SOME CHARACTERIZATIONS AND PROPERTIES OF THE "DISTANCE TO ILL-POSEDNESS" AND THE CONDITION MEASURE OF A CONIC LINEAR SYSTEM [1]

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

Jorge R. Vera[3]
University of Chile
and
CORE, Catholic University of Louvain

October, 1995

Abstract

A conic linear system is a system of the form

$$P: \quad \text{find } x \text{ that solves } b - Ax \in C_Y,\ x \in C_X,$$

where C_X and C_Y are closed convex cones, and the data for the system is $d = (A, b)$. This system is"well-posed" to the extent that (small) changes in the data (A, b) do not alter the status of the system (the system remains solvable or not). Intuitively, the more well-posed the system is, the easier it should be to solve the system or to demonstrate its infeasibility via a theorem of the alternative. Renegar defined the "distance to ill-posedness," $\rho(d)$, to be the smallest distance of the data $d = (A, b)$ to other data $\bar{d} = (\bar{A}, \bar{b})$ for which the system P is "ill=posed," i.e., $\bar{d} = (\bar{A}, \bar{b})$ is in the intersection of the closure of feasible and infeasible instances $d' = (A', b')$ of P. Renegar also defined the "condition measure" of the data instance d as $\mathcal{C}(d) \underline{\underline{\Delta}} \|d\|/\rho(d)$, and showed that this measure is a natural extension of the familiar condition measure associated with systems of linear equation. This study presents two categories of results related to $\rho(d)$, the distance to ill-posedness, and $\mathcal{C}(d)$, the condition measure of d. The first category of results involves the approximation of $\rho(d)$ as the optimal value of certain mathematical programs. We present ten different mathematical programs each of whose optimal values provides an approximation of $\rho(d)$ to within certain constant factors, depending on whether P is feasible or not. The second category of results involves the existence of certain inscribed and intersecting balls involving the feasible region of P or the feasible region of its alternative system, in the spirit of the ellipsoid algorithm. These results roughly state that the feasible region of P (or its alternative system when P is not feasible) will contain a ball of radius r that is itself no more than a distance R from the origin, where the ratio R/r satisfies $R/r \le O(n\ \mathcal{C}(d))$, and such that $r \ge \Omega\left(\frac{1}{n\ \mathcal{C}(d)}\right)$ and $R \le O(n\ \mathcal{C}(d))$, where n is the dimension of the feasible region. Therefore the condition measure $\mathcal{C}(d)$ is a relevant tool in proving the existence of an inscribed ball in the feasible region of P that is not too far from the origin and whose radius is not too small.

AMS Subject Classification: 90C, 90C05, 90C60
Keywords: Complexity of Linear Programming, Infinite Programming, Interior Point Methods, Conditioning, Error Analysis

---

[1]This research has been partially supported through a grant from FONDECYT (project number 1930948), and by a research fellowship from CORE, Catholic University of Luovain.

[2]Sloan School of Management, M.I.T., 50 Memorial Drive, Cambridge, MA 02142, 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

This paper is concerned with characterizations and properties of the "distance to ill-posedness" and of the condition measure of a conic linear system, i.e., a system of the form:

P:

$$b - Ax \in C_Y \ , \ x \in C_X \qquad , \tag{1.1}$$

where $C_X \subset X$ and $C_Y \subset Y$ are each a closed convex cone in the (finite) n-dimensional normed 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 $A : X \longrightarrow Y$. At the moment, we make no assumptions on C_X and C_Y except that each is a closed convex cone. The reader will recognize immediately that when $X = R^n$ and $Y = R^m$, and either (i) $C_X = \{x \in R^n | x \geq 0\}$ and $C_Y = \{y \in R^m | y \geq 0\}$, (ii) $C_X = \{x \in R^n | x \geq 0\}$ and $C_Y = \{0\} \subset R^m$ or (iii), $C_X = R^n$ and $C_Y = \{y \in R^m | y \geq 0\}$, then (1.1) is a linear inequality system of the format (i) $Ax \leq b, x \geq 0$, (ii) $Ax = b, x \geq 0$, or (iii) $Ax \leq b$, respectively.

The problem P is a very general format for studying the feasible region of a mathematical program, and even lends itself to analysis by interior-point methods, see Nesterov and Nemirovskii[9] and Renegar [13].

The concept of the "distance to ill-posedness" and a closely related condition measure for problems such as P was introduced by Renegar in [11] in a more specific setting, but then generalized more fully in [12] and in [13]. We now present the development of these two concepts in detail.

We denote by $d = (A, b)$ the "data" for the problem (1.1). That is, we regard the cones C_X and C_Y as fixed and given, and the data for the problem is the linear operator A together with the vector b. We denote the set of solutions of P as X_d to emphasize the dependence on the data d, i.e.,

$$X_d = \{x \in X | b - Ax \in C_Y, x \in C_X\}.$$

We define

$$\mathcal{F} = \{(A,b) \in L(X,Y) \times Y | \text{ there exists x satisying } b - Ax \in C_Y, x \in C_X\} \ . \tag{1.2}$$

Then $\mathcal{F}$ corresponds to those data instances (A, b) for which P is consistent, i.e., (1.1) has a solution.

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

$$\|d\| = \|(A,b)\| = \max\{\|A\|, \|b\|\} \tag{1.3}$$

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

$$\begin{aligned} \|A\| = \quad & \max \quad \|Ax\| \\ & s.t. \quad \|x\| \leq 1 \quad . \end{aligned} \tag{1.4}$$

We denote the complement of $\mathcal{F}$ by $\mathcal{F}^{\mathcal{C}}$. Then $\mathcal{F}^{\mathcal{C}}$ consists precisely of those data instances $d = (A, b)$ for which P is inconsistent.

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

$$\mathcal{B} = cl(\mathcal{F}) \cap cl(\mathcal{F}^{\mathcal{C}}) \tag{1.5}$$

where $cl(S)$ is the closure of a set S. Note that if $d = (A, b) \in \mathcal{B}$, then (1.1) is ill-posed in the sense that arbitrary small changes in the data $d = (A, b)$ will yield consistent instances of (1.1) as well as inconsistent instances of (1.1).

For any $d = (A, b) \in L(X, Y) \times Y$, we define

$$\begin{array}{rllcll} \rho(d) = & \inf\limits_{\bar{d}} & \|d - \bar{d}\| & = & \inf\limits_{\bar{A}, \bar{b}} & \|(A, b) - (\bar{A}, \bar{b})\| \\ & s.t. & \bar{d} \in \mathcal{B} & & s.t. & (\bar{A}, \bar{b}) \in cl(\mathcal{F}) \cap cl(\mathcal{F}^{\mathcal{C}}) \quad . \end{array} \tag{1.6}$$

Then $\rho(d)$ is the "distance to ill-posedness" of the data d, i.e., $\rho(d)$ is the distance of d from the set $\mathcal{B}$ of ill-posedness instances. In addition to the work of Renegar cited earlier, further analysis of the distance to ill-posedness has been studied by Vera [16], [17], [18], Filipowski [6], [7], and recently by Nunez and Freund [10].

In addition to the general case (1.1), we will also be interested in two special cases when one of the cones is either the entire space or only the zero-vector. Specifically, if $C_Y = \{0\}$, then (1.1) specifies to

$$Ax = b, \ x \in C_X \quad . \tag{1.7}$$

When $C_X = X$, then (1.1) specifies to

$$b - Ax \in C_Y, \ x \in X \quad . \tag{1.8}$$

One of the purposes of this paper is to explore characterizations of the distance to ill-posedness $\rho(d)$ as the optimal value of a mathematical program whose solution is relatively easy to obtain. By "relatively easy," we roughly mean that such a program is either a convex program or is solveable through $O(m)$ or $O(n)$ convex programs. Vera [16] and [17] explored such characterizations for linear programming problems, and the results herein expand the scope of this line of research in two ways: first by expanding the problem context from linear equations and linear inequalities to conic linear systems, and second by developing more efficient mathematical programs that characterize $\rho(d)$. Renegar [13] presents a characterization of the distance to ill-posedness as the solution of a certain mathematical program, but this characterization is not in general easy to solve. There are a number of reasons for exploring characterizations of $\rho(d)$, not the least of which is to better understand the underlying nature of $\rho(d)$. There is the intellectual issue of the complexity of computing $\rho(d)$ or an approximation thereof, and there is also the prospect for using such characterizations to further understand the behavior of the underlying problem P. Finally, as is shown in [17], when $\rho(d)$ can be computed efficiently, then there is promise that the problem of deciding the feasibility of P or the infeasibility of P can be processed with a "fully efficient" algorithm, see [17] or Renegar [12] for details of the concept of a fully efficient algorithm. In Section 3 of this paper, we present ten different mathematical programs each of whose optimal values provides an approximation of $\rho(d)$ to within certain constant factors, depending on whether P is feasible or not, and where the constants depend only on the "structure" of the cones C_X and C_Y and not on the dimension or on the data $d = (A, b)$.

The second purpose of this paper is to prove the existence of certain inscribed and intersecting balls involving the feasible region of P (or the feasible region of the alternative system of P if P is infeasible), in the spirit of the ellipsoid algorithm and in order to set the stage for an analysis of the ellipsoid algorithm, hopefully in a subsequent paper. Recall that when P is specified to the case of non-degenerate linear inequalities and the data $d = (A, b)$ is an array of rational numbers of bitlength L, that the feasible region of P will intersect a ball of radius R centered at the origin, and will contain a ball of radius r where $r = (1/n)2^{-L}$ and $R = n2^L$. Furthermore, the ratio R/r is of critical importance in the analysis of the complexity of using the ellipsoid algorithm to solve the system P in this particular case. (For the general case of P the Turing machine model of computation is not very appropriate for analyzing issues of complexity, and indeed other models of computation have been proposed (see Blum et.al. [3], also Smale [14].))

By analogy to the properties of rational non-degenerate linear inequalities mentioned above, Renegar [13] has shown that the feasible region X_d must intersect a ball of radius R centered at the origin where $R \leq \|d\|/\rho(d)$. Renegar [12] defines the *condition measure* of the data $d = (A, b)$ to be $\mathcal{C}(d)$:

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

and so $R \leq \mathcal{C}(d)$. Here we see the value 2^L has been replaced by the condition measure $\mathcal{C}(d)$.

For the problem P considered herein in (1.1), the feasible region is the set X_d. In Sections 4 and 5 of this paper, we utilize the characterization results of Section 3 to prove that the feasible region X_d (or the feasible region of the alternative system when P is infeasible) must contain an inscribed ball of radius r that is no more than a distance R from the origin, and where the ratio R/r must satisfy $R/r \leq O(n\,\mathcal{C}(d))$. Furthermore, we prove that $r \geq \Omega\left(\frac{1}{n\,\mathcal{C}(d)}\right)$ and $R \leq O(n\,\mathcal{C}(d))$ (and where n is replaced by m for the alternative system for the case when P is infeasible). Note that by analogy to rational non-degenerate linear inequalities, that the quantity 2^L is replaced by $\mathcal{C}(d)$. Therefore the condition measure $\mathcal{C}(d)$ is a very relevant tool in proving the existence of an inscribed ball in the feasible region of P that is not too far from the origin and whose radius is not too small. This should prove effective in the analysis of the ellipsoid algorithm as applied to solving P.

The paper is organized as follows. Section 2 contains preliminary results, definitions, and analysis. Section 3 contains the ten different mathematical programs each of whose optimal values provides an approximation of $\rho(d)$ to within certain constant factors, as discussed earlier. Section 4 contains four Lemmas that give partial or full characterizations of certain inscribed and intersecting balls related to the feasible region of P (or its alternative region in the case when P is infeasible). Section 5 presents a synthesis of all of the results in the previous two sections into theorems that give a complete treatment both of the characterization results and of the inscribed and intersecting ball results.

2 Preliminaries and Some More Notation

We will 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 $\bar{y} \in Y$.

For $\bar{d} = (\bar{A}, \bar{b}) \in L(X, Y) \times Y$, we define the ball

$$B(\bar{d}, r) = \{d = (A, b) \in L(X, Y) \times Y \mid \|d - \bar{d}\| \le r\}.$$

With this additional notation, it is easy to see that the definition of $\rho(d)$ given in (1.6) is equivalent to:

$$\rho(d) = \begin{cases} \sup\{\delta | B(d, \delta) \subset \mathcal{F}\} & \text{if } d \in \mathcal{F} \\ \\ \sup\left\{\delta | B(d, \delta) \subset \mathcal{F}^C\right\} & \text{if } d \in \mathcal{F}^C \end{cases} \tag{2.1}$$

We associate with X and Y the dual spaces X^* and Y^* of linear functionals defined on X and Y, respectively, and whose (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 C is a convex cone in X, C^* will denote the dual convex cone defined by

$$C^* = \{z \in X^* | z^T x \ge 0 \text{ for any } x \in C\} \quad .$$

Remark 2.1 *If we identify $(X^*)^*$ with X, then $(C^*)^* = C$ whenever C is a closed convex cone.*

Remark 2.2 *If $C_X = X$, then $C_X^* = \{0\}$. If $C_X = \{0\}$, then $C_X^* = X$.*

We will say that a cone C is *regular* if C is a closed convex cone, has a nonempty interior and is pointed (i.e., contains no line).

Remark 2.3 *C is regular if and only if C^* is regular.*

We denote the set of real numbers by R and the set of nonnegative real numbers by R_+.

Regarding the consistency of (1.1), we have the following partial "theorem of the alternative," the proof of which is a straightforward exercise using a separating hyperplane argument.

Proposition 2.1 *If (1.1) has no solution, then the system (2.2) has a solution:*

$$\begin{array}{c} A^T y \in C_X^* \\ y \in C_Y^* \\ y^T b \leq 0 \\ y \neq 0. \end{array} \tag{2.2}$$

If the system (2.3) has a solution:

$$\begin{array}{c} A^T y \in C_X^* \\ y \in C_Y^* \\ y^T b < 0, \end{array} \tag{2.3}$$

then (1.1) has no solution. ∎

Using Proposition 2.1, it is elementary to prove the following:

Lemma 2.1 *Consider the set of ill-posed instances $\mathcal{B}$. Then $\mathcal{B}$ can be characterized as:*

$$\begin{aligned} \mathcal{B} = \quad & \{d = (A, b) \in L(X, Y) \times Y \mid \text{there exists } (x, r) \in X \times R \text{ with} \\ & (x, r) \neq 0 \text{ and } y \in Y^* \text{ with } y \neq 0 \text{ satisfying } br - Ax \in C_Y, x \in C_X, \\ & y \in C_Y^*, A^T y \in C_X^*, \text{ and } y^T b \leq 0\} \end{aligned}$$ ∎

We now recall some facts about norms. Given a finite dimensional linear vector space X endowed with a norm $\|x\|$ for $x \in X$, the dual norm induced on the space X^* is denoted by $\|z\|_*$ for $z \in X^*$, and is defined as:

$$\begin{aligned} \|z\|_* = \quad & \max_{x} \quad z^T x \\ & s.t. \quad \|x\| \leq 1 \quad . \end{aligned} \tag{2.4}$$

If we denote the unit balls in X and X^* by B and B^*, then it is straightforward to verify that

$$B = \{x \in X \mid \|x\| \leq 1\} = \{x \in X \mid z^T x \leq 1 \text{ for all } z \text{ with } \|z\|_* \leq 1\}, \tag{2.5}$$

and

$$B^* = \{z \in X^* \mid \|z\|_* \leq 1\} = \{z \in X^* \mid z^T x \leq 1 \text{ for all } x \text{ with } \|x\| \leq 1\}. \tag{2.6}$$

Furthermore,

$$z^T x \leq \|z\|_* \|x\| \text{ for any } x \in X \text{ and } z \in X^*, \tag{2.7}$$

which is the Hölder inequality. Finally, note that if $A = uv^T$, then it is easy to derive that $\|A\| = \|v\|_* \|u\|$ using (2.4) and (1.4).

If X and V are finite-dimensional normed linear vector spaces with norm $\|x\|$ for $x \in X$ and norm $\|v\|$ for $v \in V$, then for $(x, v) \in X \times V$, the function $f(x, v)$ defined by

$$f(x, v) = \|(x, v)\| \triangleq \|x\| + \|v\| \tag{2.8}$$

defines a norm on $X \times V$, whose dual norm is given by

$$\|(w,u)\|_* \triangleq \max\left\{\|w\|_*, \|u\|_*\right\} \text{ for } (w,u) \in (X \times V)^* = X^* \times V^*. \tag{2.9}$$

The following result is a special case of the Hahn-Banach Theorem, see for example [19]:

Proposition 2.2 *For every $x \in X$, there exists $z \in X^*$ with the property that $\|z\|_* = 1$ and $\|x\| = z^T x$.*

Proof: If $x = 0$, then any $z \in X^*$ with $\|z\|_* = 1$ will satisfy the statement of the proposition. Therefore, we suppose that $x \neq 0$. Consider $\|x\|$ as a function of x, i.e., $f(x) = \|x\|$. Then $f(\cdot)$ is a real-valued convex function, and so the subdifferential operator $\partial f(x)$ is non-empty for all $x \in X$, see [2]. Consider any $x \in X$, and let $z \in \partial f(x)$. Then

$$f(w) \geq f(x) + z^T(w - x) \text{ for any } w \in X. \tag{2.10}$$

Substituting $w = 0$ we obtain $\|x\| = f(x) \leq z^T x$. Substituting $w = 2x$ we obtain $2f(x) = f(2x) \geq f(x) + z^T(2x - x)$, and so $f(x) \geq z^T x$, whereby $f(x) = z^T x$. From (2.7) it then follows that $\|z\|_* \geq 1$. Now if we let $u \in X$ and set $w = x + u$, we obtain from (2.10) that $f(u) + f(x) \geq f(u + x) = f(w) \geq f(x) + z^T(w - x) = f(x) + z^T(u + x - x) = f(x) + z^T u$. Therefore, $z^T u \leq f(u) = \|u\|$, and so from (2.4) we obtain $\|z\|_* \leq 1$. Therefore, $\|z\|_* = 1$. ∎

Because X and Y are are normed linear vector spaces of finite dimension, all norms are equivalent. Thus we can specify a norm for X and a norm for Y if we so desire. If $X = R^n$, the L_p norm is given by

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}, \tag{2.11}$$

for $p \geq 1$. 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 $p \to +\infty$.

A critical component of our analysis concerns the extent to which the norm function $\|x\|$ can be approximated by a linear function $u^T x$ over the cone C_X for some $u \in C_X^*$, and the extent to which the norm function $\|y\|_*$ can be approximated by a linear function $z^T y$ over the cone C_Y^* for some $z \in C_Y$. We now define two important constants that relate the extent to which these norms $\|x\|$ and $\|y\|_*$ can be approximated by linear functions over the convex cones C_X and C_Y^*, respectively.

Definition 2.1 *(i) If C_X is regular, let*

$$\beta = \sup_{\substack{u \in X^* \\ \|u\|_* = 1}} \inf_{\substack{x \in C_X \\ \|x\| = 1}} u^T x \;. \tag{2.12}$$

(ii) If C_Y is regular, let

$$\beta^* = \sup_{\substack{z \in Y \\ \|z\| = 1}} \inf_{\substack{y \in C_Y^* \\ \|y\|_* = 1}} y^T z \;. \tag{2.13}$$

Examining (2.12) in detail, let $\bar{u}$ denote that value of $u \in X^*$ that achieves the supremum in (2.12). Then for all $x \in C_X$, $\beta\|x\| \le \bar{u}^T x \le \|x\|$, and so $\|x\|$ is approximated by the linear function $\bar{u}^T x$ to within the factor β over the cone C_X. Therefore, β measures the extent to which $\|x\|$ can be approximated by a linear function $\bar{u}^T x$ on the cone C_X. Furthermore, $\bar{u}^T x$ is the "best" such linear approximation of $\|x\|$ over this cone. If we let $\bar{z}$ denote that value of $z \in Y$ that achieves the supremum in (2.13), then similar remarks pertain, and so $\bar{z}^T y$ is the "best" linear approximation of $\|y\|_*$ on the cone C_Y^*. (It is easy to see that $\beta \le 1$ and $\beta^* \le 1$, since, for example, $u^T x \le \|u\|_* \|x\| = 1$ for u and x as in (2.12), and $y^T z \le \|y\|_* \|z\| = 1$ for y and z as in (2.13).) The larger the value of β, the more closely that $\|x\|$ is approximated by a linear function $u^T x$ over $x \in C_X$. Similar remarks pertain to β^*. We have the following properties of β and β^*:

Proposition 2.3 *(i) If C_X is regular, then $0 < \beta \le 1$, and there exists $\bar{u} \in int\ C_X^*$ such that*

$$\|\bar{u}\|_* = 1 \text{ and } \beta = min\{\bar{u}^T x | x \in C_X,\ \|x\| = 1\}.$$

(ii) If C_Y is regular, then $0 < \beta^* \le 1$, and there exists $\bar{z} \in int\ C_Y$ such that

$$\|\bar{z}\| = 1 \text{ and } \beta^* = min\{\bar{z}^T y | y \in C_Y^*,\ \|y\|_* = 1\}.$$

The proof of Proposition 2.3 follows easily from the following observation:

Remark 2.4 *Suppose K is a closed convex cone. Then $z \in \text{ int } K^*$ if and only if $z^T x > 0$ for all $x \in K/\{0\}$. Also, if $z \in \text{ int } K^*$, the set $\{x \in K | z^T x = 1\}$ is a closed and bounded convex set.*

For the remainder of this study, it is assumed that $\bar{u}$ and $\bar{z}$ of Proposition 2.3 are known and given, whenever C_X and/or C_Y are regular.

Corollary 2.1 *(i) If C_X is regular, then*

$$\beta\|x\| \le \bar{u}^T x \le \|x\| \text{ for any } x \in C_X. \tag{2.14}$$

(ii) If C_Y is regular, then

$$\beta^*\|y\|_* \le \bar{z}^T y \le \|y\|_* \text{ for any } y \in C_Y^*. \tag{2.15}$$

We also define

Definition 2.2 *(i) If C_X is regular, let*

$$\bar{\beta} = \inf_{\substack{w \in C_X^* \\ \|w\|_* = 1}} \ \sup_{\substack{x \in C_X \\ \|x\| \le 1}} w^T x \quad . \tag{2.16}$$

(ii) If C_Y is regular, let

$$\bar{\beta}^* = \inf_{\substack{w \in C_Y \\ \|w\| = 1}} \ \sup_{\substack{v \in C_Y^* \\ \|v\|_* \le 1}} v^T w \quad . \tag{2.17}$$

We offer the following interpretation of $\bar{\beta}$ and $\bar{\beta}^*$. Note that $\bar{\beta} \leq 1$ since $w^T x \leq \|w\|_* \|x\| \leq 1$ for any w and x given in (2.16). Now, from (2.4) we have that for any $w \in C_X^*$ with $\|w\|_* = 1$, that

$$1 = \|w\|_* = \sup_{\substack{x \in X \\ \|x\| \leq 1}} w^T x \geq \sup_{\substack{x \in C_X \\ \|x\| \leq 1}} w^T x \ .$$

Thus $\bar{\beta}$ represents the extent to which $\|w\|_*$ is approximated by optimizing over C_X instead of over X in the construction of $\|w\|_*$, as w ranges over all values in C_X^*. A similar interpretation for $\bar{\beta}^*$ can also be developed. We have:

Proposition 2.4

$$(i) \text{ If } C_X \text{ is regular, then } 0 < \bar{\beta} \leq 1 \ .$$

$$(ii) \text{ If } C_Y \text{ is regular, then } 0 < \bar{\beta}^* \leq 1. \ \blacksquare$$

Finally, we note that the constants, β, β^*, $\bar{\beta}$, and $\bar{\beta}^*$ depend only on the norms $\|x\|$ and $\|y\|$ and the cones C_X and C_Y, and are independent of the data (A, b) defining the problem (1.1).

We now present two families of examples that illustrate the constructions above. For the first example, let $X = R^n$ and $C_X = \{x \in R^n | x \geq 0\}$. Then we can identify X^* with X and in so doing, $C_X^* = \{x \in R^n | x \geq 0\}$ as well. If $\|x\| = \|x\|_p$, then for $x \in C_X$, it is straightforward to show that $\bar{u} = \left(n^{\frac{1}{p}-1}\right) e$, where $e = (1, \ldots, 1)^T$, i.e., the linear function given by $\bar{u}^T x$ is the "best" linear approximation of the function $\|x\|$ on the set C_X. Furthermore, straightforward calculation yields that $\beta = n^{\frac{1}{p}-1}$. Then if $p = 1$, $\beta = 1$, but if $p > 1$ then $\beta < 1$. However, regardless of the value of p, it will be true that $\bar{\beta} = 1$. To see this, note that if $w \in R^n$ and $w \geq 0$ (i.e., $w \in C_X^*$) and $\|w\|_* = 1$, then with $x = (x_1, \ldots, x_n)^T$ where

$$x_j = w_j^{\left(\frac{1}{p-1}\right)} \quad ,$$

that $w^T x = \|w\|_q = \|w\|_* = 1$ $\left(\text{where } \frac{1}{p} + \frac{1}{q} = 1\right)$ and $\|x\| = 1$. Note that $x \geq 0$, so that $\bar{\beta} = 1$.

The second example concerns the cone of symmetric positive semi-definite matrices, which has been shown to be of enormous importance in mathematical programming (see Alizadeh [1] and Nesterov and Nemiroskii [9]). Let X denote the set of real $n \times n$ symmetric matrices, and let $C_X = \{x \in X | x \text{ is positive semi} - \text{definite}\}$. Then C_X is a closed convex cone. We can identify X^* with X, and in so doing it is elementary to derive that $C_X^* = \{y \in X^* | y \text{ is postive semi} - \text{definite}\}$ i.e., C_X is self-dual. Foe $x \in X$, let $\lambda(x)$ denote the n-vector of ordered eigenvalues of x. That is, $\lambda(x) = (\lambda_1(x), \ldots, \lambda_n(x))^T$ where $\lambda_i(x)$ is the i$^{\text{th}}$ largest eigenvalue of X. For any $p \in [1, \infty)$, let the norm of x be defined by

$$\|x\| = \|x\|_p \triangleq \left(\sum_{j=1}^{n} |\lambda_j(x)|^p \right)^{\frac{1}{p}} \quad ,$$

i.e., $\|x\|_p$ is the p-norm of the vector of eigenvalues of X. (see Lewis [8] for a proof that $\|x\|_p$ is a norm.)

When $p = 2$, $\|x\|_2$ corresponds precisely to the Fröbenius norm of x. When $p = 1$, $\|x\|_1$ is the sum of the absolute values of the eigenvalues of x. Therefore, when $x \in C_X$, $\|x\|_1 = tr(x) = \sum_{i=1}^{n} x_{ii}$ where x_{ii} is the i$^{\text{th}}$ diagonal entry of the real matrix x, and so is linear on C_X. It is easy to show for the norm $\|x\|_p$ over C_X that $\bar{u} = \left(n^{\frac{1}{p}-1}\right) I$ has $\|\bar{u}\|_* = \|\bar{u}\|_q = 1$ and that $\beta = n^{\frac{1}{p}-1}$. Thus, for the Fröbenius norm we have $\beta = \frac{1}{\sqrt{n}}$ and for the L_1-norm, we have $\beta = 1$. Just as in the case of the non-negative orthant above, it can easily be shown that $\bar{\beta} = 1$ for any value of $p \in [1, \infty)$.

We conclude this section with the following:

Remark 2.5 *If C_X is regular, then it is possible to choose the norm on X in such a way that $\beta = 1$. If C_Y is regular, then it is possible to choose the norm on Y in such a way that $\beta^* = 1$.*

To see why this remark is true, recall that for finite dimensional linear vector spaces, that all norms are equivalent. Now suppose that C_X is regular. Pick any $\bar{u} \in$ int C_X^*. Then define the following norm on X:

$$\|x\| \triangleq \text{minimum} \ \left\{\bar{u}^T x' - \bar{u}^T x'' | x = x' + x'', \ x' \in C_X, \ x'' \in -C_X\right\} \ .$$

It can then easily be verified that $\|\cdot\|$ is a norm and that for all $x \in C_X$, that $\|x\| = \bar{u}^T x$, whereby $\beta = 1$. A parallel construction works for $\|y\|_*$ and shows that $\|\cdot\|_*$ on Y^* can be chosen so that $\beta^* = 1$.

3 Characterization Results for $\rho(d)$

Given a data instance $d = (A, b) \in L(X, Y) \times Y$, we now present characterizations of $\rho(d)$ for the feasibility problem P given in (1.1) .

The characterizations of $\rho(d)$ will depend on whether $d \in \mathcal{F}$ or $d \in \mathcal{F}^{\mathcal{C}}$ (recall (1.2)), i.e., whether (1.1) is consistent or not. We first study the case when $d \in \mathcal{F}$ ((1.1) is consistent), followed by the case when $d \in \mathcal{F}^{\mathcal{C}}$ ((1.1) is not consistent).

3.1 Characterization Results when P is consistent

In this subsection, we present five different mathematical programs and we prove that the optimal value of each of these mathematical programs provides an approximation of the value of $\rho(d)$, in the case when P is consistent. For each of these five mathematical programs, the nature of the approximation of $\rho(d)$ is specified in a theorem stating the result.

We motivate the development of these programs on intuitive grounds as follows. If P is consistent, i.e., $d \in \mathcal{F}$, then it is elementary to see that:

$$\begin{array}{rll} \rho(d) = & \underset{\bar{d}}{\text{infimum}} & \|d - \bar{d}\| \\ & & \\ & s.t. & \bar{d} = (\bar{A}, \bar{b}) \in \mathcal{F}^{\mathcal{C}} \end{array} ,$$

that is, $\rho(d)$ measures how close the data d lies to the set of data for which P has no solution. As $d \notin \mathcal{F}^{\mathcal{C}}$, then there is no solution $y \in Y^*$ to the system:

$$\begin{array}{c} A^T y \in C_X^* \\ -b^T y > 0 \\ y \in C_Y^* \\ \|y\|_* = 1 \end{array} ,$$

see Proposition 2.1, and notice in the above that we have added a normalizing constraint "$\|y\|_* = 1$." However, when C_X is regular, then the vector $\bar{u} \in intC_X^*$ (see Proposition 2.3), and we can measure how far away $d = (A, b)$ is from admitting a solution to the above system (i.e., how far away d is from the set $\mathcal{F}^{\mathcal{C}}$) by solving the following program:

$P_\alpha(d)$:

$$\begin{array}{rll} \alpha(d) = & \underset{y,\gamma}{\text{minimum}} & \gamma \\ & & \\ & s.t. & A^T y + \gamma \bar{u} \in C_X^* \\ & & -b^T y + \gamma \geq 0 \\ & & \|y\|_* = 1 \\ & & y \in C_Y^* \end{array} . \tag{3.1}$$

Notice that $\alpha(d) \geq 0$, since otherwise $d \in \mathcal{F}^{\mathcal{C}}$ via Proposition 2.1, which would violate the hypothesis of this subsection. Also notice that $\alpha(d) < +\infty$, since $\bar{u} \in intC_X^*$, and so $P_\alpha(d)$ is feasible for any $y \in C_Y^*$ with $\|y\|_* = 1$ and γ chosen sufficiently large. The smaller the value of $\alpha(d)$ is, the closer the conditions (2.3) of infeasibility are to being satisfied, and so the smaller the value of $\rho(d)$ should be. These arguments are obviously imprecise, but we will prove their validity in the following:

Theorem 3.1 *If $d \in \mathcal{F}$ and C_X is regular, then*

$$\beta \cdot \alpha(d) \leq \rho(d) \leq \alpha(d) .$$

This theorem states that $\alpha(d)$ approximates $\rho(d)$ to within the factor β, where recall the definition of β in Definition 2.1. The proof of this theorem and all of the results of this subsection are deferred to the end of the subsection.

Remark 3.1 *It should be pointed out that $P_\alpha(d)$ is not in general a convex program, due to the non-convex constraint "$\|y\|_* = 1$". However, in the case when $Y = R^m$, (then Y^* can also be identified with R^m), if we choose the norm on Y^* to be the L_∞ norm (so the norm on Y is the L_1 norm), then $P_\alpha(d)$ can be solved by solving $2m$ convex programs. To see this, observe that the* *constraint "$\|y\|_* = 1$" can be written as $-e \le y \le e$ and $y_i = \pm 1$ for some $i \in \{1, \ldots, m\}$, where e is the vector of ones. Then $\alpha(d) = minimum\left\{\alpha^{'}(d),\ \alpha^{''}(d)\right\}$, where :*

$$
\begin{array}{rll}
\alpha^{'}(d) = \displaystyle\min_{i=1,\ldots,m}\ \min_{y,\gamma} & \gamma & \\
\text{s.t.} & A^T y + \gamma\bar{u} \in C_X^* & \\
 & -b^T y + \gamma \ge 0 & \\
 & -e \le y \le e & \\
 & y_i = 1 & \\
 & y \in C_Y^* & ,
\end{array}
$$

and

$$
\begin{array}{rll}
\alpha^{''}(d) = \displaystyle\min_{i=1,\ldots,m}\ \min_{y,\gamma} & \gamma & \\
\text{s.t.} & A^T y + \gamma\bar{u} \in C_X^* & \\
 & -b^T y + \gamma \ge 0 & \\
 & -e \le y \le e & \\
 & y_i = -1 & \\
 & y \in C_Y^* & .
\end{array}
$$

The next mathematical program is obtained by considering the following homogenization of P:

$$
\begin{array}{r}
br - Ax \in C_Y \\
x \in C_X \\
r \ge 0 \quad ,
\end{array}
$$

which can be normalized in the case when C_X is regular as follows:

H :

$$
\begin{array}{r}
br - Ax \in C_Y \\
x \in C_X \\
r \ge 0 \\
r + \bar{u}^T x = 1 \quad .
\end{array}
$$

One can think of $\bar{u}^T x$ as a linear approximation of $\|x\|$ over the cone C_X, see Corollary 2.1. (In fact, the construction of β and $\bar{u}$ in Definition 2.1 and Proposition 2.3 is such that $\bar{u}^T x$ is the "best" linear approximation of $\|x\|$ over the cone C_X.) Now, an "internal view" of $\rho(d)$ is that $\rho(d)$ measures the extent to which the data $d = (A, b)$ can be altered and yet (1.1) will still be feasible for the new system. A modification of this view is that $\rho(d)$ measures the extent to which the system (1.1) can be modified while ensuring its feasibility. Consider the following program:

$P_w(d)$:

$$
\begin{array}{rlll}
w(d) = & \underset{v}{\text{minimum}} & \underset{r,x,\theta}{\text{maximum}} & \theta \\
& \|v\| \le 1 & s.t. & br - Ax - v\theta \in C_Y \\
& & & x \in C_X \\
& & & r + \bar{u}^T x = 1 \\
& & & r \ge 0 \quad .
\end{array}
\tag{3.2}
$$

Then $w(d)$ is the largest scaling factor θ such that for any v with $\|v\| \le 1$, $v\theta$ can be added to the first inclusion of H without affecting the feasibility of the system. We will prove:

Theorem 3.2 *If $d \in \mathcal{F}$ and C_X is regular, then $\alpha(d) = w(d)$, and so*

$$\beta \cdot w(d) \le \rho(d) \le w(d) \quad .$$

Three remarks are in order here. First, the theorem asserts that $\alpha(d) = w(d)$, and in fact the proof of the theorem will show that $P_w(d)$ can be obtained from $P_\alpha(d)$ by dualizing on a subset of the variables and constraints of $P_\alpha(d)$, i.e., $P_\alpha(d)$ and $P_w(d)$ are partial duals of one another. Second, there is an underlying geometry in $P_w(d)$. To see this, let

$$S = \{v \in Y \mid \textit{there exists } r \ge 0 \textit{ and } x \in C_X \textit{ satisfying } br - Ax - v \in C_Y,\ \bar{u}^T x + r = 1\}. \tag{3.3}$$

Then $P_w(d)$ can alternatively be written as:

$$
\begin{array}{rll}
w(d) = & \underset{\theta}{\sup} & \theta \\
& s.t. & \{v \in Y \mid \|v\| \le \theta\} \subset S \quad .
\end{array}
\tag{3.4}
$$

Then $w(d)$ is the radius of the largest ball centered at the origin and contained in the set S. Third, if we replace the normalizing linear constraint "$r + \bar{u}^T x = 1$" of $P_w(d)$ by the norm constraint "$r + \|x\| = 1$," then the modified program is analogous to Renegar's characterization of the distance to ill-posedness (see Theorem 3.5 in [13]) when P is consistent. The modified program is:

$P_r(d)$:

$$
\begin{array}{rlll}
r(d) = & \underset{v}{\text{minimum}} & \underset{r,x,\theta}{\text{maximum}} & \theta \\
& \|v\| \le 1 & s.t. & br - Ax - v\theta \in C_Y \\
& & & x \in C_X \\
& & & r + \|x\| = 1 \\
& & & r \ge 0 \quad ,
\end{array}
$$

and Renegar shows that $r(d) = \rho(d)$ when $d \in \mathcal{F}$. We will not use this fact, nor duplicate the proof of this fact here; rather it is our intent to show the connection to the results in [13].

One problem with $P_\alpha(d)$ is that $P_\alpha(d)$ is generally nonconvex, due to the constraint "$\|y\|_* = 1$." When C_Y is also regular, then from Corollary 2.1 the linear function $\bar{z}^T y$ is a "best" linear approximation of $\|y\|_*$ on C_Y^*, and if we replace "$\|y\|_* = 1$" by "$\bar{z}^T y = 1$" in $P_\alpha(d)$ we obtain:

$P_{\tilde{\alpha}}(d)$:

$$\begin{array}{rll} \tilde{\alpha}(d) = & \underset{y,\gamma}{\text{minimum}} & \gamma \\ & s.t. & A^T y + \gamma \bar{u} \in C_X^* \\ & & -b^T y + \gamma \geq 0 \\ & & \bar{z}^T y = 1 \\ & & y \in C_Y^* \end{array} \qquad (3.5)$$

Replacing the norm constraint by its linear approximation will reduce (by a constant) the extent to which the program computes an approximation of $\rho(d)$, and the analog of Theorem 3.1 becomes:

Theorem 3.3 *If $d \in \mathcal{F}$ and both C_X and C_Y are regular, then*

$$\beta^* \beta \cdot \tilde{\alpha}(d) \leq \rho(d) \leq \tilde{\alpha}(d).$$

Notice that a very nice feature of $P_{\tilde{\alpha}}(d)$ is that it is a convex program.

The fourth mathematical program is derived by once again measuring the extent to which the data $d = (A, b)$ does not admit a solution of (2.3). In the case when C_Y is regular, consider the program:

$P_u(d)$:

$$\begin{array}{rlll} u(d) = & \underset{y,q,g}{\text{minimum}} & \text{maximum} & \left\{ \|A^T y - q\|_*, \ |b^T y + g| \right\} \\ & s.t. & & y \in C_Y^* \\ & & & q \in C_X^* \\ & & & g \geq 0 \\ & & & \bar{z}^T y = 1 \end{array} \qquad (3.6)$$

If $d = (A, b)$ were in $\mathcal{F}^{\mathcal{C}}$, then form Proposition 2.1 it would be true that $u(d) = 0$. The nonnegative quantity $u(d)$ measures the extent to which (2.3) is not feasible. The smaller the value of $u(d)$ is, the closer the conditions (2.3) are to being satisfied, and so the smaller the value of $\rho(d)$ should be. These arguments are imprecise, but the next theorem validates the intuition of this line of thinking:

Theorem 3.4 *If $d \in \mathcal{F}$ and C_Y is regular, then*

$$\beta^* u(d) \leq \rho(d) \leq u(d) \quad .$$

Therefore, $u(d)$ approximates $\rho(d)$ to within the factor β^*, where β^* is defined in Definition 2.1.

Notice that $P_u(d)$ is also a convex program, which is convenient. If the constraint "$\bar{z}^T y = 1$" in $P_u(d)$ is replaced with the norm constraint "$\|y\|_* = 1$," one obtains the nonconvex program:

$P_j(d)$:

$$
\begin{array}{rlll}
j(d) = & \underset{y,q,g}{\text{minimum}} & \text{maximum} & \left\{ \|A^T y - q\|_*, \ |b^T y + g| \right\} \\
 & \text{s.t.} & & y \in C_Y^* \\
 & & & q \in C_X^* \\
 & & & g \geq 0 \\
 & & & \|y\|_* = 1 \quad ,
\end{array}
$$

and one can prove (although we will not do this here) that $P_j(d)$ can be derived as a partial dual of $P_r(d)$, and so $j(d) = r(d)$, whereby $j(d) = \rho(d)$ as in Theorem 3.5 of [13].

Returning to $P_u(d)$, notice that the feasible region of this program is a convex set, and that the objective function is a gauge function, i.e., a nonnegative convex function that is positively homogeneous of degree 1, see [15]. A mathematical program that minimizes a gauge function over a convex set is called a gauge program, and corresponding to every gauge program is a dual gauge program that also minimizes a (dual) gauge function over a (dual) convex set, see [5]. For the program $P_u(d)$, its dual gauge program is given by:

$P_v(d)$:

$$
\begin{array}{rlrl}
v(d) = & \underset{x,r}{\text{minimum}} & & \|x\| + |r| \\
 & & br - & Ax - \bar{z} \in C_Y \\
 & & & x \in C_X \\
 & & & r \geq 0 \quad .
\end{array}
\tag{3.7}
$$

In general, dual gauge programs will have the product of their optimal values equal to 1, as the last theorem of this subsection indicates:

Theorem 3.5 *If $d \in \mathcal{F}$ and C_Y is regular, then $u(d) \cdot v(d) = 1$ whenever $u(d) > 0$, and*

$$
\frac{\beta^*}{v(d)} \leq \rho(d) \leq \frac{1}{v(d)} \quad .
$$

Note that $P_v(d)$ is also a convex program. One can interpret $P_v(d)$ as measuring the extent to which P has a solution that is interior the cone C_Y. To see this, note from Proposition 2.3 that $\bar{z} \in intC_Y$, and so $P_v(d)$ will only be feasible if P has a solution interior to C_Y. The more interior a solution there is, the smaller (r, x) can be scaled and still satisfy $br - Ax - \bar{z} \in C_Y$. One would then expect $\rho(d)$ to be inversely proportional to $v(d)$, as Theorem 3.5 indicates.

The proofs of these five theorems are given below.

Proof of Theorem 3.1: Suppose $\gamma > \alpha(d)$. Then there exist $\bar{y}, \bar{\gamma}$, with $\alpha(d) \leq \bar{\gamma} < \gamma$ and $(\bar{y}, \bar{\gamma})$ is feasible for $P_\alpha(d)$. Therefore, $A^T\bar{y} + \bar{\gamma}\bar{u} \in C_X^*$, $-b^T\bar{y} + \bar{\gamma} \geq 0$, $\|\bar{y}\|_* = 1$, and $\bar{y} \in C_Y^*$. From Proposition 2.2, there exists $\bar{v} \in Y$ that satisfies $\|\bar{v}\| = 1$ and $\bar{y}^T\bar{v} = \|\bar{y}\|_* = 1$. Let $\bar{d} = (\bar{A}, \bar{b}) = (A + \gamma\bar{v}\bar{u}^T, \ b - \gamma\bar{v})$. Then $\|\bar{A} - A\| = \gamma\|\bar{v}\bar{u}^T\| = \gamma\|\bar{v}\| \, \|\bar{u}\|_* = \gamma$, and $\|\bar{b} - b\| = \gamma\|\bar{v}\| = \gamma$. Thus $\|\bar{d} - d\| = max\left\{\|\bar{A} - A\|, \ \|\bar{b} - b\|\right\} = \gamma$. Next note that $\bar{A}^T\bar{y} = A^T\bar{y} + \gamma\bar{u}\bar{v}^T\bar{y} = A^T\bar{y} + \gamma\bar{u} = A^T\bar{y} + \bar{\gamma}\bar{u} + (\gamma - \bar{\gamma})\bar{u} \in C_X^*$, and $-\bar{b}^T\bar{y} = -b^T\bar{y} + \gamma\bar{v}^T\bar{y} = -b^T\bar{y} + \gamma > -b^T\bar{y} + \bar{\gamma} \geq 0$. Therefore, by

Proposition 2.1, $(\bar{A}, \bar{b}) \in \mathcal{F}^C$. Therefore, $\rho(d) \le \|d - \bar{d}\| = \gamma$. As this is true for any $\gamma > \alpha(d)$, we must have $\rho(d) \le \alpha(d)$, proving the second inequality of the theorem.

To prove the first inequality of the theorem, let $\alpha > \rho(d)$ be chosen. Then there exists $(\bar{A}, \bar{b})$ for which $\|d - \bar{d}\| \le \alpha$ and $\bar{d} \in \mathcal{F}^C$. Then, from Proposition 2.1, there exists $\bar{y} \in C_Y^*$ satisfying $\bar{A}^T\bar{y} \in C_X^*$, $\bar{b}^T\bar{y} \le 0$, and $\|\bar{y}\|_* = 1$ (without loss of generality). Then for any $x \in C_X$ that satisfies $\|x\| = 1$, we will have

$$\begin{aligned} x^T A^T \bar{y} + \tfrac{\alpha}{\beta}\bar{u}^T x &= x^T(A - \bar{A})^T\bar{y} + x^T\bar{A}^T\bar{y} + \tfrac{\alpha}{\beta}\bar{u}^T x \\ &\ge -\|x\|\,\|A - \bar{A}\|\,\|\bar{y}\|_* + 0 + \alpha \qquad \text{(from Proposition 2.3)} \\ &\ge -\alpha + \alpha = 0 \qquad . \end{aligned}$$

Therefore, $A^T\bar{y} + \frac{\alpha}{\beta}\bar{u} \in C_X^*$. Also, $-b^T\bar{y} + \frac{\alpha}{\beta} = -\bar{b}^Ty + (\bar{b} - b)^T\bar{y} + \frac{\alpha}{\beta} \ge -\|\bar{b} - b\|\,\|\bar{y}\|_* + \frac{\alpha}{\beta} \ge -\alpha + \frac{\alpha}{\beta} \ge 0$, since $\beta \le 1$ from Proposition 2.3. Thus $\gamma = \frac{\alpha}{\beta}$ is feasible for $P_{\alpha(d)}$ with $y = \bar{y}$. Therefore, $\alpha(d) \le \gamma = \frac{\alpha}{\beta}$. As this is true for any $\alpha > \rho(d)$, it is also true that $\alpha(d) \le \frac{\rho(d)}{\beta}$, completing the proof. ∎

Proof of Theorem 3.2: Consider the program $P_\alpha(d)$ given in (3.1). Because $d \in \mathcal{F}$, i.e., P has a solution, there can be no feasible solution (y, γ) of $P_\alpha(d)$ with $\gamma < 0$ (for if so, then $y \in C_Y^*$, $b^Ty \le \gamma < 0$, and $A^Ty = A^Ty + \gamma\bar{u} - \gamma\bar{u} \in intC_X^*$ since $\bar{u} \in intC_X^*$, violating Proposition 2.1). Therefore, we can amend $P_\alpha(d)$ by replacing the constraint "$\|y\|_* = 1$" by "$\|y\|_* \ge 1$", so that

$$\begin{aligned} \alpha(d) = \quad \underset{y,\gamma}{\text{minimum}} \quad & \gamma \\ s.t. \quad & A^Ty + \gamma\bar{u} \in C_X^* \\ & -b^Ty + \gamma \ge 0 \\ & \|y\|_* \ge 1 \\ & y \in C_Y^* \qquad . \end{aligned}$$

However, $\|y\|_* \ge 1$ if and only if there exists $v \in Y$ that satisfies $\|v\| \le 1$ and $y^Tv \ge 1$ (see (2.4)), and so we can write:

$$\begin{aligned} \alpha(d) = \quad \underset{v,y,\gamma}{\text{minimum}} \quad & \gamma \\ s.t. \quad & A^Ty + \gamma\bar{u} \in C_X^* \\ & -b^Ty + \gamma \ge 0 \\ & y^Tv \ge 1 \\ & y \in C_Y^* \\ & v \in Y, \|v\| \le 1 \qquad . \end{aligned}$$

We then separate this problem into the two-level optimization problem:

$$\begin{array}{rllr} \alpha(d) = & \underset{\substack{v \in Y \\ \|v\| \le 1}}{\text{minimum}} & \underset{y,\gamma}{\text{minimum}} & \gamma \\ & & \text{s.t.} & A^T y + \gamma \bar{u} \in C_X^* \\ & & & -b^T y + \gamma \ge 0 \\ & & & y^T v \ge 1 \\ & & & y \in C_Y^* \quad . \end{array}$$

Next note that the right-most (or bottom-level) problem is an instance of ND of the Appendix given in (A.9), where $K_1^* = C_Y^*$, $w = v$, and $K_2^* = C_X^* \times R_+$, $M^T y = (-A^T y,\ +b^T y)$, and $f = (\bar{u}, 1)$. The dual problem associated with ND is given by NP in (A.8), which translates to

$$\begin{array}{rllr} v_P = & \underset{x,r,\theta}{\text{maximum}} & & \theta \\ & \text{s.t.} & br - & Ax - \theta v \in C_Y \\ & & & x \in C_X \\ & & & r \ge 0 \\ & & \bar{u}^T x & + \ r = 1 \\ & & & \theta \ge 0 \quad . \end{array}$$

Furthermore, since $d \in \mathcal{F}$, this program is feasible (set $\theta = 0$, and rescale any solution x of P), and then since all of the hypotheses of Theorem A.2 are satisfied, $v_P = v_D$, and so we can replace the bottom-level program by its dual, whereby

$$\begin{array}{rlllr} \alpha(d) = & \underset{\substack{v \in Y \\ \|v\| \le 1}}{\text{minimum}} & \underset{y,r,\theta}{\text{maximum}} & & \theta \\ & & \text{s.t.} & br - & Ax - \theta v \in C_Y \\ & & & & x \in C_X \\ & & & & r \ge 0 \\ & & & \bar{u}^T x & + \ r = 1 \\ & & & & \theta \ge 0 \quad , \end{array}$$

which is precisely $w(d)$, so $\alpha(d) = w(d)$. ▮

Proof of Theorem 3.3: Suppose $\alpha > \tilde{\alpha}(d)$. Then there exists $(\bar{y},\ \bar{\gamma})$ feasible for $P_{\tilde{\alpha}}(d)$ such that $\bar{\gamma} < \alpha$. Therefore, $A^T \bar{y} + \bar{\gamma}\bar{u} \in C_X^*$, $-b^T \bar{y} + \bar{\gamma} \ge 0$, and $\bar{z}^T \bar{y} = 1$ and $\bar{y} \in C_Y^*$. From Corollary 2.1, $1 \le \|\bar{y}\|_* \le \frac{1}{\beta^*}$. From Proposition 2.2, there exists $\bar{v} \in Y$ such that $\bar{v}^T \bar{y} = \|\bar{y}\|_*$ and $\|\bar{v}\| = 1$. Now let $\bar{d} = (\bar{A}, \bar{b}) = (A + \alpha \bar{v}\bar{u}^T,\ b - \alpha \bar{v})$. Then $\|\bar{A} - A\| = \alpha\|\bar{v}\|\ \|\bar{u}\|_* = \alpha$, and $\|\bar{b} - b\| = \alpha\|\bar{v}\| = \alpha$, so that $\|\bar{d} - d\| = \alpha$. Now $\bar{b}^T \bar{y} = b^T \bar{y} - \alpha \bar{v}^T \bar{y} = b^T \bar{y} - \alpha \|\bar{y}\|_* \le b^T \bar{y} - \alpha < b^T \bar{y} - \bar{\gamma} \le 0$, so that $\bar{b}^T \bar{y} < 0$ (the first inequality here being an instance of (2.15)). Also, for any $x \in C_X$ satisfying $\|x\| = 1$, we have $x^T \bar{A}^T \bar{y} = x^T A^T \bar{y} + \alpha x^T \bar{u}\bar{v}^T \bar{y} + \bar{\gamma}\bar{u}^T x - \bar{\gamma}\bar{u}^T x \ge \alpha x^T \bar{u}\bar{v}^T \bar{y} - \gamma \bar{u}^T x = \bar{u}^T x(\alpha \bar{v}^T \bar{y} - \bar{\gamma}) \ge \bar{\gamma}\bar{u}^T x(\bar{v}^T \bar{y} - 1) \ge 0$. Thus $\bar{A}^T \bar{y} \in C_X^*$. From Proposition 2.1, then $\bar{d} \in \mathcal{F}^C$, and so $\rho(d) \le \|\bar{d} - d\| = \alpha$. As this is true for any $\alpha > \tilde{\alpha}(d)$, then $\rho(d) \le \tilde{\alpha}(d)$, which proves the second inequality of the theorem.

To prove the first inequality of the theorem, suppose that $\alpha > \rho(d)$. Then there exists a data instance $\bar{d} = (\bar{A}, \bar{b})$ with $\|\bar{d} - d\| \le \alpha$ for which $\bar{d} \in \mathcal{F}^C$. From Proposition 2.1, there exists $\bar{y} \in C_Y^*$ satisfying $\bar{A}^T\bar{y} \in C_X^*$, $\bar{b}^T\bar{y} \leq 0$, and without loss of generality, $\bar{z}^T\bar{y} = 1$. We now show that $(y,\gamma) = \left(\bar{y}, \frac{\alpha}{\beta\beta^*}\right)$ is feasible for $P_{\tilde{\alpha}}(d)$. First observe that $\bar{y} \in C_Y^*$ and $\bar{z}^T\bar{y} = 1$. Also $-b^T\bar{y} + \frac{\alpha}{\beta\beta^*} = -\bar{b}^T\bar{y} + (\bar{b}-b)^T\bar{y} + \frac{\alpha}{\beta\beta^*} \geq -\|\bar{b}-b\|\|\bar{y}\|_* + \frac{\alpha}{\beta\beta^*} \geq -\frac{\alpha}{\beta^*} + \frac{\alpha}{\beta\beta^*} \geq 0$. It remains to show that $A^T\bar{y} + \frac{\alpha}{\beta\beta^*}\bar{u} \in C_X^*$. To see this, observe that for any $x \in C_X$ with $\|x\| = 1$, that

$$\begin{aligned}
x^TA^T\bar{y} + \frac{\alpha}{\beta\beta^*}\bar{u}^Tx &= x^T\bar{A}^T\bar{y} + x^T(A-\bar{A})^T\bar{y} + \frac{\alpha\bar{u}^Tx}{\beta\beta^*} \\
&\geq -\|x\|\,\|A-\bar{A}\|\,\|\bar{y}\|_* + \frac{\alpha\bar{u}^Tx}{\beta\beta^*} \\
&\geq -\alpha\|\bar{y}\|_* + \frac{\alpha\bar{u}^Tx}{\beta\beta^*} \\
&\geq -\frac{\alpha\bar{y}^T\bar{z}}{\beta^*} + \frac{\alpha}{\beta^*}\|x\| \qquad \text{(from Proposition 2.1)} \\
&= \frac{-\alpha}{\beta^*} + \frac{\alpha}{\beta^*} = 0 \qquad .
\end{aligned}$$

Thus $A^T\bar{y} + \frac{\alpha\bar{u}^Tx}{\beta\beta^*}\bar{u} \in C_X^*$, showing that $(y,\gamma) = \left(\bar{y}, \frac{\alpha}{\beta\beta^*}\right)$ is feasible for $P_{\tilde{\alpha}}(d)$. Therefore, $\frac{\alpha}{\beta\beta^*} \geq \tilde{\alpha}(d)$. As this is true for any $\alpha > \rho(d)$, then $\rho(d) \geq \beta\beta^*\tilde{\alpha}(d)$, completing the proof. ∎

Proof of Theorem 3.4: Suppose that $\alpha > \rho(d)$. Then there exists $\bar{d} = (\bar{A}, \bar{b})$, with the property that $\|\bar{d} - d\| \leq \alpha$ and $\bar{d} \in \mathcal{F}^C$. From Proposition 2.1, this implies the existence of a vector $\bar{y} \in C_Y^*$ that satisfies $\bar{A}^T\bar{y} \in C_X^*$, $\bar{b}^T\bar{y} \leq 0$, and without loss of generality, $\bar{z}^T\bar{y} = 1$. Let $q = \bar{A}^T\bar{y}$. Then $A^T\bar{y} - q = (A-\bar{A})^T\bar{y} + \bar{A}^T\bar{y} - q = (A-\bar{A})^T\bar{y}$, whereby $\|A^T\bar{y} - q\|_* \leq \|(A-\bar{A})\|\;\|\bar{y}\|_* \leq \alpha\|\bar{y}\|_* \leq \frac{\alpha\bar{z}^T\bar{y}}{\beta^*} = \frac{\alpha}{\beta^*}$, from Corollary 2.1. Let $g = -\bar{b}^T\bar{y}$. Then $g \geq 0$ and $|b^T\bar{y} + g| = |b^T\bar{y} - \bar{b}^T\bar{y}| \leq \|b-\bar{b}\|\;\|\bar{y}\|_* \leq \frac{\alpha\bar{z}^T\bar{y}}{\beta^*} = \frac{\alpha}{\beta^*}$. Thus $(\bar{y}, q, g)$ is feasible for $P_u(d)$ with objective value $max\left\{\|A^T\bar{y} - q\|_*,\ |b^T\bar{y} + g|\right\} \leq \frac{\alpha}{\beta^*}$, whereby $u(d) \leq \frac{\alpha}{\beta^*}$. As this is true for any $\alpha > \rho(d)$, it is also true that $u(d) \leq \frac{\rho(d)}{\beta^*}$, which proves the first inequality of the theorem.

Next, suppose that $\alpha > u(d)$. Then there exists $\bar{y} \in C_Y^*$, $q \in C_X^*$, and $g \geq 0$ such that $\bar{z}^T\bar{y} = 1$ and $\|A^T\bar{y} - q\|_* < \alpha$ and $|b^T\bar{y} + g| < \alpha$. Let $\epsilon = \alpha - |b^T\bar{y} + g|$, and note that $\epsilon > 0$. Define $\bar{d} = (\bar{A}, \bar{b}) = \left(A - \bar{z}\left(\bar{y}^TA - q^T\right),\ b - \left(b^T\bar{y} + g + \epsilon\right)\bar{z}\right)$. Then $\|\bar{A} - A\| = \|\bar{z}\|\;\|A^T\bar{y} - q\|_* < \alpha$, and $\|\bar{b} - b\| = \|\bar{z}\|\;|b^T\bar{y} + g + \epsilon| \leq \|\bar{z}\|\left(|b^T\bar{y} + g| + |\epsilon|\right) = \|\bar{z}\|\alpha = \alpha$. Therefore, $\|\bar{d} - d\| \leq \alpha$. However, $\bar{A}^T\bar{y} = A^T\bar{y} - \left(A^T\bar{y} - q\right)\bar{z}^T\bar{y} = q \in C_X^*$, and $\bar{b}^T\bar{y} = b^T\bar{y} - \left(b^T\bar{y} + g + \epsilon\right)\bar{z}^T\bar{y} = -g - \epsilon < 0$. Thus from Proposition 2.1, $\bar{d} \in \mathcal{F}^C$, and so $\rho(d) \leq \|d - \bar{d}\| \leq \alpha$. As this is true for any $\alpha > u(d)$, it is true that $\rho(d) \leq u(d)$, completing this proof. ∎

Proof of Theorem 3.5: Note that $P_u(d)$ in (3.6) is an instance of the program GP of the Appendix given in (A.2). To see this, we make the following associations. We let $U = Y^* \times X^* \times R$, and $V = X^* \times R$, where R is the set of real numbers. We let $K = C_Y^* \times C_X^* \times R_+$, where R_+ is the cone of non-negative real numbers. We let $d = (\bar{z}, 0, 0)$ and let the gauge function $f(\cdot):\ V \to R$ be defined by $F(w,\alpha) = \ max\{\|w\|_*, |\alpha|\}$ for $(w,\alpha) \in X^* \times R$. Then with the linear operator M given by $M(y,q,g) = \left(-A^Ty + q,\ b^Ty + g\right)$, (3.6) is seen to be an instance of GP. The gauge dual GD of GP for this instance is precisely $P_v(d)$ given in (3.7), and so we can apply Theorem A.1. If $u(d) = 0$, then $v(d) = +\infty$, proving the result. If $u(d) > 0$, we cannot have $u(d) = +\infty$, since $P_u(d)$ always has a feasible solution. It remains to show the result when $u(d) > 0$, and from Theorem A.1 of the

Appendix, we need only to verify that the four hypotheses of part (III) of Theorem A.1 are satisfied. Hypothesis (i) translates to the assertion that all projections of $\{y \in C_Y^* | \bar{z}^T y = 1\} \times C_X^* \times R_+$ are closed sets. This is true because C_X^* and R_+ are closed cones (whose projections will be closed sets), and $\{y \in C_Y^* | \bar{z}^T y = 1\}$ is a closed bounded set, and so its projections will be closed sets. Hypothesis (ii) translates to $sup\{\bar{z}^T y | y \in C_Y^*\} = +\infty$, which is true because C_Y is regular and $\bar{z} \in intC_Y$ from Proposition 2.3. Hypothesis (iii) is true because $f(\cdot) = f(w, \alpha) = max\,\{\|w\|_*,\ |\alpha|\}$ is a norm, and so is a closed gauge function, and hypothesis (iv) is true because $f(w, \alpha)$ is a norm and so its level sets are bounded. Since the hypotheses of Theorem A.1 are satisfied for the gauge program $P_u(d)$, then $u(d) > 0$ implies $u(d)v(d) = 1$, proving the result. ▮

3.2 Characterization Results when P is not consistent

In this subsection, we parallel the results of the previous subsection for the case when P is not consistent. That is, we present five different mathematical programs and we prove that the optimal value of each of these mathematical programs provides an approximation of the value of $\rho(d)$, in the case when P is not consistent. For each of these five mathematical programs, the nature of the approximation of $\rho(d)$ is specified in a theorem stating the result.

We motivate the development of these programs on intuitive grounds as follows. If P is not consistent, i.e.,$d \in \mathcal{F}^{\mathcal{C}}$, then it is elementary to see that:

$$\begin{array}{rll} \rho(d) = & \underset{\bar{d}}{\text{infimum}} & \|d - \bar{d}\| \\ & & \\ & s.t. & \bar{d} = (\bar{A}, \bar{b}) \in \mathcal{F} \quad , \end{array}$$

that is, $\rho(d)$ measures how close the data d lies to the set of data for which P has a solution. As $d \notin \mathcal{F}$, then there is no solution $x \in X$ and $r > 0$ to the homogenized system:

$$\begin{array}{c} br - Ax \in C_Y \\ x \in C_X \\ r > 0 \\ r + \|x\| = 1 \quad , \end{array}$$

and notice in the above that we have added a normalizing constraint "$r + \|x\| = 1$." However, when C_Y is regular, then the vector $\bar{z} \in intC_Y$ (see Proposition 2.3), and we can measure how close $d = (A, b)$ is to being in the set $\mathcal{F}$ by solving the following program:

$P_\sigma(d)$:

$$\begin{array}{rll} \sigma(d) = & \underset{r, x, \gamma}{\text{minimum}} & \gamma \\ & & \\ & s.t. & br - Ax + \bar{z}\gamma \in C_Y \\ & & r + \|x\| = 1 \\ & & r \geq 0 \\ & & x \in C_X \quad . \end{array} \tag{3.8}$$

Notice that $\sigma(d) \geq 0$, since otherwise it would be true that $d \in \mathcal{F}$ (by perturbing r to make r positive if need be in $P_\sigma(d)$), which would violate the hypothesis of this subsection. Also notice that $\sigma(d) < +\infty$, since $\bar{z} \in intC_Y$, and so $P_\sigma(d)$ is feasible for any $x \in C_X$ and $r \geq 0$ with $\|x\| + r = 1$ and γ chosen sufficiently large. The smaller the value of $\sigma(d)$ is, the closer the conditions (1.1) are to being satisfied, and so the smaller the value of $\rho(d)$ should be. These arguments are obviously imprecise, but we will prove their validity in the following:

Theorem 3.6 *If $d \in \mathcal{F}^C$ and C_Y is regular, then*

$$\beta^* \cdot \sigma(d) \leq \rho(d) \leq \sigma(d) \quad .$$

This theorem states that $\sigma(d)$ approximates $\rho(d)$ to within the factor β^*, where recall the definition of β^* in Definition 2.1. The proof of this theorem and all of the results of this subsection are deferred to the end of the subsection.

Remark 3.2 *It should be pointed out that $P_\sigma(d)$ is not in general a convex program, due to the non-convex constraint "$r + \|x\| = 1$". However, in the case when $X = R^n$, if we choose the norm on X to be the L_∞ norm, then $P_\sigma(d)$ can be solved by solving $2n$ convex programs, where the construction exactly parallels that given for $P_\alpha(d)$ earlier in this section. One can easily show that $\sigma(d) = min\left\{\sigma'(d),\ \sigma''(d)\right\}$, where :*

$$\sigma'(d) = \underset{j=1,\ldots,n}{\text{minimum}} \ \underset{r,x,\gamma}{\text{minimum}} \ \gamma$$

$$\begin{array}{ll} s.t. & br - Ax + \bar{z}\gamma \in C_Y \\ & -(1-r)e \leq x \leq (1-r)e \\ & x_j = (1-r) \\ & r \geq 0 \\ & x \in C_X \end{array} \quad ,$$

and

$$\sigma''(d) = \underset{j=1,\ldots,n}{\text{minimum}} \ \underset{r,x,\gamma}{\text{minimum}} \ \gamma$$

$$\begin{array}{ll} s.t. & br - Ax + \bar{z}\gamma \in C_Y \\ & -(1-r)e \leq x \leq (1-r)e \\ & x_j = -(1-r) \\ & r \geq 0 \\ & x \in C_X \end{array} \quad .$$

The next mathematical program is obtained by considering the following normalized version of the infeasibilty conditions (2.2):

HD :

$$\begin{array}{c} A^T y \in C_X^* \\ -b^T y \geq 0 \\ y \in C_Y^* \\ \bar{z}^T y = 1 \end{array} \quad .$$

One can think of $\bar{z}^T y$ as a linear approximation of $\|y\|_*$ over the cone C_Y^*, see Corollary 2.1. (In fact, the construction of β^* and $\bar{z}$ in Definition 2.1 and Proposition 2.3 is such that $\bar{z}^T y$ is the "best" linear approximation of $\|y\|_*$ over the cone C_Y^*.) Now, an "internal view" of $\rho(d)$ is that $\rho(d)$ measures the extent to which the data $d = (A, b)$ can be altered and yet (2.2) will still be feasible for the new system. A modification of this view is that $\rho(d)$ measures the extent to which the system (2.2) can be modified while still ensuring the feasibility of the system. Consider the following program:

$P_\delta(d)$:

$$\begin{array}{rcccl} \delta(d) = & \underset{v \in X^*}{\text{minimum}} & \underset{y,\theta}{\text{maximum}} & \theta & \\ & \|v\|_* \le 1 & s.t. & A^T y - v\theta \in C_X^* & \\ & & & -b^T y - \theta \ge 0 & \\ & & & y \in C_Y^* & \\ & & & \bar{z}^T y = 1 & . \end{array} \tag{3.9}$$

Then $\delta(d)$ is the largest scaling factor θ such that for any v with $\|v\|_* \le 1$, $(-v\theta)$ can be added to the first inclusion of HD and θ can be added to the inequality in HD while still ensuring its feasibility. We will prove:

Theorem 3.7 *If $d \in \mathcal{F}^\mathcal{C}$ and C_Y is regular, then $\sigma(d) = \delta(d)$, and so*

$$\beta^* \cdot \delta(d) \le \rho(d) \le \delta(d) \quad .$$

Three remarks are in order here. First, the theorem asserts that $\sigma(d) = \delta(d)$, and in fact the proof of the theorem will show that $P_\delta(d)$ can be obtained from $P_\sigma(d)$ by dualizing on a subset of the variables and constraints of $P_\sigma(d)$, i.e., $P_\sigma(d)$ and $P_\delta(d)$ are partial duals of one another. Second, there is an underlying geometry in $P_\sigma(d)$. To see this, let

$$T = \{(v,\gamma) \in X^* \times R | \; there\ exists\ y \in C_Y^*\ satisfying\ A^T y - v \in C_X^*,\ -b^T y - \gamma \ge 0,\ \bar{z}^T y = 1\}. \tag{3.10}$$

Then $P_\delta(d)$ can alternatively be written as:

$$\begin{array}{rll} \delta(d) = & \underset{\theta}{\text{maximum}} & \theta \\ & s.t. & \{(v,\gamma) \in X^* \times R | max\{\|v\|_*,\ |\gamma|\} \le \theta\} \subset T \quad , \end{array} \tag{3.11}$$

and we see that $\delta(d)$ is the radius of the largest ball centered at $(v, \gamma) = (0, 0)$ and contained in T, where the norm is $\max\{\|v\|_*, |\gamma|\}$. Third, if we replace the normalizing linear constraint "$\bar{z}^T y = 1$" of $P_\delta(d)$ by the norm constraint "$\|y\|_* = 1$," then the modified program is analogous to the construction of Renegar (as applied to the system (2.2)) for characterizing the distance to ill-posedness (see Theorem 3.5 of [13]) when P is not consistent. The modified program is:

$P_\pi(d)$:

$$\begin{array}{rlll} \pi(d) = & \underset{v \in X^*}{\text{minimum}} & \underset{y,\theta}{\text{maximum}} & \theta \\ & \|v\|_* \leq 1 & s.t. & A^T y - v\theta \in C_X^* \\ & & & -b^T y - \theta \geq 0 \\ & & & y \in C_Y^* \\ & & & \|y\|_* = 1 \end{array},$$

and one can in fact show that $\pi(d) = \rho(d)$ when $d \in \mathcal{F}^\mathcal{C}$. We will not prove this here; rather it is our intent to show the connection to the results in [13].

One problem with $P_\sigma(d)$ is that $P_\sigma(d)$ is generally nonconvex, due to the constraint "$r + \|x\| = 1$." When C_X is also regular, then from Corollary 2.1 the linear function $\bar{u}^T x$ is a "best" linear approximation of $\|x\|$ on C_X, and if we replace "$r + \|x\| = 1$" by "$r + \bar{u}^T x = 1$" in $P_\sigma(d)$ we obtain:

$P_{\tilde{\sigma}}(d)$:

$$\begin{array}{rll} \tilde{\sigma}(d) = & \underset{r,x,\gamma}{\text{minimum}} & \gamma \\ & s.t. & br - Ax + \bar{z}\gamma \in C_Y \\ & & r + \bar{u}^T x = 1 \\ & & r \geq 0 \\ & & x \in C_X \end{array} \quad . \tag{3.12}$$

Replacing the norm constraint by its linear approximation will reduce (by a constant) the extent to which the program computes an approximation of $\rho(d)$, and the analog of Theorem 3.6 becomes:

Theorem 3.8 *If $d \in \mathcal{F}^\mathcal{C}$ and both C_X and C_Y are regular, then*

$$\beta^* \beta \cdot \tilde{\sigma}(d) \leq \rho(d) \leq \tilde{\sigma}(d).$$

Notice that a very nice feature of $P_{\tilde{\sigma}}(d)$ is that it is a convex program.

The fourth mathematical program of this subsection is derived by once again measuring the extent to which the data $d = (A, b)$ does not admit a solution of (1.1). In the case when C_X is regular, consider the program:

$P_g(d)$:

$$\begin{array}{rll} g(d) = & \underset{x,r,w}{\text{minimum}} & \|br - Ax - w\| \\ & s.t. & x \in C_X \\ & & r \geq 0 \\ & & w \in C_Y \\ & & \bar{u}^T x + r = 1 \end{array} \quad . \tag{3.13}$$

If $d = (A, b)$ were in $\mathcal{F}$, then it would be true that $g(d) = 0$. The nonnegative quantity $g(d)$ measures the extent to which (1.1) is not feasible. The smaller the value of $g(d)$ is, the closer the conditions

(1.1) are to being satisfied, and so the smaller the value of $\rho(d)$ should be. These arguments are imprecise, but the next theorem validates the intuition of this line of thinking:

Theorem 3.9 *If $d \in \mathcal{F}^{\mathcal{C}}$ and C_X is regular, then*

$$\beta \cdot g(d) \leq \rho(d) \leq g(d) \quad .$$

Therefore, $g(d)$ approximates $\rho(d)$ to within the factor β, where β is defined in Definition 2.1.

Notice that $P_g(d)$ is also a convex program, which is convenient. If the constraint "$\bar{u}^T x + r = 1$" in $P_g(d)$ is replaced with the norm constraint "$\|x\| + r = 1$," one obtains the nonconvex program:

$P_k(d):$

$$\begin{array}{rll} k(d) = & \underset{x,r,w}{\text{minimum}} & \|br - Ax - w\| \\ & s.t. & x \in C_X \\ & & r \geq 0 \\ & & w \in C_Y \\ & & \|x\| + r = 1 \quad , \end{array}$$

and one can prove (although we will not do this here) that $P_k(d)$ can be derived as a partial dual of $P_\pi(d)$, and so $k(d) = \pi(d) = \rho(d)$ as mentioned earlier.

Returning to $P_g(d)$, notice that $P_g(d)$ is a gauge program. For the program $P_g(d)$, its dual gauge program is given by:

$P_h(d):$

$$\begin{array}{rll} h(d) = & \underset{y}{\text{minimum}} & \|y\|_* \\ & s.t. & A^T y - \bar{u} \in C_X^* \\ & & -b^T y - 1 \geq 0 \\ & & y \in C_Y^* \quad . \end{array} \tag{3.14}$$

In general, dual gauge programs will have the product of their optimal values equal to 1, as the last theorem of this subsection indicates:

Theorem 3.10 *If $d \in \mathcal{F}^{\mathcal{C}}$ and C_X is regular, then $g(d) \cdot h(d) = 1$ whenever $g(d) > 0$, and*

$$\frac{\beta}{h(d)} \leq \rho(d) \leq \frac{1}{h(d)} \quad .$$

Note that $P_h(d)$ is also a convex program. One can interpret $P_h(d)$ as measuring the extent to which (2.2) has a solution that is interior to the cone C_Y^* and that satisfies $b^T y \leq 0$ strictly. To see this, note from Proposition 2.3 that $\bar{u} \in intC_X^*$, and so $P_h(d)$ will only be feasible if the first and the third conditions of (2.2) are satisfied in their interior. The more interior a solution there is, the smaller y can be scaled and still satisfy $A^T y - \bar{u} \in C_Y$ and $-b^T y - 1 \geq 0$. One would then expect $\rho(d)$ to be inversely proportional to $h(d)$, as Theorem 3.5 indicates.

The proofs of these last five theorems are given below.

Proof of Theorem 3.6: Suppose that $\alpha > \sigma(d)$ is given. Then there exists a feasible solution $(\bar{r}, \bar{x}, \bar{\gamma})$ of $P_\sigma(d)$ with $\bar{\gamma} < \alpha$, i.e., $b\bar{r} - A\bar{x} + \bar{z}\bar{\gamma} \in C_Y$, $\bar{r} + \|\bar{x}\| = 1$, $\bar{r} \geq 0$, and $\bar{x} \in C_X$. From Proposition 2.2, there exists $(v,t) \in X^* \times R$ satisfying $v^T\bar{x} + t\bar{r} = \bar{r} + \|\bar{x}\| = 1$, and $max\,\{\|v\|_*,\ |t|\} = 1$. Let us then define $\bar{d} = (\bar{A}, \bar{b}) = (A - \alpha\bar{z}v^T,\ b + \alpha\bar{z}t)$. Then $\|\bar{A} - A\| = \alpha\|\bar{z}\|\ \|v\|_* \leq \alpha$ and $\|\bar{b} - b\| = \alpha|t|\ \|\bar{z}\| \leq \alpha$, so $\|\bar{d} - d\| \leq \alpha$. However, $\bar{b}\bar{r} - \bar{A}\bar{x} = b\bar{r} + \alpha\bar{z}t\bar{r} - A\bar{x} + \alpha\bar{z}v^T\bar{x} = b\bar{r} - A\bar{x} + \alpha\bar{z} = b\bar{r} - A\bar{x} + \bar{\gamma}\bar{z} + (\alpha - \bar{\gamma})\bar{z} \in int\ C_Y$. Thus $\bar{b}(\bar{r} + \epsilon) - A\bar{x} \in C_Y$ for $\epsilon > 0$ and sufficiently small, and so after dividing this expression by $\bar{r} + \epsilon$, we see that $\bar{d} = (\bar{A}, \bar{b}) \in \mathcal{F}$. Therefore, $\rho(d) \leq \alpha$. Since this is true for any $\alpha > \sigma(d)$, it is also true that $\rho(d) \leq \sigma(d)$, proving the second inequality of the theorem.

To prove the first inequality of the theorem, let $\alpha > \rho(d)$ be chosen. Then there exists $\bar{d} = (\bar{A}, \bar{b})$ for which $\|d - \bar{d}\| \leq \alpha$, and $\bar{d} \in \mathcal{F}$, whereby there exists $\bar{x} \in C_X$ and $\bar{r} > 0$ satisfying $\bar{b}\bar{r} - \bar{A}\bar{x} \in C_Y$, and $\|\bar{x}\| + \bar{r} = 1$. Thus for any $y \in C_Y^*$ that satisfies $\|y\|_* = 1$,

$$
\begin{aligned}
y^T\left(b\bar{r} - A\bar{x} + \tfrac{\alpha}{\beta^*}\bar{z}\right) &= y^T\left(\bar{b}\bar{r} - \bar{A}\bar{x} + (b - \bar{b})\bar{r} - (A - \bar{A})\bar{x} + \tfrac{\alpha}{\beta^*}\bar{z}\right) \\
&\geq -\|b - \bar{b}\|\bar{r} - \|\bar{A} - A\|\ \|\bar{x}\| + \tfrac{\alpha}{\beta^*}\bar{z}^T y \\
&\geq -\alpha(\bar{r} + \|\bar{x}\|) + \alpha \qquad\qquad \text{(from Corollary 2.1)} \\
&= 0 \qquad .
\end{aligned}
$$

Therefore, $\left(b\bar{r} - A\bar{x} + \frac{\alpha}{\beta^*}\bar{z}\right) \in C_Y$, $\|\bar{x}\| + \bar{r} = 1$, and $\bar{x} \in C_X$ and $\bar{r} \geq 0$. Thus $(x, r, \gamma) = \left(\bar{x}, \bar{r}, \frac{\alpha}{\beta^*}\right)$ is feasible for $P_\sigma(d)$, whereby $\sigma(d) \leq \frac{\alpha}{\beta^*}$. As this is true for any $\alpha > \rho(d)$, it must be true that $\sigma(d) \leq \frac{\rho(d)}{\beta^*}$, completing the proof. ∎

Proof of Theorem 3.7: Consider the program $P_\sigma(d)$ given in (3.8). Because $d \in \mathcal{F}^C$, i.e., (1.1) has no solution, there can be no feasible solution (x, r, γ) of $P_\sigma(d)$ with $\gamma < 0$ (for if so, then $br - Ax \in int\ C_Y$, $r \geq 0$, $x \in C_X$, and so for $\epsilon > 0$ and sufficiently small $b(r + \epsilon) - Ax \in C_Y$, and rescaling would give a solution $\frac{x}{r+\epsilon}$ of (1.1)). Therefore, we can amend $P_\sigma(d)$ by replacing the constraint "$\|x\| + r = 1$" by "$\|x\| + r \geq 1$", so that

$$
\begin{aligned}
\sigma(d) = \ \ &\underset{r, x, \gamma}{\text{minimum}} \quad \gamma \\
&s.t. \qquad br - Ax + \bar{z}\gamma \in C_Y \\
&\qquad\qquad \|x\| \geq 1 - r \\
&\qquad\qquad r \geq 0 \\
&\qquad\qquad x \in C_X \qquad .
\end{aligned}
$$

However, $\|x\| \geq 1 - r$ if and only if there exists $v \in X^*$ that satisfies $v^Tx \geq 1 - r$ and $\|v\|_* \leq 1$ (see

(2.4)), and so we can write:

$$\begin{array}{rll} \sigma(d) = & \underset{v,r,x,\gamma}{\text{minimum}} & \gamma \\ & \text{s.t.} & br - Ax + \bar{z}\gamma \in C_Y \\ & & v^T x \geq 1 - r \\ & & r \geq 0 \\ & & x \in C_X \\ & & v \in X^* \\ & & \|v\|_* \leq 1 \end{array} .$$

We then separate this problem into the two-level optimization problem:

$$\begin{array}{rlll} \sigma(d) = & \underset{\substack{v \in X^* \\ \|v\|_* \leq 1}}{\text{minimum}} & \underset{r,x,\gamma}{\text{minimum}} & \gamma \\ & & \text{s.t.} & br - Ax + \bar{z}\gamma \in C_Y \\ & & & r + v^T x \geq 1 \\ & & & r \geq 0 \\ & & & x \in C_X \end{array} .$$

Next note that the right-most (or bottom level) problem is an instance of **ND** of the Appendix given in (A.9), where $K_2^* = C_Y$, $K_1^* = C_X \times R_+$, $y = (x, r)$, $M^T y = Ax - br$, $f = \bar{z}$, and $w = (v, 1)$. The dual problem associated with **ND** is given by **NP** in (A.8), which translates to:

$$\begin{array}{rll} v_P = & \underset{y,\theta}{\text{maximum}} & \theta \\ & \text{s.t.} & A^T y - \theta v \in C_X^* \\ & & -b^T y - \theta \geq 0 \\ & & \bar{z}^T y = 1 \\ & & y \in C_Y^* \end{array} .$$

Furthermore, since $d \in \mathcal{F}^C$, this program is feasible (set $\theta = 0$ and apply Proposition 2.1), whereby the hypotheses of Theorem A.2 are satisfied, and so $v_P = v_D$ and we can replace the bottom-level program by its dual, so

$$\begin{array}{rlll} \sigma(d) = & \underset{\substack{v \in X^* \\ \|v\|_* \leq 1}}{\text{minimum}} & \underset{y,\theta}{\text{maximum}} & \theta \\ & & \text{s.t.} & A^T y - \theta v \in C_X^* \\ & & & -b^T y - \theta \geq 0 \\ & & & y \in C_Y^* \\ & & & \bar{z}^T y = 1 \end{array} ,$$

which is precisely $\delta(d)$, so $\sigma(d) = \delta(d)$. ▮

Proof of Theorem 3.8: Suppose that $\alpha > \rho(d)$ is given. Then there exists a data instance $\bar{d} = (\bar{A}, \bar{b})$ with $\|d - \bar{d}\| \leq \alpha$ for which $\bar{d} \in \mathcal{F}$. Then by rescaling if necessary, there exists $\bar{x} \in C_X$

and $\bar{r} > 0$ such that $\bar{b}\bar{r} - \bar{A}\bar{x} \in C_Y$ and $\bar{u}^T\bar{x} + \bar{r} = 1$. We now show that $\left(b\bar{r} - A\bar{x} + \frac{\alpha}{\beta^*\beta}\bar{z}\right) \in C_Y$. To see this, note that for any $y \in C_Y^*$ that satisfies $\|y\|_* = 1$, then

$$\begin{aligned} y^T\left(b\bar{r} - A\bar{x} + \left(\tfrac{\alpha}{\beta^*\beta}\right)\bar{z}\right) &= y^T\left(\bar{b}\bar{r} - \bar{A}\bar{x} + (b-\bar{b})\bar{r} - (A-\bar{A})\bar{x} + \tfrac{\alpha}{\beta^*\beta}\bar{z}\right) \\ &\geq -\|b-\bar{b}\|\bar{r} - \|A-\bar{A}\|\,\|\bar{x}\| + \tfrac{\alpha}{\beta^*\beta}\bar{z}^T y \\ &\geq -\alpha(\bar{r} + \|\bar{x}\|) + \tfrac{\alpha\|y\|_*\beta^*}{\beta^*\beta} \qquad \text{(from Corollary 2.1)} \\ &\geq -\alpha\left(\bar{r} + \tfrac{\bar{u}^T\bar{x}}{\beta}\right) + \tfrac{\alpha}{\beta} \\ &\geq -\tfrac{\alpha}{\beta}(\bar{r} + \bar{u}^T\bar{x}) + \tfrac{\alpha}{\beta} = 0 \qquad . \end{aligned}$$

Therefore $\left(b\bar{r} - A\bar{x} + \left(\frac{\alpha}{\beta^*\beta}\right)\bar{z}\right) \in C_Y$. Thus $(x, r, \gamma) = \left(\bar{x}, \bar{r}, \frac{\alpha}{\beta^*\beta}\right)$ is feasible for $P_{\tilde{\sigma}}(d)$, and so $\tilde{\sigma}(d) \leq \frac{\alpha}{\beta^*\beta}$. As this is true for any $\alpha > \rho(d)$, then $\tilde{\sigma}(d) \leq \frac{\rho(d)}{\beta^*\beta}$, proving the first inequality of the theorem.

Next, suppose that $\alpha > \tilde{\sigma}(d)$ is given. Then there exists a feasible solution $(\bar{x}, \bar{r}, \bar{\gamma})$ of $P_{\tilde{\sigma}}(d)$ with $\bar{\gamma} < \alpha$, i.e., $\bar{x} \in C_X$, $\bar{r} \geq 0$, $\bar{\gamma} < \alpha$, $b\bar{r} - A\bar{x} + \bar{\gamma}\bar{z} \in C_Y$, and $\bar{r} + \bar{u}^T\bar{x} = 1$. Let $\bar{d} = (\bar{A}, \bar{b}) = (A - \alpha\bar{z}\bar{u}^T,\ b + \alpha\bar{z})$. then $\|\bar{A} - A\| = \alpha\|\bar{z}\|\,\|\bar{u}\|_* = \alpha$, and $\|\bar{b} - b\| = \alpha\|\bar{z}\| = \alpha$, whereby $\|\bar{d} - d\| = \alpha$. However, $\bar{b}\bar{r} - \bar{A}\bar{x} = b\bar{r} - A\bar{x} + \alpha\bar{z}(\bar{r} + \bar{u}^T\bar{x}) = b\bar{r} - A\bar{x} + \alpha\bar{z} = b\bar{r} - A\bar{x} + \bar{\gamma}\bar{z} + (\alpha - \bar{\gamma})\bar{z} \in int\ C_Y$. Then $\bar{b}(\bar{r} + \epsilon) - \bar{A}\bar{x} \in C_Y$ for $\epsilon > 0$ and sufficiently small, so $x = \frac{\bar{x}}{(\bar{r}+\epsilon)}$ satisfies for (1.1) for the data $\bar{d} = (\bar{A}, \bar{b})$, i.e., $\bar{d} \in \mathcal{F}$. Thus $\rho(d) \leq \|d - \bar{d}\| = \alpha$. Since this is true for any $\alpha > \tilde{\sigma}(d)$, it is also true that $\rho(d) \leq \tilde{\sigma}(d)$, completing the proof. ▮

Proof of Theorem 3.9: Suppose that $\alpha > \rho(d)$ is given. Then there exists $\bar{d} = (\bar{A}, \bar{b}) \in \mathcal{F}$ that satisfies $\|\bar{d} - d\| \leq \alpha$. Thus, by rescaling if necessary, there exists $\bar{x} \in C_X$ and $\bar{r} > 0$ such that $\bar{b}\bar{r} - \bar{A}\bar{x} \in C_Y$ and $\bar{u}^T\bar{x} + \bar{r} = 1$. Let $w = \bar{b}\bar{r} - \bar{A}\bar{x}$. Then $w \in C_Y$ and $b\bar{r} - A\bar{x} - w = \bar{b}\bar{r} - \bar{A}\bar{x} - w + (b - \bar{b})\bar{r} - (A - \bar{A})\bar{x} = (b - \bar{b})\bar{r} - (A - \bar{A})\bar{x}$, so $\|b\bar{r} - A\bar{x} - w\| \leq \|b - \bar{b}\|\bar{r} + \|A - \bar{A}\|\,\|\bar{x}\| \leq \alpha\bar{r} + \alpha\|\bar{x}\| \leq \alpha\bar{r} + \alpha\left(\frac{\bar{u}^T\bar{x}}{\beta}\right) \leq \frac{\alpha}{\beta}$, from Corollary 2.1. Therefore, $(\bar{x}, \bar{r}, w)$ is feasible for $P_g(d)$ with objective value at most $\frac{\alpha}{\beta}$, and so $g(d) \leq \frac{\alpha}{\beta}$. As this is true for any $\alpha > \rho(d)$, it must be true that $g(d) \leq \frac{\rho(d)}{\beta}$, which proves the first part of the theorem.

Next, suppose that $\alpha > g(d)$ is given. Then there exists (x, r, w) feasible for $P_g(d)$ with objective value $\|br - Ax - w\| < \alpha$. Then from (3.13), (x, r, w) satisfies $x \in C_X$, $r \geq 0$, $\bar{u}^T x + r = 1$, and $w \in C_Y$. For $\epsilon > 0$, define $(\bar{x}, \bar{r}) = \left(\frac{x}{1+\epsilon}, \frac{r+\epsilon}{1+\epsilon}\right)$. Then $(\bar{x}, \bar{r}, w)$ is feasible for $P_g(d)$, and we can always choose ϵ sufficiently small that $\|b\bar{r} - A\bar{x} - w\| \leq \alpha$. Let $\bar{d} = (\bar{A}, \bar{b}) = \left(A - (A\bar{x} - b\bar{r} + w)\bar{u}^T,\ b + (A\bar{x} - b\bar{r} + w)\right)$. Then $\|\bar{A} - A\| \leq \alpha$ and $\|\bar{b} - b\| \leq \alpha$, so that $\|\bar{d} - d\| \leq \alpha$. However, $\bar{b}\bar{r} - \bar{A}\bar{x} = b\bar{r} - A\bar{x} + (A\bar{x} - b\bar{r} + w)\left(\bar{u}^T\bar{x} + \bar{r}\right) = w \in C_Y$. Since $\bar{r} > 0$, then $\bar{b} - \bar{A}\left(\frac{\bar{x}}{\bar{r}}\right) \in C_Y$, and we see that $\bar{d} = (\bar{A}, \bar{b}) \in \mathcal{F}$. This implies that $\rho(d) \leq \|d - \bar{d}\| \leq \alpha$. As this is true for any $\alpha > g(d)$, it is also true that $\rho(d) \leq g(d)$, proving the theorem. ▮

Proof of Theorem 3.10: Note that $P_g(d)$ of (3.13) is an instance of the program GP of the Appendix given in (A.2). To see this, first change the notation of "x" in GP to "$\tilde{x}$" and then $\tilde{x} = (x, r, w)$, $M\tilde{x} = Ax - br + w$, $f(\cdot) = \|\cdot\|$, $K = C_X \times R_+ \times C_Y$, and $d = (\bar{u}, 1, 0)$. The gauge dual GD of GP for this instance is precisely $P_h(d)$ given in (3.14), and so we can apply Theorem

A.2. If $g(d) = 0$, then $h(d) = +\infty$, proving the result. If $g(d) > 0$, we cannot have $g(d) = +\infty$, since $P_g(d)$ always has a feasible solution. It remains to show the result when $g(d) > 0$, and from Theorem A.1 of the Appendix, we need only to verify that the four hypotheses of Theorem A.1 are satisfied. Hypothesis (i) translates to the assertion that all projections of $\{(x, r, w) \in C_X \times R_+ \times C_Y | \bar{u}^T x + r = 1\}$ are closed sets, which is true because $\{(x, r) \in C_X \times R_+ | \bar{u}^T x + r = 1\}$ is a closed and bounded set (and hence its projections will be closed sets), and C_Y is a closed convex cone (and so its projections will be closed sets). Hypothesis (ii) translates to $sup\{\bar{u}^T x + r | x \in C_X, r \in R_+\} = +\infty$, which is true since one can set $x = 0$ and r arbitrarily large. Hypotheses (iii) and (iv) are true since $f(\cdot) = \|\cdot\|$ is a closed gauge function with bounded level sets because $\|\cdot\|$ is a norm. Since the hypotheses of Theorem A.1 are satisfied, then $g(d) > 0$ implies $g(d) \cdot h(d) = 1$, proving the theorem. ▮

4 Bounds on Radii of Contained and Intersecting Balls

In this section, we develop four results concerning the radii of certain inscribed balls in the feasible region of the system (1.1) or, in the case when P is not consistent, of the alternative system (2.2). These results are stated as Lemmas 4.1, 4.2, 4.3, and 4.4 of this section. While these results are of an intermediate nature, it is nevertheless useful to motivate them, which we do now, by thinking in terms of the ellipsoid algorithm for finding a point in a convex set.

Consider the ellipsoid algorithm for finding a feasible point in a convex set C. Roughly speaking, the main ingredients that are needed to apply the ellipsoid algorithm and to produce a complexity bound on the number of iterations of the ellipsoid algorithm are the existence of:

(i) *a ball* $B(\hat{x}, r)$ *with the property that* $B(\hat{x}, r) \subset C$,

(ii) *a ball* $B(0, R)$ *with the property that* $B(\hat{x}, r) \subset B(0, R)$, *and*

(iii) *an upper bound on the ratio* R/r .

With these three ingredients, it is then possible to produce a complexity bound on the number of iterations of the ellipsoid algorithm, which will be $O(n^2\ ln(R/r))$. In addition, it is also convenient to have the following:

(iv) *a lower bound on the radius* r *of the contained ball* $B(\hat{x}, r)$, *and*

(v) *an upper bound on the radius* R *of the initial ball* $B(0, R)$.

In the bit model of complexity as applied to linear inequality systems, one is usually able to set $r = (1/n)2^{-L}$ and $R = n2^L$, where L is the number of bits needed to represent the system. (Of course, these values of r and R break down when the system is degenerate (in our parlance, "ill=posed") in which case the system must be perturbed first.)

By analogy for the problem P considered herein in (1.1), the convex set in mind is the set X_d, which is the feasible region of the problem P, and 2^L is generally replaced by the *condition* *measure* of $d = (A, b)$, denoted $\mathcal{C}(d)$, which is defined to be

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

see Renegar [13]. The results in this section will be used in Section 5 to demonstrate in general that we can find a point $\hat{x} \in X_d$ and radii r and R with the five properties below, that are analogs of the five properties listed above:

$$(i) \quad B(\hat{x}, r) \subset X_d$$

$$(ii) \quad B(\hat{x}, r) \subset B(0, R)$$

$$(iii) \quad R/r \leq O(n\,\mathcal{C}(d)) = O\left(\tfrac{n\|d\|}{\rho(d)}\right)$$

$$(iv) \quad r \geq 1/(O(n\,\mathcal{C}(d))) = \Omega\left(\tfrac{\rho(d)}{n\|d\|}\right), \textit{and}$$

$$(v) \quad R \leq O(n\,\mathcal{C}(d)) = O\left(\tfrac{n\|d\|}{\rho(d)}\right) \quad .$$

Here the quantity 2^L is roughly replaced by $\mathcal{C}(d)$.

The above remarks pertain to the the case when P is consistent, i.e., when (1.1) has a solution. When (1.1) is not consistent, then the convex set in mind is the feasible region for the alternative system (2.2), denoted by Y_d. The results in this section will also be used in Section 5 to demonstrate in general that we can find a point $\hat{y}$ in Y_d and radii r and R with the three properties below, that are analogs of the first three properties listed above:

$$(i) \quad B(\hat{y}, r) \subset Y_d$$

$$(ii) \quad B(\hat{y}, r) \subset B(0, R)$$

$$(iii) \quad R/r \leq O(m\,\mathcal{C}(d)) = O\left(\tfrac{m\|d\|}{\rho(d)}\right) \quad .$$

Because the system (2.2) is homogeneous, it makes little sense to bound r from below or R from above, as all constructions can be scaled by any positive quantity. Therefore properties (iv) and (v) are not relevant.

The results in this section are rather technical, and their proofs are unfortunately quite long. The reader may first want to read Section 5 before pondering the results in this Section in detail.

We first examine the case when (1.1) is consistent, in which case the feasible region $X_d = \{x \in X | b - Ax \in C_Y, x \in C_X\}$ is nonempty.

Lemma 4.1 *Suppose that $d \in \mathcal{F}$ and C_Y is regular. If $\rho(d) > 0$, then there exists $\hat{x} \in X_d$ and* *positive scalars* r_1 *and* R_1 *satisfying:*

$$(i) \quad B(\hat{x}, r_1) \subset \{x \in X | b - Ax \in C_Y\} \tag{4.1}$$

$$(ii) \quad B(\hat{x}, r_1) \subset B(0, R_1) \tag{4.2}$$

$$(iii) \quad \frac{R_1}{r_1} \leq \frac{3\|d\|}{\beta^* \rho(d)}$$

$$(iv) \quad r_1 \geq \frac{\beta^* \rho(d)}{3\|d\|}$$

$$and\ (v) \quad R_1 \leq \frac{3\|d\|}{\beta^* \rho(d)} \quad .$$

In the case when $C_X = X$, (4.1) states that the feasible region X_d contains a ball of radius r_1, and (4.2) states that this ball does not lie more than the distance R_1 from the origin. Furthermore, $\frac{R_1}{r_1} \leq \frac{3\|d\|}{\beta^* \rho(d)}$. In order to prove Lemma 4.1, we first prove:

Proposition 4.1 $\rho(d) \leq \|d\|$.

Proof: If $d \in \mathcal{F}$ (respectively, $\mathcal{F}^{\mathcal{C}}$), then $\theta d \in \mathcal{F}$ (respectively, $\mathcal{F}^{\mathcal{C}}$) for all $\theta > 0$. Therefore, $\bar{d} = (\bar{A}, \bar{b}) = (0,0) \in \mathcal{B} = cl(\mathcal{F}) \cap cl\left(\mathcal{F}^{\mathcal{C}}\right)$, and so $\rho(d) \leq \|d - \bar{d}\| = \|d - 0\| = \|d\|$. ∎

Proof of Lemma 4.1: For any $w \in C_Y^*$ with $\|w\|_* = 1$, we have

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

so that $\frac{1}{2}\left(\bar{z} + \frac{\beta^*}{\|d\|} b\right) \in C_Y$. Now let $(\tilde{x}, \tilde{r})$ solve $P_v(d)$, and let

$$\hat{x} = \frac{\tilde{x}}{\tilde{r} + \frac{\beta^*}{\delta \|d\|}} = \frac{\tilde{x}}{\delta} \quad ,$$

where $\delta = \tilde{r} + \frac{\beta^*}{2\|d\|}$. Let $q = b\tilde{r} - A\tilde{x} - \bar{z}$. Then $q \in C_Y$ and we have $b\tilde{r} - A\tilde{x} + \frac{\beta^*}{2\|d\|} b - \frac{1}{2}\bar{z} = \bar{z} + q + \frac{\beta^*}{2\|d\|} b - \frac{1}{2}\bar{z} = \frac{1}{2}\left(\bar{z} + \frac{\beta^*}{\|d\|} b\right) + q \in C_Y$, so that $\delta b - A\tilde{x} - \frac{1}{2}\bar{z} \in C_Y$, whereby $b - A\hat{x} - \frac{1}{2\delta}\bar{z} \in C_Y$. Thus $\hat{x} \in C_X$ and $b - A\hat{x} \in C_Y$, so $\hat{x} \in X_d$. Let $r_1 = \frac{\beta^*}{2\delta\|d\|}$. Then if $\|x - \hat{x}\| \leq r_1$, we have

$$\begin{aligned} b - Ax = b - A\hat{x} + A(\hat{x} - x) \quad &= \tfrac{1}{\delta}\left(\delta b - A\tilde{x} - \tfrac{1}{2}\bar{z}\right) + \tfrac{1}{2\delta}\bar{z} + A(\hat{x} - x) \\ &= y + \tfrac{1}{2\delta}\bar{z} + A(\hat{x} - x) \end{aligned}$$

where $y \in C_Y$. Thus for any $w \in C_Y^*$ with $\|w\|_* = 1$,

$$\begin{aligned} w^T(b - Ax) \quad &\geq \tfrac{1}{2\delta}\bar{z}^T w + w^T A(\hat{x} - x) \quad \geq \tfrac{\beta^*}{2\delta} - \|w\|_* \|A\| \; \|\hat{x} - x\| \\ &\geq \tfrac{\beta^*}{2\delta} - \|d\| r_1 \qquad\qquad\qquad = 0 \ . \end{aligned}$$

Therefore $b - Ax \in C_Y$, proving (i).

Next, let $R_1 = \frac{v(d)}{\delta} + r_1$.

Then

$$\frac{R_1}{r_1} = \frac{v(d)}{\delta r_1} + 1 \quad \leq \frac{v(d)}{\delta r_1} + \frac{\|d\|}{\rho(d)} \qquad \text{(from Proposition 4.1)}$$

$$= \frac{2\|d\|}{\beta^* u(d)} + \frac{\|d\|}{\rho(d)} \qquad \text{(from Theorem 3.5)}$$

$$\leq \frac{3\|d\|}{\beta^* \rho(d)} \qquad , \qquad \text{(from Theorem 3.4)}$$

proving (iii). To compute the bounds in (iv) and (v), notice first that $\delta = \tilde{r} + \frac{\beta^*}{2\|d\|} \geq \frac{\beta^*}{2\|d\|}$, so that $r_1 \leq 1$. Therefore,

$$R_1 = \frac{v(d)}{\delta} + r_1 \quad \leq \frac{2\|d\| v(d)}{\beta^*} + 1$$

$$\leq \frac{2\|d\|}{\beta^* u(d)} + \frac{\|d\|}{\rho(d)} \qquad \text{(from Theorem 3.5 and Proposition 4.1)}$$

$$\leq \frac{3\|d\|}{\beta^* \rho(d)} \qquad , \qquad \text{(from Theorem 3.4)}$$

proving (v). We also have:

$$\delta = \tilde{r} + \frac{\beta^*}{2\|d\|} \quad \leq v(d) + \frac{\beta^*}{2\|d\|} \qquad \text{(from (3.7))}$$

$$\leq v(d) + \frac{1}{2\rho(d)} \qquad \text{(from Proposition 4.1)}$$

$$\leq \frac{3}{2\rho(d)} \qquad . \qquad \text{(from Theorem 3.5)}$$

Therefore, $r_1 = \frac{\beta^*}{2\delta\|d\|} \geq \frac{\beta^* \rho(d)}{3\|d\|}$, proving (iv). Finally, observe that $\|\hat{x}\| + r_1 = \frac{\|\tilde{x}\|}{\delta} + r_1 \leq \frac{v(d)}{\delta} + r_1 = R_1$, which proves (ii). ∎

We next have:

Lemma 4.2 *Suppose that $d \in \mathcal{F}$ and C_X is regular. If $\rho(d) > 0$, then there exists $\hat{x} \in X_d$ and positive scalars r_2 and R_2 satisfying:*

$$(i) \quad B(\hat{x}, r_2) \subset C_X \tag{4.3}$$

$$(ii) \quad B(\hat{x}, r_2) \subset B(0, R_2) \tag{4.4}$$

$$(iii) \quad \frac{R_2}{r_2} \leq \frac{(2n+5)\|d\|}{\beta^2 \rho(d)}$$

$$(iv) \quad r_2 \geq \frac{\beta \bar{\beta} \rho(d)}{2(n+2)\|d\|}$$

$$\textit{and } (v) \quad R_2 \leq \frac{(2n+6)\|d\|}{\beta^2 \rho(d)} \quad .$$

In the case when $C_Y = \{0\}$, we can intersect both sides of (4.3) with the affine subspace of $x \in X$ satisfying $Ax = b$. Then (4.3) will imply that the feasible region X_d contains a ball of radius

r_2 intersected with affine subspace of $x \in X$ satisfying $Ax = b$. Furthermore, (4.4) states that this ball does not lie more than a distance R_2 from the origin. Furthermore, $\frac{R_2}{r_2} \le \frac{(2n+5)\|d\|}{\rho(d)\beta^2\beta}$.

Proof of Lemma 4.2: Let us consider the set

$$C = \left\{(x, r) \in X \times R | x \in C_X,\ r \ge 0, \bar{u}^T x + r = 1,\ br - Ax \in C_Y\right\} \quad . \tag{4.5}$$

Then C is closed, bounded and convex, and there is Löwner-John pair of ellipsoids for C (see [4]). Therefore, letting $(\tilde{x}, \tilde{r})$ be the common center of both ellipsoids, then $(\tilde{x}, \tilde{r}) \in C$, and it will be true that for any $w \in X^*$ and any scalar g, that

$$w^T\tilde{x} + g\tilde{r} \;-\; \begin{array}{c} minimum \;\; w^T x + gr \\ s.t. \;\; (x,r) \in C \end{array} \;\ge \frac{1}{n+1}\left(\begin{array}{c} maximum \;\; w^T x + gr \\ s.t. \;\; (x,r) \in C \end{array} - \left(w^T\tilde{x} + g\tilde{r}\right)\right). \tag{4.6}$$

Let $v = \frac{-bw(d)}{\|d\|}$. Then $\|v\| \le w(d)$, and from (3.3) and (3.4) there exists (x', r') satisfying $x' \in C_X,\ r' \ge 0,\ br' - Ax' + \frac{bw(d)}{\|d\|} = q \in C_Y$, and $\bar{u}^T x' + r' = 1$. Let

$$\left(x'', r''\right) = \frac{\left(x',\ r' + \frac{w(d)}{\|d\|}\right)}{1 + \frac{w(d)}{\|d\|}} \quad . \tag{4.7}$$

Then $\left(x'',\ r''\right) \in C$, and $r'' = \frac{r' + \frac{w(d)}{\|d\|}}{1 + \frac{w(d)}{\|d\|}} > 0$, since $\rho(d) > 0$ and applying Theorem 3.2. If we then set $w = 0$ and $g = 1$ in (4.6), we obtain

$$\tilde{r} \ge \frac{1}{n+1}\left(r'' - \tilde{r}\right), \tag{4.8}$$

using the facts that $minimum\{\ r | (x,r) \in C\}\ \ge 0$ and $maximum\{\ r | (x,r) \in\ C\} \ge\ r''$. Therefore,

$$\tilde{r} \ge \frac{r''}{n+2} > 0 \quad . \tag{4.9}$$

We now let $\hat{x} = \frac{\tilde{x}}{\tilde{r}}$. Then $\hat{x} \in C_X$ and $b - A\hat{x} \in C_Y$ from (4.5). Thus $\hat{x} \in X_d$.

We next demonstrate the following:

$$\text{for any } w \in C_X^* \text{ with } \|w\|_* = 1,\ w^T\tilde{x} \ge \frac{w(d)\bar{\beta}}{(\|d\| + w(d))(n+2)}. \tag{4.10}$$

In order to prove (4.10), we choose any $w \in C_X^*$ with $\|w\|_* = 1$. Then, from (2.16), there exists $\bar{x} \in C_X$ such that $\|\bar{x}\| = 1$ and $w^T\bar{x} \ge \bar{\beta}$. If we set $v = \frac{A\bar{x}w(d)}{\|A\bar{x}\|}$, then $\|v\| = w(d)$ and from (3.4), there exists $\left(\bar{x}', \bar{r}'\right)$ such that $\bar{x}' \in C_X,\ \bar{r}' \ge 0,\ \bar{u}^T\bar{x}' + \bar{r}' = 1$, and $b\bar{r}' - A\bar{x}' - \frac{A\bar{x}w(d)}{\|A\bar{x}\|} = q \in C_Y$. If we then set

$$\left(\bar{x}'', \bar{r}''\right) = \frac{\left(\bar{x}' + \frac{\bar{x}w(d)}{\|A\bar{x}\|},\quad \bar{r}'\right)}{1 + \frac{w(d)\bar{x}^T\bar{u}}{\|A\bar{x}\|}} \quad , \tag{4.11}$$

then $b\bar{r}'' - A\bar{x}'' \in C_Y$, $\bar{x}'' \in C_X$, $\bar{r}'' \geq 0$, and $\bar{u}^T\bar{x}'' + \bar{r}'' = 1$, so that $\left(\bar{x}'', \bar{r}''\right) \in C$. Therefore, if we choose $g = 0$ in (4.6), we obtain

$$\begin{aligned} w^T\tilde{x} \geq w^T\tilde{x} - \underset{(x,r)\in C}{minimum}\ w^T x &\geq \left(\tfrac{1}{n+1}\right)\ \left(\underset{(x,r)\in C}{maximum}\ w^T x\ -\ w^T\tilde{x}\right) \\ &\geq \left(\tfrac{1}{n+1}\right)\ \left(w^T\bar{x}'' - w^T\tilde{x}\right) \qquad (4.12) \\ &= \left(\tfrac{1}{n+1}\right)\ \left(\frac{\left(w^T\bar{x}' + \frac{w^T\bar{x}w(d)}{\|A\bar{x}\|}\right)}{1 + \frac{w(d)\bar{x}^T\bar{u}}{\|A\bar{x}\|}} - w^T\tilde{x}\right), \end{aligned}$$

where the first inequality follows since $w^Tx \geq 0$ for any $x \in C_X$ since $w \in C_X^*$. Rearranging (4.12) we obtain

$$\begin{aligned} w^T\tilde{x} \geq\ & \frac{w^T\bar{x}w(d)}{(n+2)(\|A\bar{x}\| + w(d)\bar{x}^T\bar{u})} \qquad (since\ w^T\bar{x}' \geq 0) \\ \geq\ & \frac{\bar{\beta}w(d)}{(n+2)(\|d\| + w(d))} \qquad ,\ (from\ (2.14)) \end{aligned}$$

which demonstrates (4.10).

We now set $r_2 = \frac{\bar{\beta}w(d)}{\tilde{r}(n+2)(\|d\|+w(d))}$ and prove (4.3). It suffices to prove that if $\|x - \hat{x}\| \leq r_2$, then $w^Tx \geq 0$ for any $w \in C_X^*$ that satisfies $\|w\|_* = 1$. To verify this, suppose $w \in C_X^*$ and $\|w\|_* = 1$. Then $w^Tx = w^T(x - \hat{x}) + w^T\hat{x} \geq -\|x - \hat{x}\| + \frac{w^T\tilde{x}}{\tilde{r}} \geq -r_2 + \frac{\bar{\beta}w(d)}{(n+2)(\|d\|+w(d))\tilde{r}} = 0$, which shows (4.3).

To prove (iv), note that since $\tilde{r} \leq 1$, then

$$\begin{aligned} r_2 \geq\ & \frac{\bar{\beta}w(d)}{(n+2)(\|d\|+w(d))} \\ \geq\ & \frac{\bar{\beta}\rho(d)}{(n+2)(\|d\| + \frac{\|d\|}{\beta})} \qquad \text{(from Theorem 3.2 and Proposition 4.1)} \\ =\ & \frac{\beta\bar{\beta}\rho(d)}{(n+2)(\|d\|+\beta\|d\|)} \\ \geq\ & \frac{\beta\bar{\beta}\rho(d)}{2(n+2)\|d\|} \qquad \text{(since } \beta \leq 1) \end{aligned}$$

which proves (iv).

Now let $R_2 = \frac{1}{\beta\tilde{r}} + r_2$. Then

$$\begin{aligned} \frac{R_2}{r_2} = \frac{1}{\beta\tilde{r}r_2} + 1\ &= \frac{(n+2)(\|d\|+w(d))}{\beta\bar{\beta}w(d)} + 1 \\ &\leq \frac{(n+2)\left(\|d\| + \frac{\|d\|}{\beta}\right)}{\beta\bar{\beta}w(d)} + \frac{\|d\|}{\beta w(d)} \qquad \text{(from Proposition 4.1 and Theorem 3.2)} \\ &\leq \frac{(2n+5)\|d\|}{\beta^2\bar{\beta}w(d)} \\ &\leq \frac{(2n+5)\|d\|}{\beta^2\bar{\beta}\rho(d)} \qquad , \qquad \text{(from Theorem 3.2)} \end{aligned}$$

proving (iii). To prove (ii), note that

$$\begin{aligned} \|\hat{x}\| + r_2 = \tfrac{\|\tilde{x}\|}{\tilde{r}} + r_2 &\le \tfrac{\bar{u}^T\tilde{x}}{\beta\tilde{r}} + r_2 \qquad \text{(from (2.14))} \\ &\le \tfrac{1}{\beta\tilde{r}} + r_2 \\ &= R_2 \qquad , \end{aligned}$$

which proves (ii). As a means to proving (v), first observe that

$$\begin{aligned} (n+2)\tilde{r} &\ge r'' && \text{(from (4.9))} \\ &= \frac{r' + \frac{w(d)}{\|d\|}}{1 + \frac{w(d)}{\|d\|}} \\ &\ge \frac{w(d)}{\|d\| + w(d)} && \text{(since r}' \ge 0) \\ &\ge \frac{w(d)}{\|d\| + \frac{\|d\|}{\beta}} && \text{(from Proposition 4.1 and Theorem 3.2)} \\ &\ge \frac{w(d)\beta}{2\|d\|} && (\text{since } \beta \le 1) \\ &\ge \frac{\rho(d)\beta}{2\|d\|} \qquad . && \text{(from Theorem 3.2)} \end{aligned}$$

Therefore,

$$\begin{aligned} R_2 = \tfrac{1}{\beta\tilde{r}} + r_2 &\le \tfrac{2(n+2)\|d\|}{\beta^2\rho(d)} + r_2 && \text{(from above)} \\ &\le \tfrac{2(n+2)\|d\|}{\beta^2\rho(d)} + \tfrac{1}{\tilde{r}(n+2)} \\ &\le \tfrac{2(n+2)\|d\|}{\beta^2\rho(d)} + \tfrac{2\|d\|}{\beta\rho(d)} && \text{(from above)} \\ &\le \tfrac{(2n+6)\|d\|}{\beta^2\rho(d)} \qquad , \end{aligned}$$

which proves (v). ▮

We now turn to the case when (P) is inconsistent, i.e., (1.1) has no solution. In this case, from Proposition 2.1, the system (2.2) has a solution, and let us then examine the set of all solutions to (2.2), which we denote by Y_d to emphasize the dependence on the data $d = (A, b)$:

$$Y_d = \{y \in Y^* | A^T y \in C_X^*,\ y \in C_Y^*,\ y^T b \le 0\} \quad . \tag{4.13}$$

Lemma 4.3 *Suppose that $d \in \mathcal{F}^{\mathcal{C}}$ and C_X is regular. If $\rho(d) > 0$, then there exists $\hat{y} \in Y_d$ and positive scalars r_3 and R_3 satisfying $\frac{R_3}{r_3} \le \frac{\|d\|}{\beta\rho(d)}$, and that satisfy:*

$$B(\hat{y}, r_3) \subset \{y \in Y^* | A^T y \in C_X^*,\ b^T y \le 0\} \tag{4.14}$$

and

$$\|\hat{y}\|_* \le R_3 \quad . \tag{4.15}$$

In the case when $C_Y = \{0\}$, then $C_Y^* = Y^*$, and (4.14) states that the "dual" feasible region Y_d contains a ball of radius r_3, and (4.15) states that the center of this ball does not lie more than the distance R_3 from the origin. Furthermore, $\frac{R_3}{r_3} \leq \frac{\|d\|}{\beta\rho(d)}$.

Proof of Lemma 4.3: Let $\tilde{y}$ solve $P_h(d)$. Then $A^T\tilde{y} - \bar{u} \in C_X^*$, $-b^T\tilde{y} \geq 1$, and $\tilde{y} \in C_Y^*$. Then since $\tilde{y} \neq 0$ (otherwise $-\bar{u} \in C_X^*$ and so C_X is not regular, via Proposition 2.3), let $\hat{y} = \frac{\tilde{y}}{\|\tilde{y}\|_*}$. Let $r_3 = \frac{\beta g(d)}{\|d\|}$, and let $R_3 = 1$.

To prove (4.14) it suffices to show that if $\|y - \hat{y}\|_* \leq r_3$, then $A^T y \in C_X^*$ and $y^T b \leq 0$. We have that $q = A^T\hat{y} - \frac{\bar{u}}{\|\tilde{y}\|_*} \in C_X^*$. For any $x \in C_X$ with $\|x\| = 1$,

$$\begin{aligned}
x^T A^T y &= x^T A^T(y - \hat{y}) + x^T A^T\hat{y} - \tfrac{\bar{u}^T x}{\|\tilde{y}\|_*} + \tfrac{\bar{u}^T x}{\|\tilde{y}\|_*} \\
&\geq -\|x\|\,\|A\|\,\|y - \hat{y}\|_* + x^T q + \tfrac{\bar{u}^T x}{\|\tilde{y}\|_*} \\
&\geq -\|A\| r_3 + \tfrac{\bar{u}^T x}{\|\tilde{y}\|_*} && (since\ x^T q \geq 0) \\
&\geq -\beta g(d) + \tfrac{\bar{u}^T x}{\|\tilde{y}\|_*} && (since\ \|d\| \geq \|A\|) \\
&\geq -\beta g(d) + \tfrac{\beta}{\|\tilde{y}\|_*} && (from\ (2.14)) \\
&\geq -\beta \left(g(d) - \tfrac{1}{h(d)}\right) \\
&= 0 \qquad . && (from\ Theorem\ 3.10)
\end{aligned}$$

Therefore, $A^T y \in C_X^*$. Similarly,

$$\begin{aligned}
-b^T y &= -b^T(y - \hat{y}) - b^T\hat{y} \\
&\geq -\|d\|\,\|y - \hat{y}\|_* - \tfrac{b^T\tilde{y}}{\|\tilde{y}\|_*} \\
&\geq -\beta g(d) + \tfrac{1}{\|\tilde{y}\|_*} \\
&\geq -\beta g(d) + \tfrac{1}{h(d)} = g(d)(1-\beta) && (from\ Theorem\ 3.10) \\
&\geq 0. && (from\ Proposition\ 2.3)
\end{aligned}$$

Therefore, $A^T y \in C_X^*$ and $b^T y \leq 0$, which proves (4.14).

To prove (4.15), note that $\|\hat{y}\| = 1 = R_3$, which demonstrates (4.15). Finally note that $\frac{R_3}{r_3} = \frac{1}{r_3} = \frac{\|d\|}{\beta g(d)} \leq \frac{\|d\|}{\beta\rho(d)}$ from Theorem 3.9. ▮

We next prove:

Lemma 4.4 *Suppose that $d \in \mathcal{F}^{\mathcal{C}}$ and C_Y is regular. If $\rho(d) > 0$, then there exists $\hat{y} \in Y_d$ and* *positive scalars r_4 and R_4 satisfying $\frac{R_4}{r_4} \le \frac{2m\|d\|}{(\beta^*)^2\bar{\beta}^*\rho(d)}$, and that satisfy:*

$$B(\hat{y}, r_4) \subset C_Y^* \tag{4.16}$$

and

$$\|\hat{y}\|_* \le R_4 \quad . \tag{4.17}$$

In the case when $C_X = X$, then $C_X^* = \{0\}$, and we can intersect both sides of (4.16) with the affine subspace of $y \in Y$ that satisfy $A^T y = 0$. Then (4.16) will imply that the feasible region Y_d contains a ball of radius r_4 intersected with the affine subspace of $y \in Y$ that satisfy $A^T y = 0$. Furthermore, (4.17) states that the center of the ball does not lie more than a distance R_4 from the origin. Furthermore, $\frac{R_4}{r_4} \le \frac{2m\|d\|}{(\beta^*)^2\bar{\beta}^*\rho(d)}$.

Proof of Lemma 4.4: Let $S = Y_d \cap \{y \in Y^* | \bar{z}^T y = 1\}$. Because S is bounded, closed and convex, there is a Löwner-John pair of ellipsoids for S (see [4]). Therefore, letting $\hat{y}$ be the common center of both ellipsoids, then $\hat{y} \in S$, and it will be true that for any $w \in Y$, that

$$w^T\hat{y} - \underset{y \in S}{minimum}\ w^T y \ \ge \frac{1}{(m-1)} \left(\underset{y \in S}{maximum}\ w^T y \ - w^T\hat{y}\right). \tag{4.18}$$

(Note that $dim(S) \le m - 1$ from (4.13) and the definition of S.)

We first demonstrate the following fact:

$$\text{for any } w \in C_Y \text{ with } \|w\| = 1,\ \hat{y}^T w \ge \frac{\bar{\beta}^*\delta(d)}{m\left(\|d\| + \delta(d)\right)} \quad . \tag{4.19}$$

In order to prove (4.19), we choose any $w \in C_Y$ with $\|w\| = 1$. Then, from (2.17), there exists $\bar{v} \in C_Y^*$ with $\|\bar{v}\|_* = 1$ and $\bar{v}^T w \ge \bar{\beta}^*$. Let $M = max\{\|A^T\bar{v}\|_*,\ |b^T\bar{v}|\}$, and set $v = -\frac{A^T\bar{v}}{M}$. Then $\|v\|_* \le 1$. From (3.10) and (3.11), there exists (y, θ) with $\theta = \delta(d)$ such that $A^T y - v\theta \in C_X^*$, $-b^T y - \theta \ge 0$, $y \in C_Y^*$, and $\bar{z}^T y = 1$. Let $y' = y + \frac{\bar{v}\theta}{M}$. Then $A^T y' = A^T y + \frac{A^T\bar{v}\theta}{M} = A^T y - v\theta \in C_X^*$. Also $b^T y' = b^T y + \frac{b^T\bar{v}\theta}{M} \le b^T y + \theta \le 0$. Furthermore, since $y \in C_Y^*$ and $\bar{v} \in C_Y^*$, $y' \in C_Y^*$. We also have $y'^T w = y^T w + \frac{\bar{v}^T w\theta}{M} \ge \frac{\bar{v}^T w\theta}{M} \ge \frac{\bar{\beta}^*\theta}{M} = \frac{\bar{\beta}^*\delta(d)}{M}$. Note that $y'^T\bar{z} = y^T\bar{z} + \frac{\bar{v}^T\bar{z}\theta}{M} = 1 + \frac{\bar{v}^T\bar{z}\theta}{M} \le 1 + \frac{\theta}{M} = 1 + \frac{\delta(d)}{M}$.

Now let $y'' = \frac{y'}{(y'^T\bar{z})}$. Then $y'' \in S$, and $w^T y'' \ge \frac{\bar{\beta}^*\delta(d)}{M\left(1+\frac{\delta(d)}{M}\right)} = \frac{\bar{\beta}^*\delta(d)}{(M+\delta(d))} \ge \frac{\bar{\beta}^*\delta(d)}{\|d\|+\delta(d)}$, since $\|v\|_* = 1$ implies $M \le max\{\|A\|,\ \|b\|\} = \|d\|$.

We now apply (4.18) to assert that

$$w^T\hat{y} \ge w^T\hat{y} - \underset{y \in S}{minimum}\ w^T y \ \ge \left(\frac{1}{m-1}\right) \left(\underset{y \in S}{maximum}\ w^T y - w^T\hat{y}\right)$$

$$\ge \left(\frac{1}{m-1}\right) \left(w^T y'' - w^T\hat{y}\right)$$

which after rearranging yields

$$w^T\hat{y} \ge \frac{\bar{\beta}^*\delta(d)}{m(\|d\| + \delta(d))} \quad , \quad \text{which proves (4.19).}$$

Let $r_4 = \frac{\delta(d)\bar{\beta}^*}{m(\|d\|+\delta(d))}$. To prove (4.16), it suffices to show that $\|y - \hat{y}\|_* \leq r_4$ implies $y \in C_Y^*$. To demonstrate this, it suffices to show that for any $w \in C_Y$ with $\|w\| = 1$, that $w^T y \geq 0$. Let $w \in C_Y$ and $\|w\| = 1$. Then if $\|y - \hat{y}\|_* \leq r_4$, $w^T y = w^T(y - \hat{y}) + w^T \hat{y} \geq -r_4 + \frac{\bar{\beta}^*\delta(d)}{m(\|d\|+\delta(d))} \geq 0$. Thus $y \in C_Y^*$, proving (4.16).

Let $R_4 = \frac{1}{\beta^*}$. Then $\|\hat{y}\|_* \leq \frac{\bar{z}^T\hat{y}}{\beta^*} = \frac{1}{\beta^*} = R_4$, from (2.15), which proves (4.17). Finally, note that

$$\begin{aligned} \frac{R_4}{r_4} &= \frac{1}{\beta^* r_4} = \frac{m(\|d\|+\delta(d))}{\delta(d)\bar{\beta}^*\beta^*} \\ &\leq \frac{m\left(\|d\|+\frac{\rho(d)\|}{\beta^*}\right)}{\delta(d)\bar{\beta}^*\beta^*} \qquad \text{(from Theorem 3.7)} \\ &\leq \frac{2m\|d\|}{\rho(d)(\beta^*)^2\bar{\beta}^*} \qquad , \text{ from Proposition 4.1} \quad . \blacksquare \end{aligned}$$

5 Synthesis of Results

In this section, we synthesize the results of the previous two sections into theorems that characterize aspects of the distance to ill-posedness for the three particular cases of problem P of (1.1), namely

(i) *Case* 1 : C_X *and* C_Y *are both regular,*

(ii) *Case* 2 : C_X *is regular and* $C_Y = \{0\}$,

(iii) *Case* 3 : $C_X = X$ *and* C_Y *is regular,*

and for the status of solvability of P of (1.1), namely

(a) P *is consistent, i.e.,* (1.1) *has a solution, and*

(b) P *is inconsistent, i.e.,* (2.2) *has a solution.*

Each of the six theorems of this section synthesizes our results of the previous two sections, as applied to the one of the three cases above and one of the two status' of the solvability of P. Each theorem summarizes the applicable approximation characterizations of $\rho(d)$ of Section 3, and also synthesizes the appropriate bounds on radii of contained and intersecting balls developed in Section 4. For a motivation of the importance of these bounds on radii of contained and intersecting balls contained herein, the reader is referred to the opening discussion at the beginning of Section 4.

Each case is treated as a separate subsection, and all proofs are deferred to the end of the section.

5.1 Case 1: C_X and C_Y are both regular.

Theorem 5.1 *Suppose that C_X and C_Y are both regular. If P is consistent, i.e., $d \in \mathcal{F}$, then* *(i)* $\beta \cdot \alpha(d) \leq \rho(d) \leq \alpha(d)$

(ii) $\alpha(d) = w(d)$, *and so* $\beta \cdot w(d) \leq \rho(d) \leq w(d)$

(iii) $\beta^* \beta \cdot \tilde{\alpha}(d) \leq \rho(d) \leq \tilde{\alpha}(d)$

(iv) $\beta^* \cdot u(d) \leq \rho(d) \leq u(d)$

(v) $\frac{\beta^*}{v(d)} \leq \rho(d) \leq \frac{1}{v(d)}$.

(vi) If $\rho(d) > 0$, *then there exists* $\hat{x} \in X_d$ *and positive scalars* r *and* R *satisfying:*

$$(a) \quad B(\hat{x}, r) \subset X_d$$

$$(b) \quad B(\hat{x}, r) \subset B(0, R)$$

$$(c) \quad \frac{R}{r} \leq \frac{(4n+10)\|d\|}{\beta^2 \bar{\beta} \beta^* \rho(d)}$$

$$(d) \quad r \geq \frac{\beta \bar{\beta} \beta^* \rho(d)}{4(n+2)\|d\|}$$

$$(e) \quad R \leq \frac{(2n+6)\|d\|}{\beta^2 \bar{\beta} \beta^* \rho(d)} \quad .$$

Theorem 5.2 *Suppose that* C_X *and* C_Y *are both regular. If* P *is not consistent, i.e.,* $d \in \mathcal{F}^{\mathcal{C}}$, *then*

(i) $\beta^* \cdot \sigma(d) \leq \rho(d) \leq \sigma(d)$

(ii) $\sigma(d) = \delta(d)$, *and so* $\beta^* \cdot \delta(d) \leq \rho(d) \leq \delta(d)$

(iii) $\beta^* \beta \cdot \tilde{\sigma}(d) \leq \rho(d) \leq \tilde{\sigma}(d)$

(iv) $\beta \cdot g(d) \leq \rho(d) \leq g(d)$

(v) $\frac{\beta}{h(d)} \leq \rho(d) \leq \frac{1}{h(d)}$.

(vi) If $\rho(d) > 0$, *then there exists* $\hat{y} \in Y_d$ *and positive scalars* r *and* R *satisfying*

$$\frac{R}{r} \leq \frac{(4m+1)\|d\|}{(\beta^*)^2 \beta \bar{\beta}^* \cdot \rho(d)}$$

such that

$$\{y \in Y^* | \|y - \hat{y}\|_* \leq r\} \subset Y_d$$

and

$$\{y \in Y^* | \|y - \hat{y}\|_* \leq r\} \subset \{y \in Y^* | \|y\|_* \leq R\} \quad .$$

5.2 Case 2: C_X is regular and $C_Y = \{0\}$.

Theorem 5.3 *Suppose that* C_X *is regular and* $C_Y = \{0\}$. *If* P *is consistent, i.e.,* $d \in \mathcal{F}$, *then*

(i) $\beta \cdot \alpha(d) \leq \rho(d) \leq \alpha(d)$

(ii) $\alpha(d) = w(d)$, *and so* $\beta \cdot w(d) \leq \rho(d) \leq w(d)$

(iii) If $\rho(d) > 0$, *then there exists* $\hat{x} \in X_d$ *and positive scalars* r *and* R *satisfying:*

$(a) \quad \{x \in X \mid \|x - \hat{x}\| \leq r,\ Ax = b\} \subset X_d$

$(b) \quad \{x \in X \mid \|x - \hat{x}\| \leq r,\ Ax = b\} \subset B(0, R)$

$(c) \quad \frac{R}{r} \leq \frac{(2n+5)\|d\|}{\rho(d)\beta^2\bar{\beta}}$

$(d) \quad r \geq \frac{\beta\bar{\beta}\rho(d)}{2(n+2)\|d\|}$

$(e) \quad R \leq \frac{(2n+6)\|d\|}{\beta^2\bar{\beta}\rho(d)}$.

Theorem 5.4 *Suppose that* C_X *is regular and* $C_Y = \{0\}$. *If* P *is not consistent, i.e.,* $d \in \mathcal{F}^{\mathcal{C}}$, *then*

(i) $\beta \cdot g(d) \leq \rho(d) \leq g(d)$

(ii) $\frac{\beta}{h(d)} \leq \rho(d) \leq \frac{1}{h(d)}$

(iii) If $\rho(d) > 0$, *then there exists* $\hat{y} \in Y_d$ *and positive scalars* r *and* R *satisfying*

$$\frac{R}{r} \leq \frac{2\|d\|}{\beta \cdot \rho(d)}$$

such that

$$\{y \in Y^* \mid \|y - \hat{y}\|_* \leq r\} \subset Y_d$$

and

$$\{y \in Y^* \mid \|y - \hat{y}\|_* \leq r\} \subset \{y \in Y^* \mid \|y\|_* \leq R\} \quad .$$

5.3 Case 3: $C_X = X$ and C_Y is regular.

Theorem 5.5 *Suppose that* $C_X = X$ *and* C_Y *is regular. If* P *is consistent, i.e.,* $d \in \mathcal{F}$, *then*

(i) $\beta^* \cdot u(d) \leq \rho(d) \leq u(d)$

(ii) $\frac{\beta^*}{v(d)} \leq \rho(d) \leq \frac{1}{v(d)}$

(iii) If $\rho(d) > 0$, then there exists $\hat{x} \in X_d$ and positive scalars r and R satisfying:

$$\begin{array}{ll} (a) & \{x \in X \mid \|x - \hat{x}\| \leq r\} \subset X_d \\ (b) & \{x \in X \mid \|x - \hat{x}\| \leq r\} \subset B(0, R) \\ (c) & \frac{R}{r} \leq \frac{3\|d\|}{\beta^* \rho(d)} \\ (d) & r \geq \frac{\beta^* \rho(d)}{3\|d\|} \\ (e) & R \leq \frac{3\|d\|}{\beta^* \rho(d)} \end{array} .$$

Theorem 5.6 *Suppose that $C_X = X$ and C_Y is regular. If P is not consistent, i.e., $d \in \mathcal{F}^{\mathcal{C}}$, then*

(i) $\beta^* \cdot \sigma(d) \leq \rho(d) \leq \sigma(d)$

(ii) $\sigma(d) = \delta(d)$, and so $\beta^* \cdot \delta(d) \leq \rho(d) \leq \delta(d)$

(iii) If $\rho(d) > 0$, then there exists $\hat{y} \in Y_d$ and positive scalars r and R satisfying

$$\frac{R}{r} \leq \frac{(4m+1)\|d\|}{(\beta^*)^2 \bar{\beta}^* \cdot \rho(d)}$$

such that

$$\{y \in Y^* \mid \|y - \hat{y}\|_* \leq r,\ A^T y = 0\} \subset Y_d$$

and

$$\{y \in Y^* \mid \|y - \hat{y}\|_* \leq r\} \subset \{y \in Y^* \mid \|y\| \leq R\} \quad .$$

Proof of Theorem 5.1: Parts (i), (ii), (iii), (iv) and (v) follow directly from Theorems 3.1, 3.2, 3.3, 3.4 and 3.5, respectively. It remains to prove part (vi).

Let $S = \{x \in X | b - Ax \in C_Y\}$ and $T = C_X$. Then $S \cap T = X_d$. From Lemma 4.1, there exists $\hat{x}_1 \in X_d$ and r_1, R_1 satisfying conditions $(i) - (v)$ of Lemma 4.1. From Lemma 4.2, there exists $\hat{x}_2 \in X_d$ and r_2, R_2 satisfying conditions $(i) - (v)$ of Lemma 4.2. Then the conditions of Proposition A.2 of the Appendix are satisfied, and so there exists $\hat{x}$ and $r, \hat{R}$ satisfying the five conditions of Proposition A.2. Therefore, (i) $B(\hat{x}, r) \subset S \cap T = X_d$, which is (a). Also from (ii), $B(\hat{x}, r) \subset B(0, \hat{R})$, which is (b). From (iii), we have

$$\frac{\hat{R}}{r} \leq 2max\left\{\frac{R_1}{r_1}, \frac{R_2}{r_2}\right\} \leq \frac{(4n+10)\|d\|}{\beta^2 \bar{\beta} \beta^* \rho(d)}$$

(invoking Lemma 4.1 (iii) and Lemma 4.2 (iii)), which is (c). Similarly applying Lemma 4.1 and 4.2 and Proposition A.2 in parts (iv) and (v) yields

$$r \geq \frac{1}{2} min\{r_1, r_2\} \geq \frac{\beta \bar{\beta} \beta^* \rho(d)}{4(n+2)\|d\|}$$

and

$$\hat{R} \leq max\{R_1, R_2\} \leq \frac{(2n+6)\|d\|}{\beta^2 \beta^* \rho(d)} \quad . \ \blacksquare$$

Proof of Theorem 5.2: Parts (i), (ii), (iii), (iv) and (v) follow directly from Theorems 3.6, 3.7, 3.8, 3.9 and 3.10, respectively. It remains to prove part (vi).

Let $S = \{y \in Y^* | A^T y \in C_X^*,\ b^T y \le 0\}$ and $T = C_Y^*$. Then $S \cap T = Y_d$. From Lemma 4.3, there exists $\hat{y}_3,\ r_3,\ R_3$ satisfying $\hat{y}_3 \in S \cap T,\ B(\hat{y}_3,\ r_3) \subset S$ and $\|\hat{y}_3\|_* \le R_3$, and $\frac{R_3}{r_3} \le \frac{\|d\|}{\beta\rho(d)}$. From Lemma 4.4 there exists $\hat{y}_4,\ r_4,\ R_4$ satisfying $\hat{y}_4 \in S \cap T,\ B(\hat{y}_4,\ r_4) \subset T$, and $\|\hat{y}_4\|_* \le R_4$, and $\frac{R_4}{r_4} \le \frac{2m\|d\|}{(\beta^*)^2\beta^*\rho(d)}$. Then from Proposition A.1 of the Appendix, there exists $\hat{y}$ and $r, \hat{R}$ satisfying $B(\hat{y},\ r) \subset S \cap T = Y_d$, and $\|\hat{y}\| \le \hat{R}$, and

$$\begin{aligned}\frac{\hat{R}}{r} &\le 2max\left\{\frac{R_3}{r_3},\ \frac{R_4}{r_4}\right\} \le 2max\left\{\frac{\|d\|}{\beta\rho(d)},\ \frac{2m\|d\|}{(\beta^*)^2\beta^*\rho(d)}\right\} \\ &\le \frac{4m\|d\|}{\beta(\beta^*)^2\beta^*\rho(d)} \quad .\end{aligned}$$

Now let $R = \hat{R} + r$. Then for any $y \in B(\hat{y},\ r)$, $\|y\|_* \le \|\hat{y}\|_* + r \le \hat{R} + r = R$, and

$$\begin{aligned}\frac{R}{r} = \frac{\hat{R}}{r} + 1 &\le \frac{4m\|d\|}{\beta(\beta^*)^2\beta^*\rho(d)} + \frac{\|d\|}{\rho(d)} \quad \text{(from Proposition 4.1)} \\ &\le \frac{(4m+1)\|d\|}{\beta(\beta^*)^2\beta^*\rho(d)} \quad .\end{aligned}$$

∎

Proof of Theorem 5.3: Parts (i) and (ii) follow directly from Theorems 3.1 and 3.2, respectively. To prove (iii) we apply Lemma 4.2; there exists $\hat{x} \in X_d$ and $r_2,\ R_2$ satisfying the five conditions of Lemma 4.2. Let $r = r_2$ and $R = R_2 + r_2$. Then (b), (c), (d), and (e) follow directly. To prove (a), observe that from Lemma 4.2 (i) that

$$\{x \in X|\ \|x - \hat{x}\| \le r\} \subset C_X \quad ,$$

and intersecting both sides with the affine set $\{x \in X | Ax = b\}$ gives

$$\{x \in X|\ \|x - \hat{x}\| \le r,\ Ax = b\} \subset C_X \cap \{x \in X | Ax = b\} = X_d \quad .$$

∎

Proof of Theorem 5.4: Parts (i) and (ii) follow directly from Theorems 3.9 and 3.10, respectively. To prove (iii) we apply Lemma 4.3; there exists $\hat{y} \in Y_d$ and $r_3,\ R_3$ satisfying $\frac{R_3}{r_3} \le \frac{\|d\|}{\beta\rho(d)}$ and (4.14) and (4.15). Let $r = r_3$ and $R = R_3 + r_3$. Then from (4.14) we obtain

$$\{y \in Y^*|\ \|y - \hat{y}\|_* \le r\} \subset \{y \in Y^* | A^T y \in C_X^*,\ b^T y \le 0\} = Y_d \quad .$$

Also, for any y satisfying $\|y - \hat{y}\| \le r$, $\|y\| \le \|\hat{y}\| + r \le R_3 + r_3 = R$. Finally, note that

$$\begin{aligned}\frac{R}{r} = \frac{R_3}{r_3} + 1 &\le \frac{\|d\|}{\beta\rho(d)} + 1 \le \frac{\|d\|}{\beta\rho(d)} + \frac{\|d\|}{\rho(d)} \quad \text{(from Proposition 4.1)} \\ &\le \frac{2\|d\|}{\beta\rho(d)} \quad .\end{aligned}$$

∎

Proof of Theorem 5.5: Parts (i) and (ii) follow directly from Theorems 3.4 and 3.5, respectively. To prove (iii) we apply Lemma 4.1; there exists $\hat{x} \in X_d$ and $r_1,\ R_1$ satisfying the five conditions of Lemma 4.1. Let $r = r_1$ and $R = R_1$. Then (b), (c), (d), and (e) following directly. To prove (a), observe from Lemma 4.1(i) that

$$\{x \in X|\ \|x - \hat{x}\| \le r\} \subset \{x \in X | b - Ax \in C_Y\} = X_d \quad .$$

∎

Proof of Theorem 5.6: Parts (i) and (ii) follow from Theorems 3.6 and 3.7, respectively. It remains to prove (iii).

Let $S = \{y \in Y^* | b^T y \leq 0\}$. If we let $v = 0$ in (3.9), we see that there exists $\hat{y}_1 \in C_Y^*$ satisfying $A^T\hat{y}_1 = 0, \bar{z}^T\hat{y}_1 = 1$ and $-b^T\hat{y}_1 \geq \delta(d) \geq \rho(d)$, from Theorem 3.7. Therefore, if we set $r_1 = \frac{\rho(d)}{\|d\|}$ and $R_1 = \frac{1}{\beta^*}$, we have $\|\hat{y}_1\|_* \leq \frac{\bar{z}^T\hat{y}_1}{\beta^*} = R_1$ (from 2.15), and for any y satisfying $\|y - \hat{y}_1\|_* \leq r_1$, we have $b^T y = b^T(y - \hat{y}_1) + b^T\hat{y}_1 \leq \|b\| \, \|\hat{y}_1 - y\|_* - \rho(d) \leq \|d\| r_1 - \rho(d) = 0$, and so $B(\hat{y}_1, \ r_1) \subset S$. If we let $T = C_Y^*$, we have $S \cap T \cap \{y \in Y^* | A^T y = 0\} = Y_d$, and $\hat{y}_1 \in Y_d$. From Lemma 4.4, there exists $\hat{y}_2$ and r_4, R_4 satisfying $\frac{R_4}{r_4} \leq \frac{2m\|d\|}{(\beta^*)^2\bar{\beta}^*\rho(d)}$, and (4.16) and (4.17). Then from Proposition A.1 of the Appendix, there exists $\hat{y}$ and r, $\hat{R}$ satisfying $B(\hat{y}, r) \subset S \cap T$ and $\|\hat{y}\| \leq \hat{R}$, and

$$\frac{\hat{R}}{r} \leq 2max\left\{\frac{\|d\|}{\beta^*\rho(d)}, \ \frac{2m\|d\|}{(\beta^*)^2\bar{\beta}^*\rho(d)}\right\} = \frac{4m\|d\|}{(\beta^*)^2\bar{\beta}^*\rho(d)} \quad .$$

Note also that

$$\{y \in Y^* | \ \|y - \hat{y}\|_* \leq r, \ A^T y = 0\} \subset S \cap T \cap \{y \in Y^* | A^T y = 0\} = Y_d \ .$$

Let $R = \hat{R} + r$. Then for any $y \in B(\hat{y}, r)$, $\|y\|_* \leq \|\hat{y}\|_* + r = \hat{R} + r = R$, and

$$\begin{aligned} \frac{R}{r} = \frac{\hat{R}}{r} + 1 \quad &\leq \frac{4m\|d\|}{(\beta^*)^2\bar{\beta}^*\rho(d)} \quad + \frac{\|d\|}{\rho(d)} \quad \text{(from Proposition 4.1)} \\ &\leq \frac{(4m+1)\|d\|}{(\beta^*)^2\bar{\beta}^*\rho(d)} \quad . \end{aligned}$$

∎

APPENDIX

This appendix contains four results that are used in the body of the paper. The first result is an application of gauge duality theory from [5]. The second result is a duality theorem for a class of mathematical programs, and the last two results are simple constructions with balls on the intersection of two sets.

1. Nonlinear gauge duality for closed gauge functions

Let V be a finite-dimensional normed linear vector space with norm $\|v\|$ for $v \in V$. A function $f(\cdot) : V \to R \cup \{+\infty\}$ is a *gauge* function if $f(\cdot)$ is a nonnegative convex function that is positively homogeneous of degree one, and $f(0) = 0$. The level sets of $f(\cdot)$ are those sets of the form $S_\alpha = \{v \in V | f(v) \leq \alpha\}$. If all of the level sets of $f(\cdot)$ are closed sets, then $f(\cdot)$ is a closed gauge function.

Corresponding to every gauge function $f(\cdot)$ is a *polar* gauge function $f^*(y)$ defined for all y in the dual space V^*. When $f(\cdot)$ is a closed gauge function, this polar function takes on the convenient definitional form:

$$f^*(y) = \sup_{v} \; y^T v \quad \text{s.t.} \quad f(v) \leq 1 \quad , \tag{A.1}$$

for all $y \in V^*$. Also, when $f(\cdot)$ is a closed gauge function, $f^{**}(\cdot) = f(\cdot)$, provided that we identify V^{**} with V. An excellent treatment of polarity for gauge functions is given in [15].

Consider the following optimization problem:

GP :

$$\begin{aligned} s^* = \; & \underset{x}{\text{infimum}} \quad f(Mx) \\ & \text{s.t.} \quad x \in K \\ & \qquad\; d^T x = 1 \quad , \end{aligned} \tag{A.2}$$

and its <u>gauge dual</u>, defined as

GD :

$$\begin{aligned} t^* = \; & \underset{y}{\text{infimum}} \quad f^*(y) \\ & \text{s.t.} \quad M^T y - d \in K^* \\ & \qquad\; y \in V^* \quad , \end{aligned} \tag{A.3}$$

where M is a linear operator from a finite-dimensional normed linear vector space U to the finite-dimensional normed linear vector space V, i.e., $M \in L(U, V)$, d is a linear functional, i.e., $d \in U^*$, and K is a convex cone in U. The following duality theorem is a special instance of Theorem 2A in [5].

Theorem A.1 : (I) If $s^* = 0$, then $t^* = +\infty$.
(II) If $t^* = 0$, then $s^* = +\infty$.
(III) If $s^* > 0$ and $t^* > 0$, then $s^* t^* = 1$ under the following hypotheses :
(i) all projections of $\{x \in K | d^T x = 1\}$ are closed sets,
(ii) $\sup\{d^T x | x \in K\} = +\infty$,
(iii) $f(\cdot)$ is a closed gauge function, and
(iv) $\{v \in V | f(v) \leq 1\}$ is a closed and bounded set.

Before proving the theorem, we first prove the following:

Lemma A.1 *Under the four hypotheses of Theorem A.1 (III), suppose that $\bar{y} \in V^*$ is given. Then*

$$\bar{y}^T M x > 1 \ for\ all\ x \in K \cap \{x \in U | d^T x = 1\} \tag{A.4}$$

if and only if

$$there\ exists\ \pi > 1\ satisfying\ M^T \bar{y} - d\pi \in K^* \quad . \tag{A.5}$$

Proof : First suppose (A.5) holds. Then, if $x \in K$ and $d^T x = 1$, we have $x^T (M^T \bar{y} - d\pi) \geq 0$, so that $\bar{y}^T M x \geq \pi d^T x = \pi > 1$, so that (A.4) is true. Conversely, suppose that (A.4) holds. Let

$$\begin{array}{rll} \epsilon = & \underset{x}{minimum} & \bar{y}^T M x \\ & s.t. & x \in K \\ & & d^T x = 1 \quad . \end{array}$$

From (A.4) and the hypothesis (i), it must be true that $\epsilon > 1$. Let $S = \{(\alpha, \delta) \in R \times R \mid there\ exists\ x \in K\ satisfying\ 1 - d^T x = \alpha,\ \bar{y}^T M x - \epsilon < \delta\}$. Then S is a nonempty convex set and $(0,0) \notin S$. Thus, there exists a hyperplane that separates $(0,0)$ from S, i.e., there exists $(\pi, \gamma) \neq (0,0)$ with the property that $\alpha\pi + \delta\gamma \geq 0$ for any $(\alpha, \gamma) \in S$. Therefore $\gamma \geq 0$. Also, for any $x \in K$ and any $\mu > 0$, $\pi\left(1 - d^T x\right) + \left(\bar{y}^T M x - \epsilon + \mu\right)\gamma \geq 0$. If $\gamma = 0$, then $\pi \neq 0$ and $\pi \geq \pi d^T x$ for any $x \in K$. Thus $\pi \geq 0$ and $\pi = 1$ without loss of generality. Thus $d^T x \leq 1$ for any $x \in K$, violating the second hypothesis of the theorem. As this is a contradiction, it must be true that $\gamma > 0$, and so $\gamma = 1$ without a loss of generality. Then $\pi - \pi d^T x + \bar{y}^T M x - \epsilon + \mu \geq 0$ for any $x \in K$, and $\mu > 0$. Therefore, $M^T \bar{y} - \pi d \in K^*$. Also, upon setting $x = 0$ we obtain $\pi \geq \epsilon - \mu$ for any $\mu > 0$. Thus $\pi \geq \epsilon > 1$, proving that (A.5) holds. ∎

Proof of Theorem A.1 : If s^* and t^* are both finite, then GP and GD have feasible solutions. Note that for any feasible solutions x of GP and y of GD, that $x^T M^T y - x^T d \geq 0$. Therefore, $1 = d^T x \leq y^T M x \leq f(Mx) f^*(y)$, where the last inequality follows from (A.1). Therefore $s^* t^* \geq 1$, which is "weak duality" for these dual programs. If either $s^* = 0$ or $t^* = 0$, then $t^* = +\infty$ or $s^* = +\infty$, respectively, demonstrating (I) and (II) of the theorem.

To prove (III), we suppose that $s^* t^* > 1$ and derive a contradiction. Therefore $s^* > \frac{1}{t^*}$, and so the two sets

$$S_1 = \left\{v \in V | v = Mx\ for\ some\ x \in K\ satisfying\ d^T x = 1\right\}$$

and

$$S_2 = \left\{ v \in V | f(v) \leq \frac{1}{t^*} \right\}$$

are nonempty and disjoint. Now from hypothesis (i) we have S_1 is a closed convex set, and from hypotheses (iii) and (iv) we have that S_2 is a closed and bounded convex set. Since $S_1 \cap S_2 = \phi$, there exists a hyperplane that strictly separates S_1 and S_2, due to the boundedness of S_2. Thus there exists $\bar{y} \in V^*$ and β such that

$$\bar{y}^T Mx > \beta \text{ for all } x \in K \text{ satisfying } d^T x = 1 \tag{A.6}$$

and

$$\bar{y}^T v < \beta \text{ for all } v \text{ that satisfy } f(v) \leq \frac{1}{t^*} \quad . \tag{A.7}$$

With $v = 0$ in (A.7) we see that $\beta > 0$, and so $\beta = 1$ without loss of generality. From (A.7) and (A.1), we conclude that $f^*(\bar{y}) \leq t^*$. Next we observe that with $\beta = 1$ that (A.6) corresponds to (A.4), so by Lemma A.1, there exists $\pi > 1$ satisfying $M^T \bar{y} - d\pi \in K^*$. With $y = \frac{\bar{y}}{\pi}$, we have $M^T y - d \in K^*$, $y \in V^*$, and $f^*(y) = \frac{1}{\pi} f^*(\bar{y}) < f^*(\bar{y}) \leq t^*$. However, as y is feasible for GD, $t^* \leq f^*(y)$, which yields a contradiction. Thus $s^* t^* = 1$ is proved. ▮

2. A Strong Duality Theorem

Consider the following pair of optimization problems:

NP :

$$\begin{array}{rll} v_P = & \underset{x,\theta}{\text{maximum}} & \theta \\ & s.t. & Mx - \theta w \in K_1 \\ & & x \in K_2 \\ & & f^T x = 1 \\ & & \theta \geq 0 \quad , \end{array} \tag{A.8}$$

ND :

$$\begin{array}{rll} v_D = & \underset{y,\gamma}{\text{minimum}} & \gamma \\ & s.t. & -M^T y + \gamma f \in K_2^* \\ & & y^T w \geq 1 \\ & & y \in K_1^* \quad , \end{array} \tag{A.9}$$

where K_1, K_2 are convex cones in finite-dimensional normed linear vector spaces V_1 and V_2, respectively, $M \in L(V_2, V_1)$, $f \in V_2^*$, and $w \in V_1$.

Theorem A.2: *Suppose that NP has a feasible solution. If K_1 and K_2 are closed convex cones and $f \in int\ K_2^*$, then $v_P = v_D$.*

Proof: First suppose that (x, θ) and (y, γ) are feasible for NP and ND, respectively. Then $y^T Mx - \theta y^T w \geq 0$, $-x^T M^T y + \gamma f^T x \geq 0$, and $\theta y^T w \geq \theta$, all follow by combining feasibility conditions of NP and ND. Summing these inequalities we obtain $\gamma = \gamma f^T x \geq \theta$, whereby $v_D \geq v_P$.

By the hypothesis of the theorem, $v_P > -\infty$. If $v_P = +\infty$, then $v_D = +\infty$ from the above inequality. Therefore it remains to consider the case when v_P is finite. We proceed as follows:

Let us define the set $S = \Big\{(s,t,z) \in V_2^* \times R \times R | there\ exists\ (y,\gamma,v)\ satisfying\ y \in K_1^*,$ $v \in K_2^*, -M^T y + \gamma f + s = v, 1 - y^T w \le t, \gamma \le z\Big\}$. Then $S \ne \phi$, and for any given $\epsilon > 0$, the point $(0,0,v_D - \epsilon) \notin S$. Therefore, since S is a convex set, there exists a hyperplane separating S from $(0,0,v_D-\epsilon)$. Thus, there exists $(x,\theta,\delta) \ne 0$ and g such that

$$\delta\,(v_D - \epsilon) \le g \le x^T s + \theta t + \delta z\ for\ any\ (s,t,z) \in S\quad .$$

This implies that $\theta \ge 0$. Also, for any $v \in K_2^*$ and any $y \in K_1^*$, and any γ,

$$x^T\left(v + M^T y - \gamma f\right) + \theta\left(1 - y^T w\right) + \delta\gamma \ge \delta\,(v_D - \epsilon)\quad .$$

In particular, this implies that $x \in K_2$ and $Mx - \theta w \in K_1$, $f^T x = \delta$, and $\theta \ge \delta\,(v_D - \epsilon)$. Also, since $f \in int\ K_2^*$ from the hypothesis of the theorem, then $\delta \ge 0$. If $\delta > 0$, without loss of generality $\delta = 1$, so that $f^T x = 1$ and $\theta \ge v_D - \epsilon$. Then (x,θ) is feasible for NP, and $v_P \ge \theta \ge v_D - \epsilon$. As this is true for any $\epsilon > 0$, then $v_P \ge v_D$, which proves that $v_P = v_D$.

It only remains to examine the case when $\delta = 0$. Then $f^T x = 0$, which implies that $x = 0$, since $f \in int\ K_2^*$ and $x \in K_2$. Therefore, $\theta \ne 0$, and since $\theta \ge 0$, $\theta = 1$ without loss of generality. Then $1 - y^T w \ge 0$ for any $y \in K_1^*$, whereby $-w \in K_1$. Because NP is feasible, there exists $(\bar{x},\bar{\theta})$ feasible for NP. And since $-w \in K_1$, $(\bar{x},\bar{\theta}+\theta)$ is feasible for NP for all $\theta \ge 0$, whereby $v_P = +\infty$, which contradicts the supposition of the case. Thus $\delta > 0$ as desired. ▮

3. A Construction Using Inscribed Balls and Intersecting Sets

Proposition A.1 *Let X be a finite-dimensional normed linear vector space with norm $\|\cdot\|$ and let S and T be 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 $\|\hat{x}_1\| \le R_1$, and

(ii) $\hat{x}_2 \in S \cap T$, $B(\hat{x}_2, r_2) \subset T$, where $r_2 > 0$, and $\|\hat{x}_2\| \le R_2$.

Let $\alpha = \frac{r_2}{r_1+r_2}$, and $r = \frac{r_1 r_2}{r_1+r_2}$, and $\hat{R} = \alpha R_1 + (1-\alpha)R_2$.

Then the point $\hat{x} = \alpha\hat{x}_1 + (1-\alpha)\hat{x}_2$ will satisfy:

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

$$(ii)\quad \|\hat{x}\| \le \hat{R}\ ,$$

$$and\quad (iii)\quad \frac{\hat{R}}{r} \le 2max\left\{\frac{R_1}{r_1},\ \frac{R_2}{r_2}\right\}\quad .$$

Proof: First note that $0 \le \alpha \le 1$. Because $B(\hat{x}_1, r_1) \subset S$ and $\hat{x}_2 \in S$, $B(\alpha\hat{x}_1 + (1-\alpha)\hat{x}_2, \alpha r_1) \subset S$. Similarly, because $B(\hat{x}_2, r_2) \subset T$ and $\hat{x}_1 \in T$, $B(\alpha\hat{x}_1 + (1-\alpha)\hat{x}_2,\ (1-\alpha)r_2) \subset T$. Noticing that $\alpha r_1 = (1-\alpha)r_2 = r$, we have $B(\hat{x}, r) = B(\alpha\hat{x}_1 + (1-\alpha)\hat{x}_2,\ r) \subset S \cap T$. Also $\|\hat{x}\| \le \alpha\|\hat{x}_1\| + (1-\alpha)\|\hat{x}_2\| \le \alpha R_1 + (1-\alpha)R_2 = \hat{R}$. Finally, to show (iii), suppose first that $\frac{R_1}{r_1} \ge \frac{R_2}{r_2}$. Then

$$\frac{\hat{R}}{r} = \frac{\alpha R_1 + (1-\alpha)R_2}{r} = \frac{r_2 R_1 + r_1 R_2}{r_1 r_2} \le \frac{r_2 R_1 + r_2 R_1}{r_1 r_2} = \frac{2R_1}{r_1} = 2max\left\{\frac{R_1}{r_1},\ \frac{R_2}{r_2}\right\}\quad .$$

A similar argument holds when $\frac{R_1}{r_1} \leq \frac{R_2}{r_2}$. ∎

Proposition A.2 *Let X be a finite-dimensional normed linear vector space with norm $\|\cdot\|$ and let S and T be 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 $\alpha = \frac{r_2}{r_1+r_2}$, and $r = \frac{r_1 r_2}{r_1+r_2}$, and $\hat{R} = \alpha R_1 + (1-\alpha) R_2$. Then the point $\hat{x} = \alpha \hat{x}_1 + (1-\alpha)\hat{x}_2$ will satisfy:

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

$$(ii) \quad B(\hat{x}, r) \subset B(0, \hat{R}) \ ,$$

$$(iii) \quad \frac{\hat{R}}{r} \leq 2max\left\{\frac{R_1}{r_1}, \ \frac{R_2}{r_2}\right\} \quad ,$$

$$(iv) \quad r \geq \frac{1}{2} min\{r_1, r_2\} \quad ,$$

$$and \quad (v) \quad \hat{R} \leq max\{R_1, R_2\} \quad .$$

Proof: Parts (i) and (iii) follow identically the proof of Proposition A.1. To see (iv), note that by definition of r, $r \geq \frac{min\{r_1,r_2\}max\{r_1,r_2\}}{2max\{r_1,r_2\}} = \frac{1}{2}min\{r_1, r_2\}$. Part (v) follows from the fact that $\hat{R}$ is a convex combination of R_1 and R_2. To prove (ii), note that for any $x \in B(\hat{x}, r)$, we have $\|x\| \leq \|\hat{x}\| + r \leq \alpha\|\hat{x}_1\| + (1-\alpha)\|\hat{x}_2\| + r$. However, $\|\hat{x}_i\| + r_i \leq R_i$, $i = 1, 2$, so that $\|x\| \leq \alpha(R_1 - r_1) + (1-\alpha)(R_2 - r_2) + r = R - r \leq R$, which completes the proof. ∎

References

[1] F. Alizadeh, *Interior Point Methods in Semidefinite Programming with Applications to Combinatorial Optimization*, SIAM Journal on Optimization 5 (1995), pp. 13-51.

[2] M. Bazaraa, Sherali, H.D., and Shetty, C.M., Nonlinear Programming Theory and Applications, Wiley, 1970, New York.

[3] L. Blum, M. Shub, and S. Smale, *On a Theory of Computation and Complexity over the Real Numbers: NP-completeness, Recursive Functions and Universal Machines*, Bull. Amer. Math. Soc., 21 (1989), pp. 1-46.

[4] L. Danzer, Grünbaum, B., and Klee, V., *Helly's Theorem and its relatives*, in Convexity, B. Klee, ed., amer. Math. Soc., Providentce, RI, pp. 101-180.

[5] R.M. Freund, *Dual Gauge Programs, with Applications to Quadratic Programming and the Minimum-Norm Problem*, Mathematical Programming 38 (1987), pp. 47-67.

[6] S. Filipowski, *On the Complexity of Solving Linear Programs Specified with Approximate Data and Known to be Feasible*, Technical Report, Dept. of Industrial and Manufacturing Systems Engineering, Iowa State Univeristy, May 1994, Ames, Iowa.

[7] S. Filipowski, *On the Complexity of Solving Sparse Symmetric Linear Programs Specified with Approximate Data*, Technical Report, Dept. of Industrial and Manufacturing Systems Engineering, Iowa State Univeristy, December 1994, Ames, Iowa.

[8] A.S. Lewis, *Group invariance and convex matrix analysis*, Technical Report, Dept. of Combinatorics and Optimization, University of Waterloo, February 1995, Waterloo, Canada.

[9] Nesterov, Y., A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming. SIAM Studies in Applied Mathematics No. 13, SIAM, Philadelphia, 1994.

[10] M. Nunez, and R.M. Freund, *Condition Measures and Properties of the Central Trajectory of a Linear Program*, in preparation.

[11] J. Renegar, *Some Perturbation Theory for Linear Programming*, Mathematical Programming 65 (1994) 1, pp. 73-92.

[12] J. Renegar, *Incorporating Condition Measures into the Complexity Theory of Linear Programming*, to apperar in *SIAM Journal on Optimization.*

[13] J. Renegar, *Linear Programming, Complexity Theory and Elementary Functional Analysis.* To appear in Mathematical Programming.

[14] S. Smale, *Some Remarks on the Foundations of Numerical Analysis*, SIAM Review, 32 (1990) pp.211-220.

[15] T. Rockafellar, *Convex Analysis*, Princeton Univeristy Press, 1970, Princeton, N.J.

[16] J. Vera, Ill-Posedness in Mathematical Programming and Problem Solving with Approximate Data. Ph.D. Dissertation, Cornell University, Ithaca, N.Y. 1992.

[17] J. Vera, *Ill-Posedness and the Complexity of Deciding Existence of Solutions to Linear Programs*, to apperar in SIAM Journal on Optimization.

[18] J. Vera, *On the Complexity of Linear Programming under Finite Precision Arithmetic*, working paper, Dept. of Industrial Engineering, University of Chile, 1994.

[19] K. Yosida, *Functional Analysis*, Springer, Berlin, 1987.