

# ***OPERATIONS RESEARCH CENTER***

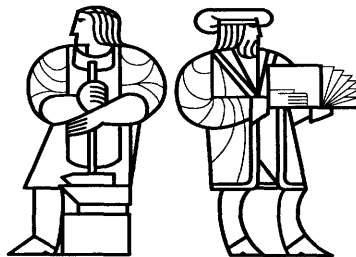
## ***Working Paper***

*Pre-Conditioners and Relations between Different Measures of  
Conditioning for Conic Linear Systems*

by  
M. Epelman  
R.M. Freund

OR 344-00

June 2000



***MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY***



# Pre-Conditioners and Relations between Different Measures of Conditioning for Conic Linear Systems

Marina Epelman<sup>\*†</sup> and Robert M. Freund<sup>‡</sup>

June, 2000

## Abstract

In recent years, new and powerful research into “condition numbers” for convex optimization has been developed, aimed at capturing the intuitive notion of problem behavior. This research has been shown to be important in studying the efficiency of algorithms, including interior-point algorithms, for convex optimization as well as other behavioral characteristics of these problems such as problem geometry, deformation under data perturbation, etc. This paper studies measures of conditioning for a conic linear system of the form  $(\text{FP}_d)$ :  $Ax = b$ ,  $x \in C_X$ , whose data is  $d = (A, b)$ . We present a new measure of conditioning, denoted  $\mu_d$ , and we show implications of  $\mu_d$  for problem geometry and algorithm complexity, and demonstrate that the value of  $\mu = \mu_d$  is independent of the specific data representation of  $(\text{FP}_d)$ . We then prove certain relations among a variety of condition measures for  $(\text{FP}_d)$ , including  $\mu_d$ ,  $\sigma_d$ ,  $\bar{\chi}_d$ , and  $\mathcal{C}(d)$ . We discuss some drawbacks of using the condition number  $\mathcal{C}(d)$  as the sole measure of conditioning of a conic linear system, and we then introduce the notion of a “pre-conditioner” for  $(\text{FP}_d)$  which results in an equivalent formulation  $(\text{FP}_{\tilde{d}})$  of  $(\text{FP}_d)$  with a better condition number  $\mathcal{C}(\tilde{d})$ . We characterize the best such pre-conditioner and provide an algorithm for constructing an equivalent data instance  $\tilde{d}$  whose condition number  $\mathcal{C}(\tilde{d})$  is within a known factor of the best possible.

**AMS Subject Classification:** 90C, 90C05, 90C25, 90C60

**Keywords:** Complexity of Convex Programming, Conditioning, Pre-conditioners

---

<sup>\*</sup>University of Michigan, Industrial and Operations Engineering, 1205 Beal Avenue Ann Arbor, MI 48109-2117. e-mail: mepelman@umich.edu

<sup>†</sup>This research has been partially supported through NSF Graduate Research Fellowship.

<sup>‡</sup>M.I.T. Sloan School of Management, 50 Memorial Drive, Cambridge, MA 02142-1347. email: rfreund@mit.edu

## 1 Introduction

The subject of this paper is the further study and development of a new measure of conditioning for the convex feasibility problem in conic linear form:

$$(\text{FP}_d) : Ax = b, \quad x \in C_X, \quad (1)$$

where  $A \in \mathcal{L}(X, Y)$  is a linear operator between  $n$ - and  $m$ -dimensional spaces  $X$  and  $Y$ ,  $b \in Y$ , and  $C_X \subset X$  is a closed convex cone,  $C_X \neq X$ . We denote by  $d = (A, b)$  the data for the problem  $(\text{FP}_d)$  (the cone  $C_X$  is regarded as fixed and given), and the set of solutions of  $(\text{FP}_d)$  by

$$X_d \triangleq \{x \in X : Ax = b, \quad x \in C_X\}.$$

The problem  $(\text{FP}_d)$  is an important tool in mathematical programming. It provides a very general format for studying the feasible regions of convex optimization problems (in fact, any convex feasibility problem can be modeled as a conic linear system), and includes linear programming and semi-definite programming feasibility problems as special cases. Over the last decade many important developments in linear programming, most notably the theory of interior-point methods, have been extended to convex problems in this form. In recent years, largely prompted by these developments, researchers have developed new and powerful theories of condition numbers for convex optimization, aimed at capturing the intuitive notion of problem behavior; this body of research has been shown to be important in studying the efficiency of algorithms, including interior-point algorithms, for convex optimization as well as other behavioral characteristics of these problems such as problem geometry, deformation under data perturbation, etc.

In this paper, we (i) establish the connection of the condition number  $\mathcal{C}(d)$  to some of the measures of conditioning arising in recent linear programming literature, (ii) we develop a new measure of conditioning  $\mu_d$  that is invariant under equivalent data representations of the problem, and (iii) we develop a theory of “pre-conditioners” for improving the condition number of  $(\text{FP}_d)$ . We begin by briefly reviewing the developments in the theory of measures of conditioning in recent literature as well as provide an overview of the issues addressed in this paper.

The study of the computational complexity of linear programming originated with the analysis of the simplex algorithm, which, while extremely efficient in practice, was shown by Klee and Minty [17] to have worst-case complexity exponential in the number of variables. Khachiyan [16] demonstrated that linear programming problems were in fact polynomially solvable via the ellipsoid algorithm. Under the assumption that the problem data is rational, the ellipsoid algorithm requires at most  $O(n^2 L)$  iterations, where  $n$  is the number of variables, and  $L$  is the *problem size*, which is roughly equal to the number of bits required to represent the problem data. The development of interior-point methods gave rise to algorithms that are theoretically efficient as well as efficient in practice (unlike the ellipsoid algorithm). The first such algorithm, developed by Karmarkar [15], has a complexity bound of  $O(nL)$  iterations, and the algorithm introduced by Renegar [25] has a complexity bound

of  $O(\sqrt{n}L)$  iterations, which is the currently best known bound for linear programming. Many interior-point algorithms have also proven to be extremely efficient computationally, and are often superior to the simplex algorithm.

Despite the importance of the above results, there are several serious drawbacks in analyzing algorithm performance in the bit-complexity framework. One such drawback is the fact that computers use floating point arithmetic, rather than integer arithmetic, in performing computations. As a result, two problems can have data that are extremely close, but have drastically different values of  $L$ . The analysis of the performance of algorithms for solving these problems will yield different performance estimates, yet actual performance of the algorithms will likely be similar due to their similar numerical properties. See Wright [40] for a detailed discussion. Secondly, the complexity analysis of linear programming algorithms in terms of  $L$  largely relies on the combinatorial structure of the linear program, in particular, the fact that the set of feasible solutions is a polyhedron and the solution is attained at one of the extreme points of this polyhedron.

A relevant way to measure the intuitive notion of conditioning of a convex optimization (or feasibility) problem via the so-called *distance to ill-posedness* and the closely related *condition number* was developed by Renegar in [26] in a more specific setting, but then generalized more fully in [27] and in [28] to convex optimization and feasibility problems in conic linear form. Recall that  $d = (A, b)$  is the data for the problem  $(\text{FP}_d)$  of (1). The *condition number*  $\mathcal{C}(d)$  of  $(\text{FP}_d)$  is essentially a scale invariant reciprocal of the smallest data perturbation  $\Delta d = (\Delta A, \Delta b)$  for which the system  $(\text{FP}_{d+\Delta d})$  changes its feasibility status. The problem  $(\text{FP}_d)$  is well-conditioned to the extent that  $\mathcal{C}(d)$  is small; when the problem  $(\text{FP}_d)$  is “ill-posed” (i.e., arbitrarily small perturbations of the data can yield both feasible and infeasible problem instances), then  $\mathcal{C}(d) = +\infty$ .

One of the important issues addressed by researchers is the relationship between the condition number  $\mathcal{C}(d)$  and the geometry of the feasible region of  $(\text{FP}_d)$ . Renegar [26] demonstrated that when a feasible instance of  $(\text{FP}_d)$  is well-posed ( $\mathcal{C}(d) < \infty$ ), there exists a point  $x$  feasible for  $(\text{FP}_d)$  which satisfies  $\|x\| \leq \mathcal{C}(d)$ . Furthermore, it is shown in [9] that under the above assumption the set of feasible solutions contains a so-called “reliable” solution: a solution  $\hat{x}$  of  $(\text{FP}_d)$  is reliable if, roughly speaking, (i) the distance from  $\hat{x}$  to the boundary of the cone  $C_X$ ,  $\text{dist}(\hat{x}, \partial C_X)$ , is not excessively small, (ii) the norm of the solution  $\|\hat{x}\|$  is not excessively large, and (iii) the ratio  $\frac{\|\hat{x}\|}{\text{dist}(\hat{x}, \partial C_X)}$  is not excessively large. The importance of reliable solutions is motivated in part by considerations of finite-precision computations. The results in [9] also demonstrate that when the system  $(\text{FP}_d)$  is feasible, there exists a feasible point  $\hat{x}$  such that

$$\frac{\|\hat{x}\|}{\text{dist}(\hat{x}, \partial C_X)} = c_1 O(\mathcal{C}(d)), \quad \text{dist}(\hat{x}, \partial C_X) = c_2 \Omega\left(\frac{1}{\mathcal{C}(d)}\right), \quad \|\hat{x}\| = c_3 O(\mathcal{C}(d)), \quad (2)$$

where the constants  $c_1$ ,  $c_2$ , and  $c_3$  depend only on the “width” of the cone  $C_X$  (to be formally defined shortly), and are independent of the data  $d$  of the problem  $(\text{FP}_d)$  (but may depend on  $n$ ).

The condition number  $\mathcal{C}(d)$  was also shown to be crucial for analyzing the complexity

of algorithms for solving  $(\text{FP}_d)$ . Renegar [28] presented an interior-point algorithm for solving  $(\text{FP}_d)$  with the complexity bound of  $O(\sqrt{\vartheta} \ln(\vartheta \mathcal{C}(d)))$  iterations, where  $\vartheta$  is the complexity parameter of a self-concordant barrier for the cone  $C_X$ . In [10] it was shown that a suitably modified version of the ellipsoid algorithm will solve  $(\text{FP}_d)$  in  $O(n^2 \ln(\mathcal{C}(d)))$  iterations. (The constants in both complexity bounds depend on the width of  $C_X$ .) In [5], a generalization of a row-action algorithm is shown to compute a reliable solution of  $(\text{FP}_d)$  in the sense of (2). The complexity of this algorithm is also closely tied to  $\mathcal{C}(d)$ .

The recent literature has explored many other important properties of the problem  $(\text{FP}_d)$  tied to the distance to ill-posedness and the condition number  $\mathcal{C}(d)$ . Renegar [26] studied the relation of  $\mathcal{C}(d)$  to sensitivity of solutions of  $(\text{FP}_d)$  under perturbations in the problem data. Peña and Renegar [24] discussed the role of  $\mathcal{C}(d)$  in the complexity of computing approximate solutions of  $(\text{FP}_d)$ . [8] and Peña [23] addressed the theoretical complexity and practical aspects of computing the distance to ill-posedness. Vera [39] considered the numerical properties of an interior-point method for solving  $(\text{FP}_d)$  (and in fact, a more general problem of optimizing a linear function over the feasible region of  $(\text{FP}_d)$ ) in the case when  $(\text{FP}_d)$  is a linear programming problem. He considered the algorithm in the floating point arithmetic model, and demonstrated that the algorithm will approximately solve the optimization problem in polynomial time, while requiring roughly  $O(\ln(\mathcal{C}(d)))$  significant digits of precision for computation. For additional discussion of ill-posedness and the condition number, see Filipowski [6, 7], Nunez and Freund [21], Nunez [20], Peña [22, 23], and Vera [36, 37, 38].

As the above discussion hopefully conveys, the condition number  $\mathcal{C}(d)$  is a relevant and important measure of conditioning of the problem  $(\text{FP}_d)$ . Note that when  $(\text{FP}_d)$  is in fact a linear programming feasibility problem,  $\mathcal{C}(d)$  provides a measure of conditioning that, unlike  $L$ , does not rely on the assumption that the problem data is rational, and is relevant in the floating point model of computation.

Nevertheless, there are some potential drawbacks in using  $\mathcal{C}(d)$  as a sole measure of conditioning of the problem  $(\text{FP}_d)$ . To illustrate this point, note that the problem  $(\text{FP}_d)$  of (1) can be interpreted as the problem of finding a point  $x$  in the intersection of the cone  $C_X$  with an affine subspace  $\mathcal{A} \subset X$ , defined as

$$\mathcal{A} \triangleq \{x : Ax = b\} = \{x : x = x_0 + x_N, x_N \in \text{Null}(A)\},$$

where  $x_0 \in X$  is an arbitrary point satisfying  $Ax_0 = b$ , and  $\text{Null}(A)$  is the null space of  $A$ . Notice that the description of the affine subspace  $\mathcal{A}$  by the data instance  $d = (A, b)$  is not unique. It easy to find an equivalent data instance  $\tilde{d} = (\tilde{A}, \tilde{b})$  such that

$$\{x : \tilde{A}x = \tilde{b}\} = \{x : Ax = b\} = \mathcal{A}$$

(take, for example,  $\tilde{b} = Bb$  and  $\tilde{A} = BA$ , where  $B$  is any nonsingular linear operator  $B : Y \rightarrow Y$ ). Then the problem

$$(\text{FP}_{\tilde{d}}) : \tilde{A}x = \tilde{b}, x \in C_X$$

is equivalent to the problem  $(\text{FP}_d)$  in the sense that their feasible regions are identical; we can think of the systems  $(\text{FP}_d)$  and  $(\text{FP}_{\tilde{d}})$  as different but equivalent formulations of the same feasibility problem

$$(\text{FP}): \text{ find } x \in \mathcal{A} \cap C_X.$$

Since the condition number  $\mathcal{C}(d)$  is, in general, different from  $\mathcal{C}(\tilde{d})$ , analyzing many of the properties of the problem (FP) above in terms of the condition number will lead to different results, depending on which formulation,  $(\text{FP}_d)$  or  $(\text{FP}_{\tilde{d}})$ , is being used. This observation is somewhat disconcerting, since many of these properties are of purely geometric nature. For example, the existence of a solution of small norm and the existence of a reliable solution depend only on the geometry of the feasible region, i.e., of the set  $\mathcal{A} \cap C_X$ , and do not depend on a specific data instance  $d$  used to “represent” the affine space  $\mathcal{A}$ .

An important research direction, therefore, is the development of relevant measures of conditioning of the problem  $(\text{FP}_d)$  that depend on the affine space  $\mathcal{A}$  rather than on a particular data instance  $d$  used to represent it, and allow us to analyze some of the properties of the problem independently of the data used to represent the problem. The recent literature contains some results on developing such measures when  $(\text{FP}_d)$  is a linear programming feasibility problem. In particular, two condition measures,  $\tilde{\chi}_d$  and  $\sigma_d$ , were used in the analysis of interior-point algorithms for linear programming (Vavasis and Ye [33, 34, 35]). These measures, discussed in detail in Section 4, provide a new perspective on the analysis of linear programming problems; for example, like the condition number  $\mathcal{C}(d)$ , they do not require the data for the problem to be rational. Also, they have the desired property that they are independent of the specific data instance  $d$  used to describe the problem, and can be defined considering only the affine subspace  $\mathcal{A}$ . Further analysis of these measures in the setting of linear programming feasibility problems can be found in Ho [12], Todd, Tunçel and Ye [30], and Tunçel [32].

In this paper we define a new measure of conditioning,  $\mu_d$ , for feasible instances of the problem  $(\text{FP}_d)$  of (1), which is independent of the specific data representation of the problem. We explore the relationship between  $\mu_d$  and measures  $\tilde{\chi}_d$ ,  $\sigma_d$ , and  $\mathcal{C}(d)$  (in particular, we demonstrate that the measure  $\sigma_d$  is directly related to  $\mu_d$  in the special case of linear programming). We show that  $\mu_d \leq \mathcal{C}(d)$ , i.e.,  $\mu_d$  is less conservative, and that for any data instance  $\tilde{d}$  equivalent to  $d$ ,  $\mu_d \leq \mathcal{C}(\tilde{d})$ . We also demonstrate that many important properties of the system  $(\text{FP}_d)$  previously analyzed in terms of  $\mathcal{C}(d)$  can be analyzed through  $\mu_d$  (independently of the data representation).

On the other hand, some properties of  $(\text{FP}_d)$  are not purely geometric and depend on the data  $d$ . Therefore, it might be beneficial, given a data instance  $d$ , to construct a data instance  $\tilde{d}$  which is equivalent to  $d$ , but is better conditioned in the sense that  $\mathcal{C}(\tilde{d}) < \mathcal{C}(d)$ . We develop a characterization of all equivalent data instances  $\tilde{d}$  by introducing the concept of a *pre-conditioner* and provide an upper bound on the condition number  $\mathcal{C}(\tilde{d})$  of the “best” equivalent data instance  $\tilde{d}$ . We also analyze the complexity of computing an equivalent data instance whose resulting condition number is within a known factor of this bound. To this end, we construct an algorithm for computing such a data instance and analyze its complexity.

An outline of the paper is as follows. Section 2 contains notation, definitions, assumptions, and preliminary results. In Section 3 we introduce the new measure of conditioning  $\mu_d$  for  $(\text{FP}_d)$ , establish several results relating  $\mu_d$  to geometric properties of the feasible region of  $(\text{FP}_d)$  and analyze the performance of several algorithms for solving  $(\text{FP}_d)$  in terms of  $\mu_d$ . In Section 4 we study the relationship between  $\mu_d$  and other measures of conditioning, completely characterizing the relationship between  $\mathcal{C}(d)$  and  $\mu_d$ , as well as  $\sigma_d$  and  $\bar{\chi}_d$  in the linear programming setting. In Section 5, we develop the notion of a pre-conditioner for the problem  $(\text{FP}_d)$ , establish an upper bound on the condition number  $\mathcal{C}(\tilde{d})$  of the best equivalent data instance  $\tilde{d}$ , and construct and analyze an algorithm for computing an equivalent data instance whose condition number is within a known factor of this bound. Section 6 contains some final conclusions and indicates potential topics of future research.

## 2 Preliminaries

We work in the setup of finite dimensional normed linear vector spaces. Both  $X$  and  $Y$  are normed linear spaces of finite dimension  $n$  and  $m$ , respectively, endowed with norms  $\|x\|$  for  $x \in X$  and  $\|y\|$  for  $y \in Y$ . For  $\bar{x} \in X$ , let  $B(\bar{x}, r)$  denote the ball centered at  $\bar{x}$  with radius  $r$ , i.e.,  $B(\bar{x}, r) = \{x \in X : \|x - \bar{x}\| \leq r\}$ , and define  $B(\bar{y}, r)$  analogously for  $\bar{y} \in Y$ . We denote the set of real numbers by  $\mathbb{R}$  and the set of nonnegative real numbers by  $\mathbb{R}_+$ . The set of real  $k \times k$  symmetric matrices is denoted by  $S^{k \times k}$ . The set  $S^{k \times k}$  is a closed linear space of dimension  $n = \frac{k(k+1)}{2}$ . We denote the set of symmetric positive semi-definite  $k$ -by- $k$  matrices by  $S_+^{k \times k}$ .  $S_+^{k \times k}$  is a closed convex cone in  $S^{k \times k}$ . The interior of the cone  $S_+^{k \times k}$  is precisely the set of  $k$ -by- $k$  positive definite matrices, and is denoted by  $S_{++}^{k \times k}$ .

We associate with  $X$  and  $Y$  the dual spaces  $X^*$  and  $Y^*$  of linear functionals defined on  $X$  and  $Y$ , respectively. 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  $f \in Y^*$  we denote  $f(y)$  by  $f^t y$ . Let  $A : X \rightarrow Y$  be a linear operator. We denote  $A(x)$  by  $Ax$ , and we denote the adjoint of  $A$  by  $A^t : Y^* \rightarrow X^*$ .

The dual norm induced on  $c \in X^*$  is defined as

$$\|c\|_* \triangleq \max\{c^t x : x \in X, \|x\| \leq 1\}, \quad (3)$$

and the Hölder inequality  $c^t x \leq \|c\|_* \|x\|$  follows easily from this definition. The dual norm induced on  $f \in Y^*$  is defined similarly.

We now present the development of the concepts of condition numbers and data perturbation for  $(\text{FP}_d)$  in detail. Recall that  $d = (A, b)$  is the data for the problem  $(\text{FP}_d)$ . Let

$$\mathcal{D} = \{d = (A, b) : A \in L(X, Y), b \in Y\}$$

denote the space of all data  $d = (A, b)$  for  $(\text{FP}_d)$ . For  $d = (A, b) \in \mathcal{D}$  we define the product



norm on the Cartesian product  $L(X, Y) \times Y$  to be

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

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

$$\|A\| = \max\{\|Ax\| : \|x\| \leq 1\}.$$

We define

$$\mathcal{F} = \{(A, b) \in \mathcal{D} : \text{there exists } x \text{ satisfying } Ax = b, x \in C_X\}$$

to be the set of data instances  $d$  for which  $(\text{FP}_d)$  is feasible. Its complement is denoted by  $\mathcal{F}^c$ , the set of data instances for which  $(\text{FP}_d)$  is infeasible. The boundary of  $\mathcal{F}$  and of  $\mathcal{F}^c$  is precisely the set  $\mathcal{B} = \partial\mathcal{F} = \partial\mathcal{F}^c = \text{cl}(\mathcal{F}) \cap \text{cl}(\mathcal{F}^c)$ , where  $\partial S$  denotes the boundary and  $\text{cl}(S)$  denotes the closure of a set  $S$ . Note that if  $d = (A, b) \in \mathcal{B}$ , then  $(\text{FP}_d)$  is ill-posed in the sense that arbitrarily small changes in the data  $d = (A, b)$  can yield instances of  $(\text{FP}_d)$  that are feasible, as well as instances of  $(\text{FP}_d)$  that are infeasible. Also, note that  $\mathcal{B} \neq \emptyset$ , since  $d = 0 \in \mathcal{B}$ .

For a data instance  $d = (A, b) \in \mathcal{D}$ , the *distance to ill-posedness* is defined to be:

$$\rho(d) \triangleq \inf\{\|\Delta d\| : d + \Delta d \in \mathcal{B}\} = \begin{cases} \inf\{\|d - \bar{d}\| : \bar{d} \in \mathcal{F}^c\} & \text{if } d \in \mathcal{F}, \\ \inf\{\|d - \bar{d}\| : \bar{d} \in \mathcal{F}\} & \text{if } d \in \mathcal{F}^c, \end{cases} \quad (4)$$

see Renegar [26, 27, 28]. The *condition number*  $\mathcal{C}(d)$  of the data instance  $d$  is defined to be:

$$\mathcal{C}(d) = \frac{\|d\|}{\rho(d)} \quad (5)$$

when  $\rho(d) > 0$ , and  $\mathcal{C}(d) = \infty$  when  $\rho(d) = 0$ . The condition number  $\mathcal{C}(d)$  is a measure of the relative conditioning of the data instance  $d$ , and can be viewed as a scale-invariant reciprocal of  $\rho(d)$ , as it is elementary to demonstrate that  $\mathcal{C}(d) = \mathcal{C}(\alpha d)$  for any positive scalar  $\alpha$ . It is easy to show that  $\rho(0) = 0$ , and hence  $\mathcal{C}(d) \geq 1$ .

If  $C$  is a convex cone in  $X$ , the dual cone of  $C$ , denoted by  $C^*$ , is defined by

$$C^* = \{z \in X^* : z^t x \geq 0 \text{ for any } x \in C\}. \quad (6)$$

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). If  $C$  is a closed convex cone, then  $C$  is regular if and only if  $C^*$  is regular.

We will use the following definition of the *width* of a regular cone  $C$ :

**Definition 1** *If  $C$  is a regular cone in the normed linear vector space  $X$ , the width of  $C$  is given by:*

$$\tau_C = \max_{x,r} \left\{ \frac{r}{\|x\|} : B(x, r) \subset C \right\}.$$

$\tau_C$  measures the maximum ratio of the radius to the norm of the center of an inscribed ball in  $C$ , and so larger values of  $\tau_C$  correspond to an intuitive notion of greater width of  $C$ . Note that  $\tau_C \in (0, 1]$ , since  $C$  is pointed and has a nonempty interior, and  $\tau_C$  is attained for some  $(\bar{x}, \bar{r})$  as well as along the ray  $(\alpha\bar{x}, \alpha\bar{r})$  for all  $\alpha > 0$ . By choosing the value of  $\alpha$  appropriately, we can find  $u \in C$  such that

$$\|u\| = 1 \text{ and } \tau_C \text{ is attained for } (x, r) = (u, \tau_C). \quad (7)$$

We now pause to illustrate the above notions on two relevant instances of the cone  $C$ , namely the nonnegative orthant  $\mathfrak{R}_+^n$  and the positive semi-definite cone  $S_+^{k \times k}$ . We first consider the nonnegative orthant. Let  $X = \mathfrak{R}^n$  and  $C = \mathfrak{R}_+^n$ . Then we can identify  $X^*$  with  $X$  and in so doing,  $C^* = \mathfrak{R}_+^n$  as well. If  $\|x\|$  is given by the  $L_\infty$  norm  $\|x\| = \max_{j=1, \dots, n} |x_j|$ , then the width of  $C$  is  $\tau_C = 1$  and  $u = e$ . If instead of the  $L_\infty$  norm, the norm  $\|x\|$  is the  $L_p$  norm defined by

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

for  $p \geq 1$ , then for  $x \in C$  it is straightforward to show that  $u = \left(n^{-\frac{1}{p}}\right) e$  and  $\tau_C = n^{-\frac{1}{p}}$ .

Now consider the positive semi-definite cone, which has been shown to be of enormous importance in mathematical programming (see Alizadeh [1] and Nesterov and Nemirovskii [19]). Let  $X = S^{k \times k}$ , and so  $n = \frac{k(k+1)}{2}$ , and let  $C = S_+^{k \times k}$ . We can identify  $X^*$  with  $X$ , and in so doing it is elementary to derive that  $C^* = S_+^{k \times k}$ , i.e.,  $C$  is self-dual. For  $x \in X$ , let  $\lambda(x)$  denote the  $k$ -vector of eigenvalues of  $x$ . For any  $p \in [1, \infty)$ , let the norm of  $x$  be defined by

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

(see [13], for example, for a proof that  $\|x\|_p$  is a norm). It is easy to show that  $u = \left(k^{-\frac{1}{p}}\right) I$  and the width of  $C$  is  $\tau_C = k^{-\frac{1}{p}}$ . If  $\|x\|$  is given by  $\|x\| = \max_{i=1, \dots, k} |\lambda_i|$ , then  $u = I$  and  $\tau_C = 1$ .

Another important parameter of a regular cone  $C$  is the *norm approximation coefficient*:

**Definition 2** *If  $C$  is a regular cone in the normed linear space  $X$ , define the norm approximation coefficient by*

$$\delta_C \triangleq \text{dist}(0, \partial \text{conv}(C(1), -C(1))), \quad (8)$$

where  $C(1) \triangleq \{x \in C : \|x\| \leq 1\}$ .

The norm approximation coefficient  $\delta_C$  measures the extent to which the unit ball  $B(0, 1) \subset X$  can be approximated by the set  $\text{conv}(C(1), -C(1))$ . As a consequence, it measures the extent to which the norm of a linear operator can be approximated over the set  $C(1)$ :

**Proposition 3** Suppose  $A \in L(X, Y)$ . Then  $\|A\| \leq \frac{1}{\delta_C} \max\{\|Ax\| : x \in C(1)\}$ .

**Proof:** As a direct consequence of Definition 2,  $B(0, \delta_C) \subseteq \text{conv}(C(1), -C(1))$ . Therefore,

$$\|A\| = \max\{\|Ax\| : x \in B(0, 1)\} \leq \frac{1}{\delta_C} \max\{\|Ax\| : x \in \text{conv}(C(1), -C(1))\}.$$

Since  $\|Ax\|$  is a convex function of  $x$ , the maximum above is attained at an extreme point of the set  $\text{conv}(C(1), -C(1))$  (see Theorem 3.4.7 of [2]), and we can assume without loss of generality that it is attained at some point  $x \in C(1)$ , establishing the proposition. ■

**Lemma 4** Suppose  $C$  is a regular cone with width  $\tau_C$ . Then

$$\delta_C \geq \frac{\tau_C}{1 + \tau_C} \geq \frac{\tau_C}{2}. \quad (9)$$

**Proof:** Let  $\bar{x} \in X$  be an arbitrary vector satisfying  $\|\bar{x}\| \leq \frac{\tau_C}{1 + \tau_C}$ . To establish the lemma we need to show that  $\bar{x} \in \text{conv}(C(1), -C(1))$ .

Let  $x = \frac{\bar{x}(1 + \tau_C)}{\tau_C}$ . If  $u$  is as in (7),  $u + \tau_C x \in C$  and  $u - \tau_C x \in C$ . Furthermore,

$$\frac{u + \tau_C x}{1 + \tau_C} \in C(1) \quad \text{and} \quad \frac{-u + \tau_C x}{1 + \tau_C} \in -C(1),$$

and so

$$\bar{x} = \frac{\tau_C}{1 + \tau_C} x = \frac{1}{2} \left( \frac{u + \tau_C x}{1 + \tau_C} \right) + \frac{1}{2} \left( \frac{-u + \tau_C x}{1 + \tau_C} \right) \in \text{conv}(C(1), -C(1)). \quad \blacksquare$$

We will assume throughout this paper that the system  $(\text{FP}_d)$  of (1) is feasible. At this point, we make no further assumptions on the cone  $C_X$  and the norms on the spaces  $X$  and  $Y$ , unless stated otherwise (we will make some additional assumptions in Sections 4 and 5).

When  $(\text{FP}_d)$  is feasible,  $\rho(d)$  can be expressed via the following characterization:

$$\rho(d) = \max\{r : B(0, r) \in \mathcal{H}_d\}, \quad (10)$$

where

$$\mathcal{H}_d \triangleq \{b\theta - Ax : \theta \geq 0, x \in C_X, |\theta| + \|x\| \leq 1\} \subset Y. \quad (11)$$

Note that  $0 \in \mathcal{H}_d$  whenever  $(\text{FP}_d)$  is feasible, and  $\rho(d) > 0$  precisely when  $0 \in \text{int } \mathcal{H}_d$ . This interpretation, presented by Renegar in [28], will serve as an important tool in developing further understanding of the properties of the system  $(\text{FP}_d)$ .

The next result follows from the definition of  $\mathcal{H}_d$  and Proposition 3.

**Corollary 5** Suppose  $d = (A, b) \in \mathcal{D}$  and  $C_X$  is regular. Then  $\|d\| \leq \frac{1}{\delta_{C_X}} \max\{\|h\| : h \in \mathcal{H}_d\}$ .

### 3 The symmetry measure $\mu_d$

In this section we define a new measure of conditioning of  $(\text{FP}_d)$ ,  $\mu_d$ , which we refer to as the “symmetry measure”, and we establish some of its properties relevant in the analysis of  $(\text{FP}_d)$ . We begin by recalling the *symmetry* of a set with respect to a point:

**Definition 6** *Let  $D \subset Y$  be a bounded convex set. For  $y \in \text{int } D$  we define  $\text{sym}(D, y)$  to be the symmetry of  $D$  about  $y$ , i.e.,*

$$\text{sym}(D, y) \triangleq \sup\{t \mid y + v \in D \Rightarrow y - tv \in D\}.$$

*If  $y \in \partial D$ , we define  $\text{sym}(D, y) = 0$ .*

This definition of symmetry is similar to that given in [28]. Observe that  $\text{sym}(D, y) \in [0, 1]$ , with  $\text{sym}(D, y) = 1$  if  $D$  is perfectly symmetric about  $y$ , and  $\text{sym}(D, y) = 0$  precisely when  $y \in \partial D$ . Moreover, the definition of  $\text{sym}(D, y)$  is independent of the norm on the space  $Y$ . The following lemma will be used later on:

**Lemma 7** *Suppose  $D$  is a compact convex set with a non-empty interior, and let  $y \in \text{int } D$ . Then there exists an extreme point  $w$  of  $D$  such that  $\text{sym}(D, y) = \text{sym}_w(D, y) \triangleq \sup\{t \mid y - t(w - y) \in D\}$ .*

**Proof:** Define  $f(w) = \text{sym}_w(D, y) = \sup\{t \mid y - t(w - y) \in D\}$ . It follows that  $f(w)$  is a quasi-concave function on  $D$ . This implies that the minimum of  $f(w)$  is attained at an extreme point of  $D$ , see for example [2], Section 3.5.3. ■

To define the *symmetry measure* of the problem  $(\text{FP}_d)$  recall that if  $(\text{FP}_d)$  is feasible,  $0 \in \mathcal{H}_d$ , where  $\mathcal{H}_d$  is defined in (11). Hence, the following quantity is well-defined:

**Definition 8** *Suppose the system  $(\text{FF}_d)$  is feasible. We define*

$$\mu_d \triangleq \frac{1}{\text{sym}(\mathcal{H}_d, 0)}. \tag{12}$$

*when  $\text{sym}(\mathcal{H}_d, 0) > 0$  and  $\mu_d = +\infty$  when  $\text{sym}(\mathcal{H}_d, 0) = 0$ .*

From the above definition,  $\mu_d \geq 1$  and  $\mu_d = +\infty$  precisely when  $0 \in \partial \mathcal{H}_d$ , i.e., precisely when  $(\text{FP}_d)$  is ill-posed.

#### 3.1 The symmetry measure and geometric properties of solutions of $(\text{FP}_d)$

We now establish two results that characterize geometric properties of the feasible region  $X_d$  of the system  $(\text{FP}_d)$  in terms of  $\mu_d$ . Theorem 9 establishes a bound on the size of a

solution of  $(\text{FP}_d)$  in terms of  $\mu_d$ ; this result is similar to the bound in terms of the condition number  $\mathcal{C}(d)$  in [26]. Theorem 10 demonstrates existence of a reliable solution of  $(\text{FP}_d)$ . This is similar to the result (2) presented in [9], however, here the bounds on the size of the solution, its distance to the boundary of the cone  $C_X$  and the ratio of the above quantities are established in terms of  $\mu_d$  rather than  $\mathcal{C}(d)$ . Also, unlike for the condition number  $\mathcal{C}(d)$ , we can establish a converse result for  $\mu_d$ , namely, if the feasible region possesses nice geometry, i.e., contains a reliable solution, then  $\mu_d$  can be nicely bounded by a function of the parameters associated with the reliable solution. This result is proven in Theorem 11.

**Theorem 9** *Suppose  $\mu_d < \infty$ . Then there exists  $x \in X_d$  such that  $\|x\| \leq \mu_d$ .*

**Proof:** By the definition of  $\mu_d$ ,  $-\frac{1}{\mu_d}b = -\text{sym}(\mathcal{H}_d, 0)b \in \mathcal{H}_d$ , since  $b \in \mathcal{H}_d$ . Therefore there exists  $(\theta, x)$  satisfying  $\theta \geq 0$ ,  $x \in C_X$ ,  $|\theta| + \|x\| \leq 1$ , and  $b\theta - Ax = -\frac{1}{\mu_d}b$ . Let  $\hat{x} = \frac{x}{\theta + \frac{1}{\mu_d}}$ .

Then  $\hat{x} \in X_d$  and  $\|\hat{x}\| = \frac{\|x\|}{\theta + \frac{1}{\mu_d}} \leq \mu_d$ . ■

**Theorem 10** *Suppose  $C_X$  is a regular cone with width  $\tau$ , and that  $\mu_d < \infty$ . Then there exist  $\hat{x}$  and  $r > 0$  such that*

1.  $\hat{x} \in X_d$ ,
2.  $\|\hat{x}\| \leq 2\mu_d + 1$ ,
3.  $\text{dist}(\hat{x}, \partial C_X) \geq r \geq \frac{\tau}{2\mu_d + 1}$ ,
4.  $\frac{\|\hat{x}\|}{r} \leq \frac{2\mu_d + 1}{\tau}$ ,

where  $\tau$  is the width of the cone  $C_X$ .

**Proof:** Let  $u$  be as in (7). Then  $\frac{1}{2}b - \frac{1}{2}Au \in \mathcal{H}_d$ . From the definition of  $\mu_d$  we conclude that  $-\frac{1}{\mu_d}(\frac{1}{2}b - \frac{1}{2}Au) \in \mathcal{H}_d$ , whereby there exists  $(\bar{\theta}, \bar{x}) \in \mathbb{R}_+ \times C_X$ ,  $|\bar{\theta}| + \|\bar{x}\| \leq 1$ , satisfying  $b\bar{\theta} - A\bar{x} = -\frac{1}{\mu_d}(\frac{1}{2}b - \frac{1}{2}Au)$ .

Let  $\hat{x} = \frac{2\mu_d\bar{x} + u}{2\mu_d\bar{\theta} + 1}$ . It is easy to verify that  $\hat{x} \in X_d$ , so that condition 1 of the theorem is satisfied. Moreover,  $\|\hat{x}\| = \frac{\|2\mu_d\bar{x} + u\|}{2\mu_d\bar{\theta} + 1} \leq 2\mu_d + 1$ , establishing condition 2.

Next, let  $r = \frac{\tau}{2\mu_d\bar{\theta} + 1}$ . Since  $B(u, \tau) \subset C_X$  and  $\bar{x} \in C_X$ , we conclude that  $B(u + 2\mu_d\bar{x}, \tau) \subset C_X$ , and therefore  $B\left(\frac{u + 2\mu_d\bar{x}}{2\mu_d\bar{\theta} + 1}, \frac{\tau}{2\mu_d\bar{\theta} + 1}\right) = B(\hat{x}, r) \subset C_X$ . Also, since  $\bar{\theta} \leq 1$ ,  $r \geq \frac{\tau}{2\mu_d + 1}$ , establishing condition 3. Finally,

$$\frac{\|\hat{x}\|}{r} = \frac{\|2\mu_d\bar{x} + u\|}{2\mu_d\bar{\theta} + 1} \cdot \frac{2\mu_d\bar{\theta} + 1}{\tau} \leq \frac{2\mu_d + 1}{\tau},$$

implying 4 and concluding the proof of the theorem. ■

We conclude from Theorems 9 and 10 that, much like for the condition number  $\mathcal{C}(d)$ , if the symmetry measure  $\mu_d$  is small, the feasible region  $X_d$  possesses nice geometry, namely, it contains a point of small norm, and also contains a reliable solution. We now establish a converse result: if the feasible region  $X_d$  possesses nice geometry, namely it contains a reliable solution, then  $\mu_d$  is small.

**Theorem 11** *Suppose  $C_X$  is a regular cone and there exists  $\hat{x} \in X_d$  and  $r > 0$  such that  $\text{dist}(\hat{x}, \partial C_X) \geq r$ . Let  $\gamma = \max \left\{ \|\hat{x}\|, \frac{1}{r}, \frac{\|\hat{x}\|}{r} \right\}$ . Then  $\mu_d \leq 1 + 2\gamma$ .*

**Proof:** Let  $\delta = \|\hat{x}\| + 1$  and  $\pi = \min\{r, 1\}$ . We first show that  $\text{sym}(\mathcal{H}_d, 0) \geq \frac{\pi}{\delta + \pi}$ .

Let  $y \in \mathcal{H}_d$ . From the definition of  $\mathcal{H}_d$ ,  $y = b\bar{\theta} - A\bar{x}$  for some  $(\bar{\theta}, \bar{x}) \in \mathfrak{R}_+ \times C_X$ ,  $|\bar{\theta}| + \|\bar{x}\| \leq 1$ . Therefore

$$\frac{\pi}{\delta + \pi}(-y) = \frac{\pi}{\delta + \pi}(-b\bar{\theta} + A\bar{x}) + \frac{1}{\delta + \pi}(b - A\hat{x}) = b \left( \frac{-\pi\bar{\theta} + 1}{\delta + \pi} \right) - A \left( \frac{-\pi\bar{x} + \hat{x}}{\delta + \pi} \right).$$

Let  $\check{\theta} = \frac{-\pi\bar{\theta} + 1}{\delta + \pi}$  and  $\check{x} = \frac{-\pi\bar{x} + \hat{x}}{\delta + \pi}$ . Since  $\pi \leq 1$  and  $\bar{\theta} \leq 1$  we have  $\check{\theta} \geq 0$ . Moreover, since  $\pi \leq r$  and  $\|\bar{x}\| \leq 1$  we have  $\check{x} \in C_X$ . Finally,  $|\check{\theta}| + \|\check{x}\| \leq \frac{1}{\delta + \pi} (1 + \pi\|\bar{x}\| + \|\hat{x}\|) \leq 1$ , and therefore  $-\frac{\pi}{\delta + \pi}y \in \mathcal{H}_d$  for an arbitrary  $y \in \mathcal{H}_d$ , establishing that  $\text{sym}(\mathcal{H}_d, 0) \geq \frac{\pi}{\delta + \pi}$ . Hence,

$$\mu_d = \frac{1}{\text{sym}(\mathcal{H}_d, 0)} \leq \frac{\delta + \pi}{\pi} = 1 + \frac{1}{\min\{r, 1\}} + \frac{\|\hat{x}\|}{\min\{r, 1\}} \leq 1 + \max\{\gamma, 1\} + \gamma \leq 1 + 2\gamma.$$

The last inequality follows from the observation that  $r \leq \|\hat{x}\|$  (since  $C_X$  is pointed and so  $\|\hat{x}\| \geq \text{dist}(\hat{x}, \partial C_X) \geq r$ ) and so  $\gamma \geq \frac{\|\hat{x}\|}{r} \geq 1$ .  $\blacksquare$

The result in Theorem 11 is quite specific to  $\mu_d$ ; no such result is possible for the condition number  $\mathcal{C}(d)$ . In fact, the example following Remark 19 in Section 4 shows that  $\mathcal{C}(d)$  can be arbitrarily large even when  $\gamma$  is fixed.

### 3.2 The symmetry measure and the complexity of computing a solution of (FP<sub>d</sub>)

In this subsection we present complexity bounds for solving (FP<sub>d</sub>) via an interior-point algorithm and via the ellipsoid algorithm, and we show that the complexity of solving (FP<sub>d</sub>) depends on  $\ln(\mu_d)$  as well as on other naturally-appearing quantities. For this subsection, we assume that the space  $X$  is an  $n$ -dimensional Euclidean space with Euclidean norm  $\|x\| = \|x\|_2 = \sqrt{x^t x}$  for  $x \in X$ . All of the developments of this section can in fact be adapted for the case when the norm in the space  $X$  is an arbitrary inner product norm. We have chosen the Euclidean norm to simplify the exposition. We also assume that  $C_X$  is a regular cone with width  $\tau$ , and the vector  $u$  of (7) is known.

When the cone  $C_X$  is represented as the closure of the domain of a self-concordant barrier function, a solution of (FP<sub>d</sub>) can be found using the barrier method developed by

Renegar, based on the theory of self-concordant functions of Nesterov and Nemirovskii [19]. Below we briefly review the barrier method as articulated in [29], and then state the main complexity result.

The version of the barrier method that we will use is designed to approximately solve a problem of the form

$$z^* = \inf\{c^t \omega : \omega \in S \cap L\}, \quad (13)$$

where  $S$  is a bounded set whose interior is convex and is the domain of a self-concordant barrier function  $f(\omega)$  with complexity parameter  $\vartheta_f$  (see [19] and [29] for details), and  $L$  is a closed subspace (or a translate of a closed subspace). The barrier method takes as input a point  $\omega' \in \text{int } S \cap L$ , and proceeds by approximately following the *central path*, i.e., the sequence of solutions of the problems

$$z(\eta) = \inf_{\omega \in L} \eta \cdot c^t \omega + f(\omega),$$

where  $\eta > 0$  is the *barrier parameter*. In particular, after the initialization stage, the method generates an increasing sequence of barrier parameters  $\eta_k > 0$  and iterates  $\omega_k \in \text{int } S \cap L$  that satisfy

$$c^t \omega_k - \frac{6\vartheta_f}{5\eta_k} \leq z^* \leq c^t \omega_k, k = 0, 1, 2, \dots \quad (14)$$

It follows from the analysis in [29] that if the barrier method is initialized at the point  $\omega' \in \text{int } S \cap L$ , then it will take at most

$$O\left(\sqrt{\vartheta_f} \ln\left(\frac{\vartheta_f(z_* - z^*)}{\text{sym}(S \cap L, \omega')} \cdot \bar{\eta}\right)\right) \quad (15)$$

iterations to bring the value of the barrier parameter  $\eta$  above the threshold of  $\bar{\eta} \geq \eta_0$  while maintaining (14) (here,  $z_* = \sup\{c^t \omega : \omega \in S \cap L\}$ ). This implies the main convergence result for the barrier method:

**Theorem 12 ([29], Theorem 2.4.10)** *Assume  $S$  is a bounded set whose interior is convex and is the domain of a self-concordant barrier function  $f(\omega)$  with complexity parameter  $\vartheta_f$ , and  $L$  is a closed subspace (or a translate of a closed subspace). Assume the barrier method is initialized at a point  $\omega' \in \text{int } S \cap L$ . If  $0 < \epsilon < 1$ , then within*

$$O\left(\sqrt{\vartheta_f} \ln\left(\frac{\vartheta_f}{\epsilon \text{sym}(S \cap L, \omega')}\right)\right)$$

*iterations of the method, all points  $\omega$  computed thereafter satisfy  $\omega \in \text{int } S \cap L$  and*

$$\frac{c^t \omega - z^*}{z_* - z^*} \leq \epsilon.$$

In order to find a solution of  $(\text{FP}_d)$  we will construct a closely related problem of the form (13) and apply the barrier method to this problem. This construction was carried

out in [28], where the complexity of solving  $(\text{FP}_d)$  was analyzed in terms of  $\mathcal{C}(d)$ . The optimization problem we consider is

$$\begin{aligned} z^* = \inf_{\theta, x, t} \quad & t \\ \text{s.t.} \quad & b\theta - Ax = t(\tfrac{1}{2}b - \tfrac{1}{2}Au) \\ & x \in \text{int } C_X \\ & \|x\| < 1 \\ & 0 < \theta < 1 \\ & -1 < t < 2. \end{aligned} \tag{16}$$

We will use the barrier method to find a feasible solution  $(\hat{\theta}, \hat{x}, \hat{t})$  of (16) such that  $\hat{t} \leq 0$ , and use the transformation  $x = \frac{\hat{x} - \frac{1}{2}\hat{t}u}{\hat{\theta} - \frac{1}{2}\hat{t}}$  to obtain a solution of  $(\text{FP}_d)$ .

Let  $z_*$  be the optimal value of the problem obtained from (16) by replacing “inf” with “sup”. Let  $\tilde{f}(x)$  be the self-concordant barrier function defined on  $\text{int } C_X$  and let  $\vartheta_{\tilde{f}}$  be the complexity parameter of  $\tilde{f}(x)$ . Then the set  $S \triangleq \{(\theta, x, t) : x \in \text{int } C_X, \|x\| < 1, 0 < \theta < 1, -1 < t < 2\}$  is convex and bounded, and is the domain of the self-concordant barrier function

$$f(\omega) = f(\theta, x, t) = \tilde{f}(x) - \ln(1 - \|x\|^2) - \ln \theta - \ln(1 - \theta) - \ln(t + 1) - \ln(2 - t)$$

with complexity parameter  $\vartheta_f \leq \vartheta_{\tilde{f}} + 5$  (see, for example, [28] or [29] for details). If we define  $L \triangleq \{(\theta, x, t) : b\theta - Ax = t(\tfrac{1}{2}b - \tfrac{1}{2}Au)\}$ , then the problem (16) is of the form (13), and we can apply the barrier method initialized at the point  $\omega' = (\theta', x', t') = (\tfrac{1}{2}, \tfrac{1}{2}u, 1)$ . The following proposition provides bounds on all of the parameters necessary in the analysis of the complexity of the barrier method via Theorem 12:

**Proposition 13**  $z_* \leq 2$ ,  $-1 \leq z^* \leq -\frac{1}{\mu_d}$ ,  $\text{sym}(S \cap L, \omega') \geq \frac{1}{12\tau}$ .

**Proof:** The upper bound on  $z_*$  and the lower bound on  $z^*$  follow from the last constraint of (16).

Let  $y = \tfrac{1}{2}b - \tfrac{1}{2}Au \in \mathcal{H}_d$ . From the definition of  $\mu_d$  we conclude that  $-\frac{y}{\mu_d} \in \mathcal{H}_d$ , so there exists  $(\theta, x)$  such that  $\theta \geq 0$ ,  $x \in C_X$ ,  $|\theta| + \|x\| \leq 1$ ,  $b\theta - Ax = -\frac{1}{\mu_d}(\tfrac{1}{2}b - \tfrac{1}{2}Au)$ . Therefore  $(\theta, x, -1/\mu_d)$  is in the closure of the feasible set of (16), and so  $z^* \leq -\frac{1}{\mu_d}$ .

To establish the last statement of the proposition, we appeal to Proposition 3.3 of Renegar [28], where it is shown that  $\omega'$  defined above satisfies

$$\text{sym}(S \cap L, \omega') \geq \frac{1}{4} \text{sym}\left(C_X(1), \tfrac{1}{2}u\right), \text{ where } C_X(1) = \{x : x \in C_X, \|x\| \leq 1\}.$$

Since  $B(\tfrac{1}{2}u, \tfrac{1}{2}\tau) \subset C_X(1)$ , it is easy to verify that  $\text{sym}(C_X(1), \tfrac{1}{2}u) \geq \frac{\tau}{3}$ , establishing the proposition. ■



**Theorem 14** *Suppose the barrier method for problem (16) is initialized at the point  $(\frac{1}{2}, \frac{1}{2}u, 1)$ . Then within*

$$O\left(\sqrt{\vartheta_{\hat{f}}}\ln\left(\frac{\vartheta_{\hat{f}}\mu_d}{\tau}\right)\right)$$

*iterations any iterate  $(\hat{\theta}, \hat{x}, \hat{t})$  of the algorithm will satisfy  $\hat{t} \leq 0$ , and therefore  $x = \frac{\hat{x} - \frac{1}{2}\hat{t}u}{\hat{\theta} - \frac{1}{2}\hat{t}}$  is a solution of  $(FP_d)$ .*

**Proof:** First note that for any iterate  $(\hat{\theta}, \hat{x}, \hat{t})$  of the algorithm,  $\hat{\theta} > 0$  and  $\hat{x} \in \text{int } C_X$ . Therefore, it is easy to check that when  $\hat{t} \leq 0$ ,  $x$  is well-defined and is a solution of  $(FP_d)$ .

It remains to verify the number of iterations needed to generate an iterate such that  $\hat{t} \leq 0$ . Let  $\epsilon = \frac{1}{3\mu_d}$ . Applying Theorem 12 and substituting the bounds of Proposition 13 into the complexity bound, we conclude that after at most

$$O\left(\sqrt{\vartheta_f}\ln\left(\frac{\vartheta_f}{\epsilon_{\text{sym}(S \cap L, \omega')}}\right)\right) = O\left(\sqrt{\vartheta_{\hat{f}}}\ln\left(\frac{\vartheta_{\hat{f}}\mu_d}{\tau}\right)\right)$$

iterations of the barrier method, any iterate  $(\hat{\theta}, \hat{x}, \hat{t})$  will satisfy

$$\hat{t} \leq \epsilon(z_* - z^*) + z^* \leq \frac{1}{3\mu_d}(2 - (-1)) - \frac{1}{\mu_d} = 0,$$

from which the theorem follows.  $\blacksquare$

When the cone  $C_X$  is represented via a separation oracle, a solution of  $(FP_d)$  can be found using a version of the ellipsoid algorithm (see, for example, [3] and [11]). Below is a generic theorem for analyzing the ellipsoid algorithm for finding a point  $\omega$  in a convex set  $S \subset \mathbb{R}^k$  given by a separation oracle.

**Theorem 15** *Suppose that a convex set  $S \subset \mathbb{R}^k$  given by a separation oracle contains a Euclidean ball of radius  $r$  centered at some point  $\hat{\omega}$ , and that an upper bound  $R$  on the quantity  $(\|\hat{\omega}\|_2 + r)$  is known. Then if the ellipsoid algorithm is initiated with a Euclidean ball of radius  $R$  centered at  $\omega^0 = 0$ , the algorithm will compute a point in  $S$  in at most*

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

*iterations, where each iteration must perform a feasibility cut on  $S$ .*

The main problem with trying to apply Theorem 15 directly to  $(FP_d)$  is that one needs to know the upper bound  $R$  in advance. Because such an upper bound is generically unknown in advance for  $(FP_d)$ , we approach solving  $(FP_d)$  by considering finding a point in the following set:

$$S \triangleq \{(\theta, x) : \theta > 0, x \in C_X, b\theta - Ax = 0\}, \quad (17)$$

which is a convex set in the linear subspace  $T \triangleq \{(\theta, x) : b\theta - Ax = 0\}$  of dimension  $k = n + 1 - m$ . Observe that it is easy to construct a separation oracle for  $S$  in the linear subspace  $T$  provided that one has a separation oracle for  $C_X$ . We will use the ellipsoid algorithm to find a point  $(\hat{\theta}, \hat{x}) \in S$  (working in the linear subspace  $T$ ), and we use the obvious transformation  $x = \frac{\hat{x}}{\hat{\theta}}$  to transform the output of the algorithm into a solution of  $(FP_d)$ .

**Proposition 16** *Let  $S$  be as in (17). Then there exists a point  $(\hat{\theta}, \hat{x}) \in S$  and  $\hat{r} > 0$  such that*

$$B((\hat{\theta}, \hat{x}), \hat{r}) \cap \{(\theta, x) : b\theta - Ax = 0\} \subset S, \quad \|(\hat{\theta}, \hat{x})\| + \hat{r} \leq 3, \quad \text{and} \quad \hat{r} \geq \frac{\tau}{2\mu_d}.$$

**Proof:** Let  $y = \frac{1}{2}b - \frac{1}{2}Au \in \mathcal{H}_d$ . From the definition of  $\mu_d$  we conclude that  $-\frac{y}{\mu_d} \in \mathcal{H}_d$ , whereby there exists  $(\bar{\theta}, \bar{x})$  such that

$$|\bar{\theta}| + \|\bar{x}\| \leq 1, \quad \bar{\theta} \geq 0, \quad \bar{x} \in C_X, \quad b\bar{\theta} - A\bar{x} = -\frac{1}{\mu_d} \left( \frac{1}{2}b - \frac{1}{2}Au \right).$$

Let  $\hat{\omega} = (\hat{\theta}, \hat{x}) \triangleq \left( \bar{\theta} + \frac{1}{2\mu_d}, \bar{x} + \frac{1}{2\mu_d}u \right)$  and  $\hat{r} = \frac{\tau}{2\mu_d}$ . Then  $\hat{\omega} \in S$ ,  $B(\hat{\omega}, \hat{r}) \cap \{(\theta, x) : b\theta - Ax = 0\} \subset S$ , and  $\|\hat{\omega}\|_2 + \hat{r} = \sqrt{\left( \bar{\theta} + \frac{1}{2\mu_d} \right)^2 + \left\| \bar{x} + \frac{1}{2\mu_d}u \right\|^2} + \frac{\tau}{2\mu_d} \leq |\bar{\theta}| + \|\bar{x}\| + \frac{1}{2\mu_d} + \frac{\|u\|}{2\mu_d} + \frac{\tau}{2\mu_d} \leq 3$ .  
■

The following theorem is an immediate consequence of Theorem 15 and Proposition 16:

**Theorem 17** *Suppose that the ellipsoid algorithm is applied in the linear subspace  $T$  to find a point in the set  $S$ , initialized with the Euclidean ball (in the space  $T$ ) of radius  $R = 3$  centered at  $(\theta^0, x^0) = (0, 0)$ . Then the ellipsoid algorithm will find a point in  $S$  (and hence, by transformation, a solution of  $(FP_d)$ ) in at most*

$$\left\lceil 2(n - m + 1)(n - m + 2) \ln \left( \frac{6\mu_d}{\tau} \right) \right\rceil$$

*iterations.*

## 4 Symmetry measure and other measures of conditioning for $(FP_d)$

### 4.1 Symmetry measure and the condition number

In this subsection we establish a relationship between  $\mu_d$  and  $\mathcal{C}(d)$ . As demonstrated in Theorem 18, if an instance of  $(FP_d)$  is “well-conditioned” in the sense that  $\mathcal{C}(d)$  is small,

then  $\mu_d$  is also small. This relationship, however, is one-sided, since  $\mu_d$  may carry no upper-bound information about  $\mathcal{C}(d)$ . In particular, in Remark 19 we exhibit a sequence of instances of  $(\text{FP}_d)$  with  $\mathcal{C}(d)$  becoming arbitrarily large, while  $\mu_d$  remains fixed.

**Theorem 18**  $\mu_d \leq \mathcal{C}(d)$ .

**Proof:** If  $\rho(d) = 0$ , then  $\mathcal{C}(d) = \infty$ , and the statement of the theorem holds trivially. Suppose  $\rho(d) > 0$ . Since  $B(0, \rho(d)) \subseteq \mathcal{H}_d$ , we conclude that for any  $v \in \mathcal{H}_d$ ,  $-\frac{\rho(d)}{\|v\|}v \in \mathcal{H}_d$ . Therefore

$$\frac{1}{\mu_d} = \text{sym}(\mathcal{H}_d, 0) \geq \inf_{v \in \mathcal{H}_d} \frac{\rho(d)}{\|v\|} \geq \frac{\rho(d)}{\|d\|} = \frac{1}{\mathcal{C}(d)},$$

proving the theorem. ■

**Remark 19**  $\mu_d$  may carry no upper-bound information about  $\mathcal{C}(d)$ .

To see why this is true, consider the parametric family of problems  $(\text{FP}_{d_\epsilon})$ , where  $d_\epsilon = (A_\epsilon, b)$ :

$$b = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } A_\epsilon = \begin{bmatrix} 1 & 1 & -1 & -1 \\ \epsilon & -\epsilon & \epsilon & -\epsilon \end{bmatrix},$$

$C_X = \mathbb{R}_+^4$  and  $\|x\| \triangleq \|x\|_1$  for  $x \in X$  and  $\|y\| = \|y\|_2$  for  $y \in Y$ . Consider the values of the parameter  $\epsilon \in (0, 1]$ . The set  $\mathcal{H}_{d_\epsilon}$  is symmetric about 0, so  $\mu_{d_\epsilon} = 1$  for any value of  $\epsilon$ . On the other hand,  $\rho(d_\epsilon) = \epsilon$  and  $\|d_\epsilon\| = \sqrt{1 + \epsilon^2}$ . Therefore,

$$\mathcal{C}(d_\epsilon) = \frac{\sqrt{1 + \epsilon^2}}{\epsilon} \geq \frac{1}{\epsilon},$$

and so  $\mathcal{C}(d)$  can be arbitrarily large, while  $\mu_d$  remains constant. Furthermore, letting  $\hat{x} = (1, 1, 1, 1)$  and  $r = 1$ , we see that  $\gamma$  in Theorem 11 has fixed value  $\gamma = 4$  for any  $\epsilon \in (0, 1]$ .

So far, we have made no assumptions on the norm on the space  $Y$ ; in fact, it can be easily seen that  $\mu_d$  is invariant under changes in the norm on  $Y$  (this is not true for  $\mathcal{C}(d)$ ). We conclude this section by providing another interpretation of the relationship between the measures  $\mu_d$  and  $\mathcal{C}(d)$ . As Theorem 20 indicates, when the space  $Y$  is endowed with the appropriate norm, then  $\mu_d$  and  $\mathcal{C}(d)$  are within a constant factor of each other. To see this, denote

$$\mathcal{T}_d \triangleq -\mathcal{H}_d \cap \mathcal{H}_d. \quad (18)$$

Then  $\mathcal{T}_d$  is a convex set that is symmetric about 0, and  $0 \in \text{int } \mathcal{T}_d$  when  $\mu_d < \infty$ . Therefore we can define the norm  $\|\cdot\|$  on  $Y$  to be that norm induced by considering  $\mathcal{T}_d$  to be the unit ball, namely:

$$\|y\| \triangleq \min \{\alpha : y \in \alpha \mathcal{T}_d\}. \quad (19)$$

**Theorem 20** Suppose  $C_X$  is regular and  $\mu_d < \infty$ . If the norm on  $Y$  is given by (19), then  $\rho(d) = 1$  and  $\mathcal{C}(d) \leq \frac{\mu_d}{\delta}$ , where  $\delta$  is the norm approximation coefficient of the cone  $C_X$ .

**Proof:** The characterization of  $\rho(d)$  in (10) easily implies that  $\rho(d) = 1$ . It remains to establish the bound on the condition number  $\mathcal{C}(d)$ . We have

$$\mathcal{C}(d) = \frac{\|d\|}{\rho(d)} = \|d\| \leq \frac{1}{\delta} \max\{\|y\| : y \in \mathcal{H}_d\} \leq \frac{\mu_d}{\delta}.$$

The first inequality above follows from Corollary 5. To verify the second inequality above, suppose that  $y \in \mathcal{H}_d$ . Then  $\frac{1}{\mu_d}y \in \mathcal{H}_d$  because  $\mu_d \geq 1$ , and  $-\frac{1}{\mu_d}y \in \mathcal{H}_d$  by the definition of  $\mu_d$ . Therefore,  $\frac{1}{\mu_d}y \in \mathcal{T}_d$ , and so  $\left\|\frac{1}{\mu_d}y\right\| \leq 1$ , which implies that  $\max\{\|y\| : y \in \mathcal{H}_d\} \leq \mu_d$ . This inequality is sufficient to prove the theorem; one can however show that  $\max\{\|y\| : y \in \mathcal{H}_d\} = \mu_d$ . ■

## 4.2 Relationships between the symmetry measure and other measures of conditioning for linear programming

In the special case when  $C_X = \mathfrak{R}_+^n$ , the problem  $(\text{FP}_d)$  becomes a linear feasibility problem, and can be written as follows:

$$(\text{FP}_d) : Ax = b, x \geq 0, \quad (20)$$

where  $x \in \mathfrak{R}^n$ ,  $b \in \mathfrak{R}^m$  and  $A \in \mathfrak{R}^{m \times n}$ . We assume in this subsection that  $(\text{FP}_d)$  has a strictly positive solution  $x^0$ , i.e.,  $Ax^0 = b$  and  $x^0 > 0$ , that the norm on  $X$  is  $\|x\| \triangleq \|x\|_1$ , and that the norm on  $Y$  is  $\|y\| \triangleq \|y\|_2$ .

Complexity analysis of linear programming sometimes relies on the complexity measures  $\sigma(\cdot)$  and  $\bar{\chi}(\cdot)$ . These measures are quite specific to the special case of linear programming, as opposed to  $\mathcal{C}(d)$  and  $\mu_d$  which apply to more general conic problems. In this subsection we state both previously known as well as new results relating all of these condition measures, which in total provide a complete characterization of the relationship between these four measures of conditioning.

For simplicity of notation, we define an “expanded” matrix  $\tilde{A} \triangleq [b; -A] \in \mathfrak{R}^{m \times (n+1)}$ . Notice that  $\|\tilde{A}\| \triangleq \max\{\|b\theta - Ax\| : \|(\theta, x)\|_1 \leq 1\} = \|d\|$ .

We first review a slight variant on  $\sigma(\cdot)$  called  $\sigma_d$ , which was introduced and used in the complexity analysis of an interior-point algorithm for solving  $(\text{FP}_d)$  by Vavasis and Ye [33]:

$$\sigma_d \triangleq \min_{j=1, \dots, n+1} \max_w \{e_j^t w : \tilde{A}w = 0, e^t w = 1, w \geq 0\},$$

where  $e_j$ ,  $j = 1, \dots, n+1$  denotes the  $j$ th unit vector and  $e \in \mathfrak{R}^{n+1}$  is the vector of all ones. Note that while the above does not coincide with the usual definition of  $\sigma$ , it does under our assumption that  $(\text{FP}_d)$  has a strictly positive solution.

We also review a slight variant on  $\bar{\chi}(\cdot)$  called  $\bar{\chi}_d$ , which has been used by Vavasis and Ye [34, 35] and Megiddo et al. [18] in the complexity analysis of another interior-point algorithm:

$$\bar{\chi}_d \triangleq \sup\{\|\tilde{A}^t(\tilde{A}D\tilde{A}^t)^{-1}\tilde{A}D\| : D \in S_{++}^{(n+1) \times (n+1)}, D \text{ diagonal}\}.$$

An alternative characterization of  $\bar{\chi}_d$  is

$$\bar{\chi}_d = \max\{\|B^{-1}\tilde{A}\| : B \in \mathcal{B}(\tilde{A})\}, \quad (21)$$

where  $\mathcal{B}(\tilde{A})$  is the set of all bases (i.e.,  $m \times m$  non-singular sub-matrices) of  $\tilde{A}$  (see [30] for the proof of the equivalence of these characterizations).

It has been established by Vavasis and Ye [33] that  $\sigma_d$  and  $\bar{\chi}_d$  are related by the inequality

$$\sigma_d \geq \frac{1}{\bar{\chi}_d + 1}.$$

On the other hand, Tunçel in [31] established that in general  $\sigma_d$  may carry no upper-bound information about  $\bar{\chi}_d$ . Specifically, he provided a family of data instances  $d_\epsilon$  such that for any  $\epsilon > 0$ ,  $\sigma_{d_\epsilon} = \frac{1}{2}$ , but  $\bar{\chi}_{d_\epsilon} \geq \frac{1}{\epsilon}$ , and so  $\bar{\chi}_{d_\epsilon}$  can be arbitrarily large.

Theorem 18 and Remark 19 established a relationship between  $\mu_d$  and  $\mathcal{C}(d)$ . Below we establish relationships between the other pairs of measures  $\mu_d$ ,  $\mathcal{C}(d)$ ,  $\bar{\chi}_d$ , and  $\sigma_d$ , or provide examples that show that no such relationship exists, in the spirit of [31].

**Remark 21**  $\mathcal{C}(d)$  and  $\bar{\chi}_d$  may carry no upper-bound or lower-bound information about each other.

To establish the above result, we provide two parametric families of matrices  $\tilde{A}_\epsilon$  such that by varying the value of the parameter  $\epsilon > 0$  we can make one of the above measures arbitrarily bad while keeping the other measure constant or bounded.

First consider the family of matrices  $\tilde{A}_\epsilon = \begin{bmatrix} \epsilon & 0 & -\epsilon \\ 0 & 1 & -1 \end{bmatrix}$ . For  $\epsilon > 0$  and sufficiently small,  $\rho(d_\epsilon) = \frac{\epsilon}{\sqrt{\epsilon^2 + 4}}$ . Furthermore,  $\|d_\epsilon\| = \sqrt{1 + \epsilon^2}$ , and so

$$\mathcal{C}(d_\epsilon) = \sqrt{\frac{\epsilon^2 + 1}{\epsilon^2 + 4}} \cdot \frac{1}{\epsilon} \rightarrow +\infty \text{ as } \epsilon \rightarrow 0.$$

On the other hand, it is easy to establish using (21) that  $\bar{\chi}(d_\epsilon) = \sqrt{2}$  for any  $\epsilon > 0$ .

To establish the second claim of the remark, consider the family  $\tilde{A}_\epsilon = [1 \ \epsilon \ -1]$  with  $0 < \epsilon < 1$ . We have:  $\|d_\epsilon\| = 1$ ,  $\rho(d_\epsilon) = 1$ , and so  $\mathcal{C}(d_\epsilon) = 1$  for any  $\epsilon$  as above. On the other hand it is easy to establish using (21) that for any  $\epsilon \in (0, 1)$ ,  $\bar{\chi}_{d_\epsilon} = \frac{1}{\epsilon} \rightarrow +\infty$  as  $\epsilon \rightarrow 0$ .

In the following proposition we establish an exact relationship between the measures  $\mu_d$  and  $\sigma_d$ .

**Proposition 22** *Suppose the system (FP<sub>d</sub>) of (20) has a positive solution. Then  $\sigma_d = \frac{1}{1+\mu_d}$ .*

**Proof:** Observe that we can redefine  $\sigma_d$  as follows:

$$\sigma_d = \min_{j=1,\dots,n+1} \sigma_j, \text{ where } \sigma_j \triangleq \max\{e_j^t w : \tilde{A}w = 0, e^t w = 1, w \geq 0\}.$$

From Lemma 7, there exists an extreme point  $\bar{w}$  of

$$\mathcal{H}_d = \{b\theta - Ax : (\theta, x) \geq 0, \|(\theta, x)\|_1 \leq 1\} = \{\tilde{A}w : w \geq 0, e^t w \leq 1\}$$

such that  $\frac{1}{\mu_d} = \text{sym}_{\bar{w}}(\mathcal{H}_d, 0) = \sup\{t : -t\bar{w} \in \mathcal{H}_d\}$ . Since the set of extreme points of the set  $\mathcal{H}_d$  is contained in the set  $\{\tilde{A}_1, \dots, \tilde{A}_{n+1}\}$ , where  $\tilde{A}_j \in \mathbb{R}^m$  is the  $j$ th column of the matrix  $\tilde{A}$ , we can characterize  $\mu_d$  as

$$\frac{1}{\mu_d} = \min_{j=1,\dots,n+1} \frac{1}{\mu_j}, \text{ where } \frac{1}{\mu_j} \triangleq \sup\{t : -t\tilde{A}_j \in \mathcal{H}_d\}.$$

We will now show that for any  $j$ ,

$$\sigma_j = \frac{1}{1 + \mu_j}. \quad (22)$$

Without loss of generality we can consider  $j = 1$  and the corresponding column  $\tilde{A}_1$  of  $\tilde{A}$ . If  $\tilde{A}_1 = 0$ , then  $\sigma_1 = 1$  and  $\frac{1}{\mu_1} = +\infty$ , and (22) holds as a limiting relationship. Suppose that  $\tilde{A}_1 \neq 0$ , and therefore  $\mu_1 > 0$  and  $\sigma_1 < 1$ .

By definition of  $\mu_1$ ,  $-\frac{1}{\mu_1}\tilde{A}_1 \in \mathcal{H}_d$ , i.e., there exists a point  $p \geq 0$ ,  $e^t p = 1$  such that  $-\frac{1}{\mu_1}\tilde{A}_1 = \tilde{A}p$ . Define  $w \triangleq \frac{\mu_1 p + e_1}{1 + \mu_1}$ . Then  $w \geq 0$ ,  $e^t w = 1$  and  $\tilde{A}w = 0$ . Therefore,  $\sigma_1 \geq w_1 \geq \frac{1}{1 + \mu_1}$ .

Suppose now that  $w$  is a solution of the linear program defining  $\sigma_1$ . Then  $w_1 = \sigma_1$ . Let  $p = \frac{1}{1 - \sigma_1}(w - \sigma_1 e_1)$ . Then  $p \geq 0$ ,  $e^t p = 1$  and  $\tilde{A}p = \frac{-\tilde{A}_1 \sigma_1}{1 - \sigma_1}$ . Therefore,  $\frac{1}{\mu_1} \geq \frac{\sigma_1}{1 - \sigma_1}$ , and so  $\sigma_1 \leq \frac{1}{\mu_1 + 1}$ . Combining this with the bound in the previous paragraph, we conclude that  $\sigma_1 = \frac{1}{\mu_1 + 1}$ , and by similar argument,  $\sigma_j = \frac{1}{\mu_j + 1}$ ,  $j = 1, \dots, n + 1$ .

Suppose now that  $\sigma_d = \sigma_j$  for some  $j$ . That means that  $\sigma_j \leq \sigma_i$  for any index  $i$ , or, equivalently,  $\frac{1}{\mu_j + 1} \leq \frac{1}{\mu_i + 1}$  and hence  $\mu_j \geq \mu_i$  for any index  $i$ . Therefore,  $\mu_d = \mu_j$  and hence  $\sigma_d = \frac{1}{1 + \mu_d}$ . ■

The following two remarks, which are easy consequences of Proposition 22, establish the remaining relationships between the four measures of conditioning.

**Remark 23**  $\mu_d \leq \bar{\chi}_d$ . However,  $\mu_d$  may carry no upper-bound information about  $\bar{\chi}_d$ .

**Remark 24**  $\sigma_d \geq \frac{1}{\mathcal{C}(d)+1}$ . However,  $\sigma_d$  may carry no upper-bound information about  $\mathcal{C}(d)$ .

In light of Proposition 22,  $\mu_d$  can in fact be viewed as a generalization of the Vavasis-Ye measure  $\sigma_d$  to a general conic linear system. Related to this, Ho in [12] provides an argument indicating that extending  $\bar{\chi}_d$  to general conic systems is not possible.

## 5 Pre-conditioners for conic linear systems

In this section we continue the development of the properties of the measure of conditioning  $\mu_d$  introduced in Section 3. We return to the issue of measures of conditioning of  $(\text{FP}_d)$  that are “data-independent,” first mentioned in Section 1, and we show that similar to  $\bar{\chi}_d$  and  $\sigma_d$ ,  $\mu_d$  is in fact independent of the particular data  $d$  used to characterize the feasible region  $X_d$  of the problem  $(\text{FP}_d)$ . We also present a characterization of all data instances  $\tilde{d}$  equivalent to  $d$  (in the sense that  $X_d = X_{\tilde{d}}$ ), by introducing the concept of a *pre-conditioner*, and we provide an upper bound on the condition number  $\mathcal{C}(\tilde{d})$  of the “best” equivalent data instance  $\tilde{d}$ . We conclude by analyzing the complexity of computing an equivalent data instance whose condition number is within a known factor of this bound, by constructing an algorithm for computing such an instance and analyzing its complexity.

Recall that for a feasible instance of the system  $(\text{FP}_d)$  we have defined the symmetry measure  $\mu_d = \frac{1}{\text{sym}(\mathcal{H}_d, 0)}$ , where  $d = (A, b) \in \mathcal{D}$  is the data instance defining the system  $(\text{FP}_d)$ , and

$$\mathcal{H}_d = \{b\theta - Ax : \theta \geq 0, x \in C_X, |\theta| + \|x\| \leq 1\}.$$

Let  $B \in \mathbb{R}^{m \times m}$  be a given non-singular matrix, and consider the data instance  $Bd \triangleq B \cdot d = (BA, Bb)$ , which gives rise to the system

$$(\text{FP}_{Bd}) : BAx = Bb, x \in C_X. \quad (23)$$

Notice that since  $B$  is non-singular, the systems  $(\text{FP}_d)$  and  $(\text{FP}_{Bd})$  are equivalent in the sense that  $X_d = X_{Bd}$ ; for this reason we say that the data instances  $d$  and  $Bd$  are equivalent as well. We can view the systems  $(\text{FP}_d)$  and  $(\text{FP}_{Bd})$  as different formulations of the same feasibility problem (FP): find  $x \in \mathcal{A} \cap C_X$ , where  $\mathcal{A}$  is the affine subspace

$$\mathcal{A} \triangleq \{x : Ax = b\} = \{x : BAx = Bb\}. \quad (24)$$

However the condition numbers of the two systems,  $\mathcal{C}(d)$  and  $\mathcal{C}(Bd)$ , are, in general, not equal.

On the other hand, consider the symmetry measures of the two systems,  $\mu_d$  and  $\mu_{Bd}$ . Observe that

$$\mathcal{H}_{Bd} \triangleq \{Bb\theta - BAx : \theta \geq 0, x \in C_X, |\theta| + \|x\| \leq 1\} = B(\mathcal{H}_d),$$

i.e., the set  $\mathcal{H}_{Bd}$  is the image of the set  $\mathcal{H}_d$  under the linear transformation defined by  $B$ . Therefore,  $\text{sym}(\mathcal{H}_{Bd}, 0) = \text{sym}(\mathcal{H}_d, 0)$ , and  $\mu_d = \mu_{Bd}$ , since the symmetry of a set is preserved under non-singular linear transformation, and so we can think of  $\mu_d$  as depending on the affine space  $\mathcal{A}$  defined in (24), which we write as  $\mu_{\mathcal{A}}$ . We record this formally as:

**Proposition 25** *Let  $d = (A, b) \in \mathcal{D}$ , let  $B \in \mathbb{R}^{m \times m}$  be a non-singular matrix, and define  $\mathcal{A} \triangleq \{x : Ax = b\}$ . Then  $\mu_d = \mu_{Bd} = \mu_{\mathcal{A}}$ .*

The next proposition shows that so long as the data instance  $d$  is not ill-posed, then any other data instance  $\tilde{d}$  giving rise to the same feasible region can be characterized as  $\tilde{d} = Bd$  for some non-singular matrix  $B$ .

**Proposition 26** *Suppose  $C_X$  is a regular cone. Let  $d = (A, b) \in \mathcal{D}$  and  $\tilde{d} = (\tilde{A}, \tilde{b}) \in \mathcal{D}$  be such that  $X_d = X_{\tilde{d}}$ . If  $\mathcal{C}(d) < \infty$ , then there exists a pre-conditioner  $B$  such that  $\tilde{d} = Bd$ .*

**Proof: (outline)** Let  $x_0 \in X_d = X_{\tilde{d}}$  be a feasible point such that  $x_0 \in \text{int } C_X$ . (Such a point must exist via Theorem 10 and Theorem 18.) Consider the two affine subspaces:

$$\mathcal{A}_d \triangleq \{x : x = x_0 + x_N, x_N \in \text{Null}(A)\} \text{ and } \mathcal{A}_{\tilde{d}} \triangleq \{x : x = x_0 + x_N, x_N \in \text{Null}(\tilde{A})\}.$$

Then  $\mathcal{A}_d \cap C_X = X_d = X_{\tilde{d}} = \mathcal{A}_{\tilde{d}} \cap C_X$ , and since both affine spaces  $\mathcal{A}_d$  and  $\mathcal{A}_{\tilde{d}}$  contain a point in the interior of the cone  $C_X$ , we conclude that  $\mathcal{A}_d = \mathcal{A}_{\tilde{d}}$ , and therefore there exists a non-singular matrix  $B \in \mathbb{R}^{m \times m}$  such that  $\tilde{A} = BA$ . It is also easy to establish that  $\tilde{b} = Bb$ , and so  $\tilde{d} = Bd$ . ■

In Section 3 we analyzed some of the properties of the problem  $(\text{FP}_d)$  in terms of  $\mu_d = \mu_{\mathcal{A}}$ . In particular, in the analysis of reliable solutions of  $(\text{FP}_d)$ , as well as the complexity of an interior-point algorithm and the ellipsoid algorithm for solving  $(\text{FP}_d)$ , we essentially replaced the dependence on  $\mathcal{C}(d)$  by the data-independent measure  $\mu_{\mathcal{A}}$ . Now suppose a feasibility problem can be represented via two equivalent data instances  $d$  and  $\tilde{d}$ , and suppose that  $\mathcal{C}(d) \ll \mathcal{C}(\tilde{d})$ . If one were to predict, for example, the performance of the interior-point algorithm for solving  $(\text{FP}_d)$  by analyzing its complexity in terms of the condition number, the bounds would be overly conservative if the problem is described by the data instance  $\tilde{d}$ . However, the analysis of the performance of the algorithm in terms of  $\mu_{\mathcal{A}}$  would yield the same bound, regardless of the data instance used.

On the other hand, the condition number  $\mathcal{C}(d)$  is a crucial parameter for analyzing properties of  $(\text{FP}_d)$  which depend on the representation of the problem  $(\text{FP}_{(\cdot)})$  by a specific data instance  $d$ , such as sensitivity of the feasible region to data perturbations, numerical properties of computations in algorithms for solving  $(\text{FP}_d)$ , etc. Therefore, it might be beneficial to *pre-condition* the system  $(\text{FP}_d)$ , i.e., find another data instance  $\tilde{d} = Bd$  for which  $\mathcal{C}(\tilde{d}) < \mathcal{C}(d)$ , and work with the corresponding system  $(\text{FP}_{\tilde{d}})$  which is better behaved. In this light, we can view the matrix  $B$  above as a pre-conditioner for the system  $(\text{FP}_d)$ ,



yielding the pre-conditioned system  $(FP_{\tilde{d}})$  with  $\tilde{d} = Bd$ , and Proposition 26 implies that any data instance  $\tilde{d}$  for which  $X_{\tilde{d}} = X_d$  can be obtained by pre-conditioning  $d$  with an appropriate  $B$ .

In the remainder of this section, we characterize a so-called best pre-conditioner, which is a pre-conditioner that gives rise to a condition number that is within a constant factor of the best possible, and we construct and analyze an algorithm for computing a pre-conditioner that yields a condition number that is within a known factor of this bound. For the remainder of this section, we assume that the space  $Y$  is the  $m$ -dimensional Euclidean space  $\mathbb{R}^m$  with Euclidean norm  $\|y\| = \|y\|_2 = \sqrt{y^t y}$ . We assume that the cone  $C_X$  is a regular cone with width  $\tau$  and norm approximation coefficient  $\delta$ . We also assume that  $m \geq 2$  (in fact, the case  $m = 1$  is trivial since in this case  $\mu_{\mathcal{A}}$  and  $\mathcal{C}(d)$  are within a factor of  $\delta$  of each other, and the issue of pre-conditioning is essentially irrelevant).

### 5.1 Best pre-conditioners and $\alpha$ -roundings

The main result of this subsection, Theorem 30, demonstrates the existence of a pre-conditioner  $\bar{B}$  such that  $\mathcal{C}(\bar{B}d)$  is within the factor  $\frac{\sqrt{m}}{\delta}$  of the lower bound of  $\mu_{\mathcal{A}}$ . In order to develop the tools to prove this result, we first review the concept of an  $\alpha$ -rounding of a set.

For any matrix  $Q \in S_{++}^{m \times m}$  we define  $E_Q$  to be the ellipsoid  $E_Q \triangleq \{y \in Y : y^t Q^{-1} y \leq 1\}$ .

**Definition 27** Let  $S \subset Y$  be a bounded set with a non-empty convex interior. For  $\alpha \in (0, 1]$ , we call an ellipsoid  $E_Q$  an  $\alpha$ -rounding of  $S$  if

$$\alpha E_Q \subseteq S \subseteq E_Q.$$

We refer to the parameter  $\alpha$  as the *tightness* of the rounding  $E_Q$ .

If the set  $S$  above satisfies  $S = -S$  (i.e., is symmetric about 0) then  $S$  possesses a  $\frac{1}{\sqrt{m}}$ -rounding, i.e., there exists an ellipsoid  $E_Q$  such that  $\frac{1}{\sqrt{m}} E_Q \subseteq S \subseteq E_Q$  (see John [14]). In particular, the ellipsoid of minimum volume containing  $S$  (often referred to as the Löwner-John ellipsoid of  $S$ ) is a  $\frac{1}{\sqrt{m}}$ -rounding of  $S$ .

The following lemma allows us to interpret pre-conditioning of the system  $(FP_d)$  by  $B$  as constructing a  $\frac{1}{\mathcal{C}(Bd)}$ -rounding of the set  $\mathcal{H}_d$ .

**Lemma 28** Let  $B \in \mathbb{R}^{m \times m}$  be a (non-singular) pre-conditioner for the system  $(FP_d)$ , and let  $\mathcal{C}(Bd)$  denote the condition number of the pre-conditioned system  $(FP_{Bd})$ . Let  $Q = \|Bd\|^2 (B^t B)^{-1}$ . Then

$$\frac{1}{\mathcal{C}(Bd)} E_Q \subseteq \mathcal{H}_d \subseteq E_Q.$$

**Proof:** First, observe that  $Q \in S_{++}^{m \times m}$ , since  $B$  is non-singular. To prove the first inclusion, let  $h \in \frac{1}{\mathcal{C}(Bd)} E_Q$ , i.e.,  $h^t Q^{-1} h \leq \frac{1}{\mathcal{C}(Bd)^2}$ . Using the definition of  $Q$  we have:  $h^t (B^t B) h \leq \frac{\|Bd\|^2}{\mathcal{C}(Bd)^2} = \rho(Bd)^2$ , that is,  $\|Bh\| \leq \rho(Bd)$ . This implies that  $Bh \in \mathcal{H}_{Bd}$ , and hence,  $h \in \mathcal{H}_d$ .

Next, suppose  $h \in \mathcal{H}_d$ , and so  $Bh \in \mathcal{H}_{Bd}$ . Then  $\|Bh\| \leq \|Bd\|$ , and therefore

$$h^t Q^{-1} h = h^t \left( \|Bd\|^2 (B^t B)^{-1} \right)^{-1} h = \frac{\|Bh\|^2}{\|Bd\|^2} \leq 1,$$

i.e.,  $h \in E_Q$ .  $\blacksquare$

**Lemma 29** Let  $Q \in S_{++}^{m \times m}$  be such that  $E_Q$  is an  $\alpha$ -rounding of the set  $\mathcal{T}_d$  of (18). Let  $B = Q^{-\frac{1}{2}}$ . Then  $B$  is a pre-conditioner for the system  $(FP_d)$  such that

$$\mathcal{C}(Bd) \leq \frac{\mu_A}{\alpha \delta} \leq \frac{2\mu_A}{\alpha \tau}.$$

**Proof:** We establish the result by providing bounds on the distance to infeasibility  $\rho(Bd)$  and the size of the data  $\|Bd\|$  of the system  $(FP_{Bd})$ . First, we will show that  $\rho(Bd) \geq \alpha$ . Let  $v \in Y$  satisfy  $\|v\| \leq \alpha$ . Then

$$(B^{-1}v)^t Q^{-1} B^{-1}v = (B^{-1}v)^t (B \cdot B) (B^{-1}v) = \|v\|^2 \leq \alpha^2,$$

and therefore  $B^{-1}v \in \alpha E_Q \subseteq \mathcal{T}_d \subseteq \mathcal{H}_d$ . Therefore,  $v \in \mathcal{H}_{Bd}$ , and so  $\rho(Bd) \geq \alpha$ .

Next, recall from Corollary 5 that  $\|Bd\| \leq \frac{1}{\delta} \max\{\|v\| : v \in \mathcal{H}_{Bd}\}$ . Let  $v \in \mathcal{H}_{Bd}$ . Then  $y = B^{-1}v \in \mathcal{H}_d$ , and  $\frac{1}{\mu_A} y \in -\mathcal{H}_d \cap \mathcal{H}_d = \mathcal{T}_d \subseteq E_Q$ . Hence  $\|v\|^2 = y^t B^t B y = y^t Q^{-1} y \leq \mu_A^2$ , whereby  $\|Bd\| \leq \frac{\mu_A}{\delta}$ .

Combining the obtained results,  $\mathcal{C}(Bd) = \frac{\|Bd\|}{\rho(Bd)} \leq \frac{\mu_A}{\delta \alpha} \leq \frac{2\mu_A}{\tau \alpha}$ .  $\blacksquare$

**Theorem 30** Suppose  $(FP_d)$  is feasible and  $\mathcal{C}(d) < +\infty$ . Then there exists a pre-conditioner  $\bar{B}$  such that

$$\mu_A \leq \mathcal{C}(\bar{B}d) \leq \frac{\sqrt{m}}{\delta} \cdot \mu_A. \quad (25)$$

**Proof:** By definition,  $\mathcal{T}_d$  is a bounded convex set symmetric about 0. Since  $\mathcal{C}(d) < \infty$ ,  $\mathcal{T}_d$  has a non-empty interior. Therefore, there exists  $Q \in S_{++}^{m \times m}$  such that  $E_Q$  is a  $\frac{1}{\sqrt{m}}$ -rounding of  $\mathcal{T}_d$ . Applying Lemma 29 with  $\alpha = \frac{1}{\sqrt{m}}$  we obtain (25).  $\blacksquare$

**Remark 31** The upper bound in (25) is tight for any  $m$ .

We verify this remark by example. Consider the system  $(FP_d)$  with  $n = 2m$ ,  $C_X = \mathfrak{R}_+^{2m}$ ,  $\|x\| = \|x\|_1$  (so that  $\delta = 1$ ) and the data  $d = (A, b)$  as follows:

$$b = 0 \text{ and } A = [e_1, -e_1, \dots, e_m, -e_m],$$

where  $e_i$  is the  $i$ th unit vector in  $\mathbb{R}^m$ . Then  $\mathcal{H}_d = \mathcal{T}_d = \text{conv}\{\pm e_i, i = 1, \dots, m\}$ , and it can be easily verified that  $\mu_{\mathcal{A}} = 1$ ,  $\rho(d) = \frac{1}{\sqrt{m}}$ , and  $\|d\| = 1$ , and therefore  $\mathcal{C}(d) = \sqrt{m}$ . Suppose  $B$  is an arbitrary pre-conditioner. Using Lemma 28, we can construct a  $\frac{1}{\mathcal{C}(\bar{B}d)}$ -rounding of the set  $\mathcal{T}_d$ . However, it is impossible to construct an  $\alpha$ -rounding of the set  $\text{conv}\{\pm e_i, i = 1, \dots, m\}$  with  $\alpha > \frac{1}{\sqrt{m}}$ , see, for example, [11]. Therefore,  $\mathcal{C}(Bd) \geq \sqrt{m}$  for any pre-conditioner  $B$ .

## 5.2 On the complexity of computing a good pre-conditioner

In this subsection we address the complexity of computing a good pre-conditioner for the problem  $(\text{FP}_d)$ . Specifically, we present an algorithm that computes a pre-conditioner  $\tilde{B}$  for which

$$\mathcal{C}(\tilde{B}d) \leq \frac{4m\mu_{\mathcal{A}}}{\delta}. \quad (26)$$

Recall that in Lemma 29 it was shown that a tight rounding of the set  $\mathcal{T}_d$  gives rise to a good pre-conditioner for the system  $(\text{FP}_d)$ . In Theorem 30 we relied on the existence of a  $\frac{1}{\sqrt{m}}$ -rounding of the set  $\mathcal{T}_d$  to establish the existence of a pre-conditioner  $\bar{B}$  such that  $\mu_{\mathcal{A}} \leq \mathcal{C}(\bar{B}d) \leq \frac{\sqrt{m}}{\delta} \mu_{\mathcal{A}}$ , i.e.,  $\mathcal{C}(\bar{B}d)$  is within the factor of  $\frac{\sqrt{m}}{\delta}$  of the lower bound. In general, we are not able to efficiently compute a  $\frac{1}{\sqrt{m}}$ -rounding of the set  $\mathcal{T}_d$  (see [11] for commentary on the difficulty of computing an approximate  $\frac{1}{\sqrt{m}}$ -rounding of a set  $\mathcal{S}$  that does not have an efficient half-space representation). However, the algorithm presented in this subsection will compute an ellipsoid which is a  $\frac{1}{4m}$ -rounding of  $\mathcal{T}_d$  (also called a weak Löwner-John ellipsoid for  $\mathcal{T}_d$ ). In particular, the algorithm of this subsection will compute a matrix  $\tilde{Q} \in S_{++}^{m \times m}$  such that

$$\frac{1}{4m} E_{\tilde{Q}} \subseteq \mathcal{T}_d \subseteq E_{\tilde{Q}}, \quad (27)$$

which can be used to obtain a pre-conditioner  $\tilde{B}$  satisfying (26) via Lemma 29. We denote this algorithm as *Algorithm WLJ* for “Weak Löwner-John”.

In order to be able to efficiently implement the algorithm described in this section, we restrict the norm  $\|x\|$  for  $x \in X$  to be the Euclidean norm  $\|x\| = \|x\|_2$  (as well as maintain the assumption that  $\|y\| = \|y\|_2$  for  $y \in Y$ ). We further assume that the interior of the cone  $C_X^*$  is the domain of a self-concordant barrier  $f^*(\cdot)$  with complexity parameter  $\vartheta^*$ . The width of the cone  $C_X^*$  is denoted by  $\tau^*$ , and we assume that we know and are given the vector  $u_* \in C_X^*$  for which  $\|u_*\| = 1$  and  $B(u_*, \tau^*) \subset C_X^*$  as in (7). Finally, we assume that an upper bound  $\bar{d}$  on  $\|d\|$  is known and given, or is easily computable. One could, for example, take

$$\bar{d} = \sqrt{n} \max\{\|b\|_2, \|A_1\|_2, \dots, \|A_m\|_2\},$$

where  $A_j$  is the  $j$ th column of the matrix  $A$ . Then  $\bar{d}$  approximates  $\|d\|$  within the factor of  $\sqrt{n}$ , i.e.,  $\frac{1}{\sqrt{n}}\bar{d} \leq \|d\| \leq \bar{d}$ .

The algorithm WLJ is a version of the parallel-cut ellipsoid algorithm, see [11]. A generic iteration of this algorithm can be described as follows. At the start of each iteration, we have a matrix  $Q \in S_{++}^{m \times m}$  such that  $\mathcal{T}_d \subseteq E_Q$ . We compute the eigenvalue decomposition of the matrix  $Q$ . In particular, we compute the eigenvalues  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$  of the matrix  $Q$  and their corresponding (orthonormal) eigenvectors  $a_1, \dots, a_m$ . Then the axes of the ellipsoid  $E_Q$  are  $v_i = \sqrt{\lambda_i} a_i$ ,  $i = 1, \dots, m$ . We denote  $V \triangleq [v_1, \dots, v_m] \in \mathbb{R}^{m \times m}$ . It is elementary to verify that  $Q = VV^t$ .

The algorithm then checks if the scaled axes  $\pm \frac{1}{4\sqrt{m}} v_i$  are elements of  $\mathcal{T}_d$ , for  $i = 1, \dots, m$ . If so, then the algorithm correctly asserts that

$$\frac{1}{4m} E_Q = \frac{1}{\sqrt{m}} \cdot \frac{1}{4\sqrt{m}} E_Q \subset \text{conv} \left\{ \pm \frac{1}{4\sqrt{m}} v_i, i = 1, \dots, m \right\} \subseteq \mathcal{T}_d \subseteq E_Q, \quad (28)$$

and the algorithm terminates. On the other hand, if the algorithm finds an axis  $v = \pm v_j$  for some  $j$  for which  $\frac{1}{4\sqrt{m}} v_j \notin \mathcal{T}_d$ , then it finds a parallel cut separating the two points  $\pm \frac{1}{2\sqrt{m}} v_j$  from the set  $\mathcal{T}_d$ , i.e., it produces a vector  $s$  such that

$$s^t v_j = 1 \text{ for some } v_j, \text{ and } \mathcal{T}_d \subseteq \left[ E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\} \right]. \quad (29)$$

This cut is then used to find an ellipsoid  $E_{\hat{Q}}$  which satisfies

$$E_{\hat{Q}} \supset \left[ E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\} \right] \supseteq \mathcal{T}_d,$$

and for which

$$\frac{\text{vol}(E_{\hat{Q}})}{\text{vol}(E_Q)} \leq \frac{1}{2} e^{\frac{3}{8}}. \quad (30)$$

The formula for  $\hat{Q}$  is:

$$\hat{Q} = \frac{m}{m-1} \left( 1 - \frac{1}{4m\xi} \right) \left( Q - \frac{m(4\xi-1)}{4m\xi-1} \cdot \frac{Q s s^t Q}{\xi} \right), \quad (31)$$

where

$$\xi = s^t Q s = \|V^t s\|^2 \geq s^t v_j = 1, \quad (32)$$

see formula (3.1.20) of [11], for example.

In order to implement this algorithm, it is necessary to be able to check if the re-scaled axes  $\pm \frac{1}{4\sqrt{m}} v_i$  are elements of  $\mathcal{T}_d$ , for  $i = 1, \dots, m$ , and if not, it is then necessary to produce the vector  $s$  describing the parallel cut of (29). These two tasks are accomplished in a subroutine called *Weak Check*, which is outlined as follows, and for which a more complete description is furnished in Appendix A.

### Subroutine *Weak Check*

Given the axes  $v_1, \dots, v_m$  of an ellipsoid  $E_Q \supseteq \mathcal{T}_d$ , either

- (i) verify that  $\pm \frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$  for all  $i = 1, \dots, m$ , or
- (ii) find a vector  $s$  such that

$$s^t v_j = 1 \text{ for some } v_j, \text{ and } \mathcal{T}_d \subseteq \left[ E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\} \right] \quad (33)$$

The formal description of algorithm WLJ is as follows:

### Algorithms WLJ (Weak Löwner-John)

- *Initialization*: The algorithm is initialized with the matrix  $Q^0 = d^2 I$ .
- Iteration  $k \geq 1$ .

**Step 1** Let  $Q = Q^k$ . Compute the eigenvalues  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$  of  $Q$  and the corresponding (orthonormal) eigenvectors  $a_1, \dots, a_m$ . Define the axes of  $E_Q$  by  $v_i = \sqrt{\lambda_i} a_i$ ,  $i = 1, \dots, m$ .

**Step 2** Call subroutine *Weak Check* with the input  $(v_1, \dots, v_m)$ . If the subroutine verifies that  $\pm \frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$ ,  $i = 1, \dots, m$ , then return  $\tilde{B} = Q^{-\frac{1}{2}}$  and terminate. Otherwise, subroutine *Weak Check* returns a vector  $s$ . Define  $\hat{Q}$  by (31).

**Step 3** Let  $Q^{k+1} = \hat{Q}$ ,  $k \leftarrow k + 1$ , go to Step 1.

To complete the description of algorithm WLJ, one must specify the specific details of subroutine *Weak Check*. The purpose of subroutine *Weak Check* is to verify if the re-scaled axes  $\pm \frac{1}{4\sqrt{m}}v_i$ ,  $i = 1, \dots, m$ , are contained in  $\mathcal{T}_d$ , or to produce a parallel cut otherwise. This is accomplished by examining the following  $2m$  optimization problems  $(P_{\phi_v})$ , where  $v = \pm v_i$ ,  $i = 1, \dots, m$ :

$$(P_{\phi_v}) \quad \begin{aligned} \phi_v &= \max_{\phi} \quad \phi \\ \text{s.t.} \quad &\phi v \in \mathcal{H}_d. \end{aligned} \quad (34)$$

It is easy to verify that  $\pm \frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$  for all  $i = 1, \dots, m$  precisely when

$$\phi_Q \triangleq \min_{\pm v_i} \phi_v \geq \frac{1}{4\sqrt{m}}. \quad (35)$$

(Here  $\min_{\pm v_i} \phi_v$  stands for  $\min\{\phi_{v_1}, -\phi_{v_1}, \dots, \phi_{v_m}, -\phi_{v_m}\}$  in order to shorten the notation). We will therefore implement the subroutine *Weak Check* by means of approximately solving the  $2m$  optimization problems (34) and checking if condition (35) is satisfied. To solve the optimization problems (34) for every value of  $v = \pm v_i$ ,  $i = 1, \dots, m$  we will apply the barrier method of [29] to a version of the Lagrangian dual of  $(P_{\phi_v})$ . The formal description of this implementation is presented in Appendix A, where the following complexity bound is proved:

**Lemma 32** *Subroutine Weak Check will terminate in at most*

$$O\left(m\sqrt{\vartheta^*} \ln\left(\frac{m\vartheta^*}{\tau^*} \cdot \frac{\bar{d}}{\sqrt{\lambda_1}} \cdot \sqrt{\frac{\lambda_m}{\lambda_1}}\right)\right) \quad (36)$$

*iterations of the barrier method. Upon termination, it will either correctly verify that  $\pm \frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$  for all  $i = 1, \dots, m$ , or will return a vector  $s$  such that*

$$s^t v_j = 1 \text{ for some } v_j, \text{ and } \mathcal{T}_d \subseteq \left[ E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\} \right]. \quad (37)$$

Note that the skewness of the ellipsoid  $E_Q$ , which is square root of the ratio of the largest to the smallest eigenvalue of  $Q$ , comes to play in the complexity bound of subroutine *Weak Check*.

We now proceed to analyze the complexity of algorithm WLJ. We first prove the volume reduction bound of (30) in Lemma 33. We then prove the main complexity of algorithm WLJ in Theorem 34.

**Lemma 33** *Let  $Q$  be an iterate of algorithm WLJ, and let  $\hat{Q}$  be defined by (31). Then*

$$\frac{\text{vol}(E_{\hat{Q}})}{\text{vol}(E_Q)} \leq \frac{1}{2}e^{\frac{3}{8}}.$$

**Proof:** Let  $R \in \mathbb{R}^{m \times m}$  be an orthonormal matrix such that  $RQ^{\frac{1}{2}}s = \|Q^{\frac{1}{2}}s\|e_1 = \sqrt{\xi}e_1$ . Then  $\hat{Q}$  can be expressed as

$$\hat{Q} = \frac{m}{m-1} \left(1 - \frac{1}{4m\xi}\right) Q^{\frac{1}{2}} R^t \left(I - \frac{m(4\xi-1)}{4m\xi-1} e_1 e_1^t\right) R Q^{\frac{1}{2}}. \quad (38)$$

Therefore,

$$\begin{aligned} \det(\hat{Q}) &= \det\left(\frac{m}{m-1} \left(1 - \frac{1}{4m\xi}\right) Q^{\frac{1}{2}} R^t \left(I - \frac{m(4\xi-1)}{4m\xi-1} e_1 e_1^t\right) R Q^{\frac{1}{2}}\right) \\ &= \left(\frac{m}{m-1} \left(1 - \frac{1}{4m\xi}\right)\right)^m \left(1 - \frac{m(4\xi-1)}{4m\xi-1}\right) \det(Q). \end{aligned}$$

We conclude that

$$\begin{aligned} \frac{\det(\hat{Q})}{\det(Q)} &= \left(\frac{m}{m-1} \left(1 - \frac{1}{4m\xi}\right)\right)^m \left(1 - \frac{m(4\xi-1)}{4m\xi-1}\right) \\ &= \frac{m^m (4m\xi-1)^{m-1}}{(m-1)^{m-1} (4m\xi)^m} = \frac{1}{4\xi} \left(\frac{4m\xi-1}{4m\xi-4\xi}\right)^{m-1} \\ &= \frac{1}{4\xi} \left(1 + \frac{4\xi-1}{4\xi(m-1)}\right)^{m-1} \leq \frac{1}{4\xi} e^{1-\frac{1}{4\xi}} \leq \frac{1}{4} e^{\frac{3}{4}}. \end{aligned}$$

The last inequality follows since the function  $te^{1-t}$  is an increasing function for  $t \in [0, 1]$ , and from (32) we have  $0 < \frac{1}{4\xi} \leq \frac{1}{4}$ . Finally,

$$\frac{\text{vol}(E_{\hat{Q}})}{\text{vol}(E_Q)} = \frac{\sqrt{\det(\hat{Q})}}{\sqrt{\det(Q)}} \leq \frac{1}{2}e^{\frac{3}{8}}. \blacksquare$$

**Theorem 34** *Suppose  $\mathcal{C}(d) < \infty$ . Then algorithm WLJ will terminate in at most*

$$O\left(m^2\sqrt{\vartheta^*}\ln^2\left(\frac{\bar{d}}{\rho(d)}\right)\ln\left(\frac{m\vartheta^*}{\tau^*}\right)\right) \quad (39)$$

*iterations of the barrier method. It will return upon termination a pre-conditioner  $\tilde{B}$  such that*

$$\mu_{\mathcal{A}} \leq \mathcal{C}(\tilde{B}d) \leq \frac{4m\mu_{\mathcal{A}}}{\delta}.$$

**Proof:** First observe that the matrix  $Q^0 = \bar{d}^2 I$  used to initialize the algorithm is a valid iterate, since for any point  $y \in \mathcal{T}_d$ ,  $\|y\| \leq \|d\| \leq \bar{d}$ , and so  $\mathcal{T}_d \subseteq E_{Q^0}$ .

Suppose algorithm WLJ has performed  $k$  iterations, and let  $Q^k$  be the current iterate. Since  $\mathcal{T}_d \subseteq E_{Q^k}$ , we conclude that

$$\text{vol}(\mathcal{T}_d) \leq \text{vol}(E_{Q^k}) \leq \left(\frac{1}{2}e^{\frac{3}{8}}\right)^k \text{vol}(E_{Q^0}) = \left(\frac{1}{2}e^{\frac{3}{8}}\right)^k \bar{d}^m \text{vol}(B(0, 1)).$$

On the other hand, since  $B(0, \rho(d)) \subseteq \mathcal{T}_d$ , we have  $\text{vol}(\mathcal{T}_d) \geq \text{vol}(B(0, \rho(d))) = \rho(d)^m \text{vol}(B(0, 1))$ . Therefore,  $\rho(d)^m \text{vol}(B(0, 1)) \leq \bar{d}^m \left(\frac{1}{2}e^{\frac{3}{8}}\right)^k \text{vol}(B(0, 1))$ , and algorithm WLJ will perform at most

$$K \leq m \ln\left(\frac{\bar{d}}{\rho(d)}\right) \cdot \frac{1}{\ln 2 - .375} \leq \frac{10}{3}m \ln\left(\frac{\bar{d}}{\rho(d)}\right) \quad (40)$$

iterations.

To bound the skewness of the ellipsoids generated by algorithm WLJ, note that all such ellipsoids contain the set  $\mathcal{T}_d$ , and therefore, contain  $B(0, \rho(d))$ . This implies that for any ellipsoid encountered by the algorithm,  $\lambda_1 \geq \rho(d)^2$ .

We now estimate the change in the largest eigenvalue of the ellipsoid matrix  $Q^k$  from one iteration of the algorithm to the next. Suppose  $Q$  and  $\hat{Q}$  are two consecutive iterates of the algorithm. Then from (38) we conclude that

$$\hat{\lambda}_m = \|\hat{Q}\| \leq \|Q\| \frac{m}{m-1} \left(1 - \frac{1}{4m\xi}\right) = \lambda_m \frac{m}{m-1} \left(1 - \frac{1}{4m\xi}\right) \leq \lambda_m \frac{m}{m-1} \leq \lambda_m e^{\frac{1}{m-1}}.$$

Hence, at any iteration  $k$ ,

$$\lambda_m^k \leq \lambda_m^0 e^{\frac{k}{m-1}} = \bar{d}^2 e^{\frac{k}{m-1}} \leq \left(\frac{\bar{d}}{\rho(d)}\right)^{\frac{10m}{3(m-1)}} \bar{d}^2,$$

the last inequality following from (40). Therefore, throughout the algorithm, the skewness of all ellipsoids generated by the algorithm is bounded above by

$$\sqrt{\frac{\lambda_m}{\lambda_1}} \leq \sqrt{\left(\frac{\bar{d}}{\rho(d)}\right)^{\frac{10m}{3(m-1)}+2}} \leq \left(\frac{\bar{d}}{\rho(d)}\right)^5. \quad (41)$$

Using (41) we conclude from Lemma 32 that any call to subroutine *Weak Check* will perform at most  $O\left(m\sqrt{\vartheta^*} \ln\left(\frac{m\vartheta^*}{\tau^*} \cdot \frac{\bar{d}}{\rho(d)}\right)\right)$  iterations of the barrier method. Combining this with (40), we can bound the total number of iterations of the barrier method performed by algorithm WLJ by

$$O\left(m^2\sqrt{\vartheta^*} \ln^2\left(\frac{\bar{d}}{\rho(d)}\right) \ln\left(\frac{m\vartheta^*}{\tau^*}\right)\right).$$

Finally, the inequalities  $\mu_{\mathcal{A}} \leq \mathcal{C}(\tilde{B}d) \leq \frac{4m\mu_{\mathcal{A}}}{\delta}$  follow from Theorem 18, (28), and Lemma 29. ■

**Remark 35** *Note that the skewness of the ellipsoids does not necessarily degrade at every iteration. In fact, the last ellipsoid of the algorithm has the very nice property that  $\sqrt{\frac{\lambda_m}{\lambda_1}} \leq 4\sqrt{m}\mathcal{C}(d)$ .*

To see why this remark is true, notice that the axes of any ellipsoid of the algorithm will satisfy  $\|v_i\| \geq \rho(d)$  for all  $i$ , and so  $\sqrt{\lambda_1} \geq \rho(d)$ . Also, the last ellipsoid of the algorithm satisfies  $\frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d \subset B(0, \|d\|)$  for all  $i$ , and so  $\|v_i\| \leq 4\sqrt{m}\|d\|$ , whereby  $\sqrt{\lambda_m} \leq \sqrt{m}\|d\|$ .

To further interpret the complexity result of Theorem 34, suppose for simplicity that  $\bar{d} = \|d\|$ , i.e., the size of the data  $\|d\|$  is known. Then algorithm WLJ will perform at most

$$O\left(m^2\sqrt{\vartheta^*} \ln^2(\mathcal{C}(d)) \ln\left(\frac{m\vartheta^*}{\tau^*}\right)\right)$$

iterations. We see that the condition number  $\mathcal{C}(d)$  of the initial data instance  $d$  plays a crucial role in the complexity of the algorithm WLJ which aims to find an equivalent data instance whose condition number is within a given factor of the best possible. In particular, if the original data instance  $d$  is badly conditioned, i.e.,  $\mathcal{C}(d)$  is large, it might take a large number of iterations to find a “good” pre-conditioner as above. Another interesting observation is that the complexity of algorithm WLJ does not depend on  $\mu_{\mathcal{A}}$ . This result, which may seem counterintuitive at first, is actually explained by the fact that, in order to obtain a pre-conditioner, algorithm WLJ has to work with the set  $\mathcal{T}_d$ , rather than  $\mathcal{H}_d$ , which is symmetric about 0 regardless of the geometry of  $\mathcal{H}_d$ .

## 6 Conclusions

In this paper we have addressed several issues related to measures of conditioning for



convex feasibility problems. We have discussed some potential drawbacks of using the condition number  $\mathcal{C}(d)$  as the sole measure of conditioning of a conic linear system, motivating the study of data-independent measures. We have introduced the symmetry measure  $\mu_{\mathcal{A}}$  for feasible conic linear systems as one such data-independent measure, and we have studied many of its implications for problem geometry, conditioning, and algorithm complexity.

One research topic that is not addressed in this paper concerns the existence of data-independent measures of conditioning for  $(\text{FP}_d)$  that are useful when  $(\text{FP}_d)$  is infeasible and/or whether any such measures can be adapted to analyze the linear optimization version of  $(\text{FP}_d)$ . Such measures might or might not be an extension of the symmetry measure discussed in this paper.

Another potential topic of research stems from the importance of the inherent conditioning of the problem data for certain properties of  $(\text{FP}_d)$  such as sensitivity to data perturbations and numerical precision required for accurate computation in algorithms. The complexity bound for computing the good pre-conditioner in algorithm WLJ is only reassuring in theory, as it would be unthinkable to use this algorithm in practice. Instead, much as in the case for linear optimization, it would be interesting to explore heuristic methods for pre-conditioning  $(\text{FP}_d)$ . The notion of a heuristic pre-conditioning/pre-processing stage in an algorithm is well-established; most optimization software packages include some type of pre-processing options, such as variable and constraint elimination or data scaling, for improving condition numbers and other numerical measures in matrix computations. We hope that the results in this paper may suggest future research on the analysis of heuristic pre-conditioning techniques for solving linear and conic optimization problems.

## A Implementation of subroutine *Weak Check*

In this appendix we present an implementation of the subroutine *Weak Check*. Recall that each iteration of algorithm WLJ calls the subroutine *Weak Check* with input being the axes  $v_1, \dots, v_m$  of an ellipsoid  $E_Q \supseteq \mathcal{T}_d$ . The purpose of *Weak Check* is to verify if the re-scaled axes  $\pm \frac{1}{4\sqrt{m}}v_i$  are elements of  $\mathcal{T}_d$ , for  $i = 1, \dots, m$ , and if not, to produce a parallel cut vector  $s$  satisfying (33).

Consider the following  $2m$  optimization problems  $(P_{\phi_v})$ , where  $v = \pm v_i$ ,  $i = 1, \dots, m$ :

$$\begin{aligned} (P_{\phi_v}) \quad \phi_v = \max_{\phi} \quad & \phi &= \max_{\theta, x, \phi} \quad & \phi \\ \text{s.t.} \quad & \phi v \in \mathcal{H}_d & \text{s.t.} \quad & b\theta - Ax = v\phi \\ & & & |\theta| + \|x\| \leq 1 \\ & & & \theta \geq 0, \ x \in C_X. \end{aligned} \tag{42}$$

It is easy to verify that  $\pm \frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$  for all  $i = 1, \dots, m$  precisely when

$$\phi_Q \triangleq \min_{\pm v_i} \phi_v \geq \frac{1}{4\sqrt{m}}. \tag{43}$$

(Here  $\min_{\pm v_i} \phi_v$  stands for  $\min\{\phi_{v_1}, -\phi_{v_1}, \dots, \phi_{v_m}, -\phi_{v_m}\}$  in order to shorten the notation.) We will therefore implement the subroutine *Weak Check* by means of approximately solving the  $2m$  optimization problems (42) and checking if condition (43) is satisfied.

The approach we will use to solve the optimization problems (42) in the subroutine *Weak Check* uses the barrier method of Renegar described in Section 3. Since no obvious starting point is available for (42), we will solve (42) for all  $2m$  values of  $v = \pm v_i$ ,  $i = 1, \dots, m$  by considering its dual:

$$\begin{aligned}
 (\text{P}_{\gamma_v}) \quad \gamma_v = & \min_{s,q,\gamma} \gamma \\
 \text{s.t.} \quad & \|A^t s - q\| \leq \gamma \\
 & b^t s \leq \gamma \\
 & q \in C_X^* \\
 & v^t s = 1.
 \end{aligned} \tag{44}$$

It is straightforward to verify that strong duality holds for  $(\text{P}_{\phi_v})$  and  $(\text{P}_{\gamma_v})$ , and so

$$\phi_Q = \min_{\pm v_i} \phi_v = \min_{\pm v_i} \gamma_v.$$

In order to be able to apply the barrier method, we need the optimization problem at hand to have a bounded feasible region. To satisfy this condition, we consider the following modification of (44):

$$\begin{aligned}
 (\text{P}_{\tilde{\gamma}_v}) \quad \tilde{\gamma}_v = & \min_{s,q,\gamma} \gamma \\
 \text{s.t.} \quad & \|A^t s - q\| \leq \gamma \\
 & b^t s \leq \gamma \\
 & \|V^t s\| \leq 2\sqrt{m} \\
 & \gamma \leq \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} \\
 & q \in C_X^* \\
 & v^t s = 1,
 \end{aligned} \tag{45}$$

where  $\bar{d}$  is the known upper bound on the norm of the data  $\|d\|$ , and  $V = [v_1, \dots, v_m] \in \mathbb{R}^{m \times m}$ . The following two simple facts are useful in the derivation of the forthcoming results. First, for all  $i = 1, \dots, m$ , we have  $\sqrt{\lambda_1} \leq \|v_i\| \leq \sqrt{\lambda_m}$ . Also, for any vector  $s \in Y^*$ ,  $\|s\| \leq \frac{\|V^t s\|}{\sqrt{\lambda_1}}$ . In the next proposition we show that solving  $(\text{P}_{\tilde{\gamma}_v})$  instead of  $(\text{P}_{\gamma_v})$  still yields a valid estimate of  $\phi_Q$ .

**Proposition 36** *For any  $v$ ,  $\gamma_v \leq \tilde{\gamma}_v$ . Moreover,*

$$\phi_Q = \min_{\pm v_i} \gamma_v = \min_{\pm v_i} \tilde{\gamma}_v. \tag{46}$$

**Proof:** The first claim of the proposition is trivially true, since the feasible region of the program  $(\text{P}_{\tilde{\gamma}_v})$  is contained in the feasible region of the program  $(\text{P}_{\gamma_v})$ .

To establish the second claim, note that

$$\phi_Q = \min_{\pm v_i} \gamma_{\pm v_i} \leq \min_{\pm v_i} \tilde{\gamma}_{\pm v_i}.$$

Suppose the minimum on the left is attained for  $v = v_{i_0}$ , and let  $(\bar{s}, \bar{q}, \gamma_v)$  be an optimal solution of the corresponding program  $(P_{\gamma_v})$ . Then we have

$$\gamma_v = \max\{\|A^t \bar{s} - \bar{q}\|, b^t \bar{s}\}, \quad \bar{q} \in C_X^*, \quad v^t \bar{s} = 1.$$

We can further assume without loss of generality that  $\|A^t \bar{s} - \bar{q}\| \leq \|A^t \bar{s}\|$ , since  $\bar{q}$  can always be chosen to minimize the distance from  $A^t \bar{s}$  to the cone  $C_X^*$ . If the point  $(\bar{s}, \bar{q}, \gamma_v)$  is feasible for the corresponding program  $(P_{\tilde{\gamma}_v})$ , then  $\gamma_v = \tilde{\gamma}_v$ , and (46) follows. Otherwise, let  $\sigma = \max_i |v_i^t \bar{s}| \geq 1$ . We can assume without loss of generality that  $\sigma = v_j^t \bar{s}$  for some  $j$  (if  $v_j^t \bar{s} < 0$ , we can re-define the  $j$ th axis of  $E_Q$  to be  $-v_j$ ). Define  $(\tilde{s}, \tilde{q}, \tilde{\gamma}) = (\frac{1}{\sigma} \bar{s}, \frac{1}{\sigma} \bar{q}, \frac{1}{\sigma} \gamma_v)$ . Note that  $v_j^t \tilde{s} = 1$ ,  $\tilde{q} \in C_X^*$  and

$$\|V^t \tilde{s}\| = \sqrt{\sum_{i=1}^m (v_i^t \tilde{s})^2} \leq \sqrt{m} \leq 2\sqrt{m}.$$

It remains to check if the upper bound constraint on  $\tilde{\gamma}$  is satisfied. Observe that  $\|\tilde{s}\| \leq \frac{\sqrt{m}}{\sqrt{\lambda_1}}$  (since  $\|V^t \tilde{s}\| \leq \sqrt{m}$ ). Therefore

$$\tilde{\gamma} = \max\{\|A^t \tilde{s} - \tilde{q}\|, b^t \tilde{s}\} \leq \max\{\|A^t \tilde{s}\|, b^t \tilde{s}\} \leq \bar{d} \cdot \frac{\sqrt{m}}{\sqrt{\lambda_1}} < \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}.$$

Hence the vector  $(\tilde{s}, \tilde{q}, \tilde{\gamma})$  is feasible for  $(P_{\tilde{\gamma}_{v_j}})$ , and  $\tilde{\gamma}_{v_j} \leq \tilde{\gamma} \leq \gamma_v \leq \gamma_{v_j} \leq \tilde{\gamma}_{v_j}$ , which implies that  $\tilde{\gamma}_{v_j} = \gamma_v$ , from which (46) follows.  $\blacksquare$

Now define

$$S \triangleq \left\{ (s, q, \gamma) : \|A^t s - q\| \leq \gamma, \quad b^t s \leq \gamma, \quad \|V^t s\| \leq 2\sqrt{m}, \quad \gamma \leq \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}, \quad q \in C_X^* \right\}$$

and

$$L_v \triangleq \{(s, q, \gamma) : v^t s = 1\}.$$

Then  $L_v$  is a translate of a closed linear subspace and  $S$  is a bounded convex set. Recall from the assumptions in Section 5.2 that  $f^*(\cdot)$  is a self-concordant barrier for the cone  $C_X^*$  with complexity parameter  $\vartheta^*$ . Then the interior of the set  $S$  is the domain of the following self-concordant barrier  $f(s, q, \gamma)$ :

$$f(s, q, \gamma) \triangleq f^*(q) - \ln(\gamma^2 - \|A^t s - q\|^2) - \ln(\gamma - b^t s) - \ln(4m - \|V^t s\|^2) - \ln\left(\frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} - \gamma\right),$$

whose complexity parameter is  $\vartheta_f \leq \vartheta^* + 5$ .

In order to use the barrier method to solve  $(P_{\tilde{\gamma}_v})$ , we need to have a point  $(s', q', \gamma') \in \text{int } S \cap L_v$  at which to initialize the method. The next proposition indicates that such point is readily available when the vector  $u_* \in C_X^*$  of (7) is known; the second part of the proposition presents a lower bound on  $\text{sym}(S \cap L_v, (s', q', \gamma'))$ , which is important in analyzing the complexity of the barrier method.

**Proposition 37**

$$(s', q', \gamma') \triangleq \left( \frac{v}{\|v\|^2}, \frac{2\bar{d}u_*}{\|v\|}, \frac{4\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} \right) \in \text{int } S \cap L_v ,$$

and

$$\text{sym}(S \cap L_v, (s', q', \gamma')) \geq \frac{\tau^*}{13\sqrt{m}} \cdot \sqrt{\frac{\lambda_1}{\lambda_m}}.$$

**Proof:** The first claim of the proposition is easily established by verifying directly that  $(s', q', \gamma')$  strictly satisfies the constraints of (45). The derivation of the bound on the symmetry in the second claim is fairly long and tedious, and is omitted. We refer the interested reader to [4] for details. ■

We now present the formal statement of the implementation of the subroutine *Weak Check*:

**Subroutine *Weak Check***

- *Input:* Axes  $v_i$ ,  $i = 1, \dots, m$  of an ellipsoid  $E_Q \supseteq \mathcal{T}_d$ .
- for  $v = \pm v_i$ ,  $i = 1, \dots, m$ ,

**Step 1** Form the problem  $(P_{\tilde{\gamma}_v})$

**Step 2** Run the barrier method on the problem  $(P_{\tilde{\gamma}_v})$  initialized at the point

$$(s', q', \gamma') = \left( \frac{v}{\|v\|^2}, \frac{2\bar{d}u_*}{\|v\|}, \frac{4\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} \right)$$

until the value of the barrier parameter  $\eta$  first exceeds  $\bar{\eta} = \frac{24\sqrt{m}\vartheta_f}{5}$ . Let  $(s, q, \gamma)$  be the last iterate of the barrier method.

**Step 3** If  $\gamma < \frac{1}{2\sqrt{m}}$ , terminate, and return  $s$ . Otherwise, continue with the next value of  $v$ .

- Assert that  $\frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$  for all  $i = 1, \dots, m$ .

**Proof of Lemma 32:** We will use (15) to establish the number of iterations of the barrier method required by subroutine *Weak Check*. Subroutine *Weak Check* will apply the barrier method to at most  $2m$  problems of the form  $(P_{\tilde{\gamma}_v})$ . Note that

$$\min_{(s,q,\gamma) \in S \cap L_v} \gamma \geq 0 \text{ and } \max_{(s,q,\gamma) \in S \cap L_v} \gamma \leq \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}.$$

Therefore, applying (15) and Proposition 37, we see that each of the (at most)  $2m$  applications of the barrier method will terminate in at most

$$\begin{aligned} & O\left(\sqrt{\vartheta_f} \ln\left(\frac{7\sqrt{m}\bar{d}\vartheta_f}{\sqrt{\lambda_1}} \cdot \frac{\bar{\eta}}{\text{sym}(S \cap L_v, (s', q', \gamma'))}\right)\right) \\ & \leq O\left(\sqrt{\vartheta^*} \ln\left(\frac{7\sqrt{m}\bar{d}\vartheta^*}{\sqrt{\lambda_1}} \cdot \frac{24\sqrt{m}\vartheta^*}{5} \cdot \frac{13\sqrt{m}}{\tau^*} \cdot \sqrt{\frac{\lambda_m}{\lambda_1}}\right)\right) \\ & = O\left(\sqrt{\vartheta^*} \ln\left(\frac{m\vartheta^*}{\tau^*} \cdot \frac{\bar{d}}{\sqrt{\lambda_1}} \cdot \sqrt{\frac{\lambda_m}{\lambda_1}}\right)\right) \end{aligned}$$

iterations of the barrier method. Multiplying the above bound by  $2m$  we obtain (36).

Suppose the subroutine *Weak Check* has terminated in **Step 3** of an iteration in which the barrier method is applied to the problem  $(P_{\tilde{\gamma}_{v_j}})$ . (This is without loss of generality; if the termination occurs during the iteration which applies the barrier method to the problem  $(P_{\tilde{\gamma}_{-v_j}})$ , we can re-define the  $j$ th axis of  $E_Q$  to be  $-v_j$  to preserve the notation.) Then the last iterate  $(s, q, \gamma)$  of the barrier method satisfies

$$\begin{aligned} \|A^t s - q\| &\leq \gamma < \frac{1}{2\sqrt{m}} \\ b^t s &\leq \gamma < \frac{1}{2\sqrt{m}} \\ \|V^t s\| &\leq 2\sqrt{m} \\ q &\in C_X^*, \quad v_j^t s = 1. \end{aligned}$$

The vector  $s$  above yields a parallel cut that separates  $\pm \frac{v_j}{2\sqrt{m}}$  from  $\mathcal{T}_d$ . To see why this is true, let  $h \in \mathcal{T}_d$ . Then  $h \in \mathcal{H}_d$ , and hence  $h = b\theta - Ax$  for some  $(\theta, x) \in \mathbb{R}_+ \times C_X$  such that  $|\theta| + \|x\| \leq 1$ . Therefore

$$\begin{aligned} s^t h &= s^t(b\theta - Ax) = \theta(b^t s) - x^t(A^t s) = \theta(b^t s) - x^t(A^t s - q) - x^t q \\ &\leq (|\theta| + \|x\|)\gamma \leq \gamma < \frac{1}{2\sqrt{m}} = \frac{s^t v_j}{2\sqrt{m}}. \end{aligned}$$

Applying the same argument for the point  $-h \in \mathcal{H}_d$ , we conclude that  $s^t h > -\frac{s^t v_j}{2\sqrt{m}}$ , and therefore the vector  $s$  returned by the subroutine *Weak Check* satisfies (37).

Next, suppose that the barrier method applied to  $(P_{\tilde{\gamma}_v})$  has not terminated in **Step 3** of the subroutine *Weak Check*, i.e., we have  $\gamma \geq \frac{1}{2\sqrt{m}}$ . Then, using (14),

$$\tilde{\gamma}_v \geq \gamma - \frac{6\vartheta_f}{5\bar{\eta}} \geq \frac{1}{2\sqrt{m}} - \frac{6\vartheta_f}{5\bar{\eta}} \geq \frac{1}{4\sqrt{m}}.$$

Therefore, if the subroutine *Weak Check* has not terminated in **Step 3** for any  $v = \pm v_i$ ,  $i = 1, \dots, m$ , we conclude that  $\phi_Q = \min_{\pm v_i} \tilde{\gamma}_v \geq \frac{1}{4\sqrt{m}}$ , and we correctly assert that  $\pm \frac{1}{4\sqrt{m}} v_i \in \mathcal{T}_d$  for all  $i = 1, \dots, m$ . ■

## References

- [1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, 1995.
- [2] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming, Theory and Algorithms*. John Wiley & Sons, Inc, New York, second edition, 1993.
- [3] R. Bland, D. Goldfarb, and M.J. Todd. The ellipsoid method: a survey. *Operations Research*, 29:1039–1091, 1981.
- [4] M. Epelman. *Complexity, Condition Numbers, and Conic Linear Systems*. PhD thesis, Massachusetts Institute of Technology, 1999.
- [5] M. Epelman and Robert M. Freund. Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system. *Mathematical Programming* (to appear), 2000.
- [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 University, May 1994.
- [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 University, December 1994.
- [8] R. M. Freund and J. R. Vera. On the complexity of computing estimates of condition measures of a conic linear system. Technical Report, Operations Research Center, MIT, August 1999.
- [9] R. M. Freund and J. R. Vera. Some characterizations and properties of the “distance to ill-posedness” and the condition measure of a conic linear system. *Mathematical Programming*, 86:225–260, 1999.
- [10] R. M. Freund and J. R. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM Journal on Optimization*, 10(1):155–176, 2000.
- [11] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag, 1994.
- [12] J. C. K. Ho. Structure of the central path and related complexity issues for linear programming. Master’s thesis, University of Waterloo, Waterloo, Ontario, Canada, 1999.
- [13] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.
- [14] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays, presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience, New York, 1948.
- [15] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.
- [16] L.G. Khachiyan. A polynomial algorithm in linear programming. *Soviet Math. Dokl.*, 20(1):191–194, 1979.

- [17] V. Klee and G.J. Minty. How good is the simplex algorithm? In O. Shisha, editor, *Inequalities*, pages 159–175. Academic Press, New York, 1972.
- [18] N. Megiddo, S. Mizuno, and T. Tsuchiya. A modified layered-step interior-point algorithm for linear programming. Technical report, IBM, IBM Research Lab, San Jose, CA, 1996.
- [19] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994.
- [20] M. A. Nunez. A characterization of ill-posed data instances for convex programming, September 1998.
- [21] M. A. Nunez and R. M. Freund. Condition measures and properties of the central trajectory of a linear program. *Mathematical Programming*, 83(1):1–28, 1998.
- [22] J. Peña. Understanding the geometry of infeasible perturbations of a conic linear system. Preprint.
- [23] J. Peña. Computing the distance to infeasibility: theoretical and practical issues. Technical report, Cornell University, October 1997.
- [24] J. Peña and J. Renegar. Computing approximate solutions for conic systems of constraints, May 1998.
- [25] J. Renegar. A polynomial-time algorithm, based on Newton’s method, for linear programming. *Mathematical Programming*, 40:59–93, 1988.
- [26] J. Renegar. Some perturbation theory for linear programming. *Mathematical Programming*, 65(1):73–91, 1994.
- [27] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM Journal on Optimization*, 5(3):506–524, 1995.
- [28] J. Renegar. Linear programming, complexity theory, and elementary functional analysis. *Mathematical Programming*, 70(3):279–351, 1995.
- [29] J. Renegar. A mathematical view of interior-point methods for convex optimization. Lecture notes, July 1998.
- [30] M.J. Todd, L. Tunçel, and Y. Ye. Probabilistic analysis of two complexity measures for linear programming problems. Research Report CORR 98-48, Faculty of Mathematics, University of Waterloo, October 1998.
- [31] L. Tunçel. On the condition numbers for polyhedra in Karmarkar’s form. Research Report CORR 97-24, Faculty of Mathematics, University of Waterloo, November 1997.
- [32] L. Tunçel. Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard. C&O Research Report 98-51, Faculty of Mathematics, University of Waterloo, November 1998.
- [33] S. A. Vavasis and Y. Ye. Condition numbers for polyhedra with real number data. *Operations Research Letters*, 17:209–214, 1995.
- [34] S. A. Vavasis and Y. Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Mathematical Programming*, 74:79–120, 1996.

- [35] S. A. Vavasis and Y. Ye. A simplification to "A primal-dual interior point method whose running time depends only on the constraint matrix", October 30 1996.
- [36] J. R. Vera. Ill-posedness and the computation of solutions to linear programs with approximate data. Technical Report, Cornell University, May 1992.
- [37] J. R. Vera. *Ill-Posedness in Mathematical Programming and Problem Solving with Approximate Data*. PhD thesis, Cornell University, 1992.
- [38] J. R. Vera. Ill-posedness and the complexity of deciding existence of solutions to linear programs. *SIAM Journal on Optimization*, 6(3):549–569, 1996.
- [39] J. R. Vera. On the complexity of linear programming under finite precision arithmetic. *Mathematical Programming*, 80(1):91–123, 1998.
- [40] S. J. Wright. *Primal-dual Interior-point methods*. SIAM, 1997.