Condition-Based Complexity of Convex Optimization
in Conic Linear Form Via The Ellipsoid Algorithm

by

R. M. Freund and
J. R. Vera

OR 321-97   September   1997

CONDITION-BASED COMPLEXITY OF CONVEX OPTIMIZATION IN CONIC LINEAR FORM VIA THE ELLIPSOID ALGORITHM [1]

Robert M. Freund[2]
M.I.T.

Jorge R. Vera[3]
University of Chile

September, 1997

Abstract

A convex optimization problem in conic linear form is an optimization problem of the form

$$CP(d): \quad \begin{array}{ll} \text{maximize} & c^T x \\ \text{s.t.} & b - Ax \in C_Y \\ & x \in C_X, \end{array}$$

where C_X and C_Y are closed convex cones in n- and m-dimensional spaces X and Y, respectively, and the data for the system is $d = (A, b, c)$. We show that there is a version of the ellipsoid algorithm that can be applied to find an ϵ-optimal solution of $CP(d)$ in at most $O\left(n^2 \ln\left(\frac{\mathcal{C}(d)\|d\|}{c_1 \epsilon}\right)\right)$ iterations of the ellipsoid algorithm, where each iteration must either perform a separation cut on one of the cones C_X or C_Y, or must perform a related optimality cut. The quantity $\mathcal{C}(d)$ is the "condition number" of the program $CP(d)$ originally developed by Renegar, and essentially is a scale invariant reciprocal of the smallest data perturbation $\Delta d = (\Delta A, \Delta b, \Delta c)$ for which the system $CP(d + \Delta d)$ becomes either infeasible or unbounded. The scalar quantity c_1 is a constant that depends only on particular properties of the cones and the norms used, and is independent of the problem data $d = (A, b, c)$, but may depend on the dimensions m and/or n.

AMS Subject Classification: 90C, 90C05, 90C60
Keywords: Complexity of Convex Optimization, Ellipsoid Method, Conditioning, Error Analysis

[2]MIT O.R. Center, 77 Massachusetts Ave., Cambridge, MA 02139, USA. email: rfreund@mit.edu

[3]Department of Industrial Engineering, University of Chile, República 701, Casilla 2777, Santiago, Chile. e-mail: jvera@dii.uchile.cl

1 Introduction

Consider a convex program in conic linear form:

$$CP(d): \quad \text{maximize} \quad c^T x$$
$$\text{s.t.} \quad b - Ax \in C_Y \qquad (1)$$
$$x \in C_X,$$

where $C_X \subset X$ and $C_Y \subset Y$ are each a closed convex cone in the (finite) n-dimensional linear vector space X (with norm $\|x\|$ for $x \in X$) and in the (finite) m-dimensional linear vector space Y (with norm $\|y\|$ for $y \in Y$), respectively. Here $b \in Y$, and $A \in L(X, Y)$ where $L(X, Y)$ denotes the set of all linear operators $A : X \rightarrow Y$. Also, $c \in X^*$, where X^* is the space of all linear functionals defined on X, i.e., X^* is the dual space of X. In order to maintain consistency with standard linear algebra notation in mathematical programming, we consider c to be a column vector in the space X^* and we 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.

The "data" d for problem $CP(d)$ is the array $d = (A, b, c) \in \{L(X, Y), Y, X^*\}$. We call the above program $CP(d)$ rather than simply CP to emphasize the dependence of the optimization problem on the data $d = (A, b, c)$, and note that the cones C_X and C_Y are not part of the data, that is, they are considered to be given and fixed. At the moment, we make no assumptions on C_X and on C_Y except to note that each is a closed convex cone.

The format of $CP(d)$ is quite general (any convex optimization problem can be cast in the format of $CP(d)$) and has received much attention recently in the context of interior-point algorithms, see Nesterov and Nemirovskii[12] and Renegar [18], [19], as well as Nesterov and Todd [14], [13] and Nesterov, Todd, and Ye [11], among others. In contrast, this paper focuses on the complexity of solving $CP(d)$ via the ellipsoid algorithm. The ellipsoid algorithm of Nemirovskii and Yudin [10] (see also [4], [8], and [9]) and the interior-point algorithm of Nesterov and Nemirovskii [12] are the two fundamental theoretically efficient algorithms for solving general convex optimization. The ellipsoid algorithm enjoys a number of important advantages over interior-point algorithms: the ellipsoid algorithm is based on elegantly simple geometric notions, it always has excellent theoretical efficiency in the dimension of the variables n, it requires only the use of a separation oracle for its implementation, and it is important in both continuous and discrete optimization [8]. When applied to solving linear programming, interior-point algorithms typically exhibit superior practical performance over the ellipsoid algorithm, but that is not the focus of this study.

Using the constructs of Lagrangean duality, one can construct the following dual problem of $CP(d)$:

$$CD(d): \quad \text{minimize} \quad b^T y$$
$$\text{s.t.} \quad A^T y - c \in C_X^* \qquad (2)$$
$$y \in C_Y^*,$$

where C_X^* and C_Y^* are the dual convex cones associated with the cones C_X and C_Y, respectively, and where the dual cone of a convex cone K in a linear vector space X is defined by

$$K^* = \{z \in X^* | z^T x \geq 0 \text{ for any } x \in K\}.$$

The data for the program $CD(d)$ is also the array $d = (A, b, c)$.

Two special cases of $CP(d)$ deserve special mention: linear programming and semi-definite programming. Regarding linear programming, note that when $X = \Re^n$ and $Y = \Re^m$, and either (i) $C_X = \{x \in \Re^n | x \geq 0\}$ and $C_Y = \{y \in \Re^m | y \geq 0\}$, (ii) $C_X = \{x \in \Re^n | x \geq 0\}$ and $C_Y = \{0\} \subset \Re^m$ or (iii), $C_X = \Re^n$ and $C_Y = \{y \in \Re^m | y \geq 0\}$, then $CP(d)$ is a linear program of the format (i) $\max\{c^T x | Ax \leq b, x \geq 0, x \in \Re^n\}$, (ii) $\max\{c^T x | Ax = b, x \geq 0, x \in \Re^n\}$, or (iii) $\max\{c^T x | Ax \leq b, x \in \Re^n\}$, respectively.

The other special case of $CP(d)$ that we mention is semi-definite programming. Semi-definite programming has been shown to be of enormous importance in mathematical programming (see Alizadeh [1] and Nesterov and Nemiroskii [12] as well as Vandenberghe and Boyd [20]). Let X denote the set of real $k \times k$ symmetric matrices, whereby $n = k(k+1)/2$, and define the Löwner partial ordering "$\succeq$" on X as $x \succeq w$ if and only if the matrix $x - w$ is positive semi-definite. The semi-definite program in standard (primal) form is the problem $\max\{c^T x \mid Ax = b, x \succeq 0\}$. Define $C_X = \{x \in X \mid x \succeq 0\}$. Then C_X is a closed convex cone. Let $Y = \Re^m$ and $C_Y = \{0\} \subset \Re^m$. Then the standard form semi-definite program is easily seen to be an instance of $CP(d)$.

Most studies of the ellipsoid algorithm (for example, [9], [4], [8]) pertain to the case when $CP(d)$ is a linear or convex quadratic program, and focus on the complexity of the algorithm in terms of the bit length L of a binary representation of the data $d = (A, b, c)$. However, when the cones C_X and/or C_Y are not polyhedral and/or when the data $d = (A, b, c)$ is not rational, it makes little or no sense to study the complexity of the ellipsoid algorithm in terms of L. Indeed, a much more natural and intuitive measure that is relevant for complexity analysis and that captures the inherent data-dependend behavior of $CP(d)$ is the "condition number" $\mathcal{C}(d)$ of the problem $CP(d)$, which was developed by Renegar in a series of papers [16], [17], and [18]. The quantity $\mathcal{C}(d)$ is essentially a scale invariant reciprocal of the smallest data perturbation $\Delta d = (\Delta A, \Delta b, \Delta c)$ for which the system $CP(d + \Delta d)$ becomes either infeasible or unbounded. (These concepts will be reviewed in detail shortly.)

We show (in Section 4) that there is a version of the ellipsoid algorithm that can be applied to find an ϵ-optimal solution of $CP(d)$ in at most $O\left(n^2 \ln\left(\frac{\mathcal{C}(d)\|d\|}{c_1 \epsilon}\right)\right)$ iterations of the ellipsoid algorithm, where each iteration must either perform a separation cut on one of the cones C_X or C_Y, or must perform a related optimality cut. The quantity $\mathcal{C}(d)$ is the condition number of the program $CP(d)$, and $\|d\|$ is the norm of the data $d = (A, b, c)$. The scalar quantity c_1 is a constant that depends

only on particular properties of the cones and the norms used, and is independent of the problem data $d = (A, b, c)$, but may depend on the dimensions m and/or n.

The paper is organized as follows. The remainder of this introductory section discusses the condition number $\mathcal{C}(d)$ of the optimization problem $CP(d)$. Section 2 contains further notation and a discussion of the "coefficient of linearity" for a cone. Section 3 briefly reviews relevant complexity aspects of the ellipsoid algorithm, and reviews a transformation of $CP(d)$ into a homogenized form called $HP(d)$ that is more convenient for the application of the ellipsoid algorithm. Section 4 contains the complexity results of the ellipsoid algorithm for solving $CP(d)$. Section 5 discusses related issues: testing for ϵ-optimality, the complexity of solving the dual, the complexity of testing for infeasibility of $CP(d)$, and bounding the skewness of the ellipsoids computed in the ellipsoid algorithm. Section 6 contains two technical lemmas that form the basis for the results of Section 4.

The concept of the "distance to ill-posedness" and a closely related condition number for problems such as $CP(d)$ was introduced by Renegar in [16] in a more specific setting, but then generalized more fully in [17] and in [18]. We now describe these two concepts in detail.

We denote the space of all data $d = (A, b, c)$ for $CP(d)$ by $\mathcal{D}$. Then $\mathcal{D} = \{d = (A, b, c) \mid A \in L(X, Y), b \in Y, c \in X^*\}$. Because X and Y are normed linear vector spaces, we can define the following product norm on the data space $\mathcal{D}$:

$$\|d\| = \|(A, b, c)\| = \max\{\|A\|, \|b\|, \|c\|_*\} \text{ for any } d \in \mathcal{D},$$

where $\|A\|$ is the operator norm, namely

$$\|A\| = \max\{\|Ax\| \mid \|x\| \le 1\},$$

and where $\|c\|_*$ is the dual norm of c induced on $c \in X^*$, and is defined as:

$$\|c\|_* = \max\{c^T x \mid \|x\| \le 1, x \in X\},$$

with a similar definition holding for $\|v\|_*$ for $v \in Y^*$.

Consider the following subsets of the data set $\mathcal{D}$:

$$\mathcal{F}_P = \{(A, b, c) \in \mathcal{D} \mid \text{ there exists } x \text{ such that } b - Ax \in C_Y, x \in C_X\},$$

$$\mathcal{F}_D = \{(A, b, c) \in \mathcal{D} \mid \text{ there exists } y \text{ such that } A^T y - c \in C_X^*, y \in C_Y^*\},$$

and

$$\mathcal{F} = \mathcal{F}_P \cap \mathcal{F}_D.$$

The elements in $\mathcal{F}_P$ correspond to those data instances $d = (A, b, c)$ in $\mathcal{D}$ for which $CP(d)$ is feasible and the elements in $\mathcal{F}_D$ correspond to those data instances $d = (A, b, c)$ in $\mathcal{D}$ for which $CD(d)$ is feasible. Observe that $\mathcal{F}$ is the set of data instances $d = (A, b, c)$ that are both primal and dual feasible. The complement of $\mathcal{F}_P$, denoted by $\mathcal{F}_P^C$, is the set of data instances $d = (A, b, c)$ for which $CP(d)$ is infeasible, and the complement of $\mathcal{F}_D$, denoted by $\mathcal{F}_D^C$, is the set of data instances $d = (A, b, c)$ for which $CD(d)$ is infeasible.

The boundary of $\mathcal{F}_P$ and $\mathcal{F}_P^C$ is the set

$$\mathcal{B}_P = \partial \mathcal{F}_P = \partial \mathcal{F}_P^C = \text{cl}(\mathcal{F}_P) \cap \text{cl}(\mathcal{F}_P^C),$$

and the boundary of $\mathcal{F}_D$ and $\mathcal{F}_D^C$ is the set

$$\mathcal{B}_D = \partial \mathcal{F}_D = \partial \mathcal{F}_D^C = \text{cl}(\mathcal{F}_D) \cap \text{cl}(\mathcal{F}_D^C),$$

where ∂S denotes the boundary of a set S, and cl(S) is the closure of a set S. Note that $\mathcal{B}_P \neq \emptyset$ since $(0,0,0) \in \mathcal{B}_P$. The data instances $d = (A, b, c)$ in $\mathcal{B}_P$ are called the ill-posed data instances for the primal, in that arbitrarily small changes in the data $d = (A, b, c)$ can yield data instances in $\mathcal{F}_P$ as well as data instances in $\mathcal{F}_P^C$. Similarly, the data instances $d = (A, b, c)$ in $\mathcal{B}_D$ are called the ill-posed data instances for the dual.

For $d = (A, b, c) \in \mathcal{D}$, we define the ball centered at d with radius δ as:

$$B(d, \delta) = \{\bar{d} \in \mathcal{D} : \|\bar{d} - d\| \leq \delta\}.$$

For a data instance $d \in \mathcal{D}$, the "primal distance to ill-posedness" is defined as follows:

$$\rho_P(d) = \inf\{\|\Delta d\| : d + \Delta d \in \mathcal{B}_P\},$$

see [16, 17, 18], and so $\rho_P(d)$ is the distance of the data instance $d = (A, b, c)$ to the set $\mathcal{B}_P$ of ill-posed instances for the primal problem $CP(d)$. It is straightforward to show that

$$\rho_P(d) = \begin{cases} \sup\{\delta : B(d, \delta) \subset \mathcal{F}_P\} & \text{if } d \in \mathcal{F}_P, \\ \sup\{\delta : B(d, \delta) \subset \mathcal{F}_P^C\} & \text{if } d \in \mathcal{F}_P^C, \end{cases} \tag{3}$$

so that we could also define $\rho_P(d)$ by employing (3). In the typical case when $CP(d)$ is feasible, i.e., $d \in \mathcal{F}_P$, $\rho_P(d)$ is the minimum change Δd in the data d needed to create a primal-infeasible instance $d + \Delta d$, and so $\rho_P(d)$ measures how close the data instance $d = (A, b, c)$ is to the set of infeasible instances of $CP(d)$. Put another way, $\rho_P(d)$ measures how close $CP(d)$ is to being infeasible. Note that $\rho_P(d)$ measures the distance of the data d to primal infeasible instances, and so the objective function vector c plays no role in this measure.

The "primal condition number" $\mathcal{C}_P(d)$ of the data instance d is defined as

$$\mathcal{C}_P(d) = \frac{\|d\|}{\rho_P(d)}$$

when $\rho_P(d) > 0$, and $\mathcal{C}_P(d) = \infty$ when $\rho_P(d) = 0$. The primal condition number $\mathcal{C}_P(d)$ can be viewed as a scale-invariant reciprocal of $\rho_P(d)$, as it is elementary to demonstrate that $\mathcal{C}_P(d) = \mathcal{C}_P(\alpha d)$ for any positive scalar α. Observe that since $\bar{d} = (\bar{A}, \bar{b}, \bar{c}) = (0, 0, 0) \in \mathcal{B}_P$ and $\mathcal{B}_P$ is a closed set, then for any $d \notin \mathcal{B}_P$ we have

$\|d\| \geq \rho_P(d) > 0$, so that $\mathcal{C}_P(d) \geq 1$. The value of $\mathcal{C}_P(d)$ is a measure of the relative conditioning of the primal *feasibility* problem for the data instance d. For a discussion of the relevance of using $\mathcal{C}_P(d)$ as a condition number for the problem $CP(d)$, see Renegar [16], [17], and Vera [21].

These measures are not nearly as intangible as they might seem at first glance. In [7], it is shown that $\rho_P(d)$ can be computed by solving rather simple convex optimization problems involving the data $d = (A, b, c)$, the cones C_X and C_Y, and the norms $\| \cdot \|$ given for the problem. As in traditional condition numbers for systems of linear equations, the computation of $\rho_P(d)$ and hence of $\mathcal{C}_P(d)$ is roughly as difficult as solving $CP(d)$, see [7].

For a data instance $d \in \mathcal{D}$, the "dual distance to ill-posedness" is defined in a matter exactly analogous to the "primal distance to ill-posedness":

$$\rho_D(d) = \inf\{\|\Delta d\| : d + \Delta d \in \mathcal{B}_D\},$$

or equivalently,

$$\rho_D(d) = \begin{cases} \sup\{\delta : B(d, \delta) \subset \mathcal{F}_D\} & \text{if } d \in \mathcal{F}_D, \\ \sup\{\delta : B(d, \delta) \subset \mathcal{F}_D^C\} & \text{if } d \in \mathcal{F}_D^C. \end{cases} \tag{4}$$

The "dual condition number" $\mathcal{C}_D(d)$ of the data instance d is defined as

$$\mathcal{C}_D(d) = \frac{\|d\|}{\rho_D(d)}$$

when $\rho_D(d) > 0$, and $\mathcal{C}_D(d) = \infty$ when $\rho_D(d) = 0$.

The two measures of distances to ill-posed instances and condition numbers are combined as follows. Recalling the definition of $\mathcal{F}$, the elements in $\mathcal{F}$ correspond to those data instances $d = (A, b, c)$ in $\mathcal{D}$ for which both $CP(d)$ and $CD(d)$ are feasible. The complement of $\mathcal{F}$, denoted by $\mathcal{F}^C$, is the set of data instances $d = (A, b, c)$ for which $CP(d)$ is infeasible or $CD(d)$ is infeasible. The boundary of $\mathcal{F}$ and $\mathcal{F}^C$ is the set

$$\mathcal{B} = \partial \mathcal{F} = \partial \mathcal{F}^C = \text{cl}(\mathcal{F}) \cap \text{cl}(\mathcal{F}^C).$$

The data instances $d = (A, b, c)$ in $\mathcal{B}$ are called the ill-posed data instances, in that arbitrarily small changes in the data $d = (A, b, c)$ can yield data instances in $\mathcal{F}$ as well as data instances in $\mathcal{F}^C$. For a data instance $d \in \mathcal{D}$, the "distance to ill-posedness" is defined as follows:

$$\rho(d) = \inf\{\|\Delta d\| : d + \Delta d \in \mathcal{B}\},$$

or equivalently,

$$\rho(d) = \begin{cases} \sup\{\delta : B(d, \delta) \subset \mathcal{F}\} & \text{if } d \in \mathcal{F}, \\ \sup\{\delta : B(d, \delta) \subset \mathcal{F}^C\} & \text{if } d \in \mathcal{F}^C. \end{cases} \tag{5}$$

In the typical case when $CP(d)$ and $CD(d)$ are both feasible, i.e., $d \in \mathcal{F}$, $\rho(d)$ is the minimum change Δd in the data d needed to create a data instance $d + \Delta d$ that

is either primal-infeasible or is dual-infeasible. The "condition number" $\mathcal{C}(d)$ of the data instance d is defined as

$$\mathcal{C}(d) = \frac{\|d\|}{\rho(d)}$$

when $\rho(d) > 0$, and $\mathcal{C}(d) = \infty$ when $\rho(d) = 0$. The condition number $\mathcal{C}(d)$ can be viewed as a scale-invariant reciprocal of $\rho(d)$. The value of $\mathcal{C}(d)$ is a measure of the relative conditioning of the problem $CP(d)$ and its dual $CD(d)$ for the data instance d.

Let C be a regular cone in the normed linear vector space X. A critical component of our analysis concerns the extent to which the norm function $\|x\|$ can be approximated by some linear function $u^T x$ over the cone C for some particularly good choice of $u \in X^*$. Let $u \in intC^*$ be given, and suppose that u has been normalized so that $\|u\|_* = 1$. Let $f(u) = \text{minimum}\{u^T x \mid x \in C, \|x\| = 1\}$. Then it is elementary to see that $0 < f(u) \leq 1$, and also that $f(u)\|x\| \leq u^T x \leq \|x\|$ for any $x \in C$. Therefore the linear function $u^T x$ approximates $\|x\|$ over all $x \in C$ to within the factor $f(u)$. Put another way, the larger $f(u)$ is, the closer $u^T x$ approximates $\|x\|$ over all $x \in C$. Maximizing the value of $f(u)$ over all $u \in X^*$ satisfying $\|u\|_* = 1$, we are led to the following definition:

Definition 2.1 *If C is a regular cone in the normed linear vector space X, the coefficient of linearity for the cone C is given by:*

$$
\beta = \sup_{\substack{u \in X^* \\ \|u\|_* = 1}} \quad \inf_{\substack{x \in C \\ \|x\| = 1}} u^T x \quad . \tag{8}
$$

Let $\bar{u}$ denote that value of $u \in X^*$ that achieves the supremum in (8). We refer to $\bar{u}$ generically as the "norm approximation vector" for the cone C. We have the following properties of the coefficient of linearity β:

Proposition 2.1 *Suppose that C is a regular cone in the normed linear vector space X, and let β denote the coefficient of linearity for C. Then $0 < \beta \leq 1$. Furthermore, the norm approximation vector $\bar{u}$ exists and is unique, and satisfies the following properties:*

(i) $\bar{u} \in int\ C^*$,

(ii) $\|\bar{u}\|_* = 1$,

(iii) $\beta = \min\{\bar{u}^T x \mid x \in C, \|x\| = 1\}$, and

(iv) $\beta\|x\| \leq \bar{u}^T x \leq \|x\|$ for any $x \in C$.

We illustrate the construction of the coefficient of linearity on two families of cones, the nonnegative orthant $\Re^k_+$ and the positive semi-definite cone $S^{k \times k}_+$.

First consider the nonnegative orthant. Let $X = \Re^k$ with Euclidean norm $\|x\| := \|x\|_2 = \sqrt{x^T x}$ (where $\|\cdot\|_2$ denotes the L_2-norm (Euclidean) norm, and $C = \Re_+^k = \left\{ x \in \Re^k \mid x \geq 0 \right\}$. Then it is straightforward to show that the coefficient of linearity of $\Re^k$ is $\beta = 1/\sqrt{k}$ and that the norm approximation vector is $\bar{u} = \left(\frac{1}{\sqrt{k}} \right) e$, where $e = (1, \ldots, 1)^T$.

Now consider the positive semi-definite cone $S_+^{k \times k}$, which has been shown to be of enormous importance in mathematical programming (see Alizadeh [1] and Nesterov and Nemirovskii [12]). Let $X = S^{k \times k}$ denote the set of real $k \times k$ symmetric matrices with Frobenius norm $\|x\| := \sqrt{\text{trace}(x^T x)}$, and let $C = S_+^{k \times k} = \left\{ x \in S^{k \times k} \mid x \succeq 0 \right\}$, where "$\succeq$" is the Löwner partial ordering, i.e., $x \succeq w$ if $x - w$ is a positive semi-definite symmetric matrix. Then $S_+^{k \times k}$ is a closed convex cone, and it is easy to show that the coefficient of linearity of $S_+^{k \times k}$ is $\beta = \frac{1}{\sqrt{k}}$ and that the norm approximation vector is $\bar{u} = \left(\frac{1}{\sqrt{k}} \right) I$, where I is the identity matrix.

The coefficient of linearity β for the regular cone C is essentially the same as the scalar α defined in Renegar [18] on page 328.

For the remainder of this paper, we amend our notation as follows:

Definition 2.2 *Whenever the cone C_X is regular, the coefficient of linearity for C_X is denoted by β, and the coefficient of linearity for C_X^* is denoted by β^*. Whenever the cone C_Y is regular, the coefficient of linearity for C_Y is denoted by $\bar{\beta}$, and the coefficient of linearity for C_Y^* is denoted by $\bar{\beta}^*$.*

3 The Ellipsoid Algorithm, and a Transformed Problem

3.1 The Ellipsoid Algorithm for Optimization

For the purpose of concreteness, we henceforth assume that $X = \Re^n$ and that the norm $\|x\|$ on $\Re^n$ is the Euclidean norm, i.e., $\|x\| = \|x\|_2 = \sqrt{x^T x}$. No such assumption is made for the space Y.

We first review a few basic results regarding the ellipsoid algorithm for solving an optimization problem, see [10], [9], [4], [8], [3]. The ellipsoid algorithm (in optimization mode) is typically designed for an optimization problem of the form

$$P : \begin{array}{ll} \text{maximize} & f(x) \\ x & \\ \text{s.t.} & x \in S, \end{array} \qquad (9)$$

where S is a convex set (closed or not) in $\Re^k$, $f(x)$ is a quasi-concave function, and where one knows *a priori* an upper bound R on the norm of some optimal solution x^* of P. (Actually, the ellipsoid algorithm is more usually associated with the assumption that S is a closed convex set and that $f(x)$ is a concave function, but these assumptions can be relaxed slightly. One only needs that S is a convex set and that the upper level sets of $f(x)$ are convex sets on S, which is equivalent to the statement that $f(x)$ is a quasi-concave function on S (see [2]).)

In order to implement the ellipsoid algorithm to approximately solve $CP(d)$, it is necessary that one has available a separation oracle for the set S, i.e., that for any $\bar{x} \notin S$, one can perform a feasibility cut for the set S, which consists of computing a vector $h \neq 0$ for which $S \subset \{x \mid h^T x \geq h^T \bar{x}\}$. Suppose that T_1 is an upper bound on the number of operations needed to perform a feasibility cut for the set S. It is also necessary that one has available a support oracle for the upper level sets $U_\alpha = \{x \in S \mid f(x) \geq \alpha\}$ of the function $f(x)$. That is, for any $\bar{x} \in S$, it is necessary to be able to perform an optimality cut for the objective function $f(x)$. This consists of computing a vector $h \neq 0$ for which $U_{f(\bar{x})} \subset \{x \in \Re^k \mid h^T x \geq h^T \bar{x}\}$. Suppose that T_2 is an upper bound on the number of operations needed to compute an optimality cut for the function $f(x)$ on the set S.

Let z^* denote the optimal value of P, and denote the set of ϵ-optimal solutions of P by $S(\epsilon)$, i.e., $S(\epsilon) = \{x \in \Re^k \mid x \in S \text{ and } f(x) \geq z^* - \epsilon\}$. Suppose that $S(\epsilon)$ contains a Euclidean ball of radius r, i.e., there exists $\hat{x}$ for which

$$\{x \in \Re^k \mid \|x - \hat{x}\|_2 \leq r\} \subset S(\epsilon).$$

Furthermore, suppose that an upper bound R on the quantity $(\|\hat{x}\|_2 + r)$ is known in advance, i.e., we know R for which $R \geq \|\hat{x}\|_2 + r$. One then has the following generic result about the ellipsoid algorithm as applied to problem P:

Ellipsoid Algorithm Theorem *Suppose that the set of ϵ-optimal solutions of P contains a Euclidean ball of radius r centered at some point $\hat{x}$, and that an upper bound R on the quantity $(\|\hat{x}\|_2 + r)$ is known. Then if the ellipsoid algorithm is initiated with a Euclidean ball of radius R centered at $x^0 = 0$, the algorithm will compute an ϵ-optimal solution of P in at most*

$$\lceil 2k(k+1)\ln(R/r) \rceil$$

iterations, where each iteration must perform at most $(k^2 + \max\{T_1, T_2\})$ operations, where T_1 and T_2 are the number of operations needed to perform a feasibility cut on S and an optimality cut on $f(x)$, respectively.

We note that the bound on the number of operations arises from performing either a feasibility or an optimality cut (which takes $\max\{T_1, T_2\}$ operations), and then performing a rank-one update of the positive definite matrix defining the ellipsoid (see [3], for example), which takes k^2 operations.

3.2 The Transformed Problem $HP(d)$

Let $z^*(d)$ denote the optimal objective function value of $CP(d)$. Let X_d and $X_d(\epsilon)$ denote the feasible region of $CP(d)$ and the set of ϵ-optimal solutions of $CP(d)$, respectively, i.e.,

$$X_d = \{x \in X \mid b - Ax \in C_Y, x \in C_X\} \tag{10}$$

and

$$X_d(\epsilon) = \{x \in X \mid b - Ax \in C_Y, x \in C_X, c^T x \geq z^*(d) - \epsilon\}. \tag{11}$$

For the purpose of applying the ellipsoid algorithm to solve $CP(d)$, we employ a standard transformation to convert $CP(d)$ to the homogenized fractional program:

$$HP(d): \quad \underset{w, \theta}{\text{maximize}} \quad \frac{c^T w}{\theta}$$
$$\text{s.t.} \quad b\theta - Aw \in C_Y$$
$$w \in C_X, \tag{12}$$
$$\theta > 0,$$

(see [5] and [6]), with the (obvious) transformations:

$$x = \frac{w}{\theta} \tag{13}$$

and

$$(w, \theta) = \frac{(x, 1)}{\|(x, 1)\|_2} \tag{14}$$

where the norm $\|(w, \theta)\|_2$ is simply the Euclidean norm applied to the concatenated vector (w, θ), i.e., $\|(w, \theta)\|_2 = \sqrt{w^T w + \theta^2}$.

It is trivial to show that $z^*(d)$ is also the optimal objective function value of $HP(d)$. Let X_d^h and $X_d^h(\epsilon)$ denote the feasible region of $HP(d)$ and the set of ϵ-optimal solutions of $HP(d)$, respectively (where "h" is a mnemonic for "homogeneous"), i.e.,

$$X_d^h = \{(w, \theta) \in X \times \Re \mid b\theta - Aw \in C_Y, w \in C_X, \theta > 0\} \tag{15}$$

and

$$X_d^h(\epsilon) = \{(w, \theta) \in X \times \Re \mid b\theta - Aw \in C_Y, w \in C_X, \theta > 0, c^T w/\theta \geq z^*(d) - \epsilon\}. \tag{16}$$

Then X_d^h and $X_d^h(\epsilon)$ are both convex sets. Also, the objective function $c^T w/\theta$ of $HP(d)$ is easily seen to be a quasi-concave function over the feasible region X_d^h. In

fact, if $(\bar{w}, \bar{\theta})$ is a feasible solution of $HP(d)$, then

$$\left\{ (w, \theta) \in X_d^h \mid \frac{c^T w}{\theta} \geq \frac{c^T \bar{w}}{\bar{\theta}} \right\} \subset \left\{ (w, \theta) \in X \times \Re \mid c^T w - (c^T \bar{w})/\bar{\theta})\theta \geq 0 \right\},$$

and so the concatenated vector $(c, -c^T \bar{w}/\bar{\theta})$ is a support vector for the upper level set of the function $c^T w/\theta$ at the feasible point $(\bar{w}, \bar{\theta})$. Therefore, the number of operations needed to perform an optimality cut in $HP(d)$ is at most $2n$ operations.

Because any feasible solution of $HP(d)$ can be scaled by an arbitrary positive scalar without changing its objective function value or affecting its feasibility, the feasible region and all upper level sets of the objective function $c^T w/\theta$ of $HP(d)$ contain points in the unit Euclidean ball $\{(w, \theta) \in X \times \Re \mid \|(w, \theta)\|_2 \leq 1\}$. Therefore, the key advantage in using $HP(d)$ rather than $CP(d)$ in the ellipsoid algorithm is that we will be able to conveniently start the ellipsoid algorithm for solving $HP(d)$ with the unit Euclidean ball. This will be further explored in the next section.

Finally, we remark that the set X_d is mapped into X_d^h by (14) and X_d^h is mapped onto X_d by (13). Also $X_d(\epsilon)$ is mapped into $X_d^h(\epsilon)$ by (14) and $X_d^h(\epsilon)$ is mapped onto $X_d(\epsilon)$ by (13).

4   Complexity Results

For the purpose of developing complexity results, we focus on three different classes of instances of $CP(d)$, namely:

Class (i): C_X and C_Y are both regular,

Class (ii): C_X is regular and $C_Y = \{0\}$,

Class (iii): $C_X = X$ and C_Y is regular.

For these three classes of instances, $CP(d)$ can be written as (i) $\max\{c^T x \mid b - Ax \in C_Y, x \in C_X\}$, (ii) $\max\{c^T x \mid Ax = b, x \in C_X\}$, and (iii) $\max\{c^T x \mid b - Ax \in C_Y, x \in X\}$.

To simplify the notation, let $B_2\left((\bar{w}, \bar{\theta}), r\right)$ denote the Euclidean ball in $X \times \Re$ centered at $(\bar{w}, \bar{\theta})$ with radius r, i.e.,

$$B_2\left((\bar{w}, \bar{\theta}), r\right) = \left\{ (w, \theta) \in X \times \Re \mid \sqrt{(w - \bar{w})^T (w - \bar{w}) + (\theta - \bar{\theta})^2} \leq r \right\}.$$

The following three Lemmas are the main technical tools that will be used to prove the complexity results. Under the assumption that both C_X and C_Y are regular

cones, Lemma 4.1 essentially states that there exists a ball of radius $\check{r}$ that is contained in the set of ϵ-optimal solutions of $HP(d)$, and that is also contained in the Euclidean unit ball, and that $\check{r}$ is "not too small". Lemma 4.2 contains a similar result under the assumption that the cone C_X is regular but that the cone $C_Y = \{0\}$. Lemma 4.3 contains a similar result under the assumption that the cone $C_X = X$ and that C_Y is regular.

Lemma 4.1 *Suppose that C_X is a regular cone and C_Y is a regular cone, and $d \in \mathcal{F}$ and that $\mathcal{C}(d) < +\infty$. Let $\epsilon > 0$ be given. Then there exists $(\check{w}, \check{\theta}) \in X_d^h(\epsilon)$ and a positive scalar $\check{r}$ satisfying*

$$B_2\left((\check{w}, \check{\theta}), \check{r}\right) \subset X_d^h(\epsilon),$$

$$\|(\check{w}, \check{\theta})\|_2 + \check{r} \leq 1,$$

and

$$\check{r} \geq \left(\frac{\bar{\beta}^* \beta^*}{63}\right) \left(\frac{1}{\mathcal{C}(d)}\right)^4 \min\left\{\frac{\epsilon}{\|d\|}, \mathcal{C}(d)\right\}.$$

Lemma 4.2 *Suppose that C_X is a regular cone and that $C_Y = \{0\}$, and $d \in \mathcal{F}$ and that $\mathcal{C}(d) < +\infty$. Let $\epsilon > 0$ be given. Then there exists $(\check{w}, \check{\theta}) \in X_d^h(\epsilon)$ and a positive scalar $\check{r}$ satisfying*

$$B_2\left((\check{w}, \check{\theta}), \check{r}\right) \cap \{(w, \theta) \in X \times \Re \mid b\theta - Aw = 0\} \subset X_d^h(\epsilon),$$

$$\|(\check{w}, \check{\theta})\|_2 + \check{r} \leq 1,$$

and

$$\check{r} \geq \left(\frac{\beta^*}{32}\right) \left(\frac{1}{\mathcal{C}(d)}\right)^4 \min\left\{\frac{\epsilon}{\|d\|}, \mathcal{C}(d)\right\}.$$

Lemma 4.3 *Suppose that $C_X = X$ and C_Y is a regular cone, and $d \in \mathcal{F}$ and that $\mathcal{C}(d) < +\infty$. Let $\epsilon > 0$ be given. Then there exists $(\check{w}, \check{\theta}) \in X_d^h(\epsilon)$ and a positive scalar $\check{r}$ satisfying*

$$B_2\left((\check{w}, \check{\theta}), \check{r}\right) \subset X_d^h(\epsilon),$$

$$\|(\check{w}, \check{\theta})\|_2 + \check{r} \leq 1,$$

and

$$\check{r} \geq \left(\frac{\bar{\beta}^*}{32}\right)\left(\frac{1}{\mathcal{C}(d)}\right)^4 \min\left\{\frac{\epsilon}{\|d\|}, \mathcal{C}(d)\right\}.$$

The proofs of these three Lemmas are deferred to Section 6. Based on these Lemmas, we now state complexity results for the ellipsoid algorithm in the following three theorems, where each theorem states a complexity bound for the ellipsoid algorithm for solving $HP(d)$ (equivalently, $CP(d)$), for one of the three respective classes of $CP(d)$.

Theorem 4.1 *Suppose that C_X is a regular cone and C_Y is a regular cone, and that $d \in \mathcal{F}$ and that $\mathcal{C}(d) < +\infty$. Let ϵ satisfying $0 < \epsilon < \|d\|$ be given. Suppose that the ellipsoid algorithm is applied to solve $HP(d)$, and is initiated with the Euclidean unit ball centered at $(w^0, \theta^0) = (0, 0)$. Then the ellipsoid algorithm will compute an ϵ-optimal solution of $HP(d)$ (and hence, by transformation, to $CP(d)$) in at most*

$$\left\lceil 8(n+1)(n+2)\ln\left(\frac{63\mathcal{C}(d)}{\bar{\beta}^*\beta^*}\frac{\|d\|}{\epsilon}\right)\right\rceil$$

iterations, where each iteration must perform at most $((n+1)^2 + \max\{2n, S_1, m + mn + S_2\})$ operations, where S_1 and S_2 are the number of operations needed to perform a feasibility cut on C_X and C_Y, respectively.

Proof: The proof follows directly by combining the Ellipsoid Algorithm Theorem with Lemma 4.1. The dimension in which the ellipsoid algorithm is implemented is $k = n + 1$, and from Lemma 4.1 we have $R = 1$ and $r = \check{r}$ in the Ellipsoid Algorithm Theorem. Because $\epsilon \leq \|d\|$, then $\epsilon/\|d\| \leq 1 \leq \mathcal{C}(d)$, and so the minimum in the expression in Lemma 4.1 is $\epsilon/\|d\|$. Then combining the Ellipsoid Algorithm Theorem with Lemma 4.1 yields an iteration bound of

$$\begin{aligned}
\lceil 2k(k+1)\ln(R/r)\rceil &= \left\lceil 2(n+1)(n+2)\ln\left(\frac{1}{\check{r}}\right)\right\rceil \\
&\leq \left\lceil 2(n+1)(n+2)\ln\left(\frac{63\mathcal{C}(d)^4}{\bar{\beta}^*\beta}\frac{\|d\|}{\epsilon}\right)\right\rceil \\
&\leq \left\lceil 8(n+1)(n+2)\ln\left(\frac{63\mathcal{C}(d)}{\bar{\beta}^*\beta}\frac{\|d\|}{\epsilon}\right)\right\rceil.
\end{aligned}$$

The number of operations needed to perform an optimality cut is at most $2n$ as discussed earlier, and the number of operations needed to compute and test for feasibility of $b\theta - Aw \in C_Y$ is then $m + mn + S_2$ operations. ∎

Theorem 4.2 *Suppose that C_X is a regular cone and that $C_Y = \{0\}$, and that $d \in \mathcal{F}$ and that $\mathcal{C}(d) < +\infty$. Let ϵ satisfying $0 < \epsilon < \|d\|$ be given. Suppose that the ellipsoid algorithm is applied to solve $HP(d)$, and is initiated with the Euclidean unit disk centered at $(w^0, \theta^0) = (0,0)$, in the subspace $\{(w,\theta) \in X \times \Re \mid Aw - b\theta = 0\}$. Then the ellipsoid algorithm will compute an ϵ-optimal solution of $HP(d)$ (and hence, by transformation, to $CP(d)$) in at most*

$$\left\lceil 8(n - m + 1)(n - m + 2) \ln \left(\frac{32\mathcal{C}(d)}{\beta^*} \frac{\|d\|}{\epsilon} \right) \right\rceil$$

iterations, where each iteration must perform at most $((n - m + 1)^2 + \max\{2n, S_1\})$ operations, where S_1 is the number of operations needed to perform a feasibility cut on C_X.

Proof: The proof follows directly by combining the Ellipsoid Algorithm Theorem with Lemma 4.2. The dimension in which the ellipsoid algorithm is implemented is $k = n - m + 1$, and the rest of proof follows as in the proof of Theorem 4.1. ∎

Theorem 4.3 *Suppose that $C_X = X$ and C_Y is a regular cone, and that $d \in \mathcal{F}$ and that $\mathcal{C}(d) < +\infty$. Let ϵ satisfying $0 < \epsilon < \|d\|$ be given. Suppose that the ellipsoid algorithm is applied to solve $HP(d)$, and is initiated with the Euclidean unit ball centered at $(w^0, \theta^0) = (0,0)$. Then the ellipsoid algorithm will compute an ϵ-optimal solution of $HP(d)$ (and hence, by transformation, to $CP(d)$) in at most*

$$\left\lceil 8(n + 1)(n + 2) \ln \left(\frac{32\mathcal{C}(d)}{\bar{\beta}^*} \frac{\|d\|}{\epsilon} \right) \right\rceil$$

iterations, where each iteration must perform at most $((n + 1)^2 + \max\{2n, m + mn + S_2\})$ operations, where S_2 is the number of operations needed to perform a feasibility cut on C_Y.

Proof: The proof follows directly by combining the Ellipsoid Algorithm Theorem with Lemma 4.3 and follows as in the proof of Theorem 4.1. ∎

5   Other Issues: Testing for ϵ-optimality; the Dual Problem; Infeasibility; Skewness of the Ellipsoids

Testing for ϵ-optimality by Solving the Dual Problem. One uncomfortable fact about Theorems 4.1, 4.2, and 4.3 is that while the ellipsoid algorithm is guaranteed to find an ϵ-approximate solution of $CP(d)$ in the stated complexity bounds of these theorems, the quantities in the bounds may be unknown (one may know the relevant coefficients of linearity, but in all likelihood the condition number $\mathcal{C}(d)$ is unknown), and so one does not know when an ϵ-approximate solution of $CP(d)$ has been found. An obvious strategy for overcoming this difficulty is to solve the primal and the dual problem in parallel, and then test at each iteration (of each algorithm) if the best primal and dual solutions obtained so far satisfy a duality gap of at most ϵ. Because of the natural symmetry in format of the dual pair of problems $CP(d)$ and $DP(d)$, one can obtain complexity results for solving the dual problem $DP(d)$ that exactly parallel those of Theorems 4.1, 4.2, and 4.3, where the quantities n, β^*, and $\bar{\beta}^*$ are replaced by $m, \bar{\beta}$, and β, respectively, and where the cones C_X and C_Y are replaced by C_Y^* and C_X^* in the statements of the complexity results. One also must assume that $Y^* = \Re^m$ and that the norm $\|y\|_*$ on $\Re^m$ is the Euclidean norm.

Testing for Infeasibility. If one is not sure whether or not $CP(d)$ has a feasible solution, the ellipsoid algorithm can be run to test for infeasibilty of the primal problem (in parallel with attempting to solve $CP(d)$). This can be accomplished as follows. First, assume that the dual space $Y^* = \Re^m$ is endowed with the Euclidean norm $\|y\|_2$. Second, note that $CP(d)$ has no feasible solution if the "alternative" system:

$$AP(d): \quad \begin{aligned} A^T y &\in C_X^* \\ y &\in C_Y^* \\ y^T b &< 0 \end{aligned}$$

has a solution. Define the following "alternative" set

$$Y_d = \{ y \in Y^* \mid A^T y \in C_X^*, y \in C_Y^*, y^T b \leq 0 \}. \tag{17}$$

Suppose $CP(d)$ has no feasible solution. Then, as special cases of Theorems 5.2, 5.4, and 5.6 of [7], Y_d must contain an inscribed Euclidean ball $B_2(\hat{y}, r)$ (or a disk in the vector subspace $\{ y \in \Re^m \mid A^T y = 0 \}$ if $C_X = X$) such that $\|\hat{y}\|_2 + r \leq 1$ and such that

(i): $r \geq \dfrac{\beta \bar{\beta}}{5 \mathcal{C}_P(d)}$ when C_X and C_Y are both regular,

(ii): $r \geq \dfrac{\beta}{2 \mathcal{C}_P(d)}$ when C_X is regular and $C_Y = \{0\}$,

(iii): $r \geq \dfrac{\bar{\beta} \bar{\beta}^*}{5 \mathcal{C}_P(d)}$ when $C_X = X$ and C_Y is regular.

These results can then be used to demonstrate that an upper bound on the number

of iterations needed to find a solution of $AP(d)$ using the ellipsoid algorithm starting with the Euclidean unit ball in $\Re^m$ (or the unit disk in the vector subspace $\{y \in \Re^m \mid A^T y = 0\}$ if $C_X = X$) is:

(i): $O\left(m^2 \ln\left(\dfrac{\mathcal{C}_P(d)}{\beta\bar{\beta}}\right)\right)$ when C_X and C_Y are both regular,

(ii): $O\left(m^2 \ln\left(\dfrac{\mathcal{C}_P(d)}{\beta}\right)\right)$ when C_X is regular and $C_Y = \{0\}$,

(iii): $O\left((m-n)^2 \ln\left(\dfrac{\mathcal{C}_P(d)}{\bar{\beta}\bar{\beta}^*}\right)\right)$ when $C_X = X$ and C_Y is regular.

Bounding the Skewness of the Ellipsoids in the Ellipsoid Algorithm. Let $E_{\bar{x},Q} = \{x \in X \mid (x-\bar{x})^T Q^{-1}(x-\bar{x}) \le 1\}$ be an ellipsoid centered at the point $\bar{x}$, where Q is a positive-definite matrix. The skewness of $E_{\bar{x},Q}$ is defined to be the ratio of the largest to the smallest eigenvalue of the matrix Q defining $E_{\bar{x},Q}$, and so the skewness also corresponds to the traditional condition number of the matrix Q. The skewness of the ellipsoids generated in an application of the ellipsoid algorithm determines the numerical stability of the ellipsoid algorithm, since each iteration of the ellipsoid algorithm uses the current value of Q^{-1} to update the center $\bar{x}$ of the ellipsoid and to perform a rank-one update of Q^{-1}, see [3] for details. Furthermore, one can show that the logarithm of the skewness of the ellipsoid computed at a given iteration is sufficient to specify the numerical precision requirements of the ellipsoid algorithm at that iteration. Herein, we provide an upper bound on the skewness of all of the ellipsoids computed in the ellipsoid algorithm as a function of the condition number $\mathcal{C}(d)$ of $CP(d)$.

The skewness of the unit ball (which is used to initiate the ellipsoid algorithm herein) is 1. From the formula for updating the ellipsoids encountered in the ellipsoid algorithm at each iteration, the skewness increases by at most $\left(1 + \frac{2}{k-1}\right)$ at each iteration, where k is the dimension of the space in which the ellispoid algorithm is implemented, see [3] for example. Therefore the skewness of the ellipsoid at iteration j is bounded above by $\left(1 + \frac{2}{k-1}\right)^j$. Let us consider the class of instances defined for Theorem 4.1, for example, and let J be the (unrounded) iteration bound for the ellipsoid algorithm from Theorem 4.1, i.e.,

$$J = 8(n+1)(n+2)\ln\left(\frac{63\mathcal{C}(d)}{\bar{\beta}^*\beta^*}\frac{\|d\|}{\epsilon}\right), \tag{18}$$

and assume for simplicity of exposition that J is an integer. Let $(\text{Skew})_j$ denote the skewness of the ellipsoid computed in the ellipsoid algorithm at iteration j. Then for this class of instances we have $k = n+1$, whereby

$$(\text{Skew})_J \le \left(1 + \frac{2}{n}\right)^J = \left(e^{\left(\ln\left(1+\frac{2}{n}\right)\right)}\right)^J = e^{J\left(\ln\left(1+\frac{2}{n}\right)\right)} = \left(e^J\right)^{\left(\ln\left(1+\frac{2}{n}\right)\right)}. \tag{19}$$

Substituting for (18) in (19), we obtain

$$(\text{Skew})_J \le \left(\frac{63\mathcal{C}(d)}{\bar{\beta}^*\beta^*}\frac{\|d\|}{\epsilon}\right)^{8(n+1)(n+2)\ln\left(1+\frac{2}{n}\right)}.$$

However, the exponent in the above expression is bounded above by $34n$ for $n \geq 2$ (actually, it is bounded above by $17n$ for large $n \geq 33$), and we have

$$(\text{Skew})_J \leq \left(\frac{63 \mathcal{C}(d)}{\bar{\beta}^* \beta^*} \frac{\|d\|}{\epsilon} \right)^{34n}.$$

Taking logarithms, we can re-write this bound as

$$\ln (\text{Skew})_J \leq 34n \ln \left(\frac{63 \mathcal{C}(d)}{\bar{\beta}^* \beta^*} \frac{\|d\|}{\epsilon} \right). \tag{20}$$

Therefore, the logarithm of the skewness of the ellipsoids encountered in the ellipsoid algorithm grows at most linearly in the logarithm of the condition number $\mathcal{C}(d)$. Also, the bound in (20) specifies the sufficient numerical precision requirements for the ellipsoid algorithm (in terms of $\ln \mathcal{C}(d)$ and other quantities) because the logarithm of the skewness is sufficient to specify such requirements. This is similar to the results on numerical precision presented in [22] for an interior point method for linear programming.

Finally, the above logic can be used to obtain similar bounds on the skewness for the other two classes of instances of $CP(d)$.

6   Technical Lemmas and Proofs

In this Section, we first prove two technical Lemmas, that are then used prove Lemmas 4.1, 4.2, and 4.3.

Lemma 6.1 *Suppose that the feasible region of $CP(d)$ is nonempty, i.e., $X_d \neq \emptyset$,*

and that C_Y is a regular cone. If $\rho(d) > 0$, then there exists $(\hat{w}, \hat{\theta}) \in X_d^h$ and positive

scalars r_1 and R_1 satisfying:

> *(i) $B_2 \left((\hat{w}, \hat{\theta}), r_1 \right) \subset \{ (w, \theta) \in X \times \Re \mid b\theta - Aw \in C_Y, \theta > 0 \},$*

> *(ii) $\|(\hat{w}, \hat{\theta})\|_2 + r_1 \leq R_1,$*

> *(iii) $\dfrac{R_1}{r_1} \leq \dfrac{6}{\bar{\beta}^*} \mathcal{C}(d).$*

Proof: The proof relies on Theorem A.2 of the Appendix, which is Theorem 3.5 of [7]. Let $\bar{z}$ be the norm approximation vector for the cone C_Y^*. For any $y \in C_Y^*$, we

have from Proposition 2.1 that

$$\bar{z}^T y + \frac{\bar{\beta}^* b^T y}{\|d\|} \geq \|y\|_* \bar{\beta}^* - \frac{\bar{\beta}^* \|b\| \, \|y\|_*}{\|d\|} \geq 0 \; ,$$

so that $\dfrac{1}{2}\left(\bar{z} + \dfrac{\bar{\beta}^*}{\|d\|}b\right) \in C_Y$. Now let $(\tilde{w}, \tilde{\theta})$ solve $P_v(d)$ (see (41)), and let $\hat{w} = \tilde{w}$ and

$$\hat{\theta} = \tilde{\theta} + \frac{\bar{\beta}^*}{2\|d\|}.$$

Note that $\hat{\theta} > 0$ from Proposition 2.1. Let $q = b\tilde{\theta} - A\tilde{w} - \bar{z}$, and note that $q \in C_Y$ from (41). We also have

$$b\tilde{\theta} - A\tilde{w} + \frac{\bar{\beta}^*}{2\|d\|}b - \frac{1}{2}\bar{z} \;=\; \bar{z} + q + \frac{1}{2}\left(\frac{\bar{\beta}^*}{\|d\|}b + \bar{z}\right) - \bar{z}$$

$$=\; q + \frac{1}{2}\left(\frac{\bar{\beta}^*}{\|d\|}b + \bar{z}\right) \in C_Y,$$

so that

$$b\hat{\theta} - A\hat{w} - \frac{1}{2}\bar{z} \in C_Y. \tag{21}$$

In particular, (21) and Proposition 2.1 imply that $b\hat{\theta} - A\hat{w} \in C_Y$. Therefore $(\hat{w}, \hat{\theta}) \in X_d^h$, since $\hat{w} = \tilde{w} \in C_X$ from (41).

Now let

$$r_1 = \frac{\bar{\beta}^*}{2\sqrt{2}\|d\|}.$$

Let $(f, t) \in X \times \Re$ satisfy $\|(f, t)\|_2 \leq 1$, and consider $(w, \theta) = (\hat{w}, \hat{\theta}) + r_1(f, t)$. Then $\theta = \hat{\theta} + r_1 t \geq \tilde{\theta} + \dfrac{\bar{\beta}^*}{2\|d\|} - \dfrac{\bar{\beta}^*}{2\sqrt{2}\|d\|} > 0$ since $\tilde{\theta} \geq 0$ (see (41)) and $\bar{\beta}^* > 0$ (from Proposition 2.1). We also have

$$b\theta - Aw = b\hat{\theta} - A\hat{w} + r_1(bt - Af) = b\hat{\theta} - A\hat{w} - \frac{1}{2}\bar{z} + r_1(bt - Af) + \frac{1}{2}\bar{z}.$$

Therefore, for any $y \in C_Y^*$, we have

$$
\begin{aligned}
y^T(b\theta - Aw) &\geq r_1 y^T(bt - Af) + \frac{1}{2}y^T\bar{z} && \text{(from (21))} \\[2mm]
&\geq -r_1\|y\|_*\|bt - Af\| + \frac{1}{2}\bar{\beta}^*\|y\|_* && \text{(from Proposition 2.1)} \\[2mm]
&\geq \|y\|_* \left(-r_1\|d\|(\|f\|_2 + |t|) + \frac{1}{2}\bar{\beta}^* \right) \\[2mm]
&\geq \|y\|_* \left(-r_1\|d\|\sqrt{2} + \frac{1}{2}\bar{\beta}^* \right) && \text{(since } \|(f,t)\|_2 \leq 1) \\[2mm]
&= 0
\end{aligned}
$$

from the definition of r_1, thereby showing that $b\theta - Aw \in C_Y$, which proves (i).

Next, let $R_1 = \|(\hat{w},\hat{\theta})\|_2 + r_1$, which automatically proves (ii). To prove (iii), note that

$$
\begin{aligned}
\frac{R_1}{r_1} &= 1 + \frac{2\sqrt{2}\|d\|\;\|(\hat{w},\hat{\theta})\|_2}{\bar{\beta}^*} \\[3mm]
&\leq 1 + \frac{2\sqrt{2}\|d\|(\|\hat{w}\|_2 + \hat{\theta})}{\bar{\beta}^*} \\[3mm]
&= 1 + \frac{2\sqrt{2}\|d\|\left(\|\tilde{w}\|_2 + \tilde{\theta} + \dfrac{\bar{\beta}^*}{2\|d\|} \right)}{\bar{\beta}^*} \\[3mm]
&= 1 + \frac{2\sqrt{2}\|d\|\left(v(d) + \dfrac{\bar{\beta}^*}{2\|d\|} \right)}{\bar{\beta}^*} && \text{(from (41))} \\[3mm]
&\leq 1 + \frac{2\sqrt{2}\|d\|\left(\dfrac{1}{\rho_P(d)} + \dfrac{\bar{\beta}^*}{2\|d\|} \right)}{\bar{\beta}^*} && \text{(from Theorem A.2)} \\[3mm]
&\leq 1 + \frac{2\sqrt{2}\mathcal{C}(d)}{\bar{\beta}^*} + \sqrt{2} \\[3mm]
&\leq \frac{2\sqrt{2}\mathcal{C}(d)}{\bar{\beta}^*} + \mathcal{C}(d)(1 + \sqrt{2}) && \text{(since } \mathcal{C}(d) \geq 1) \\[3mm]
&\leq \frac{(1 + 3\sqrt{2})\mathcal{C}(d)}{\bar{\beta}^*} \leq \frac{6\mathcal{C}(d)}{\bar{\beta}^*}, && \text{(since } \bar{\beta}^* \leq 1)
\end{aligned}
$$

which proves (*iii*). ∎

We next have:

Lemma 6.2 *Suppose that the feasible region of $CP(d)$ is nonempty, i.e., $X_d \neq \emptyset$, and that C_X is a regular cone. If $\rho(d) > 0$, then there exists $(\hat{w}, \hat{\theta}) \in X_d^h$ and positive scalars r_2 and R_2 satisfying:*

$$(i) \quad B_2\left((\hat{w}, \hat{\theta}), r_2\right) \subset \{(w, \theta) \in X \times \Re \mid w \in C_X, \theta > 0\},$$

$$(ii) \quad \|(\hat{w}, \hat{\theta})\|_2 + r_2 \leq R_2,$$

$$(iii) \quad \frac{R_2}{r_2} \leq \frac{6}{\beta^*}\mathcal{C}(d).$$

Proof: The proof relies on Theorem A.1 of the Appendix, which is Theorem 3.5 of Renegar [18]. First of all, let $\bar{x}$ denote the norm approximation vector for the cone C_X^*, so that from Proposition 2.1,

$$\bar{x}^T z \geq \beta^* \|z\|_* = \beta^* \|z\|_2 \text{ for any } z \in C_X^*, \tag{22}$$

(since we assumed that $\|x\| = \|x\|_2$ and hence $\|z\|_* = \|z\|_2$ as well), and define

$$(\tilde{w}, \tilde{\theta}) = \frac{(\bar{x}, \beta^*)}{\sqrt{1 + (\beta^*)^2}} \tag{23}$$

and

$$\tau' := \frac{\beta^*}{2} < \frac{\beta^*}{\sqrt{2}} \leq \frac{\beta^*}{\sqrt{1 + (\beta^*)^2}}. \tag{24}$$

Then the following inclusion must be true:

$$B_2\left((\tilde{w}, \tilde{\theta}), \tau'\right) \subset \{(w, \theta) \mid w \in C_X, \theta > 0\}. \tag{25}$$

To see why (25) is true, let (f, t) satisfy $\|(f, t)\|_2 \leq 1$, and let $(w, \theta) = (\tilde{w}, \tilde{\theta}) + \tau'(f, t)$. Then it is sufficient to show that $\theta > 0$ and $w \in C_X$. We have

$$\theta = \tilde{\theta} + \tau' t = \frac{\beta^*}{\sqrt{1 + (\beta^*)^2}} + \tau' t > 0$$

from (24) and the fact that $|t| \leq 1$, thus showing that $\theta > 0$. Now note that for any $z \in C_X^*$,

$$z^T w = z^T \tilde{w} + \tau' z^T f$$

$$\geq \frac{z^T \bar{x}}{\sqrt{1 + (\beta^*)^2}} - \tau' \|z\|_2 \|f\|_2 \qquad \text{(from (23))}$$

$$\geq \frac{\beta^* \|z\|_2}{\sqrt{1 + (\beta^*)^2}} - \tau' \|z\|_2 \|f\|_2 \qquad \text{(from (22))}$$

$$\geq 0, \qquad \text{(from (24) and since } \|f\|_2 \leq 1\text{)}$$

and so $w \in C_X$. This demonstrates (25).

Now consider the system

$$
\begin{aligned}
b\theta - Aw + \frac{\rho(d)}{\|b\tilde{\theta} - A\tilde{w}\|} \left(b\tilde{\theta} - A\tilde{w}\right) &\in C_Y \\
w &\in C_X \\
\theta &\geq 0 \\
|\theta| + \|w\|_2 &\leq 1.
\end{aligned}
\tag{26}
$$

From (39) and Theorem A.1 of the Appendix, (26) must have a solution (w', θ'). Let

$$(\hat{w}, \hat{\theta}) = \frac{\|b\tilde{\theta} - A\tilde{w}\|}{\rho(d) + \|b\tilde{\theta} - A\tilde{w}\|} \left((w', \theta') + \frac{\rho(d)}{\|b\tilde{\theta} - A\tilde{w}\|} (\tilde{w}, \tilde{\theta})\right). \tag{27}$$

Then $b\hat{\theta} - A\hat{w} \in C_Y$ (from (26)), and $w' \in C_X$, and $\theta' \geq 0$ (from (26)). Therefore $\hat{w} \in C_X$ and $\hat{\theta} > 0$ (from (26) and (27)). Re-write (27) as

$$(\hat{w}, \hat{\theta}) = \frac{\|b\tilde{\theta} - A\tilde{w}\|}{\rho(d) + \|b\tilde{\theta} - A\tilde{w}\|} (w', \theta') + \frac{\rho(d)}{\rho(d) + \|b\tilde{\theta} - A\tilde{w}\|} (\tilde{w}, \tilde{\theta}) \tag{28}$$

and we see that $(\hat{w}, \hat{\theta})$ is a convex combination of (w', θ') and $(\tilde{w}, \tilde{\theta})$. It then follows that $\|(\hat{w}, \hat{\theta})\|_2 \leq 1$, because $\|(w', \theta')\|_2 \leq \|w'\|_2 + |\theta'| \leq 1$ from (26) and $\|(\tilde{w}, \tilde{\theta})\|_2 = 1$ from (23). From (25) and the fact that $(w', \theta') \in C_X \times \Re_+$, it follows from (28) that

$$B_2\left((\hat{w}, \hat{\theta}), r_2\right) \subset \{(w, \theta) \mid w \in C_X, \theta > 0\} \tag{29}$$

where

$$r_2 = \frac{\tau' \rho(d)}{\rho(d) + \|b\tilde{\theta} - A\tilde{w}\|}. \tag{30}$$

This is part (i) of the Lemma. Now let $R_2 = \|(\hat{w}, \hat{\theta})\|_2 + r_2$, which immediately proves (ii) of the Lemma. To prove (iii) of the Lemma, we have

$$\frac{R_2}{r_2} = 1 + \frac{\left(\|(\hat{w}, \hat{\theta})\|_2\right)\left(\rho(d) + \|b\tilde{\theta} - A\tilde{w}\|\right)}{\tau' \rho(d)}$$

$$\leq 1 + \frac{\rho(d) + \|d\|(\|\tilde{w}\|_2 + |\tilde{\theta}|)}{\tau' \rho(d)},$$

since $\|(\hat{w}, \hat{\theta})\|_2 \le 1$ and $\|b\tilde{\theta} - A\tilde{w}\| \le \max\{\|b\|, \|A\|\}\left(\|\tilde{w}\|_2 + |\tilde{\theta}|\right) \le \|d\|\left(\|\tilde{w}\|_2 + |\tilde{\theta}|\right)$. Therefore,

$$
\begin{aligned}
\frac{R_2}{r_2} &\le 1 + \frac{\rho(d) + \|d\|(\sqrt{2}\|(\tilde{w}, \tilde{\theta})\|_2)}{\tau'\rho(d)} \\[2ex]
&\le 1 + \frac{\rho(d) + \sqrt{2}\|d\|}{\tau'\rho(d)} & \text{(since } \|(\tilde{w}, \tilde{\theta})\|_2 = 1 \text{ from (23))} \\[2ex]
&\le 1 + \frac{2\rho(d) + 2\sqrt{2}\|d\|}{\beta^*\rho(d)} & \text{(from (24))} \\[2ex]
&\le \frac{\rho(d) + 2\rho(d) + 2\sqrt{2}\|d\|}{\beta^*\rho(d)} & \text{(since } \beta^* \le 1\text{)} \\[2ex]
&\le \frac{(3 + 2\sqrt{2})}{\beta^*}\mathcal{C}(d) & \text{(since } \mathcal{C}(d) \ge 1\text{)} \\[2ex]
&\le \frac{6}{\beta^*}\mathcal{C}(d),
\end{aligned}
$$

completing the proof. $\blacksquare$

As a consequence of Lemmas 6.1 and 6.2 we have:

Corollary 6.1 *Suppose that the feasible region of $CP(d)$ is nonempty, i.e., $X_d \ne \emptyset$, and that both C_X and C_Y are regular cones. If $\rho(d) > 0$, then there exists $(\hat{w}, \hat{\theta}) \in X_d^h$ and positive scalars r_3 and R_3 satisfying:*

$$
(i) \quad B_2\left((\hat{w}, \hat{\theta}), r_3\right) \subset X_d^h,
$$

$$
(ii) \quad \|(\hat{w}, \hat{\theta})\|_2 + r_3 \le R_3,
$$

$$
(iii) \quad \frac{R_3}{r_3} \le \frac{12}{\bar{\beta}^*\beta^*}\mathcal{C}(d).
$$

Proof: Let $S = \{(w, \theta) \in X \times \Re \mid b\theta - Aw \in C_Y\}$ and $T = \{(w, \theta) \in X \times \Re \mid w \in C_X, \theta > 0\}$. Then $S \cap T = X_d^h$. From Lemma 6.1, there exists $(\hat{w}, \hat{\theta})_1 \in X_d^h$ and r_1, R_1 satisfying conditions $(i) - (iii)$ of Lemma 6.1. From Lemma 6.2, there exists $(\hat{w}, \hat{\theta})_2 \in X_d^h$ and r_2, R_2 satisfying conditions $(i) - (iii)$ of Lemma 6.2. Then the conditions of Proposition A.1 of the Appendix are satisfied, and so there exists $(\hat{w}, \hat{\theta})$

and r_3, R_3 satisfying the five conditions of Proposition A.1. From (i) of Proposition A.1, $B_2\left((\hat{w}, \hat{\theta}), r_3\right) \subset S \cap T = X_d^h$, which is (i) of the Corollary. Also from (ii) of Proposition A.1, $B_2\left((\hat{w}, \hat{\theta}), r_3\right) \subset B_2(0, R_3)$, which is (ii) of the Corollary. From (iii) of Proposition A.1, we have

$$\frac{R_3}{r_3} \leq 2 \max\left\{\frac{R_1}{r_1}, \frac{R_2}{r_2}\right\} \leq \frac{12}{\bar{\beta}^* \beta^*} \mathcal{C}(d)$$

(invoking Lemma 6.1 (iii) and Lemma 6.2 (iii)), which is (iii) of the Corollary. ∎

We now prove Lemmas 4.1, 4.2, and 4.3.

Proof of Lemma 4.1: Invoking Corollary 6.1, let $(\hat{w}, \hat{\theta})$ and R_3, r_3 satisfy the conditions in Corollary 6.1. From the positive homogeneity of the set X_d^h, we can rescale $(\hat{w}, \hat{\theta})$ and R_3, r_3 and so we presume that $R_3 = 1$. Now let

$$\gamma = \frac{\beta^* \bar{\beta}^*}{12 \mathcal{C}(d)}. \tag{31}$$

Then $r_3 \geq \gamma$ from (iii) of Corollary 6.1. Let $(w, \theta) \in B_2\left((\hat{w}, \hat{\theta}), \gamma\right)$. Then $(w, \theta) = (\hat{w}, \hat{\theta}) + (f, t)$ where $\|f, t)\|_2 \leq \gamma$, and so

$$c^T w - (z^*(d) - \epsilon)\theta = c^T \hat{w} - (z^*(d) - \epsilon)\hat{\theta} + c^T f - (z^*(d) - \epsilon)t$$

$$\geq c^T \hat{w} - (z^*(d) - \epsilon)\hat{\theta} - \gamma \|(c, z^*(d) - \epsilon)\|_2,$$

and so from parts (i) and (ii) of Corollary 6.1 we have

$$B_2\left((\hat{w}, \hat{\theta}), \gamma\right) \subset$$

$$\left\{(w, \theta) \in X_d^h \mid \|(w, \theta)\|_2 \leq 1, c^T w - (z^*(d) - \epsilon)\theta \geq c^T \hat{w} - (z^*(d) - \epsilon)\hat{\theta} - \gamma \|(c, z^*(d) - \epsilon)\|_2\right\}. \tag{32}$$

Let (w^*, θ^*) be any optimal solution of $HP(d)$, and again by homogeneity, we can presume that $\|(w^*, \theta^*)\|_2 = 1$. Then $x^* := w^*/\theta^*$ is an optimal solution of $CP(d)$, and from Theorem 1.1 of Renegar [16], $\|x^*\|_2 \leq \mathcal{C}(d)^2$. It then follows that

$$(\theta^*)^2 = \frac{\|w^*\|_2^2}{\|x^*\|_2^2} \geq \frac{\|w^*\|_2^2}{\mathcal{C}(d)^4} = \frac{1 - (\theta^*)^2}{\mathcal{C}(d)^4},$$

which implies that

$$\theta^* \geq \frac{1}{\sqrt{1 + \mathcal{C}(d)^4}} \geq \frac{1}{\sqrt{2}\mathcal{C}(d)^2}, \tag{33}$$

since $\mathcal{C}(d) \geq 1$. This also implies that

$$c^T w^* - (z^*(d) - \epsilon)\theta^* = \epsilon\theta^* \geq \frac{\epsilon}{\sqrt{2}\mathcal{C}(d)^2}.$$

Therefore

$$(w^*, \theta^*) \in \left\{(w, \theta) \in X_d^h \mid \|(w, \theta)\|_2 \leq 1, c^T w - (z^*(d) - \epsilon)\theta \geq \frac{\epsilon}{\sqrt{2}\mathcal{C}(d)^2}\right\}. \tag{34}$$

Let δ_1 and δ_2 denote the scalars on the right-hand side of the inequalities in (32) and (34), respectively, i.e.,

$$\delta_1 = c^T \hat{w} - (z^*(d) - \epsilon)\hat{\theta} - \gamma \|(c, z^*(d) - \epsilon)\|_2 \quad \text{and} \quad \delta_2 = \frac{\epsilon}{\sqrt{2}\mathcal{C}(d)^2}. \qquad (35)$$

Combining (32) and (34), we have for any $\alpha \in [0, 1]$,

$$B_2\left((\alpha(\hat{w}, \hat{\theta}) + (1 - \alpha)(w^*, \theta^*), \alpha\gamma\right) \subset$$
$$\left\{(w, \theta) \in X_d^h \mid \|(w, \theta)\|_2 \leq 1, c^T w - (z^*(d) - \epsilon)\theta \geq \alpha\delta_1 + (1 - \alpha)\delta_2\right\}. \qquad (36)$$

Setting

$$\alpha = \frac{\delta_2}{\delta_1^- + \delta_2}, \qquad (37)$$

(where a^- is the negative part of a), then $\alpha \in (0, 1]$, and

$$\alpha\delta_1 + (1 - \alpha)\delta_2 = \frac{\delta_2\delta_1 + \delta_2\delta_1^-}{\delta_1^- + \delta_2} \geq 0. \qquad (38)$$

Now let $(\breve{w}, \breve{\theta}) = \alpha(\hat{w}, \hat{\theta}) + (1 - \alpha)(w^*, \theta^*)$ and $\breve{r} = \alpha\gamma$. Then from (36) and (38) we have

$$B_2\left((\breve{w}, \breve{\theta}), \breve{r}\right) \subset \left\{(w, \theta) \in X_d^h \mid \|(w, \theta)\|_2 \leq 1, c^T w - (z^*(d) - \epsilon)\theta \geq 0\right\}.$$

This proves the first part of the Lemma. For the second part, we have

$$\frac{1}{\breve{r}} = \frac{1}{\alpha}\frac{1}{\gamma} \;=\; \frac{1}{\gamma}\left(1 + \frac{\delta_1^-}{\delta_2}\right) \qquad \text{(from (37))}$$

$$= \frac{1}{\gamma}\left(1 + \frac{\sqrt{2}\mathcal{C}(d)^2}{\epsilon}\delta_1^-\right) \qquad \text{(from (35))}$$

$$\leq \frac{1}{\gamma\epsilon}\left(\epsilon + \sqrt{2}\mathcal{C}(d)^2\left[c^T\hat{w} - (z^*(d) - \epsilon)\hat{\theta} - \gamma\|(c, z^*(d) - \epsilon)\|_2\right]^-\right) \qquad \text{(from (35))}$$

$$\leq \frac{1}{\gamma\epsilon}\left(\epsilon + \sqrt{2}\mathcal{C}(d)^2\left[\|(c, z^*(d) - \epsilon)\|_2\left(\|\hat{w}, \hat{\theta}\|_2 + \gamma\right)\right]\right)$$

$$\leq \frac{1}{\gamma\epsilon}\left(\epsilon + \sqrt{2}\mathcal{C}(d)^2\|(c, z^*(d) - \epsilon)\|_2\right)$$

where the last inequality follows from (ii) of Corollary 6.1 and the fact that $\gamma \leq r_3$. From the last inequality we obtain

$$\frac{1}{\breve{r}} \leq \frac{1}{\gamma\epsilon}\left(\epsilon + \epsilon\sqrt{2}\mathcal{C}(d)^2 + \sqrt{2}\mathcal{C}(d)^2\|c\|_2 + \sqrt{2}\mathcal{C}(d)^2|z^*(d)|\right).$$

However, $\|c\|_2 \leq \|d\|$, and from Theorem 1.1 of Renegar [16] we have $|z^*(d)| \leq \|d\|\mathcal{C}(d)$, and so

$$\frac{1}{\check{r}} \leq \frac{1}{\gamma \epsilon} \left(\epsilon + \epsilon\sqrt{2}\mathcal{C}(d)^2 + \sqrt{2}\mathcal{C}(d)^2\|d\| + \sqrt{2}\mathcal{C}(d)^3\|d\| \right).$$

Using the fact that $\mathcal{C}(d) \geq 1$ yields

$$\frac{1}{\check{r}} \leq \frac{\mathcal{C}(d)^3}{\gamma \epsilon} \left(\frac{\epsilon(1+\sqrt{2})}{\mathcal{C}(d)} + 2\sqrt{2}\|d\| \right) \leq \frac{(1+3\sqrt{2})\mathcal{C}(d)^3}{\gamma} \max \left\{ \frac{1}{\mathcal{C}(d)}, \frac{\|d\|}{\epsilon} \right\},$$

and so

$$\check{r} \geq \frac{1}{(1+3\sqrt{2})} \frac{\gamma}{\mathcal{C}(d)^3} \min \left\{ \mathcal{C}(d), \frac{\epsilon}{\|d\|} \right\}.$$

Substituting the value of γ from (31) completes the proof. $\blacksquare$

Proof of Lemma 4.2: The proof follows identically as in the proof of Lemma 4.1, except that we invoke Lemma 6.2 instead of Corollary 6.1, and the value of γ in (31) is replaced by

$$\gamma = \frac{\beta^*}{6\mathcal{C}(d)}.$$

The rest of the proof is identical to that of Lemma 4.1. $\blacksquare$

Proof of Lemma 4.3: The proof also follows identically as in the proof of Lemma 4.1, except that we invoke Lemma 6.1 instead of Corollary 6.1, and the value of γ in (31) is replaced by

$$\gamma = \frac{\bar{\beta}^*}{6\mathcal{C}(d)}.$$

The rest of the proof is identical to that of Lemma 4.1. $\blacksquare$

APPENDIX

1. Characterization Results for $\rho_P(d)$

We state two previously known characterization results for $\rho_P(d)$. First, consider the following program:

$P_r(d)$:

$$
\begin{array}{lll}
r(d) = & \text{minimum} & \text{maximum} & \delta \\
& v \in Y & w, \theta, \delta \\
\end{array}
$$

$$
\begin{array}{llllll}
\|v\| \le 1 & \quad s.t. & b\theta & - & Aw & - & v\delta & \in C_Y & (39) \\
& & & & w & & & \in C_X \\
& & \theta & & & & & \ge 0 \\
& & |\theta| & + & \|w\| & & & \le 1.
\end{array}
$$

Then $r(d)$ is the largest scaling factor δ such that for any v with $\|v\| \le 1$, $v\delta$ can be added to the first inclusion of $P_r(d)$ without affecting the feasibility of the system. The following Theorem is a slightly altered restatement of Theorem 3.5 of [18].

Theorem A.1 (Theorem 3.5 of Renegar [18]) *Suppose that $d \in \mathcal{F}_P$. Then*

$$
r(d) = \rho_P(d). \tag{40}
$$

Second, suppose that the cone C_Y is a regular cone, whereby C_Y^* is also a regular cone. Let $\bar{z}$ be the norm approximation vector for the cone C_Y^* (see Proposition 2.1), and consider the following program:

$P_v(d)$:

$$
\begin{array}{ll}
v(d) = & \text{minimum} \qquad \|w\| + |\theta| \\
& w, \theta
\end{array}
$$

$$
\begin{array}{l}
b\theta - Aw - \bar{z} \in C_Y \\
w \in C_X \\
\theta \ge 0
\end{array} \tag{41}
$$

The next Theorem is a restatement of Theorem 3.5 of [7].

Theorem A.2 (Theorem 3.5 of [7]) *Suppose that $d \in \mathcal{F}_P$ and C_Y is regular, and that $\rho_P(d) > 0$. Then*

$$
\frac{\bar{\beta}^*}{v(d)} \le \rho_P(d) \le \frac{1}{v(d)}, \tag{42}
$$

where $\bar{\beta}^*$ is the coefficient of linearity for the cone C_Y^*.

2. A Construction Using Inscribed Balls and Intersecting Sets

We state the following result, which is Proposition A.2 of the Appendix of [7]:

Proposition A.1 *Let X be a finite-dimensional normed linear vector space with norm $\| \cdot \|$ and let S and T be convex subsets of X. Suppose that*

(i) $\hat{x}_1 \in S \cap T$, $B(\hat{x}_1, r_1) \subset S$, where $r_1 > 0$, and $B(\hat{x}_1, r_1) \subset B(0, R_1)$ and

(ii) $\hat{x}_2 \in S \cap T$, $B(\hat{x}_2, r_2) \subset T$, where $r_2 > 0$, and $B(\hat{x}_2, r_2) \subset B(0, R_2)$.
Let $\alpha = \dfrac{r_2}{r_1 + r_2}$, and $r_3 = \dfrac{r_1 r_2}{r_1 + r_2}$, and $R_3 = \alpha R_1 + (1 - \alpha) R_2$. Then the point $\hat{x} = \alpha \hat{x}_1 + (1 - \alpha) \hat{x}_2$ will satisfy:

$$(i) \quad B(\hat{x}, r_3) \subset S \cap T \ ,$$

$$(ii) \quad B(\hat{x}, r_3) \subset B(0, R_3) \ ,$$

$$(iii) \quad \frac{R_3}{r_3} \le 2 \max \left\{ \frac{R_1}{r_1}, \ \frac{R_2}{r_2} \right\} \ ,$$

$$(iv) \quad r_3 \ge \frac{1}{2} \min\{r_1, r_2\} \ ,$$

$$and \quad (v) \quad R_3 \le \max\{R_1, R_2\} \ .$$

References

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

[2] Mokhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. *Nonlinear Programming, Theory and Algorithms*. John Wiley & Sons, Inc, New York, second edition, 1993.

[3] Dimitris Bertsimas and John Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, Belmont, Massachusetts, 1997.

[4] R. Bland, D. Goldfarb, and M.J. Todd. The ellipsoid method: a survey. *Operations Research*, 29:1039–1091, 1981.

[5] A. Charnes and W.W. Cooper. Programming with linear fractionals. *Naval Research Quarterly*, 9:181–186, 1962.

[6] B.D. Craven and B. Mond. The dual of a fractional linear program. *Journal of Mathematical Analysis and Applications*, 42:507–512, 1973.

[7] Robert M. Freund and Jorge R. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition number of a conic linear system. Technical Report W.P. #3862-95-MSA, Sloan School of Management, Massachusetts Institute of Technology, 1995.

[8] M. Grötschel, L. Lovasz, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1:169–197, 1981.

[9] L.G. Khachian. A polynomial algorithm in linear programming. *Soviet Math Dokl.*, 20(1):191–194, 1979.

[10] A.S. Nemirovskii and D.B. Yudin. Infomational complexity and efficient methods for solution of complex extremal problems. *Matekon*, 13:25–45, 1977.

[11] Y. Nesterov, M. Todd, and Y. Ye. Primal-dual methods and infeasibility detectors for nonlinear programming problems. Technical Report TR 1156, Dept. of IE&OR, Cornell University, 1996.

[12] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994.

[13] Yurii Nesterov and M.J. Todd. Primal-dual interior-point methods for self-scaled cones. Technical Report TR1125, Dept. of IE&OR, Cornell University, 1995.

[14] Yurii Nesterov and M.J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research*, 22:1–42, 1997.

[15] Manuel Nunez. *Condition Numbers and Properties of Central Trajectories in Convex Programming*. PhD thesis, Massachusetts Institute of Technology, 1997.

[16] James Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65(1):73–91, 1994.

[17] James Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM Journal on Optimization*, 5(3):506–524, 1995.

[18] James Renegar. Linear programming, complexity theory, and elementary functional analysis. *Mathematical Programming*, 70(3):279–351, 1995.

[19] James Renegar. Condition numbers, the barrier method, and the conjugate gradient method. *SIAM Journal on Optimization*, 64(4):879–912, 1996.

[20] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[21] Jorge Vera. Ill-posedness and the complexity of deciding existence of solutions to linear programs. *SIAM Journal on Optimization*, 6(3):549–569, 1996.

[22] Jorge R. Vera. On the complexity of linear programming under finite precision arithmetic. Technical Report, University of Chile, August 1994. To appear in *Mathematical Programming*.