*$. This means that (17) and (18) are satisfied. We now verify that the other constraints of the problem are satisfied.

For (13), we have $\|A^T\bar{y} - \bar{q}\|_2 \leq \|A^T\bar{y}\|_2 + \|\bar{q}\|_2 \leq \bar{\delta} + 2\bar{\delta} = 3\bar{\delta} < 4\bar{\delta} = \bar{\gamma}$, where recall that $\bar{\delta}$ is the estimate for the norm of d; see the discussion regarding Inequality (11). To verify (14), we have that $b^T\bar{y} = \pm b^T e_i \leq \bar{\delta} < 4\bar{\delta} = \bar{\gamma}$. For (16) just observe that $\bar{\gamma} = 4\bar{\delta} < 7\bar{\delta}$. Finally, for (15) we have that $\sqrt{\bar{y}^T\bar{y}} = 1 < 2\sqrt{m}$, which completes the proof. □

The next lemma establishes that when the stopping criterion of algorithm EST-INT is satisfied in Step 2, an appropriate approximation to $\tilde{f}_{\pm i}(d)$ is obtained.

LEMMA 5.1. *Let j be the iteration index when the stopping criterion is satisfied. Then $\gamma^j_{\pm i}/2 \leq \tilde{f}_{\pm i}(d) \leq \gamma^j_{\pm i}$.*

PROOF. Suppose that the stopping criterion in Step 2 is satisfied. Then $\gamma^j_{\pm i} \geq \tilde{f}_{\pm i}(d) \geq \gamma^j_{\pm i} - 2\mu_j\vartheta \geq \gamma^j_{\pm i} - \frac{1}{2}\gamma^j_{\pm i} = \frac{1}{2}\gamma^j_{\pm i}$. Here the first inequality follows by definition of $\tilde{f}_{\pm i}(d)$, the third inequality follows from the stopping criterion, while the second inequality follows from (12). □

PROOF OF PART (i) OF THEOREM 5.1. This follows immediately from Lemma 5.1, since $\hat{\rho} = \min_{\pm i} \hat{\gamma}_{\pm i}$ and $\rho(d) = \min_{\pm i} \tilde{f}_{\pm i}(d)$. □

The next result establishes the objective function tolerance needed to satisfy the stopping criterion in Step 2 of algorithm EST-INT.

PROPOSITION 5.3. *Let $\epsilon = \rho(d)/2$. Let J be the number of iterations of the barrier method needed to achieve a guaranteed ϵ-optimal solution of $\tilde{F}_{\pm i}(d)$, i.e., a solution for which $\gamma^J_{\pm i} - (\gamma^J_{\pm i} - 2\mu_J\vartheta) \leq \epsilon$. Then the stopping criterion in Step 2 of algorithm EST-INT is satisfied on or before iteration J.*

PROOF. We have that $2\mu_J\vartheta \leq \epsilon = \rho(d)/2 \leq \tilde{f}_{\pm i}(d)/2 \leq \gamma^J_{\pm i}/2$. We conclude that the stopping criterion is satisfied at iteration J or earlier. □

We next demonstrate a lower bound on the symmetry of the starting point $\bar{w}_{\pm i}$.

PROPOSITION 5.4.

$$\operatorname{sym}(\bar{w}_{\pm i}) \geq \frac{\eta}{11 + 2\sqrt{m}}.$$

PROOF. Let (y, q, γ) be such that $(\bar{y}, \bar{q}, \bar{\gamma}) + (y, q, \gamma) \in \mathscr{D}_{\pm i}$. By construction, $\pm y_i = 0$. In order to prove the proposition, we must show that for all values of t satisfying $0 \leq t \leq \eta/(11 + 2\sqrt{m})$, that

$$(\bar{y}, \bar{q}, \bar{\gamma}) - t(y, q, \gamma) \in \mathscr{D}_{\pm i}. \tag{21}$$

The proof proceeds as follows: For each constraint defining $\mathscr{D}_{\pm i}$, we determine an appropriate upper bound on t for which (21) is satisfied. The smallest of these upper bounds provides a lower bound on $\operatorname{sym}(\bar{w})$. First note that $\bar{\gamma} + \gamma \leq 7\bar{\delta}$ implies that $\gamma \leq 7\bar{\delta} - \bar{\gamma} = 3\bar{\delta}$.

(i) For constraint (13), let $t_1 = \frac{1}{13}$. Notice that $\|A^T\bar{y} - \bar{q}\|_2 \leq \|d\| + 2\bar{\delta} \leq 3\bar{\delta}$. Therefore, $\|A^Ty - q\|_2 = \|A^T\bar{y} - \bar{q} + A^Ty - q - (A^T\bar{y} - \bar{q})\|_2 \leq \|A^T(\bar{y} + y) - (\bar{q} + q)\|_2 + \|A^T\bar{y} - \bar{q}\|_2 \leq \bar{\gamma} + \gamma + 3\bar{\delta} \leq 10\bar{\delta}$.

Now, let t satisfy $0 \leq t \leq t_1$. We have $\|A^T(\bar{y} - ty) - (\bar{q} - tq)\|_2 - (\bar{\gamma} - t\gamma) \leq \|A^T\bar{y} - \bar{q}\|_2 + t\|A^Ty - q\|_2 - \bar{\gamma} + t\gamma \leq 3\bar{\delta} + 10t\bar{\delta} - 4\bar{\delta} + 3t\bar{\delta} \leq \bar{\delta}(13t - 1) \leq 0$, and hence, $\|A^T(\bar{y} - ty) - (\bar{q} - t\gamma)\|_2 \leq \bar{\gamma} - t\gamma$.

(ii) For the constraint (14), let $t_2 = 3/(4+2\sqrt{m})$. We have that $b^T(\bar{y}+y) \le \bar{\gamma}+\gamma$. Also, $\|y\|_\infty = \|y+\bar{y}-\bar{y}\|_\infty \le \|y+\bar{y}\|_\infty + \|\bar{y}\|_\infty \le 2\sqrt{m}+1$, where the last inequality follows from the fact that $y+\bar{y}$ is feasible and so $\|y+\bar{y}\|_\infty \le \|y+\bar{y}\|_2 \le 2\sqrt{m}$. Now let t satisfy $0 \le t \le t_2$. We have that $b^T(\bar{y}-ty) - (\bar{\gamma}-t\gamma) = b^T\bar{y} - \bar{\gamma} + t(-b^Ty+\gamma) \le \bar{\delta} - 4\bar{\delta} + t(\bar{\delta}(2\sqrt{m}+1) + 3\bar{\delta}) = -3\bar{\delta} + t\bar{\delta}(2\sqrt{m}+1+3) \le 0$, and hence, $b^T(\bar{y}-ty) \le \bar{\gamma}-t\gamma$.

(iii) For the constraint (15), let $t_3 = (2\sqrt{m}-1)/(2\sqrt{m}+1)$. We have that $\|y\|_2 = \|y+\bar{y}-\bar{y}\|_2 \le \|y+\bar{y}\|_2 + \|\bar{y}\|_2 \le 2\sqrt{m}+1$. Let t satisfy $0 \le t \le t_3$. Then, $\|\bar{y}-ty\|_2 \le \|\bar{y}\|_2 + t\|y\|_2 \le 1 + t(2\sqrt{m}+1) \le 1 + 2\sqrt{m} - 1 = 2\sqrt{m}$, and hence, $\|\bar{y}-ty\|_2 \le 2\sqrt{m}$.

(iv) For the constraint (16), let $t_4 = 3/4$. We have $\bar{\gamma}+\gamma \ge \|A^T(\bar{y}+y) - (\bar{q}+q)\|_2 \ge 0$, and so $\gamma \ge -\bar{\gamma} = -4\bar{\delta}$. Let t satisfy $0 \le t \le t_4$. We have that $\bar{\gamma} - t\gamma = 4\bar{\delta} - t\gamma \le 4\bar{\delta} + 4t\bar{\delta} \le 4\bar{\delta} + 3\bar{\delta} = 7\bar{\delta}$, and so satisfies constraint (16).

(v) For the constraint (17), let $t_5 = 2\eta/(9+2\sqrt{m})$. We have that $\bar{q}+q \in C_X^*$. Now, we have that $\|q\|_2 = \|-A^T(\bar{y}+y) + (\bar{q}+q) + A^T(\bar{y}+y) - \bar{q}\|_2 \le \|A^T(\bar{y}+y) - (\bar{q}+q)\|_2 + \|A^T(\bar{y}+y)\|_2 + \|\bar{q}\|_2 \le \bar{\gamma}+\gamma+\bar{\delta}\|\bar{y}+y\|_\infty + 2\bar{\delta} \le 7\bar{\delta} + 2\sqrt{m}\bar{\delta} + 2\bar{\delta} = (9+2\sqrt{m})\bar{\delta}$. Now, we have

$$\bar{q} - t_5 q = \bar{q} - \frac{2\eta}{9+2\sqrt{m}} q = \frac{2\bar{\delta}}{\|u^0\|_2}\left(u^0 - \frac{\operatorname{dist}(u^0, \partial C_X^*)}{\bar{\delta}(9+2\sqrt{m})} q\right).$$

But

$$\left\| \frac{\operatorname{dist}(u^0, \partial C_X^*)}{\bar{\delta}(9+2\sqrt{m})} q \right\|_2 = \frac{\|q\|_2 \operatorname{dist}(u^0, \partial C_X^*)}{\bar{\delta}(9+2\sqrt{m})} \le \operatorname{dist}(u^0, \partial C_X^*),$$

and hence, $u^0 - (\operatorname{dist}(u^0, \partial C_X^*)/\bar{\delta}(9+2\sqrt{m}))q \in C_X^*$. We conclude that $\bar{q} - t_5 q \in C_X^*$, and so $\bar{q} - tq \in C_X$ for any t satisfying $0 \le t \le t_5$.

As a consequence of all cases, we see that $\operatorname{sym}(\bar{w}_{\pm i}) \ge \min\{t_1, t_2, t_3, t_4, t_5\} \ge \eta/(11+2\sqrt{m})$, proving the result. □

The next result, which is evident, will also be used in the proof of the main theorem.

LEMMA 5.2. *If* $a, b \ge 1$ *then* $\frac{1}{2}(\ln a + \ln b) \le \ln(a+b) \le \ln 2 + (\ln a + \ln b)$.

PROOF OF PART (ii) OF THEOREM 5.1. From the discussion of the barrier method, the total number of iterations will be bounded by

$$2m\left[O\left(\sqrt{\vartheta}\ln\left(\vartheta + \frac{1}{\operatorname{sym}(\bar{w}_{\pm i})}\right)\right) + O\left(\sqrt{\vartheta}\ln\left(\vartheta + \frac{R}{\epsilon}\right)\right)\right],$$

where the "$2m$" comes from the fact that the algorithm approximately solves the $2m$ problems $\tilde{F}_{\pm i}(d)$, $\pm i = 1, \dots, m$. Now, $\vartheta = \vartheta_* + 5 = O(\vartheta_*)$. Also R, which is the range of the objective function of $\tilde{F}_{\pm i}(d)$, satisfies $R \le 7\bar{\delta}$, since $0 \le \gamma \le 7\bar{\delta}$ for all feasible solutions of $\tilde{F}_{\pm i}(d)$. Also, from Proposition 5.3, we can bound ϵ from below by $\rho(d)/2$. Finally, from Proposition 5.4, we can bound $\operatorname{sym}(\bar{w}_{\pm i})$ from below by $\eta/(11+2\sqrt{m})$. We then obtain a total iteration bound of

$$O\left(m\sqrt{\vartheta_*}\ln\left(\vartheta_* + \frac{11+2\sqrt{m}}{\eta} + \frac{7\bar{\delta}}{\frac{1}{2}\rho(d)}\right)\right),$$

which is

$$O\left(m\sqrt{\vartheta_*}\ln\left(\vartheta_* + m + \frac{\|u^0\|_2}{\operatorname{dist}(u^0, \partial C_X^*)} + \frac{\bar{\delta}}{\|d\|} + C(d)\right)\right),$$

where we made use of Lemma 5.2, the definition of η in (19), and the fact that $\ln\sqrt{m} = O(\ln m)$ to obtain the final expression. □

5.2. Estimation of $\rho(d)$ using a separation oracle, via the ellipsoid algorithm. In this section, we develop a version of the ellipsoid algorithm, called algorithm EST-ELL, to compute a 2-estimate of $\rho(d)$. As in the development of the interior-point algorithm in the previous subsection, we develop and analyze the algorithm in this subsection under the choice of norms given in (4). The algorithm works by using the ellipsoid algorithm to approximately solve the $2m$ convex optimization problems $F_{\pm i}(d)$ to obtain an upper bound $\bar{\rho}$ on $\rho(d)$. However, unlike an interior-point algorithm, the ellipsoid algorithm does not furnish lower bounds on objective function values that have desirable convergence or complexity properties. Therefore, in order to generate a lower bound on $\rho(d)$, algorithm EST-ELL also uses the ellipsoid algorithm to approximately solve the $2m$ convex optimization problems $S_{\pm i}(d)$ to obtain a lower bound $\underline{\rho}$ on $\rho(d)$.

Our approach is based on the optimization version of the ellipsoid algorithm, originally developed by Yudin and Nemirovskii (1976). We refer the reader to Grötschel et al. (1988) for an expository presentation. The ellipsoid algorithm is designed to approximately solve a problem of the form

$$(P) \qquad z^* = \min_x \{f(x) \colon x \in S\},$$

where S is a convex set in a k-dimensional space X, and $f(x)$ is a quasi-convex function on S. The algorithm requires a separation oracle for the set S in order to detect infeasibility and to perform feasibility cuts. The algorithm also requires a support oracle for the (lower) level sets L_α of $f(\cdot)$ (where $L_\alpha = \{x \in S\colon f(x) \le \alpha\}$), in order to perform optimality cuts. Let $S_\epsilon := \{x \in S\colon f(x) \le z^* + \epsilon\}$ denote the set of ϵ-optimal solutions of (P), and suppose that we are interested in using the ellipsoid algorithm to compute an ϵ-optimal solution of (P), i.e., to compute a point $x \in S_\epsilon$. In order to start the algorithm, we require a known ellipsoid,

$$E_{Q, x^0, R} := \{x \in X\colon (x - x^0)^T Q (x - x^0) \le R^2\},$$

with the property that $E_{Q, x^0, R} \cap S_\epsilon \neq \varnothing$. We point out that the information inputs needed to start the ellipsoid algorithm are the triplet (x^0, Q, R). One description of the complexity of the ellipsoid algorithm is as follows:

- Suppose that there exists $\hat{x}$ and $r > 0$ with the property

$$E_{Q, \hat{x}, r} \subset (E_{Q, x^0, R} \cap S_\epsilon);$$

that is, there exists a scaled and translated version of $E_{Q, x^0, R}$ that is contained in $E_{Q, x^0, R}$ and that it is also contained in the set of ϵ-optimal solutions. Then the ellipsoid algorithm will compute a point $x \in S_\epsilon$ in at most

$$(22) \qquad O\left(k^2 \ln\left(\frac{R}{r}\right)\right)$$

iterations. Each iteration must perform either a feasibility cut or an optimality cut. In addition, each iteration also requires $O(k^2)$ operations to update the iterate representation of the ellipsoid.

We point out that the above complexity bound is by no means the most general result for the ellipsoid algorithm, but it is sufficient for our purposes. For further results on the ellipsoid algorithm, we recommend Grötschel et al. (1988).

In order to compute upper and lower bounds $\bar{\rho}$ and $\underline{\rho}$ on $\rho(d)$, we will use the ellipsoid algorithm to approximately solve both $F_{\pm i}(d)$ and $S_{\pm i}(d)$, respectively, for $\pm i = 1, \ldots, m$. However, it will be more convenient for our purposes to instead solve the following modified version of $F_{\pm i}(d)$:

$$\begin{aligned} \hat{F}_{\pm i}(d)\colon \hat{f}_{\pm i}(d) = \min_{y, q} \ & h(y, q) := \max\{\|A^T y - q\|_2, b^T y\} \\ \text{s.t.} \ & q \in C_X^*, \ \|q\|_2 \le \bar{\delta}, \qquad \pm y_i = 1, \qquad \|y\|_\infty \le 1, \end{aligned}$$

where we recall that $\bar{\delta}$ is a known upper bound on $\|d\|$. The following result can be proved easily.

PROPOSITION 5.5. *$\hat{f}_{\pm i}(d) = f_{\pm i}(d)$, and so $\rho(d) = \min_{\pm i} f_{\pm i}(d) = \min_{\pm i} \hat{f}_{\pm i}(d)$.*

In order to apply the ellipsoid algorithm to approximately solve $\hat{F}_{\pm i}(d)$ and $S_{\pm i}(d)$, we need to specify starting ellipsoids for each problem. For problem $\hat{F}_{\pm i}(d)$, we will start the ellipsoid algorithm using $\hat{E}^{\text{start}}_{\pm i} = \{(y, q)\colon \pm y_i = 1, \|(y, q) - (\pm e_i, 0)\|' \le \sqrt{2}\}$, where for notational convenience, we define the norm $\|v\|' = \|(y, q)\|'$ to be:

$$\|v\|' = \|(y, q)\|' = \sqrt{\frac{y^T y}{m} + \frac{q^T q}{\bar{\delta}^2}}.$$

Note that $\hat{E}^{\text{start}}_{\pm i}$ is an ellipsoid in the affine space $\{(y, q)\colon \pm y_i = 1\}$. For problem $S_{\pm i}(d)$, we will start the ellipsoid algorithm using $E^{\text{start}}_{\pm i} = \{(r, x, \theta)\colon br - Ax \pm e_i\theta = 0, \|(r, x, \theta)\|'' \le 2\}$, where again for notational convenience, we define the norm $\|w\|'' = \|(r, x, \theta)\|''$ to be

$$\|w\|'' = \|(r, x, \theta)\|'' := \sqrt{r^2 + x^T x + \left(\frac{\theta}{\bar{\delta}}\right)^2}.$$

Note that $E^{\text{start}}_{\pm i}$ is an ellipsoid in the subspace $\{(r, x, \theta)\colon br - Ax \pm e_i\theta = 0\}$.

We now formally state the algorithm EST-ELL for computing an α-estimate of $\rho(d)$ using the ellipsoid algorithm. Recall the notation $h(y, q) := \max\{\|A^T y - q\|_2, b^T y\}$, which is the objective function of problem $\hat{F}_{\pm i}(d)$.

Algorithm EST-ELL$(A, b, \bar{\delta})$**.**

- **For** $\pm i = 1, \ldots, m$ **do**
  - Initiate the ellipsoid algorithm for problem $\hat{F}_{\pm i}(d)$ with the ellipsoid $\hat{E}^{\text{start}}_{\pm i}$ in the affine space $\{(y, q)\colon \pm y_i = 1\}$.
  - Initiate the ellipsoid algorithm for problem $S_{\pm i}(d)$ with the ellipsoid $E^{\text{start}}_{\pm i}$ in the subspace $\{(r, x, \theta)\colon br - Ax \pm e_i\theta = 0\}$.
  - **Set** $j = 1$
  - **Iteration** j
    - **Step 1.** Compute the next iterate of the ellipsoid algorithm for $\hat{F}_{\pm i}(d)$. Let $(y^j_{\pm i}, q^j_{\pm i})$ denote the center-point of the new ellipsoid.
    - **Step 2.** Compute the next iterate of the ellipsoid algorithm for $S_{\pm i}(d)$. Let $(r^j_{\pm i}, x^j_{\pm i}, \theta^j_{\pm i})$ denote the center-point of the new ellipsoid.
    - **Step 3.** If both $(y^j_{\pm i}, q^j_{\pm i})$ and $(r^j_{\pm i}, x^j_{\pm i}, \theta^j_{\pm i})$ are feasible for their respective problems, then stop if

      $$2\theta^j \ge h\big(y^j_{\pm i}, q^j_{\pm i}\big). \tag{23}$$

      Otherwise, set $j \leftarrow j + 1$ and go to **Step 1**.
  - Let $(\hat{y}_{\pm i}, \hat{q}_{\pm i})$ and $(\hat{r}_{\pm i}, \hat{x}_{\pm i}, \hat{\theta}_{\pm i})$ denote the final iteration values of $\hat{F}_{\pm i}(d)$ and $S_{\pm i}(d)$, respectively.
- Set $\hat{\rho} = \min_{\pm i}\{h(\hat{y}_{\pm i}, \hat{q}_{\pm i})\}$.

The next theorem establishes the validity of algorithm EST-ELL (in part (i)) and provides a complexity bound for the algorithm (in part (ii)) under the choice of norms specified in (4).

THEOREM 5.2. *Suppose that d is a feasible instance of* (1) *and that the norms on X and Y are chosen via* (4). *Then*:

(i) *The value $\hat{\rho}$ produced by algorithm EST-ELL will satisfy*:

$$\frac{\hat{\rho}}{2} \le \rho(d) \le \hat{\rho}.$$

(ii) *Algorithm EST-ELL will terminate in*

$$O\left(m(m+n)^2 \ln\left(m + \frac{1}{\tau} + \frac{1}{\tau^*} + \frac{\bar{\delta}}{\|d\|} + C(d)\right)\right)$$

iterations of the ellipsoid algorithm, where τ *and* τ^* *are the width parameters of the cones* C_X *and* C_X^*, *respectively.*

Note that just like Theorem 5.1, the complexity bound in Theorem 5.2 for computing a 2-estimate of $\rho(d)$ is linear in $\ln(C(d))$. It has been shown in Freund and Vera (2000a) that computing a feasible solution of $P(d)$ using the ellipsoid algorithm requires

$$O\left((n-m)^2 \ln\left(\frac{1}{\tau} + C(d)\right)\right)$$

iterations. Again we see that, with respect to the dependency on $C(d)$, the two complexity bounds are in accord.

The work per iteration of algorithm EST-ELL can easily be estimated. For problem $\widehat{F}_{\pm i}(d)$, updating the representations of the ellipsoids takes $O((n+m)^2)$ operations, since the dimension of the problem is $k = n+m-1$. Feasibility cuts require $O(m+n+K^*)$ operations, where m arises from checking $\|y\|_\infty \le 1$, n arises from checking $\|q\|_2 \le \bar{\delta}$, and K^* is the number of operations required by the separation oracle for C_X^*. Optimality cuts require $O(mn)$ operations, since the vector $A^T y$ must be computed. For problem $S_{\pm i}(d)$, updating the representations of the ellipsoids takes $O((n-m)^2)$ operations (since the algorithm is executed in the space $\{(r, x, \theta) \in \Re \times \Re^n \times \Re\colon\ br - Ax \pm e_i\theta = 0\}$, whose dimension is $n-m+2$. Feasibility cuts require $O(n+K)$ operations, where K is the number of operations required by the separation oracle for C_X. Notice that separation oracles for both C_X and C_X^* are needed.

Proof of Theorem 5.2, Part (i). From Proposition 5.5 and from stopping criterion (23), we have that $\frac{1}{2}\hat{\rho} = \frac{1}{2}\min_{\pm i}\{h(\hat{y}_{\pm i}, \hat{q}_{\pm i})\} \le \min_{\pm i}\{\hat{\theta}_{\pm i}\} \le \min_{\pm i}\{s_{\pm i}(d)\} = \rho(d) = \min_{\pm i}\{\hat{f}_{\pm i}(d)\} \le \min_{\pm i}\{h(\hat{y}_{\pm i}, \hat{q}_{\pm i})\} = \hat{\rho}$, proving the result. □

Towards proving part (ii) of Theorem 5.2, we proceed as follows. Denote the feasible region of $\widehat{F}_{\pm i}(d)$ by $\mathcal{V}_{\pm i} = \{(y, q)\colon\ \pm y_i = 1, \|y\|_\infty \le 1, q \in C_X^*, \|q\|_2 \le \bar{\delta}\}$, and denote the set of ϵ-optimal solutions of $\widehat{F}_{\pm i}(d)$ by $\mathcal{V}_{\pm i}^{\epsilon} = \{(y, q) \in \mathcal{V}_{\pm i}\colon\ h(y, q) \le \hat{f}_{\pm i}(d) + \epsilon\}$. The following lemma will be used in the proof of part (ii) of Theorem 5.2. We defer the proof to the end of this subsection.

Lemma 5.3. *For* $\pm i = 1, \ldots, m$, *there exists* $(\tilde{y}_{\pm i}, \tilde{q}_{\pm i})$ *and* $\tilde{r}$ *with the property that the ellipsoid*

$$\widehat{E}_{\pm i}^{\text{stop}} = \{(y, q)\colon\ \pm y_i = 1, \|(y, q) - (\tilde{y}_{\pm i}, \tilde{q}_{\pm i})\|' \le \tilde{r}\}$$

satisfies

(1) $\widehat{E}_{\pm i}^{\text{stop}} \subset \widehat{E}_{\pm i}^{\text{start}}$;

(2) $\widehat{E}_{\pm i}^{\text{stop}} \subset \mathcal{V}_{\pm i}^{\rho(d)/4}$;

(3) $\tilde{r} \ge \left(\tau^*/(20C(d)\sqrt{m})\right)(\|d\|/\bar{\delta})$.

Now denote the feasible region of $S_{\pm i}(d)$ by $\mathcal{I}_{\pm i} = \{(r, x, \theta)\colon\ br - Ax \pm e_i\theta = 0,\ r \ge 0,\ x \in C_X,\ |r| + \|x\|_2 \le 1\}$, and denote the set of ϵ-optimal solutions of $S_{\pm i}(d)$ by $\mathcal{I}_{\pm i}^{\epsilon} = \{(r, x, \theta) \in \mathcal{I}_{\pm i}\colon\ \theta \ge s_{\pm i}(d) - \epsilon\}$. We also have the following lemma, which will be used in the proof of part (ii) of Theorem 5.2. Again, we defer the proof to the end of this section.

Lemma 5.4. *For* $\pm i = 1, \ldots, m$, *there exist* $(\tilde{r}_{\pm i}, \tilde{x}_{\pm i}, \tilde{\theta}_{\pm i})$ *and* r' *with the property that the ellipsoid*

$$E_{\pm i}^{\text{stop}} := \{(r, x, \theta)\colon\ br - Ax \pm e_i\theta = 0, \|(r, x, \theta) - (\tilde{r}_{\pm i}, \tilde{x}_{\pm i}, \tilde{\theta}_{\pm i})\|'' \le r'\}$$

satisfies

(1) $E_{\pm i}^{\text{stop}} \subset E_{\pm i}^{\text{start}}$;
(2) $E_{\pm i}^{\text{stop}} \subset \mathcal{F}_{\pm i}^{\rho(d)/4}$;
(3) $r' \geq \tau/(88C(d)^2)$.

PROOF OF THEOREM 5.2, PART (ii). With $\epsilon = \rho(d)/4$, the ellipsoid algorithm will generate an ϵ-optimal solution $(\hat{y}_{\pm i}, \hat{q}_{\pm i})$ of $\hat{F}_{\pm i}(d)$ in

$$O\left((n+m)^2 \ln\left(\frac{20\sqrt{2}C(d)\sqrt{m}\bar{\delta}}{\tau^* \|d\|}\right)\right)$$

iterations, according to Lemma 5.3 and the complexity bound (22) for the ellipsoid algorithm. Similarly, with $\epsilon = \rho(d)/4$ and using Lemma 5.4, the ellipsoid algorithm will generate an ϵ-optimal solution $(\hat{r}_{\pm i}, \hat{x}_{\pm i}, \hat{\theta}_{\pm i})$ of $S_{\pm i}(d)$ in

$$O\left((n-m)^2 \ln\left(\frac{2 \times 88C(d)^2}{\tau}\right)\right)$$

iterations. Therefore, after

$$O\left((n+m)^2 \ln\left(m + \frac{1}{\tau} + \frac{1}{\tau^*} + C(d) + \frac{\bar{\delta}}{\|d\|}\right)\right)$$

sequential iterations of the ellipsoid algorithm for $\hat{F}_{\pm i}(d)$ and $S_{\pm i}(d)$, algorithm EST-ELL will produce iterates $(\hat{y}_{\pm i}, \hat{q}_{\pm i})$ and $(\hat{r}_{\pm i}, \hat{x}_{\pm i}, \hat{\theta}_{\pm i})$ for $\hat{F}_{\pm i}(d)$ and $S_{\pm i}(d)$ that satisfy $\hat{\theta}_{\pm i} + \rho(d)/4 \geq s_{\pm i}(d) = f_{\pm i}(d) = \hat{f}_{\pm i}(d) \geq h(\hat{y}_{\pm i}, \hat{q}_{\pm i}) - \rho(d)/4$ (where the first equality follows from strong duality between $S_{\pm i}(d)$ and $F_{\pm i}(d)$), and so $\hat{\theta}_{\pm i} \geq h(\hat{y}_{\pm i}, \hat{q}_{\pm i}) - \rho(d)/2 \geq h(\hat{y}_{\pm i}, \hat{q}_{\pm i}) - \hat{f}_{\pm i}(d)/2 \geq h(\hat{y}_{\pm i}, \hat{q}_{\pm i}) - h(\hat{y}_{\pm i}, \hat{q}_{\pm i})/2 = h(\hat{y}_{\pm i}, \hat{q}_{\pm i})/2$. Therefore the stopping criterion (23) will be satisfied. Since the algorithm is applied to $2m$ problems, the total iteration bound is

$$O\left(m(n+m)^2 \ln\left(m + \frac{1}{\tau} + \frac{1}{\tau^*} + C(d) + \frac{\bar{\delta}}{\|d\|}\right)\right)$$

iterations. □

PROOF OF LEMMA 5.3. Let $\bar{u} \in C_X^*$ be the norm approximation vector of the cone C_X, as defined in §2. Let $\bar{v} = (\bar{y}, \bar{q}) = (\pm e_i, \frac{1}{2}\bar{\delta}\bar{u})$ and define the ellipsoid

$$\mathcal{E} := \left\{ v = (y, q)\colon\ \pm y_i = 1, \|(y, q) - (\bar{y}, \bar{q})\|' \leq \frac{\tau^*}{2\sqrt{m}} \right\}.$$

We first show that $\mathcal{E} \subset \mathcal{V}_{\pm i}$. Let $v = (y, q) \in \mathcal{E}$. Hence,

$$\|(y, q) - (\bar{y}, \bar{q})\|' = \left\| \left(y - (\pm e_i), q - \frac{1}{2}\bar{\delta}\bar{u} \right) \right\|' \leq \frac{\tau^*}{2\sqrt{m}},$$

from which it follows that

$$\sqrt{\frac{(y - (\pm e_i))^T (y - (\pm e_i))}{m}} \leq \frac{\tau^*}{2\sqrt{m}}, \tag{24}$$

and

$$\left\| q - \frac{1}{2}\bar{\delta}\bar{u} \right\|_2 \leq \frac{\tau^* \bar{\delta}}{2\sqrt{m}}. \tag{25}$$

From (24) and the fact that $\pm y_i = 1$, it follows that $\|y\|_\infty \le 1$. Furthermore,

$$\begin{aligned}\|q\|_2 &= \left\| q - \frac{1}{2}\bar{\delta}\bar{u} + \frac{1}{2}\bar{\delta}\bar{u} \right\|_2 \\ &\le \left\| q - \frac{1}{2}\bar{\delta}\bar{u} \right\|_2 + \frac{1}{2}\bar{\delta} \\ &\le \frac{\tau^*\bar{\delta}}{2\sqrt{m}} + \frac{1}{2}\bar{\delta} \\ &\le \bar{\delta},\end{aligned}$$

where in the third line above we used (25). Therefore, $\|q\|_2 \le \bar{\delta}$. It remains to prove that $q \in C_X^*$. To prove this, let $x \in C_X$, with $\|x\|_2 = 1$. Then,

$$\begin{aligned}q^T x &= \left(q - \frac{1}{2}\bar{\delta}\bar{u} + \frac{1}{2}\bar{\delta}\bar{u} \right)^T x \\ &= \left(q - \frac{1}{2}\bar{\delta}\bar{u} \right)^T x + \frac{1}{2}\bar{\delta}\bar{u}^T x \\ &\ge \left(q - \frac{1}{2}\bar{\delta}\bar{u} \right)^T x + \frac{1}{2}\tau^*\bar{\delta},\end{aligned}$$

because $\bar{u}^T x \ge \tau^*$. Next, notice that

$$\left(q - \frac{1}{2}\bar{\delta}\bar{u} \right)^T x \ge -\left\| q - \frac{1}{2}\bar{\delta}\bar{u} \right\|_2 \|x\|_2 \ge -\frac{\tau^*\bar{\delta}}{2\sqrt{m}},$$

and hence, $q^T x \ge -\tau^*\bar{\delta}/(2\sqrt{m}) + \tau^*\bar{\delta}/2 \ge 0$, which implies that $q \in C_X^*$. This implies that $(y, q) \in \mathcal{V}_{\pm i}$, as we wanted to prove.

Next, let $v^* = (y^*, q^*)$ be an optimal solution of $\hat{F}_{\pm i}(d)$, and let $r' = \tau^*/(2\sqrt{m})$. Define the following (ellipsoidal) ball: $B'(v', \gamma) := \{v = (y, q)\colon \pm y_i = 1, \|(y, q) - (y', q')\|' \le \gamma\}$ for any $v' = (y', q')$ satisfying $\pm y_i' = 1$. With $\bar{v} = (\bar{y}, \bar{q})$, we have $B'(\bar{v}, r') \subset \mathcal{V}_{\pm i}$, and $B'(v^*, 0) \subset \mathcal{V}_{\pm i}$. Therefore, for $\lambda \in [0, 1]$, we have that $B'(\lambda\bar{v} + (1-\lambda)v^*, \lambda r') \subset \mathcal{V}_{\pm i}$. It is also easy to see that

$$h(\bar{v}) \le \frac{3\bar{\delta}}{2}. \tag{26}$$

We now show that for any $v = (y, q) \in B'(\lambda\bar{v} + (1-\lambda)v^*, \lambda r')$ we have

$$h(v) \le h(v^*) + \lambda\left(\frac{5\bar{\delta}}{2} - \rho(d) \right). \tag{27}$$

In order to prove this inequality, observe that

$$v = \left(\lambda(\pm e_i) + (1-\lambda)y^* + w, \lambda\left(\frac{\bar{\delta}}{2}\right)\bar{u} + (1-\lambda)q^* + s \right),$$

where $\|(w, s)\|' \le \lambda r'$. From this, it follows that $\|w\|_2 \le \lambda\tau^*/2$, and $\|s\|_2 \le \bar{\delta}\lambda\tau^*/(2\sqrt{m})$. We have

$$\begin{aligned}\|A^T y - q\|_2 &= \left\| \lambda\left(A^T(\pm e_i) - \frac{1}{2}\bar{\delta}\bar{u} \right) + (1-\lambda)(A^T y^* - q^*) + A^T w - s \right\|_2 \\ &\le (1-\lambda)\|A^T y^* - q^*\|_2 + \lambda\left\| A^T(\pm e_i) - \frac{1}{2}\bar{\delta}\bar{u} \right\|_2 + \|A^T w - s\|_2\end{aligned}$$

$$\begin{aligned}
&\le (1-\lambda)h(v^*) + \lambda h(\bar{v}) + \bar{\delta}\|w\|_2 + \|s\|_2 \\
&\le h(v^*) - \lambda h(v^*) + \lambda \frac{3}{2}\bar{\delta} + \frac{\lambda\tau^*\bar{\delta}}{2} + \frac{\lambda\tau^*\bar{\delta}}{2\sqrt{m}} \\
&\le h(v^*) - \lambda\rho(d) + \lambda\frac{5}{2}\bar{\delta} \\
&= h(v^*) + \lambda\left(\frac{5}{2}\bar{\delta} - \rho(d)\right),
\end{aligned}$$

where in the fourth line we used (26). We also have

$$\begin{aligned}
b^T y &= \lambda b^T(\pm e_i) + (1-\lambda)b^T y^* + b^T w \\
&\le \lambda\bar{\delta} + (1-\lambda)h(v^*) + \bar{\delta}\frac{\lambda\tau^*}{2} \\
&\le h(v^*) + \lambda\left(\frac{3}{2}\bar{\delta} - \rho(d)\right) \\
&\le h(v^*) + \lambda\left(\frac{5}{2}\bar{\delta} - \rho(d)\right).
\end{aligned}$$

Combining both relations, we obtain (27).

Now let $\lambda' = \rho(d)/(10\bar{\delta} - 4\rho(d))$. We see immediately from (27) that if $v \in B'(\lambda'\bar{v} + (1-\lambda')v^*, \lambda' r')$, we have

$$h(v) \le h(v^*) + \frac{\rho(d)}{4}. \tag{28}$$

Finally, consider the set $\mathcal{V}^{\epsilon}_{\pm i}$ with $\epsilon = \rho(d)/4$. Let $\tilde{r} = \lambda' r'$ and $\tilde{v} = \lambda'\bar{v} + (1-\lambda')v^*$. From (27) and (28) we conclude that $\hat{E}^{\text{stop}}_{\pm i} := B'(\tilde{v}, \tilde{r}) \subset \mathcal{V}^{\rho(d)/4}_{\pm i}$. This shows the second part of the lemma.

To prove the third part of the lemma, observe that

$$\tilde{r} = \lambda' r' = \left(\frac{\rho(d)/2}{5\bar{\delta} - 2\rho(d)}\right)\left(\frac{\tau^*}{2\sqrt{m}}\right) \ge \frac{\rho(d)\tau^*}{20\bar{\delta}\sqrt{m}} = \frac{\tau^*}{20C(d)\sqrt{m}}\frac{\|d\|}{\bar{\delta}}.$$

Finally, the first part of the lemma follows from the inclusions: $\hat{E}^{\text{stop}}_{\pm i} \subset \mathcal{V}^{\rho(d)/4}_{\pm i} \subset \mathcal{V}_{\pm i} \subset \hat{E}^{\text{start}}_{\pm i}$. □

PROOF OF LEMMA 5.4. We begin by observing that if $\|(r, x, \theta)\|'' \le 1$, then

$$|r| \le 1, \qquad \|x\|_2 \le 1, \qquad |\theta| \le \bar{\delta} \qquad \text{and} \qquad |r| + \|x\|_2 \le \sqrt{2}. \tag{29}$$

Now, for $w' = (r', x', \theta')$ that satisfies $br' - Ax' \pm e_i\theta' = 0$, let

$$B''(w', \gamma) := \{w = (r, x, \theta) \colon br - Ax \pm e_i\theta = 0, \|w - w'\|'' \le \gamma\}$$

be the (ellipsoidal) ball in the $\|\cdot\|''$ norm centered at w' with radius γ, relative to the linear space defined by the constraint $br - Ax \pm e_i\theta = 0$. From Theorem 5.3 of Freund and Vera (2000b), there exists $\hat{x}$ such that $A\hat{x} = b$, $\hat{x} \in C_X$, and scalars $\hat{r}$ and $\hat{R}$ such that $B(\hat{x}, \hat{r}) \subset C_X$, $\|\hat{x}\|_2 + \hat{r} \le \hat{R}$, and

$$\hat{r} \ge \frac{\tau}{3C(d)}, \qquad \hat{R} \le 4C(d), \qquad \frac{\hat{R}}{\hat{r}} \le \frac{4C(d)}{\tau}. \tag{30}$$

Let

$$\bar{w} = (\bar{r}, \bar{x}, \bar{\theta}) = \left(\frac{1}{2(\|\hat{x}\|_2 + 1)}, \frac{\hat{x}}{2(\|\hat{x}\|_2 + 1)}, 0\right).$$

Notice that $\bar{w}$ is feasible for $S_{\pm i}(d)$. Let $t = \min\{1, \hat{r}\}/(2\sqrt{2}(\|\hat{x}\|_2 + 1))$, and also note that

$$t \le \frac{1}{2(\|\hat{x}\|_2 + 1)}, \qquad t \le \frac{1}{2\sqrt{2}}, \qquad \text{and} \qquad t \le \frac{\hat{r}}{2(\|\hat{x}\|_2 + 1)}. \tag{31}$$

Now define $E := B''(\bar{w}, t)$. We will now show that

$$E \subset \mathcal{I}_{\pm i}. \tag{32}$$

To see why this is true, let $(r, x, \theta) \in E$. Then, $(r, x, \theta) = (\bar{r} + v, \bar{x} + z, \bar{\theta} + p)$, with $\|(v, z, p)\|'' \le t$. From (29), we have that $|v| \le t, \|z\|_2 \le t, |p| \le t\bar{\delta}$, and $|v| + \|z\|_2 \le t\sqrt{2}$. Therefore, $r = \bar{r} + v \ge \bar{r} - |v| \ge 1/(2(\|\hat{x}\|_2 + 1)) - t \ge 0$, by (31). Furthermore, $|r| + \|x\|_2 = \bar{r} + v + \|\bar{x} + z\|_2 \le \bar{r} + |v| + \|z\|_2 + \|\bar{x}\|_2 \le \frac{1}{2} + t\sqrt{2} \le 1$, by (31). Also, $x = \bar{x} + z = 1/(2(\|\hat{x}\|_2 + 1))(\hat{x} + 2(\|\hat{x}\|_2 + 1)z) \in C_X$, since $\|2(\|\hat{x}\|_2 + 1)z\| = 2(\|\hat{x}\|_2 + 1)\|z\|_2 \le 2(\|\hat{x}\|_2 + 1)t \le \hat{r}$, which proves (32).

Next, let $w^* = (r^*, x^*, \theta^*)$ be an optimal solution of $S_{\pm i}(d)$. Then it follows immediately that if $\lambda \in [0, 1]$, then $B''(\lambda\bar{w} + (1 - \lambda)w^*, \lambda t) \subset \mathcal{I}_{\pm i}$.

We now show that if $w = (r, x, \theta) \in B''(\lambda\bar{w} + (1 - \lambda)w^*, \lambda t)$, then

$$\theta \ge \theta^* - \lambda\bar{\delta}(1 + t). \tag{33}$$

To see this, observe that $\theta = \lambda\bar{\theta} + (1 - \lambda)\theta^* + \gamma$, where $|\gamma| \le \lambda t\bar{\delta}$. Hence, $\theta = \theta^* - \lambda\theta^* + \gamma \ge \theta^* - \lambda\bar{\delta} - \lambda t\bar{\delta} = \theta^* - \lambda\bar{\delta}(1 + t)$, which proves (33).

Next, let $\lambda' = \rho(d)/(4\bar{\delta}(1 + t))$. Then it follows from (33) that for all $w = (r, x, \theta) \in B'(\lambda'\bar{w} + (1 - \lambda')w^*, \lambda' t)$ we have

$$\theta \ge \theta^* - \frac{1}{4}\rho(d). \tag{34}$$

It also follows that if $\tilde{w} = \lambda'\bar{w} + (1 - \lambda')w^*$ and $r' = \lambda' t$, we have that $E_{\pm i}^{\text{stop}} := B''(\tilde{w}, r') \subset \mathcal{I}_{\pm i}^{\rho(d)/4}$. This shows the second part of the lemma.

To prove the third part of the lemma, observe that

$$\begin{aligned}
\frac{1}{r'} &= \frac{4\|d\|(1 + t)}{\rho(d)t} \frac{\bar{\delta}}{\|d\|} \\
&= 4C(d)\left(1 + \frac{1}{t}\right)\frac{\bar{\delta}}{\|d\|} \\
&= \left(4C(d) + \frac{4C(d)}{t}\right)\frac{\bar{\delta}}{\|d\|} \\
&= \left(4C(d) + \frac{4C(d)2\sqrt{2}(\|\hat{x}\|_2 + 1)}{\min\{1, \hat{r}\}}\right)\frac{\bar{\delta}}{\|d\|} \\
&\le \left(4C(d) + 12C(d)\left(\frac{4C(d)}{\tau} + \frac{3C(d)}{\tau}\right)\right)\frac{\bar{\delta}}{\|d\|} \\
&\le \frac{88C(d)^2}{\tau}\frac{\bar{\delta}}{\|d\|},
\end{aligned}$$

where in the fifth line we have used the lower bounds from (30). Finally, the first part of the lemma follows from the inclusions: $E_{\pm i}^{\text{stop}} \subset \mathcal{I}_{\pm i}^{\rho(d)/4} \subset \mathcal{I}_{\pm i} \subset E_{\pm i}^{\text{start}}$. This completes the proof of the lemma. □

6. Extensions, relaxing the assumptions, and practical considerations.

6.1. Complexity of computing an arbitrary α-estimate of $\rho(d)$. Under the choice of norms given in (4), Theorems 5.1 and 5.2 present bounds on the complexity of computing an $\alpha = 2$-estimate of $\rho(d)$. By suitably modifying the stopping criteria in the algorithms EST-INT and EST-ELL, one can instead compute an α-estimate for any $\alpha > 1$. It is straightforward to show that the complexity bounds would then be the same as given in Theorems 5.1 and 5.2 except for an additional term $\alpha/(1-\alpha)$ inside the $\ln(\cdot)$ term. In this way, the complexity of computing an $\alpha = (1+\epsilon)$-estimate has an iteration bound whose dependency on ϵ is $O(m\sqrt{\vartheta_*}\ln(1/\epsilon))$ for ϵ small for algorithm EST-INT, for example, which grows only logarithmically in $1/\epsilon$. Note that this complexity bound is consistent with standard notions of efficiency of computation in convex optimization.

6.2. Complexity under different choices of norms. The complexity results obtained herein depended very much on the choice of norms on X and Y given in (4). And as was pointed out in §4, the computation of an α-estimate of $\rho(d)$ is co-NP complete even when $C_X = \Re^n_+$, under a particular choice of norms.

Consider an arbitrary norm $\|\cdot\|^X$ for the space X and an arbitrary norm $\|\cdot\|^Y$ for the space Y. Then because all norms are equivalent in finite-dimensional spaces, there exist constants c_1, c_2, c_3, c_4 (that typically depend monotonically on the dimensions m and n) such that $c_1\|v\|_2 \le \|v\|^X \le c_2\|v\|_2$ for all $v \in X$ and $c_3\|v\|_1 \le \|v\|^Y \le c_4\|v\|_1$ for all $v \in Y$. Then it is obvious that the complexity results we have demonstrated in Theorems 5.1 and 5.2 would still follow if the goal was to compute a $2c_2c_4/(c_1c_3)$-estimate rather than a 2-estimate of $\rho(d)$. This is satisfactory if we know the norm equivalence constants $c_1, \ldots, c_4$ (quite typical in practice) and we are satisfied with the degree of the approximation $2c_2c_4/(c_1c_3)$. However, because of the typical dependence of these constants on m and n, the resulting approximation value might be disappointing for large m and/or n.

Alternatively, let us now consider letting the norm $\|\cdot\|^X$ for the space X be arbitrary, and let the norm $\|\cdot\|^Y$ for the space Y be the convex hull of $2l$ explicitly given extreme points $\pm y^1, \ldots, \pm y^l$. Let us further presume that we have a self-concordant barrier for the interior of the cone $K_{\|\cdot\|^X} := \{(x,t) \colon \|x\|^X \le t\}$ whose complexity value is $\vartheta_{K_{\|\cdot\|^X}}$. Then the methodology developed in §§4 and 5 can be modified to yield an algorithm that will compute a 2-estimate of $\rho(d)$ in

$$O\left(l\sqrt{\vartheta_* + \vartheta_{K_{\|\cdot\|^X}} + l}\ \ln\left(\vartheta_* + \vartheta_{K_{\|\cdot\|^X}} + l + \frac{\|u^0\|}{\operatorname{dist}(u^0, \partial C_X^*)} + \frac{\bar{\delta}}{\|d\|} + C(d)\right)\right)$$

iterations of Newton steps. When $\|\cdot\|^X = \|\cdot\|_2$, then $\vartheta_{K_{\|\cdot\|^X}} = 2$ and other simplifications become possible as well.

6.3. Different formats for $P(d)$. Theorems 5.1 and 5.2 are predicated on the standard primal format of $P(d)$ given in (1). There are usually two other "standard" formats for $P(d)$, namely (i) $b - Ax \in C_Y, x \in X$, where C_Y is a regular cone, and (ii) $b - Ax \in C_Y, x \in C_X$, where C_X and C_Y are each a regular cone. Notice in both of these formats that C_Y is a regular cone. Then under the choice of norms given in (4), Problem (10) can be split into the $2m$ problems:

$$\begin{aligned}\bar{F}_{\pm i}(d) \colon \bar{f}_{\pm i}(d) = \min_{y,q}\ & \max\{\|A^T y - q\|_2, b^T y\}\\ \text{s.t.}\ & y \in C_Y^*,\\ & q \in C_X^*,\\ & \pm y_i = 1.\end{aligned}$$

By Proposition 4.1, $\rho(d) = \min_{\pm i} \bar{f}_{\pm i}(d)$, and an analysis similar to the one we did for $f(d)$ will yield similar complexity results.

6.4. Infeasible instances of $P(d)$. In this case, it is very relevant to know the distance to ill-posedness, since the distance to ill-posedness is also the distance to feasibility. From basic duality theory, it follows that problem $P(d)$ is feasible, or the alternative problem,

(35) $$D(d)\text{: find } y \neq 0 \text{ that solves } A^T y \in C_X^*, \; b^T y \leq 0,$$

has a solution. Notice that $D(d)$ has a format similar to that of $P(d)$ and so is amenable to analysis using the algorithms developed herein. Since knowledge of whether or not $P(d)$ has a solution is usually not given, one can consider an algorithm to estimate the distance to ill-posedness that will process $P(d)$ and $D(d)$ "in parallel" until the corresponding lower bounds on the estimates allow the user to correctly claim one of $P(d)$ and $D(d)$ to be feasible. The value of the corresponding estimate will provide an approximation to the distance to ill-posedness of $P(d)$.

6.5. Practical considerations. Given the potential importance of condition numbers in understanding the behavior of convex optimization, it is consequently important to address both the theoretical complexity of accurately computing condition measures as well as the practical issue of computing condition measures for real problems. In this paper, we have addressed the theoretical complexity, obtaining complexity bounds for computing a 2-estimate of $\rho(d)$ under a suitable choice of norms. Because our algorithms must solve $2m$ convex optimization problems of the same difficulty as the original problem, we would only expect our algorithms to be useful in practice when m is relatively small and/or when there are relatively fast practical algorithms for solving the original problem (such as for linear programming). In contrast, the cited work of Peña (1997) is very promising from a practical point of view. Peña (1997) presents a method for computing an $\sqrt{m}$-estimate of $\rho(d)$ (or $C(d)$) by solving a single convex optimization problem that is an analytic center problem (which is typically fairly easy to solve in practice), under a particular choice of norms (namely the L_2 norms for X and Y). In a sense, the method in Peña (1997) sacrifices some guaranteed accuracy but gains in computability: When m is not too large, such an estimate may be quite sufficient for all practical purposes. This work suggests that it is in fact possible to attain the goal of computing good estimates of the condition measure of a conic system within the context of traditional optimization algorithms, without too much additional computational overhead.

7. Acknowledgments. The authors thank the two anonymous referees and the associate editor for their observations and comments on the original versions of the paper. Their insight and careful reviewing helped us to significantly improve the presentation of our results. They also thank the funding agencies that made international collaboration possible at different stages: The NSF/Fundación Andes International Collaboration between Chile and the United States, CONICYT in Chile (through FONDECYT and FONDAP) and the Singapore-MIT Alliance.

This research has been partially supported through National Science Foundation Grant INT-9703803 and a corresponding grant from Fundación Andes/CONICYT (Chile), in the framework of an International Cooperation Agreement. Support has also been provided through the Singapore-MIT Alliance. Part of this research was conducted while the first author was a visiting scientist at Delft University of Technology, Delft, The Netherlands. The second author was also supported by a FONDECYT project (No. 1980729) and a grant from CONICYT within the FONDAP program in Applied Mathematics.

References

Epelman, M., R. M. Freund. 2000. Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system. *Math. Programming* **88**(3) 451–485.

Filipowski, S. 1997. On the complexity of solving sparse symmetric linear programs specified with approximate data. *Math. Oper. Res.* **22** 769–792.

———. 1999. On the complexity of solving linear programs specified with approximate data and known to be feasible. *SIAM J. Optim.* **9** 1010–1040.

Freund, R. M., J. B. Orlin. 1985. On the complexity of four polyhedral set containment problems. *Math. Programming* **33** 133–145.

———, J. R. Vera. 2000a. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM J. Optim.* **10**(1) 155–176.

———, ———. 2000b. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Math. Programming* **86** 225–260.

Grötschel, M., L. Lovasz, A. Schrijver. 1988. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, Berlin, Germany.

Mangasarian, O. 1987. Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *SIAM J. Control Optim.* **25**(3) 41–87.

Nesterov, Y., A. Nemirovskii. 1994. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Nunez, M. A., R. M. Freund. 1998. Condition measures and properties of the central trajectory of a linear program. *Math. Programming* **83**(1) 1–28.

Peña, J. 1997. Computing the distance to infeasibility: Theoretical and practical issues. Technical report, Cornell University, Center for Applied Mathematics, Ithaca, NY.

Renegar, J. 1994. Some perturbation theory for linear programming. *Math. Programming* **65**(1) 73–91.

———. 1995a. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.* **5**(3) 506–524.

———. 1995b. Linear programming, complexity theory, and elementary functional analysis. *Math. Programming* **70**(3) 279–351.

———. 1996. Condition numbers, the barrier method, and the conjugate gradient method. *SIAM J. Optim.* **64**(4) 879–912.

Vera, J. R. 1992. Ill-posedness in mathematical programming and problem solving with approximate data. Ph.D. thesis, Cornell University, Ithaca, NY.

———. 1996. Ill-posedness and the complexity of deciding existence of solutions to linear programs with approximate data. *SIAM J. Optim.* **6**(3) 549–569.

———. 1998. On the complexity of linear programming under finite precision arithmetic. *Math. Programming* **80**(1) 91–123.

Yudin, D. B., A. S. Nemirovskii. 1976. Informational complexity and efficient methods for solving complex extremal problems. *Ekonomika i Matem. Metody* **12** 357–369.

R. M. Freund: MIT Sloan School of Management, Cambridge, MA 02142-1347; e-mail: rfreund@mit.edu

J. R. Vera: Department of Industrial and System Engineering, Catholic University of Chile, School of Engineering, Campus San Joaquin, Vicuña Mackenna 4860, Santiago, Chile; e-mail: jvera@ing.puc.cl