Complexity, Condition Numbers, and Conic Linear Systems

by

Marina A. Epelman

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1999

Author ....
Sloan School of Management
May 7, 1999

Certified by ....
Robert M. Freund
Theresa Seley Professor of Management Science
Thesis Supervisor

Accepted by ....
James B. Orlin
Edward Pennell Brooks Professor of Operations Research,
Co-director, MIT Operations Research Center

Complexity, Condition Numbers, and Conic Linear Systems

by

Marina A. Epelman

Submitted to the Sloan School of Management
on May 7, 1999, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

The unifying theme of this thesis is the study of measures of conditioning for convex feasibility problems in conic linear form, or conic linear systems. Such problems are an important tool in mathematical programming. They provide 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 include linear programming feasibility problems as a special case. 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, a new and powerful theory of "condition numbers" for convex optimization has been developed. The condition numbers for convex optimization capture the intuitive notion of problem "conditioning" and have 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 geometry, etc.

The contribution of this thesis is twofold. We continue the research in the theory of condition numbers for convex problems by developing an elementary algorithm for solving a conic linear system, whose complexity depends on the condition number of the problem.

We also discuss some potential drawbacks in using the condition number as the sole measure of conditioning of a conic linear system, motivating the study of "data-independent" measures. We introduce a new measure of conditioning for feasible conic linear systems in special form and study its relationship to the condition number and other measures of conditioning arising in recent linear programming literature. We study many of the implications of the new measure for problem geometry, conditioning, and algorithm complexity, and demonstrate that the new measure is data-independent. We also introduce the notion of "pre-conditioning" for conic linear systems, i.e., finding equivalent formulations of the problem with better condition numbers. We characterize the best such formulation and provide an algorithm for constructing a formulation whose condition number is within a known factor of the best possible.

Thesis Supervisor: Robert M. Freund
Title: Theresa Seley Professor of Management Science

Acknowledgments[1]

I am greatly indebted to Professor Robert Freund for his contributions to all aspects of my work on this thesis, as well as my life at MIT. He was a great teacher, educator, and advisor, who provided a lot of the ideas and inspiration for the thesis, and taught me the beauty of optimization with his unparalleled intuition and excitement. He was the most enthusiastic and generous research supervisor and co-author, and without his contributions and guidance this work would not have been possible. Rob also was a great personal friend, supporter, and fellow classic rock amateur-enthusiast, who was always there for me through the good, the bad, and the ill-posed.

I would also like to thank my two committee members, Professors Thomas Magnanti and Georgia Perakis, for their many helpful comments about this thesis, and their excitement and interest in all my professional and personal undertakings.

The faculty at MIT and Cornell played an invaluable role in my education. I am especially grateful to Professor James Renegar who introduced me to the subject of nonlinear programming, provided the first glimpse into the world of condition numbers, and contributed many important insights throughout my work and studies.

I am greatly thankful to the entire Operations Research Center, to all the people who make it what it is. The co-directors, the staff, the faculty, and the students all contributed to the environment and atmosphere of the ORC, making it an exciting place to come to every day, be it for work or socializing, and often both.

I am very grateful to my friends, old and new, for keeping in touch, being interested in my work when it was on my mind, and being able to quickly change the subject when it was the source of my frustration. I would like to list many of the students — my contemporaries at the ORC, but the list is long, and besides, you know who you are! I would also like to thank my friends back in Moscow and all over the world and this country.

My deepest gratitude and love belong to Leonid, whose support and superhuman patience were with me all these years.

Last but not least, I would like to thank my family — my parents, grandparents, Galya,

Alex, and Masha — for their love, support, encouragement, advice, and occasional well-timed nagging.

Dedication

To my grandfather Lev, who would have gotten a kick out of all this!

Contents

Chapter 1

Introduction and Overview of the Thesis

The unifying theme of this thesis is the study of measures of conditioning for convex feasibility problems in conic linear form, or *conic linear systems*. Such problems are an important tool in mathematical programming. They provide 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 include linear programming feasibility problems as a special case. 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 the above breakthroughs, a new and powerful theory of condition numbers for convex optimization has been developed. The condition numbers for convex optimization capture the intuitive notion of problem "conditioning" and have 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 geometry, etc.

At the same time, continuing investigations in the field of linear programming have resulted in the development of novel approaches and new algorithms for solving linear programming problems. The complexity analysis of these algorithms has often incorporated new ways of characterizing the complexity of the problems via numerical parameters, or *measures*

of conditioning, which often provided a better understanding of the problem behavior.

The contribution of this thesis is twofold. We continue the research in the theory of condition numbers for convex problems by developing a new algorithm for solving a conic linear system, whose complexity depends on the condition number of the problem. Secondly, we explore the connection of the condition number to some of the measures of conditioning arising in recent linear programming literature, introduce a new relevant measure of conditioning, and study the pre-conditioners that improve the condition number of the problem.

In the following section we briefly review the developments in the theory of measures of conditioning in recent literature as well as give some indication of the issues that will be addressed in this thesis. Section 1.2 contains a detailed overview of the thesis.

1.1   Literature Review

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 [28] to have worst-case complexity exponential in the number of variables.

Khachiyan [27] demonstrated that linear programming problems were in fact polynomially solvable via the ellipsoid algorithm. In particular, 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. Unlike the simplex method, the ellipsoid algorithm, while providing a breakthrough in the theory of linear programming, did not prove to be computationally efficient.

The development of interior-point methods, however, gave rise to algorithms that are theoretically efficient as well as efficient in practice. The first such algorithm, developed by Karmarkar [26], has the complexity bound of $O(nL)$ iterations, and the algorithm introduced by Renegar [38] has the complexity bound of $O(\sqrt{n}L)$ iterations, which is the 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 drawback in analyz-

ing 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 give different performance estimates, yet actual performance of the algorithms will likely be similar due to their similar numerical properties. See Wright [59] 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 vertices 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 [39] in a more specific setting, but then generalized more fully in [40] and in [41] to convex optimization and feasibility problems in conic linear form. Before defining and discussing these concepts in detail in Chapter 2, we briefly review their importance for studying the properties of a convex feasibility problem in conic linear form, which is the central object considered in this thesis.

A convex feasibility problem in conic linear linear form is a problem

$$(\text{FP}_d) \quad b - Ax \in C_Y$$
$$x \in C_X,$$

where $A : X \to Y$ is a linear operator between n- and m-dimensional spaces X and Y, $b \in Y$, and $C_X \subset X$ and $C_Y \subset Y$ are each a closed convex cone. We denote by $d = (A, b)$ the "data" for the problem (FP_d). The *condition number* $\mathcal{C}(d)$ of (FP_d), developed by Renegar in a series of papers [39, 40, 41], 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 (FP_d) is well-conditioned to the extent that $\mathcal{C}(d)$ is small; when the problem (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 (FP$_d$). Renegar [39] demonstrated that when a feasible instance of (FP$_d$) is well-posed in the sense that $\mathcal{C}(d) < \infty$, then there exists a point x feasible for (FP$_d$) which satisfies

$$\|x\| \leq \mathcal{C}(d).$$

Therefore, when (FP$_d$) is well-conditioned, there exists a feasible point of small size. Furthermore, Freund and Vera [16] showed that under the above assumptions the set of feasible solutions contains a so-called "reliable" solution. We consider a solution reliable if, roughly speaking, the size R of this solution is not excessively large, the solution is contained in the relative interior of the feasible region, the distance r from this solution to the relative boundary of the feasible region is not excessively small, and the ratio R/r is not excessively large. Freund and Vera showed that when the system (FP$_d$) is feasible, there exists a feasible point $\hat{x}$ along with parameters r and R as above, such that

$$R/r = c_1 O(\mathcal{C}(d)), \; r = c_2 \Omega\left(\frac{1}{\mathcal{C}(d)}\right), \; R = c_3 O(\mathcal{C}(d)), \tag{1.1}$$

where the constants c_1, c_2, and c_3 depend only on the properties of the cones C_X and C_Y, and are independent of the data d of the problem (FP$_d$), but may depend on the dimensions n and m.

The condition number $\mathcal{C}(d)$ was also shown to be crucial for analyzing the complexity of algorithms for solving (FP$_d$). Renegar [41] presented an incredibly general interior-point (i.e., barrier) algorithm for solving (FP$_d$) and showed, roughly speaking, that the iteration complexity bound of the algorithm depends linearly on only two quantities: the barrier parameter for the underlying cones, and $\ln(\mathcal{C}(d))$, i.e., the logarithm of the condition number $\mathcal{C}(d)$. Freund and Vera [14] showed that a suitably modified version of the ellipsoid algorithm will find a point in the feasible region of (FP$_d$) in $O(n^2 \ln(\mathcal{C}(d)))$ iterations.

In this thesis, we develop a new "elementary" algorithm called "algorithm CLS" (for "Conic Linear System") which is based on a generalization of the algorithm of Von Neumann

studied by Dantzig [7, 8] (also see [10, 11]). Algorithm CLS computes a reliable solution of (FP_d) in the sense of (1.1), or demonstrates that (FP_d) is infeasible by computing a reliable solution of an alternative (i.e., dual) conic linear system. The complexity of algorithm CLS is also closely tied to the condition number $\mathcal{C}(d)$.

The recent literature has explored many other important properties of the problem (FP_d) tied to the distance to ill-posedness and the condition number $\mathcal{C}(d)$. Renegar [39] studied the relation of $\mathcal{C}(d)$ to sensitivity of solutions of (FP_d) under perturbations in the problem data. Peña and Renegar [36] discussed the role of $\mathcal{C}(d)$ in the complexity of computing approximate solutions of (FP_d). Freund and Vera [15] and Peña [35] addressed the theoretical complexity and practical aspects of computing the distance to ill-posedness.

Vera [57] considered the numerical properties of an interior-point method for solving (FP_d) (and in fact, a more general problem of optimizing a linear function over the feasible region of (FP_d)) in the case when (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 [12, 13], Nunez and Freund [33], Nunez [32], Peña [34, 35], and Vera [54, 55, 56].

As the above discussion hopefully conveys, the condition number $\mathcal{C}(d)$ is a relevant and important measure of conditioning of the problem (FP_d). Note that when (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 (FP_d). To illustrate this point, consider a problem of the form

$$
\begin{aligned}
(\mathrm{FP}_d) \quad & Ax = b \\
& x \in C_X,
\end{aligned}
\tag{1.2}
$$

i.e., assume that $C_Y = \{0\}$. The problem (FP_d) of (1.2) 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 the linear operator A. Notice that the description of the affine subspace $\mathcal{A}$ by the data instance $d = (A, b)$ is not unique. It is not hard to find an alternative 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 $B : Y \to Y$). Then the problem

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

is equivalent to the problem (FP_d) in the sense that their feasible regions are identical; we can think of the systems (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 the properties of the problem (FP) above in terms of the condition number will lead to different results, depending on which formulation, (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 existence of reliable solutions 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 used to "represent" the affine space $\mathcal{A}$.

An important research direction, therefore, is the development of relevant measures of conditioning of the problem (FP_d) that do not depend on a particular data instance d, but rather capture the geometry of the problem, 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 (FP_d) is

a linear programming feasibility problem. In particular, two measures, $\bar{\chi}_d$ and σ_d, were used in the analysis of interior-point algorithms for linear programming (see, for example, Vavasis and Ye [51, 52, 53]). These measures, discussed in detail in Chapter 5, 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 [22], Todd, Tunçel and Ye [48], and Tunçel [50].

In this thesis we define a new measure of conditioning, μ_d, for feasible instances of the problem (FP$_d$) of (1.2), which is independent of the specific data representation of the problem. We explore the relationship between μ_d and measures $\bar{\chi}_d$, σ_d, and $\mathcal{C}(d)$. In particular, we demonstrate that the measure σ_d is directly related to μ_d in the special case of linear programming. We also demonstrate that many important properties of the system (FP$_d$) previously analyzed in terms of $\mathcal{C}(d)$ can be analyzed through μ_d (independently of the data representation).

On the other hand, some properties of (FP$_d$) are not purely geometric and depend on the data d. We show how, given a data instance d, to construct a data instance $\tilde{d}$ such that the problem (FP$_{\tilde{d}}$) is equivalent to (FP$_d$) but is better conditioned in the sense that $\mathcal{C}(\tilde{d}) < \mathcal{C}(d)$ (we refer to this construction as *pre-conditioning* of (FP$_d$)).

1.2 Overview of the Thesis

Chapter 2 contains notation, definitions and some preliminary results. In Section 2.1 we formally define a convex feasibility problem in conic linear form, denoted by (FP$_d$). In Section 2.2 we define the distance to ill-posedness and the condition number $\mathcal{C}(d)$ of the problem (FP$_d$). In Section 2.3, we discuss several "theorems of the alternative" for the conic linear system (FP$_d$) in the spirit of Farkas' lemma; we denote by (SA$_d$) the "strong alternative" conic linear system of (FP$_d$). Section 2.4 motivates and formally introduces the notion of reliable solutions of (FP$_d$) and (SA$_d$). Finally, Section 2.5 contains the definition

and an extensive discussion of properties of regular cones, which play an important role throughout the thesis.

In Chapters 3 and 4 we develop an "elementary" algorithm called "algorithm CLS" (for Conic Linear System) for computing a reliable solution of (FP_d) whose complexity is bounded appropriately by the condition number $\mathcal{C}(d)$. Our motivation for developing algorithm CLS lies in instances of (FP_d) where an interior-point or other theoretically-efficient algorithm might not be an attractive choice for solving (FP_d). Such instances might arise when n is extremely large, and/or when A is a real matrix whose sparsity structure is incompatible with efficient computation in interior-point methods.

Algorithm CLS either computes a solution of the system (FP_d), or demonstrates that (FP_d) is infeasible by computing a solution of an alternative (i.e., dual) system (SA_d). In both cases algorithm CLS returns a reliable solution of the appropriate system.

Algorithm CLS is based on a generalization of the algorithm of von Neumann studied by Dantzig [7] and [8], and is part of a large class of "elementary" algorithms for finding a point in a suitably described convex set, such as reflection algorithms for linear inequality systems (see [1, 30, 9, 17]), the "perceptron" algorithm [44, 45, 46, 47], and other so-called "row-action" methods. When applied to linear inequality systems, these elementary algorithms share the two desirable properties: (i) the work per iteration is extremely low (typically involving only a few matrix-vector or vector-vector multiplications), and (ii) the algorithms fully exploit the sparsity of the original data at each iteration. The performance of these algorithms can be quite competitive when applied to certain very large problems with very sparse data, see [6]. We refer to these algorithms as "elementary" in that, unlike interior-point algorithms or the ellipsoid algorithm, the algorithms do not perform particularly sophisticated computations at each iteration.

In Chapter 3 we develop elementary algorithms for conic linear systems in special form. Section 3.1 presents a generalization of the von Neumann algorithm (appropriately called algorithm GVNA) that can be applied to conic linear systems in a special compact form (i.e., with a compactness constraint added). We analyze the properties of the iterates of algorithm GVNA under different termination criteria. Sections 3.2 and 3.3 present the development of algorithms HCI (Homogeneous Conic Inequalities) and HCE (Homogeneous

17

Conic Equalities), respectively, for resolving two essential types of homogeneous conic linear systems. Both algorithms HCI and HCE consist of calls to algorithm GVNA applied to appropriate transformations of the homogeneous systems at hand.

In Chapter 4 we indicate how to use algorithms HCI and HCE of Chapter 3 to obtain a reliable solution of a conic linear system in the most general form. We discuss three different cases of instances of (FP_d), which result from making different assumptions on the two cones appearing in the formulation of (FP_d). Due to the differences in the underlying geometry, a different algorithm has to be used in each of the three cases. However, the general framework of all of the algorithms is the same. In every case, the main algorithm CLS consists of applying algorithm HCE and/or algorithm HCI to suitable transformations of the system (FP_d), and transforming the output into a solution of the system (FP_d) or (SA_d). The complexity of algorithm CLS is slightly different for the three different cases, but in any of the cases the algorithm will perform at most

$$O(c\,\mathcal{C}(d)^2\ln(\mathcal{C}(d)))$$

iterations, where c is a constant that depends only on the properties of the cones C_X and C_Y and is independent of the data d, but may depend on the dimensions n and m.

It is interesting to compare the complexity bound of algorithm CLS to that of other algorithms for solving (FP_d). Recall from Section 1.1 that both the interior-point algorithm and the ellipsoid algorithm have an iteration complexity bound that is linear in $\ln(\mathcal{C}(d))$, and so are efficient algorithms in a sense defined by Renegar [40]. Both the interior-point algorithm and the ellipsoid algorithm are also very sophisticated algorithms, in contrast with the elementary algorithm CLS. The interior-point algorithm makes implicit and explicit use of information from a self-concordant barrier at each iteration, and uses this information in the computation of the next iterate by solving for the Newton step along the central trajectory. The work per iteration is $O(n^3)$ operations to compute the Newton step. The ellipsoid algorithm makes use of a separation oracle for the cones C_X and C_Y in order to perform a special space dilation at each iteration, and the work per iteration of the ellipsoid algorithm is $O(n^2)$ operations. Intuition strongly suggests that the sophistication

of these methods is responsible for their excellent computational complexity. In contrast, the elementary algorithm CLS relies only on relatively simple assumptions regarding the ability to work conveniently with the cones C_X and C_Y (discussed in detail in Chapter 3) and does not perform any sophisticated mathematics at each iteration. Consequently one would not expect its theoretical complexity to be nearly as good as an interior-point algorithm or the ellipsoid algorithm. However, because the work per iteration of algorithm CLS is low, and each iteration fully exploits the sparsity of the original data, it is reasonable to expect that algorithm CLS could outperform more theoretically-efficient algorithms on large structured problems that are well-conditioned.

In this vein, the recent literature contains several elementary-like algorithms for obtaining approximate solutions of certain structured convex optimization problems. For example, Grigoriadis and Khachiyan [19, 20] and Villavicencio and Grigoriadis [58] consider algorithms for block angular resource sharing problems, Plotkin, Shmoys, and Tardos [37] and Karger and Plotkin [25] consider algorithms for fractional packing problems, and Bienstock [4] and Goldberg et al. [18] discuss results of computational experiments with these methods. The many applications of such problems include multi-commodity network flows, scheduling, combinatorial optimization, etc. The dimensionality of these structured problems arising in practice is often prohibitively large for theoretically efficient algorithms such as interior-point methods to be effective. However, these problems are typically sparse and structured, which allows for efficient implementation and good performance of Lagrangian-decomposition based algorithms (see, for example, [58]), which offer a general framework for row-action methods. These algorithms can also be considered "elementary" in the exact same sense as the row-action algorithms mentioned earlier, i.e., they do not perform any sophisticated mathematics at each iteration and they fully exploit the sparsity of the original data. The complexity analysis as well as the practical computational experience of this body of literature lends more credence to the practical viability of elementary algorithms in general, when applied to large-scale, sparse (well-structured), and well-conditioned problems.

In Chapters 5 and 6 we discuss other measures of conditioning of the problem (FP$_d$). We

restrict our attention to feasible instances of (FP$_d$) of the form

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

and define a new measure of conditioning μ_d for such problems. Chapter 5 is dedicated to establishing several properties of the measure μ_d and characterizing its relationship to the condition number $\mathcal{C}(d)$ and the measures $\bar{\chi}_d$ and σ_d mentioned above. In Section 5.2, we show how to characterize some geometric properties of the feasible region of (FP$_d$) in terms of μ_d. We are able to effectively replace the condition number $\mathcal{C}(d)$ in the bound on the size of solutions (Renegar [39]) and properties of reliable solutions (Freund and Vera [16]) by μ_d. In particular, we show that there exists a solution x of (FP$_d$) which satisfies

$$\|x\| \leq \mu_d,$$

and there exists a reliable solution $\hat{x}$ along with parameters r and R such that

$$R/r = \tilde{c}_1 O(\mu_d), \;\; r = \tilde{c}_2 \Omega\left(\frac{1}{\mu_d}\right), \;\; R = \tilde{c}_3 O(\mu_d),$$

where the constants $\tilde{c}_1$, $\tilde{c}_2$, and $\tilde{c}_3$ depend only on the properties of the cone C_X (recall from Section 1.1 that R is the size of the solution $\hat{x}$ and r is the distance from $\hat{x}$ to the boundary of the feasible region of (FP$_d$)). Moreover, unlike for $\mathcal{C}(d)$, the converse is also true: μ_d is guaranteed to be small if the feasible region has nice geometry. Namely, we show that if the system (FP$_d$) has a reliable solution $\hat{x}$, then

$$\mu_d \leq 1 + 2\max\left\{R, \; \frac{1}{r}, \; \frac{R}{r}\right\}.$$

We conclude the section by establishing a "one-sided" relationship between the two measures of conditioning, namely, $\mu_d \leq \mathcal{C}(d)$, i.e., if $\mathcal{C}(d)$ is small, then so is μ_d. However, we show that μ_d may carry no upper-bound information about $\mathcal{C}(d)$.

We continue studying the relationship between μ_d and other measures of conditioning in

Section 5.3, where we completely characterize the relationship between μ_d, $\mathcal{C}(d)$, $\bar{\chi}_d$, and σ_d in the linear programming setting.

In Section 5.4 we analyze the performance of several algorithms for solving (FP$_d$) in terms of μ_d. We effectively replace $\mathcal{C}(d)$ by μ_d in the complexity analyses of the interior-point algorithm of Renegar [41] and the ellipsoid algorithm of Freund and Vera [14].

In Chapter 6 we return to the issue of the importance of measures of conditioning of (FP$_d$) that are "formulation invariant." In Section 6.1 we show that, similarly to $\bar{\chi}_d$ and σ_d, μ_d does not depend on the particular data d used to characterize the feasible region of the problem (FP$_d$), and is therefore a purely geometric measure. Therefore, the properties of (FP$_d$) and the complexity of algorithms studied in Chapter 5 can be characterized independently of the data d.

On the other hand, some properties of (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 $\tilde{d}$ equivalent to d (i.e., such that the problem (FP$_{\tilde{d}}$) has the same feasible region as (FP$_d$)), but is better conditioned in the sense that $\mathcal{C}(\tilde{d}) < \mathcal{C}(d)$. In Section 6.2 we develop a characterization of all equivalent data instances $\tilde{d}$ by considering a *pre-conditioner* for the problem (FP$_d$) in the form of a non-singular matrix $B : \Re^m \to \Re^m$. The data instance $Bd \triangleq (BA, Bb)$ is equivalent to d, and we prove that under some assumptions on the cone C_X, any equivalent data instance $\tilde{d}$ can be constructed by considering an appropriate pre-conditioner. We also give a geometric interpretation of a pre-conditioner B and show that there exists a pre-conditioner $\bar{B}$ such that

$$\mu_d \leq \mathcal{C}(\bar{B}d) \leq \tilde{c}\sqrt{m}\mu_d,$$

i.e., $\mathcal{C}(\bar{B}d)$ is within a factor $\tilde{c}\sqrt{m}$ of the lower bound. Here, $\tilde{c}$ is a constant that depends only on the cone C_X.

Finally, in Section 6.3 we address the issue of computing a "good" pre-conditioner for the problem (FP$_d$). By exploiting the geometric interpretation of pre-conditioners, we develop

21

an algorithm that computes a pre-conditioner $\tilde{B}$ such that

$$\mathcal{C}(\tilde{B}d) \leq \tilde{c} \cdot 4m\mu_d.$$

We also present a complexity analysis of our algorithm.

Chapter 7 contains some final discussion and indicates potential topics of future research arising from this thesis.

Chapter 2

Preliminaries

2.1 Conic Linear Systems

Throughout this thesis we will be working with convex feasibility problems in conic linear form:

$$\text{(FP}_d)\quad \begin{aligned} b - Ax &\in C_Y \\ x &\in C_X, \end{aligned} \tag{2.1}$$

where $C_X \subset X$ and $C_Y \subset Y$ are each a closed convex cone in the (finite) n-dimensional linear space X and the (finite) m-dimensional linear space Y, respectively. Here $b \in Y$ and $A \in L(X, Y)$, the set of all linear operators $A : X \to Y$.

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

$$X_d = \{x \in X : b - Ax \in C_Y,\ x \in C_X\}. \tag{2.2}$$

The problem (FP$_d$) is a very general format for studying the feasible regions of convex optimization problems, and has recently received much attention in the analysis of interior-point methods, see Nesterov and Nemirovskii [31] and Renegar [40] and [41], among others,

who show that interior-point methods for (FP_d) are theoretically efficient.

We will work in the setup of finite dimensional normed linear vector spaces. Both X and Y are normed linear spaces of finite dimension n and m, respectively, endowed with norms $\|x\|$ for $x \in X$ and $\|y\|$ for $y \in Y$. Unless explicitly specified, the norms on the spaces are arbitrary. For $\bar{x} \in X$, let $B(\bar{x}, r)$ denote the ball centered at $\bar{x}$ with radius r, i.e.,

$$B(\bar{x}, r) = \{x \in X : \|x - \bar{x}\| \le r\},$$

and define $B(\bar{y}, r)$ analogously for $\bar{y} \in Y$.

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 $A : X \to Y$ be a linear operator. We denote $A(x)$ by Ax, and we denote the adjoint of A by $A^t : Y^* \to X^*$.

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

$$\|c\|_* \triangleq \max\{c^t x : x \in X, \ \|x\| \le 1\}, \tag{2.3}$$

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

If X_1 and X_2 are finite-dimensional normed linear spaces with norms $\|x_1\|$ for $x_1 \in X_1$ and $\|x_2\|$ for $x_2 \in X_2$, then for $x = (x_1, x_2) \in X_1 \times X_2$,

$$\|x\| = \|(x_1, x_2)\| \triangleq \|x_1\| + \|x_2\|$$

defines a norm on the linear space $X = X_1 \times X_2$. From (2.3), the dual norm on the space $X^* = X_1^* \times X_2^*$ is then defined as

$$\|c\|_* = \|(c_1, c_2)\|_* \triangleq \max\{\|c_1\|_*, \|c_2\|_*\} \tag{2.4}$$

24

for $c = (c_1, c_2) \in X_1^* \times X_2^*$.

We denote the set of real numbers by $\Re$, the set of nonnegative real numbers by $\Re_+$, and the nonnegative orthant in $\Re^n$ by $\Re_+^n$.

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 $S_+^{k \times k} \triangleq \left\{ x \in S^{k \times k} : x \succeq 0 \right\}$, where $x \succeq 0$ is the Löwner partial ordering, i.e., $x \succeq w$ if and only if $x - w$ is a positive semi-definite symmetric matrix. $S_+^{k \times k}$ 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 $S_{++}^{k \times k}$.

2.2 Distance to Ill-Posedness and Condition Numbers

We now present the concepts of condition numbers and data perturbation for (FP$_d$) in detail. Recall that $d = (A, b)$ is the data for the problem (FP$_d$). The space of all data $d = (A, b)$ for (FP$_d$) is denoted by $\mathcal{D}$:

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

For $d = (A, b) \in \mathcal{D}$ we define the product norm on the cartesian product $L(X, Y) \times Y$ to be

$$\|d\| = \|(A, b)\| = \max\{\|A\|, \|b\|\}, \tag{2.5}$$

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

$$\|A\| = \max\{\|Ax\| : \|x\| \leq 1\}. \tag{2.6}$$

We define

$$\mathcal{F} = \{(A, b) \in \mathcal{D} : \text{ there exists } x \text{ satisfying } b - Ax \in C_Y, \ x \in C_X\}. \tag{2.7}$$

Then $\mathcal{F}$ corresponds to those data instances $d = (A, b)$ for which (FP$_d$) is feasible. The

25

complement of $\mathcal{F}$, denoted by $\mathcal{F}^C$, consists of those data instances $d = (A, b)$ for which (FP_d) is infeasible.

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

$$\mathcal{B} = \partial \mathcal{F} = \partial \mathcal{F}^C = \text{cl}(\mathcal{F}) \cap \text{cl}(\mathcal{F}^C), \qquad (2.8)$$

where ∂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 (FP_d) is ill-posed in the sense that arbitrarily small changes in the data $d = (A, b)$ can yield instances of (FP_d) that are feasible, as well as instances of (FP_d) that are infeasible. Also, note that if $C_Y \neq Y$ (i.e., the problem (FP_d) is nontrivial), $\mathcal{B} \neq \emptyset$, since $d = (0, 0) \in \mathcal{B}$ (see, for example, [16]).

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

$$\rho(d) = \inf\{\|\Delta d\| : d + \Delta d \in \mathcal{B}\}, \qquad (2.9)$$

see [39, 40, 41], and so $\rho(d)$ is the distance of the data instance $d = (A, b)$ to the set $\mathcal{B}$ of ill-posed instances for the problem (FP_d). It is straightforward to show that

$$\rho(d) = \begin{cases} \inf\{\|d - \bar{d}\| : \bar{d} \in \mathcal{F}^C\} & \text{if } d \in \mathcal{F}, \\ \inf\{\|d - \bar{d}\| : \bar{d} \in \mathcal{F}\} & \text{if } d \in \mathcal{F}^C, \end{cases} \qquad (2.10)$$

so that $\rho(d)$ can be interpreted as the "smallest" perturbation of the data d that causes the system (FP_d) to change its feasibility status . The *condition number* $\mathcal{C}(d)$ of the data instance d is defined to be:

$$\mathcal{C}(d) = \frac{\|d\|}{\rho(d)} \qquad (2.11)$$

when $\rho(d) > 0$, and $\mathcal{C}(d) = \infty$ when $\rho(d) = 0$. The condition number $\mathcal{C}(d)$ can be viewed as a scale-invariant reciprocal of $\rho(d)$, as it is elementary to demonstrate that $\mathcal{C}(d) = \mathcal{C}(\alpha d)$ for any positive scalar α. Observe that since $\tilde{d} = (\tilde{A}, \tilde{b}) = (0, 0) \in \mathcal{B}$, then for any $d \notin \mathcal{B}$ we have $\|d\| = \|d - \tilde{d}\| \geq \rho(d)$, whereby $\mathcal{C}(d) \geq 1$. The value of $\mathcal{C}(d)$ is a measure of the relative conditioning of the data instance d. For further analysis of the distance to ill-posedness,

26

see Filipowski [12, 13], Freund and Vera [16, 14, 15], Nunez [32], Nunez and Freund [33], Peña [34, 35], Peña and Renegar [36], and Vera [54, 55, 57, 56].

2.3 Theorems of the Alternative and Characterizations of the Distance to Ill-Posedness

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

$$C^* = \{z \in X^* : z^t x \geq 0 \text{ for any } x \in C\}. \tag{2.12}$$

The "strong alternative" system of (FP_d) is:

$$\begin{aligned}
(\text{SA}_d) \quad & A^t s \in C_X^* \\
& s \in C_Y^* \\
& b^t s < 0.
\end{aligned} \tag{2.13}$$

A separating hyperplane argument yields the following partial theorem of the alternative regarding the feasibility of the system (FP_d):

Proposition 2.1 *If (SA_d) is feasible, then (FP_d) is infeasible. If (FP_d) is infeasible, then the following "weak alternative" system (2.14) is feasible:*

$$\begin{aligned}
& A^t s \in C_X^* \\
& s \in C_Y^* \\
& b^t s \leq 0 \\
& s \neq 0.
\end{aligned} \tag{2.14}$$

In light of Proposition 2.1, we will often refer to solutions of the system (SA_d) as *certificates of infeasibility* of the system (FP_d).

The following four theorems provide characterizations of $\rho(d)$ for feasible and infeasible instances of (FP_d).

Theorem 2.1 ([41]) *Suppose that $d \in \mathcal{F}$. Then $\rho(d) = r(d)$, where*

$$
\begin{aligned}
r(d) = \quad &\min \quad\;\; \max \quad \phi \\
&v \in Y \quad \theta, x, \phi \\
&\|v\| \le 1 \quad \text{s.t} \quad b\theta - Ax - v\phi \in C_Y \\
&\qquad\qquad\qquad\;\; x \in C_X, \;\; \theta \ge 0 \\
&\qquad\qquad\qquad\;\; \|x\| + |\theta| \le 1.
\end{aligned}
\tag{2.15}
$$

Theorem 2.2 ([16]) *Suppose that $d \in \mathcal{F}$. Then $\rho(d) = j(d)$, where*

$$
\begin{aligned}
j(d) = \quad &\min \quad \max \quad \{\|A^t s - q\|_*, \; |b^t s + g|\} \\
&s, q, g \\
&\qquad \text{s.t.} \quad s \in C_Y'^* \\
&\qquad\qquad\quad q \in C_X^* \\
&\qquad\qquad\quad g \ge 0 \\
&\qquad\qquad\quad \|s\|_* = 1.
\end{aligned}
\tag{2.16}
$$

Theorem 2.3 ([41]) *Suppose that $d \in \mathcal{F}^C$. Then $\rho(d) = \pi(d)$, where*

$$
\begin{aligned}
\pi(d) = \quad &\min \quad\;\; \max \quad \phi \\
&v \in X^* \quad s, \phi \\
&\|v\|_* \le 1 \quad \text{s.t} \quad A^t s - v\phi \in C_X^* \\
&\qquad\qquad\qquad\quad -b^t s - \phi \ge 0 \\
&\qquad\qquad\qquad\quad s \in C_Y^* \\
&\qquad\qquad\qquad\quad \|s\|_* \le 1.
\end{aligned}
\tag{2.17}
$$

Theorem 2.4 ([16]) *Suppose that $d \in \mathcal{F}^C$. Then $\rho(d) = k(d)$, where*

$$k(d) = \quad \min_{\theta, x, y} \quad \|b\theta - Ax - y\|$$

$$\text{s.t.} \quad x \in C_X$$
$$\theta \geq 0 \qquad\qquad\qquad (2.18)$$
$$y \in C_Y$$
$$\|x\| + |\theta| = 1.$$

When the system (FP_d) is well-posed, we have the following strong theorem of the alternative:

Proposition 2.2 *Suppose $\rho(d) > 0$. Then exactly one of the systems (FP_d) and (SA_d) is feasible.*

Proof: The proof is a direct application of Proposition 2.1 and Theorem 2.3. $\blacksquare$

We denote the set of solutions of (SA_d) as A_d, i.e.,

$$A_d = \{s \in Y^* : A^t s \in C_X^*, \ s \in C_Y^*, \ b^t s < 0\}. \qquad (2.19)$$

2.4 Reliable Solutions

When studying the properties of the system (FP_d) and developing algorithms for finding solutions, we will often be interested in so-called *reliable solutions* of the system (FP_d) and *reliable solutions* of the system (SA_d), also referred to as reliable certificates of infeasibility of (FP_d). We consider a solution $\hat{x}$ of the system (FP_d) to be reliable if, roughly speaking, (i) the distance from $\hat{x}$ to the boundary of the feasible region, X_d, $\text{dist}(\hat{x}, \partial X_d)$, is not excessively small, (ii) the norm of the solution $\|\hat{x}\|$ is not excessively large, and (iii) the ratio $\frac{\|\hat{x}\|}{\text{dist}(\hat{x}, \partial X_d)}$ is not excessively large. A reliable solution of the alternative system is defined similarly: we consider a solution $\hat{s}$ of the system (SA_d) to be reliable if the ratio $\frac{\|\hat{s}\|_*}{\text{dist}(\hat{s}, \partial A_d)}$ is not excessively large. (Because the system (2.13) is homogeneous, it makes little sense to

bound $\|\hat{s}\|_*$ from above or to bound $\mathrm{dist}(\hat{s}, \partial A_d)$ from below, as all solutions can be scaled by any positive quantity.)

The importance of reliable solutions can be motivated by considerations of finite-precision computations. Suppose, for example, that a solution $\hat{x}$ of the problem (FP$_d$) (computed as an output of an algorithm involving iterates $x^1, \ldots, x^k = \hat{x}$, and/or used as input to another algorithm) has the property that $\mathrm{dist}(\hat{x}, \partial X_d)$ is very small. Then the numerical precision requirements for checking or guaranteeing feasibility of iterates will necessarily be large. Similar remarks hold for the case when $\|\hat{x}\|$ and/or the ratio $\frac{\|\hat{x}\|}{\mathrm{dist}(\hat{x}, \partial C_X)}$ is very large.

The sense of what is meant by "excessive" can be measured using the condition number $\mathcal{C}(d)$. Freund and Vera [16] show that when the system (FP$_d$) is feasible, there exists a point $\tilde{x} \in X_d$ such that

$$\|\tilde{x}\| \leq c_1 \mathcal{C}(d), \ \mathrm{dist}(\tilde{x}, \partial X_d) \geq c_2 \frac{1}{\mathcal{C}(d)}, \ \text{and} \ \frac{\|\tilde{x}\|}{\mathrm{dist}(\tilde{x}, \partial X_d)} \leq c_3 \mathcal{C}(d), \qquad (2.20)$$

where the scalar quantities c_1, c_2, and c_3 depend only on the width of the cone C_X and/or of the cone C_Y, and are independent of the data d of the problem (FP$_d$), but may depend on the dimensions n and m. (The concept of the width of a cone will be defined shortly.) Similarly, when the system (FP$_d$) is infeasible, they show that there exists a point $\tilde{s} \in A_d$ such that

$$\frac{\|\tilde{s}\|_*}{\mathrm{dist}(\tilde{s}, \partial A_d)} \leq c_4 \mathcal{C}(d), \qquad (2.21)$$

where the scalar quantity c_4 depends only on the width of the cone C_X^* and/or of the cone C_Y^*.

We end this section with the following lemmas which give a precise mathematical characterization of the problem of computing the distance from a given point to the boundary of a given convex set. Let S be a closed convex set in $\Re^m$ and let $f \in \Re^m$ be given. The distance from f to the boundary of S is defined as:

$$r = \min_z \{\|f - z\| : z \in \partial S\}. \qquad (2.22)$$

Lemma 2.1 *Let r be defined by (2.22). Suppose $f \in S$. Then*

$$r = \min_{\substack{v \\ \|v\| \le 1}} \max_{\substack{z \\ s.t. \quad f - z - \phi v = 0 \\ z \in S.}} \phi$$

Lemma 2.2 *Let r be defined by (2.22). Suppose $f \notin S$. Then*

$$r = \min_{\substack{z \\ s.t. \quad z \in S.}} \|f - z\|$$

2.5   Regular Cones

In this section we define *regular cones* and discuss three important parameters associated with regular cones, namely, *width, coefficient of linearity,* and *norm approximation coefficient.*

Definition 2.1 *We will say that a cone C is* regular *if C is a closed convex cone, has a nonempty interior, and is pointed (i.e., contains no line).*

Remark 2.1 *If C is a closed convex cone, then C is regular if and only if C^* is regular.*

Let C be a regular cone in the normed linear vector space X. We will use the following definition of the *width* of C:

Definition 2.2 *If C is a regular cone in the normed linear vector space X, the width of C is given by:*

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

τ_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 τ_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 τ_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 α appropriately, we can find $u \in C$ such that

$$\|u\| = 1 \text{ and } \tau_C \text{ is attained for } (u, \tau_C). \tag{2.23}$$

Closely related to the width is the notion of the *coefficient of linearity* of a regular cone:

Definition 2.3 *If C is a regular cone in the normed linear vector space X, the coefficient of linearity of the cone C is given by:*

$$\beta_C = \begin{array}{cc} \sup & \inf \ q^t x \\ q \in X^* & x \in C \\ \|q\|_* = 1 & \|x\| = 1. \end{array} \tag{2.24}$$

The coefficient of linearity β_C measures the extent to which the norm $\|x\|$ can be approximated by a linear function over the cone C. We have the following properties of β_C:

Remark 2.2 ([16]) $0 < \beta_C \leq 1$. *There exists $\bar{u} \in \text{int} C^*$ such that $\|\bar{u}\|_* = 1$ and $\beta_C = \min\{\bar{u}^t x : x \in C, \ \|x\| = 1\}$. For any $x \in C$, $\beta_C \|x\| \leq \bar{u}^t x \leq \|x\|$. The set $\{x \in C : \bar{u}^t x = 1\}$ is a bounded and closed convex set.*

In light of Remark 2.2 we refer to $\bar{u}$ as the norm linearization vector for the cone C. The following proposition shows that the width of C is equal to the coefficient of linearity of C^*:

Proposition 2.3 ([14]) *Suppose that C is a regular cone in the normed linear vector space X, and let τ_C denote the width of C and let β_{C^*} denote the coefficient of linearity C^*. Then $\tau_C = \beta_{C^*}$. Moreover, τ_C is attained for (u, τ_C), where u is the norm linearization vector for the cone C^*.*

We now pause to illustrate the above notions on two relevant instances of the cone C, namely the nonnegative orthant $\Re_+^n$ and the positive semi-definite cone $S_+^{k \times k}$. We first consider the nonnegative orthant. Let $X = \Re^n$ and $C = \Re_+^n$. Then we can identify X^* with X and in so doing, $C^* = \Re_+^n$ as well. If $\|x\|$ is given by the L_1 norm $\|x\| = \sum_{j=1}^n |x_j|$, then note that $\|x\| = e^t x$ for all $x \in C$ (where e is the vector of ones), whereby the coefficient

of linearity is $\beta_C = 1$ and $\bar{u} = e$. If instead of the L_1 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 $\bar{u} = \left(n^{\left(\frac{1}{p}-1\right)} \right) e$ and the coefficient of linearity is $\beta_C = n^{\left(\frac{1}{p}-1\right)}$. Also, by setting $u = \left(n^{-\frac{1}{p}} \right) e$, it is straightforward to show that the width of C is $\tau_C = n^{-\frac{1}{p}}$.

Now consider the positive semi-definite cone, which has been shown to be of enormous importance in mathematical programming (see Alizadeh [2] and Nesterov and Nemirovskii [31]). 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 ordered 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 [23], for example, for a proof that $\|x\|_p$ is a norm). When $p = 1$, $\|x\|_1$ is the sum of the absolute values of the eigenvalues of x. Therefore, when $x \in C$, $\|x\|_1 = tr(x) = \sum_{i=1}^{k} x_{ii}$ where x_{ij} is the ijth entry of the real matrix x (and $tr(x)$ is the trace of x), and so $\|x\|_1$ is a linear function on C. Therefore, when $p = 1$, we have $\bar{u} = I$ and the coefficient of linearity is $\beta_C = 1$. When $p > 1$, it is easy to show that $\bar{u} = \left(k^{\left(\frac{1}{p}-1\right)} \right) I$ and $\beta_C = k^{\left(\frac{1}{p}-1\right)}$. Also, it is easy to show by setting $u = \left(k^{-\frac{1}{p}} \right) I$ that the width of C is $\tau_C = k^{-\frac{1}{p}}$.

We will now derive the formulas for the coefficient of linearity and the width of a cartesian product of two cones. Suppose that $X = X_1 \times X_2$ (where X_1 and X_2 are normed linear spaces) with norm $\|x\| = \|(x_1, x_2)\| \triangleq \|x_1\| + \|x_2\|$ where $x_1 \in X_1$, $x_2 \in X_2$. Let $C = C_1 \times C_2$, where $C_1 \subset X_1$ and $C_2 \subset X_2$ are regular cones with norm linearization vectors $\bar{u}_1$ and $\bar{u}_2$, respectively. The corresponding coefficients of linearity are denoted by β_1 for C_1 and β_2 for C_2. Suppose further that vectors u_1 and u_2 satisfy $\|u_1\| = 1$, $\|u_2\| = 1$, and $B(u_1, \tau_1) \subset C_1$, and $B(u_2, \tau_2) \subset C_2$, where τ_1 is the width of C_1, and τ_2 is the width of C_2.

Proposition 2.4 *The norm linearization vector for the cone $C = C_1 \times C_2$ is $\bar{u} = (\bar{u}_1, \bar{u}_2)$,*

and the coefficient of linearity is $\beta_C = \min\{\beta_1, \beta_2\}$.

Proof: Observe that $\bar{u} = (\bar{u}_1, \bar{u}_2) \in \text{int}C_1^* \times \text{int}C_2^* = \text{int}C^*$, and using (2.4), $\|\bar{u}\|_* = \max\{\|\bar{u}_1\|_*, \|\bar{u}_2\|_*\} = 1$. Moreover, for any $x = (x_1, x_2) \in C$, we have:

$$\bar{u}^t x = \bar{u}_1^t x_1 + \bar{u}_2^t x_2 \leq \|x_1\| + \|x_2\| = \|x\|, \text{ and}$$

$$\bar{u}^t x = \bar{u}_1^t x_1 + \bar{u}_2^t x_2 \geq \beta_1 \|x_1\| + \beta_2 \|x_2\| \geq \min\{\beta_1, \beta_2\} \|x\|.$$

We conclude that $\beta_C \geq \min\{\beta_1, \beta_2\}$.

It remains to show that $\bar{u}$ above achieves the best approximation of the norm $\|x\|$ over the cone C. We assume without loss of generality, that $\beta_1 \leq \beta_2$.

Suppose that the norm linearization vector for the cone C is $\tilde{u} = (\tilde{u}_1, \tilde{u}_2) \in \text{int}C^*$, and the coefficient of linearity is $\beta_C > \beta_1$. This implies, in particular, that for any $x_1 \in \text{int}C_1$, $x = (x_1, 0) \in C$ and we have

$$\beta_1 \|x_1\| < \beta_C \|x_1\| = \beta_C \|x\| \leq \tilde{u}^t x = \tilde{u}_1^t x_1 \leq \|x_1\|,$$

contradicting the assumption that β_1 is the coefficient of linearity of the cone C_1. ∎

Proposition 2.5 *The width of the cone $C = C_1 \times C_2$ is $\tau_C = \frac{\tau_1 \tau_2}{\tau_1 + \tau_2}$, and is attained for (u, τ_C) where $u = \frac{1}{\tau_1 + \tau_2}(\tau_2 u_1, \tau_1 u_2)$*

Proof: It is easy to verify that u defined above satisfies $u \in C$, $\|u\| = 1$, and $B(u, \tau_C) \subset C$, where $\tau_C = \frac{\tau_1 \tau_2}{\tau_1 + \tau_2}$.

To prove that τ_C is indeed the width of the cone C, suppose that there exists a vector $\tilde{u} \in C$ such that $\|\tilde{u}\| = 1$ and $B(\tilde{u}, \tilde{\tau}) \subset C$, with $\tilde{\tau} > \tau_C$. The vector $\tilde{u}$ can be represented as $\tilde{u} = (\lambda \tilde{u}_1, (1 - \lambda)\tilde{u}_2)$, where $\tilde{u}_1 \in C_1$, $\tilde{u}_2 \in C_2$, $\|u_1\| = \|u_2\| = 1$ and $\lambda \in [0, 1]$. Let $\tilde{\tau}_1$ be the radius of the largest ball centered at $\tilde{u}_1$ inscribed in C_1 and $\tilde{\tau}_2$ be the radius of the largest ball centered at $\tilde{u}_2$ inscribed in C_2.

A simple algebraic argument yields

$$\tau = \min\{\lambda \tilde{\tau}_1, (1 - \lambda)\tilde{\tau}_2\} \leq \frac{\tilde{\tau}_1 \tilde{\tau}_2}{\tilde{\tau}_1 + \tilde{\tau}_2} = \frac{1}{\frac{1}{\tilde{\tau}_2} + \frac{1}{\tilde{\tau}_1}} \leq \frac{1}{\frac{1}{\tau_2} + \frac{1}{\tau_1}} = \tau_C$$

(the last inequality follows since $\tau_1 \geq \tilde{\tau}_1$ and $\tau_2 \geq \tilde{\tau}_2$), resulting in a contradiction. ∎

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

Definition 2.4 *If C is a regular cone in the normed linear space X, define the norm approximation coefficient by*

$$\delta_C \triangleq \mathrm{dist}(0, \partial \mathrm{conv}(C(1), -C(1)), \tag{2.25}$$

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

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

Proposition 2.6 *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 Lemma 2.1, $B(0, \delta_C) \subseteq \mathrm{conv}(C(1), -C(1))$. Therefore,

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

Since $\|Ax\|$ is a convex function of x, the maximum above is attained at an extreme point of the set $\mathrm{conv}(C(1), -C(1))$ (see Theorem 3.4.7 of [3]), and we can assume without loss of generality that it is attained at some point $x \in C(1)$, establishing the proposition. ∎

Corollary 2.1 *Suppose $d = (A, b) \in \mathcal{D}$. Then*

$$\begin{aligned}
\|d\| \leq \quad &\frac{1}{\delta_{C_X}} \max \quad \|b\theta - Ax\| \\
&\text{s.t.} \quad \theta \geq 0, \ x \in C_X \\
&\qquad\quad |\theta| + \|x\| \leq 1.
\end{aligned}$$

35

Proof: The proof follows from Proposition 2.6 by observing that

$$
\max_{\substack{\theta \geq 0, \ x \in C_X \\ |\theta| + \|x\| \leq 1}} \|b\theta - Ax\| = \max\{\|b\|, \ \max_{\substack{x \in C_X \\ \|x\| \leq 1}} \|Ax\| \ \}
$$

∎

The following lemma established a bound on δ_C in terms of τ_C:

Lemma 2.3 *Suppose C is a regular cone with width τ_C. Then*

$$
\delta_C \geq \frac{\tau_C}{1 + \tau_C} \geq \frac{\tau_C}{2}. \tag{2.26}
$$

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 $\bar{x} \in \mathrm{conv}(C(1), -C(1))$.

Let $x = \frac{\bar{x}(1+\tau_C)}{\tau_C}$. If u is the norm approximation vector of C_X^*, then, from Proposition 2.3, $u + \tau_C x \in C$ and $u - \tau_C x \in C$. Furthermore,

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

and so

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

Chapter 3

Elementary Algorithms for Special Types of Conic Linear Systems

In this chapter we develop elementary algorithms for conic linear systems in special form, which will be used in Chapter 4 as "building blocks" for algorithms for general conic linear systems. The outline of this chapter is as follows: Section 3.1 presents a generalization of the von Neumann algorithm (appropriately called algorithm GVNA) that can be applied to conic linear systems in a special compact form (i.e, with a compactness constraint added). We analyze the properties of the iterates of algorithm GVNA under different termination criteria in Lemmas 3.1, 3.2 and 3.3. Sections 3.2 and 3.3 present the development of algorithms HCI (Homogeneous Conic Inequalities) and HCE (Homogeneous Conic Equalities), respectively, for solving two essential types of homogeneous conic linear systems. Both algorithms HCI and HCE consist of calls to algorithm GVNA applied to appropriate transformations of the homogeneous systems at hand.

We make the following assumption throughout this chapter:

Assumption 3.1 $C \subset X$ *is a regular cone. The width* τ_C *of the cone* C *and the coefficient of linearity* β_C *for the cone* C*, together with vectors* $\bar{u}$ *and* u *of Remark 2.2 and Proposition 2.3 are known and given.* Y *is an* m*-dimensional Euclidean space with Euclidean norm* $\|y\| = \|y\|_2 = \sqrt{y^t y}$ *for* $y \in Y$ *(and therefore the dual norm is also* $\|s\|_* = \|s\| = \|s\|_2$ *for* $s \in Y^*$*).*

Suppose C is as in Assumption 3.1. Given any linear function $c^t x$ defined on $x \in X$, we define the following conic section optimization problem:

$$
\begin{aligned}
(\text{CSOP}_C) \quad &\min_{x} \quad c^t x \\
&\text{s.t.} \quad x \in C \\
&\qquad\quad \bar{u}^t x = 1.
\end{aligned}
\tag{3.1}
$$

Let T_C denote an upper bound on the number of operations needed to solve (CSOP_C).

For the algorithms developed in this chapter, we presume that we can work conveniently with the cone C in that we can solve (CSOP_C) easily, i.e., that T_C is not excessive, for otherwise the algorithms will not be very efficient.

We now pause to illustrate how (CSOP_C) is easily solved for two relevant instances of the cone C, namely $\Re^n_+$ and $S^{k \times k}_+$. We first consider $\Re^n_+$. As discussed in Chapter 2, when $\|x\|$ is given by L_p norm with $p \geq 1$, the norm linearization vector $\bar{u}$ is a positive multiple of the vector e. Therefore, for any c, the problem (CSOP_C) is simply the problem of finding the index of the smallest element of the vector c, so that the solution of (CSOP_C) is easily computed as $x_c = e^i$, where $i \in \text{argmin}\{c_j : j = 1, \ldots, n\}$. Thus $T_C = n$.

We now consider $S^{k \times k}_+$. Recall that for $x \in S^{k \times k}$, $\lambda(x)$ is a k-vector of ordered eigenvalues of x. As discussed in Chapter 2, when $\|x\|$ is given by

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

with $p \geq 1$, the norm linearization vector $\bar{u}$ is a positive multiple of the matrix I. For any $c \in S^{k \times k}$, the problem (CSOP_C) corresponds to the problem of finding the normalized eigenvector corresponding to the smallest eigenvalue of the matrix c, i.e., (CSOP_C) is a minimum eigenvalue problem and is solvable to within machine tolerance in $O(k^3)$ operations in practice (though not in theory).

Solving (CSOP_C) for the cartesian product of two cones is easy if (CSOP_C) is easy to solve for each of the two cones: suppose that $X = X_1 \times X_2$ with norm $\|x\| = \|(x_1, x_2)\| \triangleq$

$\|x_1\| + \|x_2\|$, and $C = C_1 \times C_2$ where $C_1 \subset X_1$ and $C_2 \subset X_2$ are regular cones with norm linearization vectors $\bar{u}_1$ and $\bar{u}_2$, respectively. Then the norm linearization vector for the cone C is $\bar{u} = (\bar{u}_1, \bar{u}_2)$, and $T_C = T_{C_1} + T_{C_2} + O(1)$.

3.1 A Generalized Von Neumann Algorithm for a Conic Linear System in Compact Form

In this section we consider a generalization of the algorithm of von Neumann studied by Dantzig in [7] and [8], see also [10, 11]. We will work with a conic linear system of the form:

$$
\begin{aligned}
\text{(P)} \quad Mx &= g \\
x &\in C \\
\bar{u}^t x &= 1,
\end{aligned}
\tag{3.2}
$$

where $C \subset X$ is a closed convex cone in the (finite) n-dimensional normed linear vector space X, and $g \in Y$ where Y is the (finite) m-dimensional linear vector space with Euclidean norm $\|y\| = \|y\|_2$, and $M \in L(X, Y)$. Assumption 3.1 is presumed valid; in particular, the norm linearization vector $\bar{u}$ of Remark 2.2 is presumed to be known and given. (The original algorithm of von Neumann presented and analyzed by Dantzig in [7] and [8] was developed for the case when $C = \Re_+^n$ and $\bar{u} = e$.) We will refer to a system of the form (3.2) as a conic linear system in compact form, or simply a compact-form system.

The "alternative" system to (P) of (3.2) is:

$$
\text{(A)} \quad M^t s - \bar{u}(g^t s) \in \text{int} C^*,
\tag{3.3}
$$

and a generalization of Farkas' Lemma yields the following duality result:

Proposition 3.1 *Exactly one of the systems (P) of (3.2) and (A) of (3.3) has a solution.*

Notice that the feasibility problem (P) is equivalent to the following optimization prob-

lem:

$$\text{(OP)} \quad \min_{x} \quad \|g - Mx\|$$

$$\text{s.t.} \quad x \in C$$

$$\bar{u}^t x = 1.$$

If (P) has a feasible solution, the optimal value of (OP) is 0; otherwise, the optimal value of (OP) is strictly positive. We will say that a point x is "admissible" if it is a feasible point for (OP), i.e., $x \in C$ and $\bar{u}^t x = 1$.

We now describe a generic iteration of our algorithm. At the beginning of the iteration we have an admissible point $\bar{x}$. Let $\bar{v}$ be the "residual" at the point $\bar{x}$, namely, $\bar{v} = g - M\bar{x}$. Notice that $\|\bar{v}\| = \|g - M\bar{x}\|$ is the objective value of (OP). The algorithm calls an oracle to solve the following instance of the conic section optimization problem (CSOP$_C$) of (3.1):

$$\begin{array}{ll}
\min_{p} \quad \bar{v}^t(g - Mp) = & \min_{p} \quad \bar{v}^t(g\bar{u}^t - M)p \\[1em]
\text{s.t.} \quad p \in C & \text{s.t.} \quad p \in C \\
\qquad \bar{u}^t p = 1 & \qquad \bar{u}^t p = 1,
\end{array} \tag{3.4}$$

where (3.4) is an instance of the (CSOP$_C$) with $c = (-M^t + \bar{u}g^t)\bar{v}$. Let $\bar{p}$ be an optimal solution to the problem (3.4), and $\bar{w} = g - M\bar{p}$.

Next, the algorithm checks whether the termination criterion is satisfied. The termination criterion for the algorithm is given in the form of a function STOP($\cdot$), which evaluates to 1 exactly when its inputs satisfy some termination criterion (some relevant examples are presented after the statement of the algorithm). If STOP($\cdot$) = 1, the algorithm concludes that the appropriate termination criterion is satisfied and stops.

On the other hand, if STOP($\cdot$) = 0, the algorithm continues the iteration. The direction $\bar{p} - \bar{x}$ turns out to be a direction of potential improvement of the objective function of (OP). The algorithm takes a step in the direction $\bar{p} - \bar{x}$ with step-size found by constrained line-search. In particular, let

$$\tilde{x}(\lambda) = \bar{x} + \lambda(\bar{p} - \bar{x}).$$

Then the next iterate $\tilde{x}$ is computed as $\tilde{x} = \tilde{x}(\lambda^*)$, where

$$\lambda^* = \text{argmin}_{\lambda \in [0,1]} \|g - M\tilde{x}(\lambda)\|$$

$$= \text{argmin}_{\lambda \in [0,1]} \|g - M(\bar{x} + \lambda(\bar{p} - \bar{x}))\| = \text{argmin}_{\lambda \in [0,1]} \|(1 - \lambda)\bar{v} + \lambda\bar{w}\|.$$

Notice that $\tilde{x}$ is a convex combination of the two admissible points $\bar{x}$ and $\bar{p}$ and therefore $\tilde{x}$ is also admissible. Also, λ^* above can be computed as the solution of the following simple constrained convex quadratic minimization problem:

$$\min_{\lambda \in [0,1]} \|(1 - \lambda)\bar{v} + \lambda\bar{w}\|^2 = \min_{\lambda \in [0,1]} \lambda^2 \|\bar{v} - \bar{w}\|^2 + 2\lambda(\bar{v}^t(\bar{w} - \bar{v})) + \|\bar{v}\|^2. \qquad (3.5)$$

The solution of the program (3.5) is easily seen to be

$$\lambda^* = \begin{cases} 1 & \text{if } \|\bar{w}\|^2 \leq \bar{w}^t\bar{v}, \\ \frac{\bar{v}^t(\bar{v} - \bar{w})}{\|\bar{v} - \bar{w}\|^2} & \text{otherwise.} \end{cases} \qquad (3.6)$$

The formal description of the algorithm is as follows:

Algorithm GVNA

- *Data:* (M, g, x^0) (where x^0 is an arbitrary admissible starting point).

- *Initialization:* The algorithm is initialized with x^0.

- *Iteration k, $k \geq 1$:* At the start of the iteration we have an admissible point x^{k-1} : $x^{k-1} \in C$, $\bar{u}^t x^{k-1} = 1$.

  Step 1 Compute $v^{k-1} = g - Mx^{k-1}$ (the residual).

  Step 2 Solve the following conic section optimization problem (CSOP$_C$):

$$\begin{array}{llll} \min & (v^{k-1})^t(g - Mp) = & \min & (v^{k-1})^t(g\bar{u}^t - M)p \\ p & & p & \\ \text{s.t.} & p \in C & \text{s.t.} & p \in C \\ & \bar{u}^t p = 1 & & \bar{u}^t p = 1. \end{array} \qquad (3.7)$$

Let p^{k-1} be an optimal solution of the optimization problem (3.7) and $w^{k-1} = g - Mp^{k-1}$. Evaluate STOP($\cdot$). If STOP($\cdot$) = 1, stop, return appropriate output.

Step 3 Else, let

$$\lambda^{k-1} = \mathrm{argmin}_{\lambda \in [0,1]}\{\|g - M(x^{k-1} + \lambda(p^{k-1} - x^{k-1}))\|\} \tag{3.8}$$

$$
\begin{cases}
1 & \text{if } \|w^{k-1}\|^2 \leq (w^{k-1})^t v^{k-1}, \\
\frac{(v^{k-1})^t(v^{k-1}-w^{k-1})}{\|v^{k-1}-w^{k-1}\|^2} & \text{otherwise.}
\end{cases}
$$

and

$$x^k = x^{k-1} + \lambda^{k-1}(p^{k-1} - x^{k-1}).$$

Step 4 Let $k \leftarrow k+1$, go to Step 1.

Note that the above description is rather generic; to apply the algorithm we have to specify the function STOP($\cdot$) to be used in Step 2. Some examples of functions STOP($\cdot$) that will be used in this and the following chapters are:

1. STOP1(v^{k-1}, w^{k-1}) = 1 if and only if $(v^{k-1})^t w^{k-1} > 0$. If the vectors v^{k-1}, w^{k-1} satisfy termination criterion STOP1, then the vector $s = -\frac{v^{k-1}}{\|v^{k-1}\|}$ is a solution to the alternative system (A), see Proposition 3.2 below. Therefore, algorithm GVNA with STOP = STOP1 will terminate only if the system (P) is infeasible.

2. STOP2(v^{k-1}, w^{k-1}) = 1 if and only if $(v^{k-1})^t w^{k-1} > \frac{\|v^{k-1}\|^2}{2}$. This termination criterion is a stronger version of STOP1.

3. STOP3(v^{k-1}, w^{k-1}, k) = 1 if and only if $(v^{k-1})^t w^{k-1} > 0$ or $k \geq I$, where I is some pre-specified integer. This termination criterion is essentially equivalent to STOP1, but it ensures finite termination (in no more that I iterations) regardless of the status of (P).

Proposition 3.2 *Suppose v^{k-1} and w^{k-1} are as defined in Steps 1 and 2 of algorithm GVNA. If $(v^{k-1})^t w^{k-1} > 0$, then the vector $s = -\frac{v^{k-1}}{\|v^{k-1}\|}$ is a solution to the alternative system(A) and so (P) is infeasible.*

Proof: By definition of w^{k-1},

$$0 < (v^{k-1})^t w^{k-1} = (v^{k-1})^t (g\bar{u}^t - M)p^{k-1} \le (v^{k-1})^t (g\bar{u}^t - M)p$$

for any $p \in C$, $\bar{u}^t p = 1$. Hence, $(g\bar{u}^t - M)^t v^{k-1} \in \text{int}C^*$ and $s = -\frac{v^{k-1}}{\|v^{k-1}\|}$ is a solution of (A). ∎

Analogous to the von Neumann algorithm of [7] and [8], we regard algorithm GVNA as "elementary" in that the algorithm does not rely on particularly sophisticated mathematics at each iteration (each iteration must perform a few matrix-vector and vector-vector multiplications and solve an instance of (CSOP_C)). Furthermore the work per iteration will be low so long as T_C (the number of operations needed to solve (CSOP_C)) is small. A thorough evaluation of the work per iteration of algorithm GVNA is presented in Remark 3.1 at the end of this section.

As was mentioned in the discussion preceding the statement of the algorithm, the size of the residual $\|v^k\|$ is decreased at each iteration. The rate of decrease depends on the termination criterion used and on the status of the system (P). In the rest of this section we present three lemmas that provide upper bounds on the size of the residual throughout the algorithm. The first result is a generalization of Dantzig's convergence result [7].

Lemma 3.1 (Dantzig [7]) *If algorithm GVNA with* $\text{STOP} = \text{STOP1}$ *(or* $\text{STOP} = \text{STOP3}$*) has performed* k *(complete) iterations, then*

$$\|v^k\| \le \frac{\|M - g\bar{u}^t\|}{\beta_C \sqrt{k}}. \tag{3.9}$$

Proof: First note that if x is any admissible point (i.e., $x \in C$ and $\bar{u}^t x = 1$), then $\|x\| \le \frac{\bar{u}^t x}{\beta_C} = \frac{1}{\beta_C}$, and so

$$\|g - Mx\| = \|(g\bar{u}^t - M)x\| \le \|M - g\bar{u}^t\| \cdot \|x\| \le \frac{\|M - g\bar{u}^t\|}{\beta_C}. \tag{3.10}$$

From the discussion preceding the formal statement of the algorithm, all iterates of the algorithm are admissible, so that $x^k \in C$ and $\bar{u}^t x^k = 1$ for all k. We prove the bound on the

43

norm of the residual by induction on k.

For $k = 1$,

$$||v^1|| = ||g - Mx^1|| \leq \frac{||M - g\bar{u}^t||}{\beta_C} = \frac{||M - g\bar{u}^t||}{\beta_C\sqrt{1}},$$

where the inequality above derives from (3.10).

Next suppose by induction that $||v^{k-1}|| \leq \frac{||M-g\bar{u}^t||}{\beta_C\sqrt{k-1}}$. At the end of iteration k we have

$$||v^k|| = ||g - Mx^k|| = ||(1 - \lambda^{k-1})(g - Mx^{k-1}) + \lambda^{k-1}(g - Mp^{k-1})||$$

$$\tag{3.11}$$

$$= ||(1 - \lambda^{k-1})v^{k-1} + \lambda^{k-1}w^{k-1}||,$$

where p^{k-1} and w^{k-1} were computed in Step 2. Recall that λ^{k-1} was defined in Step 3 as the minimizer of $||(1 - \lambda)v^{k-1} + \lambda w^{k-1}||$ over all $\lambda \in [0, 1]$. Therefore, in order to obtain an upper bound on $||v^k||$, we can substitute any $\lambda \in [0, 1]$ into (3.11). We will substitute $\lambda = \frac{1}{k}$. Making this substitution, we obtain:

$$||v^k|| \leq \left|\left|\frac{k-1}{k}v^{k-1} + \frac{1}{k}w^{k-1}\right|\right| = \frac{1}{k}||(k-1)v^{k-1} + w^{k-1}||. \tag{3.12}$$

Squaring (3.12) yields:

$$||v^k||^2 \leq \frac{1}{k^2}\left((k-1)^2||v^{k-1}||^2 + ||w^{k-1}||^2 + 2(k-1)(v^{k-1})^t(w^{k-1})\right). \tag{3.13}$$

Since the algorithm did not terminate at Step 2, the termination criterion was not met, i.e., in the case STOP = STOP1 (or STOP = STOP3), $(v^{k-1})^t w^{k-1} \leq 0$. Also, since p^{k-1} is admissible, $||w^{k-1}|| = ||g - Mp^{k-1}|| \leq \frac{||M-g\bar{u}^t||}{\beta_C}$. Combining these results with the inductive bound on $||v^{k-1}||$ and substituting into (3.13) above yields

$$||v^k||^2 \leq \frac{1}{k^2}\left((k-1)^2\frac{||M - g\bar{u}^t||^2}{\beta_C^2(k-1)} + \frac{||M - g\bar{u}^t||^2}{\beta_C^2}\right) = \frac{1}{k} \cdot \frac{||M - g\bar{u}^t||^2}{\beta_C^2}. \quad \blacksquare$$

We now develop another line of analysis of the algorithm, which will be used when the

problem (P) is "well-posed." Let

$$\mathcal{H}_M \triangleq \{Mx : x \in C, \ \bar{u}^t x = 1\}, \tag{3.14}$$

and notice that (P) is feasible precisely when $g \in \mathcal{H}_M$. Define

$$r(M, g) \triangleq \inf\{\|g - h\| : h \in \partial \mathcal{H}_M\}, \tag{3.15}$$

where $\mathcal{H}_M$ is defined above in (3.14). As it turns out, the quantity $r(M, g)$ plays a crucial role in analyzing the complexity of algorithm GVNA.

Observe that $r(M, g) = 0$ precisely when the vector g is on the boundary of the set $\mathcal{H}_M$. Thus, when $r(M, g) = 0$, the problem (P) has a feasible solution, but arbitrarily small changes in the data (M, g) can yield instances of (P) that have no feasible solution. Therefore when $r(M, g) = 0$ we can rightfully call the problem (P) unstable, or in the language of data perturbation and condition numbers, the problem (P) is "ill-posed." We will refer to the system (P) as being "well-posed" when $r(M, g) > 0$.

The following proposition gives a useful characterization of the value of $r(M, g)$.

Proposition 3.3 *Let $\mathcal{H}_M$ and $r(M, g)$ be defined as in (3.14) and (3.15). If (P) has a feasible solution, then*

$$
r(M, g) = \min_{v} \max_{h} \phi = \min_{v} \max_{x} \phi
$$
$$
\|v\| \le 1 \quad s.t. \quad g - h - \phi v = 0 \qquad \|v\| \le 1 \quad s.t. \quad g - Mx - \phi v = 0
$$
$$
h \in \mathcal{H}_M \qquad\qquad\qquad\qquad x \in C
$$
$$
\bar{u}^t x = 1.
$$

$$\tag{3.16}$$

If (P) does not have a feasible solution, then

$$
\begin{aligned}
r(M,g) = & \min_{h} \quad \|g - h\| \quad = \min_{x} \quad \|g - Mx\| \\
& \text{s.t.} \quad h \in \mathcal{H}_M \qquad \text{s.t.} \quad x \in C \\
& \qquad\qquad\qquad\qquad\qquad\quad \bar{u}^t x = 1.
\end{aligned}
\tag{3.17}
$$

Proof: The proof is a straightforward application of Lemmas 2.1 and 2.2. ∎

In light of Proposition 3.3, when (P) has a feasible solution, $r(M,g)$ can be interpreted as the radius of the largest ball centered at g and contained in the set $\mathcal{H}_M$.

We now present an analysis of the performance of algorithm GVNA in terms of the quantity $r(M,g)$.

Proposition 3.4 *Suppose that (P) has a feasible solution. Let v^k be the residual at point x^k, and let p^k be the direction found in Step 2 of the algorithm at iteration $k+1$. Then $(v^k)^t(g - Mp^k) + r(M,g)\|v^k\| \leq 0$.*

Proof: If $v^k = 0$, the result follows trivially. Suppose $v^k \neq 0$. By definition of $r(M,g)$, there exists a point $h \in \mathcal{H}_M$ such that $g - h + r(M,g)\frac{v^k}{\|v^k\|} = 0$. By the definition of $\mathcal{H}_M$, $h = Mx$ for some admissible point x. It follows that

$$
g - Mx = -r(M,g)\frac{v^k}{\|v^k\|}.
$$

Recall that $p^k \in \operatorname{argmin}_p\{(v^k)^t(g - Mp) : p \in C, \ \bar{u}^t p = 1\}$. Therefore,

$$
(v^k)^t(g - Mp^k) \leq (v^k)^t(g - Mx) = -(v^k)^t r(M,g)\frac{v^k}{\|v^k\|} = -r(M,g)\|v^k\|,
$$

so

$$
(v^k)^t(g - Mp^k) + r(M,g)\|v^k\| \leq 0. \quad \blacksquare
$$

Proposition 3.4 is used to prove the following linear convergence rate for algorithm GVNA:

46

Lemma 3.2 *Suppose the system (P) is feasible, and that $r(M, g) > 0$. If GVNA with* STOP = STOP1 *(or* STOP = STOP3*) has performed k (complete) iterations, then*

$$||v^k|| \leq ||v^0|| e^{-\frac{k}{2}\left(\frac{\beta_C r(M,g)}{||M - g\bar{u}^t||}\right)^2}. \tag{3.18}$$

Proof: Let $\bar{x}$ be the current iterate of GVNA. Furthermore, let $\bar{v} = g - M\bar{x}$ be the residual at the point $\bar{x}$, $\bar{p}$ be the solution of the problem (CSOP$_C$), and $\bar{w} = g - M\bar{p}$. Suppose that the algorithm has not terminated at the current iteration, and $\tilde{x} = \bar{x} + \lambda^*(\bar{p} - \bar{x})$ is the next iterate and $\tilde{v}$ is the residual at $\tilde{x}$. Then

$$||\tilde{v}||^2 = ||(1 - \lambda^*)\bar{v} + \lambda^*\bar{w}||^2 = (\lambda^*)^2||\bar{v} - \bar{w}||^2 + 2\lambda^*\bar{v}^t(\bar{w} - \bar{v}) + ||\bar{v}||^2, \tag{3.19}$$

where λ^* is given by (3.6). Since the algorithm has not terminated at Step 2, the termination criterion has not been satisfied, i.e., in the case of STOP = STOP1 (or STOP = STOP3), $\bar{v}^t\bar{w} \leq 0$. Consider two cases:

Case 1: $||\bar{w}||^2 > \bar{w}^t\bar{v}$. In this case $\lambda^* = \frac{\bar{v}^t(\bar{v} - \bar{w})}{||\bar{v} - \bar{w}||^2}$. Substituting this value of λ^* into (3.19) yields:

$$||\tilde{v}||^2 = \frac{||\bar{v}||^2||\bar{w}||^2 - (\bar{v}^t\bar{w})^2}{||\bar{v} - \bar{w}||^2}. \tag{3.20}$$

Recall from Proposition 3.4 that $\bar{v}^t\bar{w} \leq -r(M, g)||\bar{v}||$. Thus, $||\bar{v}||^2(||\bar{w}||^2 - r(M, g)^2)$ is an upper bound on the numerator of (3.20). Also, $||\bar{v} - \bar{w}||^2 = ||\bar{v}||^2 + ||\bar{w}||^2 - 2\bar{v}^t\bar{w} \geq ||\bar{w}||^2$. Substituting this into (3.20) yields

$$||\tilde{v}||^2 \leq \frac{||\bar{v}||^2(||\bar{w}||^2 - r(M, g)^2)}{||\bar{w}||^2} = \left(1 - \frac{r(M, g)^2}{||\bar{w}||^2}\right)||\bar{v}||^2$$

$$\leq \left(1 - \left(\frac{\beta_C r(M, g)}{||g\bar{u}^t - M||}\right)^2\right)||\bar{v}||^2$$

where the last inequality derives from (3.10). Applying the inequality $1 - t \leq e^{-t}$ for $t = \left(\frac{\beta_C r(M,g)}{||g\bar{u}^t - M||}\right)^2$, we obtain:

$$||\tilde{v}||^2 \leq ||\bar{v}||^2 e^{-\left(\frac{\beta_C r(M,g)}{||g\bar{u}^t - M||}\right)^2}.$$

47

<u>Case 2:</u> $\|\bar{w}\|^2 \leq \bar{w}^t\bar{v}$. In this case $\lambda^* = 1$. Also, since the termination criterion was not met, we conclude that $\bar{w} = 0$, and therefore $\tilde{v} = 0$.

Combining Case 1 and Case 2, we have

$$\|\tilde{v}\|^2 \leq \|\bar{v}\|^2 e^{-\left(\frac{\beta_C r(M,g)}{\|g\bar{u}^t - M\|}\right)^2},$$

or, substituting $\bar{v} = v^{k-1}$ and $\tilde{v} = v^k$,

$$\|v^k\| \leq \|v^{k-1}\| e^{-\frac{1}{2}\left(\frac{\beta_C r(M,g)}{\|g\bar{u}^t - M\|}\right)^2}. \tag{3.21}$$

Applying (3.21) inductively, we can bound the size of the residual $\|v^k\|$ by

$$\|v^k\| \leq \|v^0\| e^{-\frac{k}{2}\left(\frac{\beta_C r(M,g)}{\|g\bar{u}^t - M\|}\right)^2}. \ \blacksquare$$

We now establish a bound on the size of the residual for STOP = STOP2.

Lemma 3.3 *If GVNA with* STOP = STOP2 *has performed k (complete) iterations, then*

$$\|v^k\| \leq \frac{4\|M - g\bar{u}^t\|}{\beta_C\sqrt{k}}.$$

Proof: Let $\bar{x}$ be the current iterate of GVNA. Furthermore, let $\bar{v} = g - M\bar{x}$ be the residual at the point $\bar{x}$, $\bar{p}$ be the solution of the problem (CSOP$_C$) and $\bar{w} = g - M\bar{p}$. Suppose that the algorithm has not terminated at the current iteration, and $\tilde{x} = \bar{x} + \lambda^*(\bar{p} - \bar{x})$ is the next iterate and $\tilde{v}$ is the residual at $\tilde{x}$. Then

$$\|\tilde{v}\|^2 = \|(1 - \lambda^*)\bar{v} + \lambda^*\bar{w}\|^2 = (\lambda^*)^2\|\bar{v} - \bar{w}\|^2 + 2\lambda^*\bar{v}^t(\bar{w} - \bar{v}) + \|\bar{v}\|^2, \tag{3.22}$$

where λ^* is given by (3.6). Consider two cases:

<u>Case 1:</u> $\|\bar{w}\|^2 \leq \bar{w}^t\bar{v}$. In this case $\lambda^* = 1$. Substituting this value of λ^* into (3.22), algebraic manipulations yield

$$\|\tilde{v}\|^2 = \|\bar{w}\|^2 \leq \bar{w}^t\bar{v} \leq \frac{\|\bar{v}\|^2}{2} = \|\bar{v}\|^2 - \frac{\|\bar{v}\|^2}{2} \leq \|\bar{v}\|^2 - \frac{\|\bar{v}\|^4\beta_C^2}{16\|M - g\bar{u}^t\|^2}. \tag{3.23}$$

48

The second inequality in (3.23) follows from the assumption that the algorithm did not terminate at the present iteration. This implies that the termination criterion was not met, i.e., $\bar{v}^t \bar{w} \leq \frac{\|\bar{v}\|^2}{2}$. The last inequality follows since

$$\|\bar{v}\|^2 \leq \frac{\|M - g\bar{u}^t\|^2}{\beta_C^2} \leq \frac{8\|M - g\bar{u}^t\|^2}{\beta_C^2}.$$

The need for the last inequality may not be immediately clear at this stage, but will become more apparent later in this proof.

$\underline{\text{Case 2: }} \|\bar{w}\|^2 > \bar{w}^t \bar{v}$. In this case $\lambda^* = \frac{\bar{v}^t(\bar{v}-\bar{w})}{\|\bar{v}-\bar{w}\|^2}$. Substituting this value of λ^* into (3.22) yields:

$$\|\tilde{v}\|^2 = \|\bar{v}\|^2 - \frac{(\bar{v}^t(\bar{v}-\bar{w}))^2}{\|\bar{w}-\bar{v}\|^2}.$$

Since $\bar{v}^t \bar{w} \leq \frac{\|\bar{v}\|^2}{2}$, we have:

$$\bar{v}^t(\bar{v}-\bar{w}) \geq \frac{\|\bar{v}\|^2}{2},$$

so that

$$\|\tilde{v}\|^2 \leq \|\bar{v}\|^2 - \frac{\|\bar{v}\|^4}{4\|\bar{w}-\bar{v}\|^2} \leq \|\bar{v}\|^2 - \frac{\|\bar{v}\|^4 \beta_C^2}{16\|M - g\bar{u}^t\|^2},$$

since

$$\|\bar{w}-\bar{v}\|^2 \leq \|\bar{v}\|^2 + \|\bar{w}\|^2 + 2\|\bar{v}\| \cdot \|\bar{w}\| \leq \frac{4\|M - g\bar{u}^t\|^2}{\beta_C^2}.$$

Combining Case 1 and Case 2, we conclude that

$$\|\tilde{v}\|^2 \leq \|\bar{v}\|^2 - \frac{\|\bar{v}\|^4}{\gamma^2}, \quad \text{where } \gamma \triangleq \frac{4\|M - g\bar{u}^t\|}{\beta_C}. \tag{3.24}$$

Next, we establish (using induction) the following relation, from which the statement of the lemma will follow: if the algorithm has performed k (complete) iterations, then

$$\|v^k\|^2 \leq \frac{\gamma^2}{k}. \tag{3.25}$$

First, note that $\|v^1\|^2 \leq \frac{\|M - g\bar{u}^t\|^2}{\beta_C^2} \leq \frac{\gamma^2}{1}$, thus establishing (3.25) for $k = 1$. Suppose that (3.25) holds for $k \geq 1$. If $v^{k+1} = 0$, then (3.25) is trivially valid for v^{k+1}. Also, if $v^k = 0$, then it is easy to show that $v^{k+1} = 0$. We will therefore consider the case when v^k and v^{k+1} are

49

both non-zero vectors. Using the relationship for $\tilde{v}$ and $\bar{v}$ established above with $\tilde{v} = v^{k+1}$ and $\bar{v} = v^k$, we have:

$$\|v^{k+1}\|^2 \leq \|v^k\|^2 - \frac{\|v^k\|^4}{\gamma^2},$$

or, dividing by $\|v^{k+1}\|^2 \cdot \|v^k\|^2$,

$$\frac{1}{\|v^k\|^2} \leq \frac{1}{\|v^{k+1}\|^2} - \frac{\|v^k\|^2}{\|v^{k+1}\|^2 \gamma^2} \leq \frac{1}{\|v^{k+1}\|^2} - \frac{1}{\gamma^2}.$$

Therefore,

$$\frac{1}{\|v^{k+1}\|^2} \geq \frac{1}{\|v^k\|^2} + \frac{1}{\gamma^2} \geq \frac{k}{\gamma^2} + \frac{1}{\gamma^2},$$

and so

$$\|v^{k+1}\|^2 \leq \frac{\gamma^2}{k+1},$$

thus establishing the relation (3.25), which completes the proof of the lemma. ∎

To complete the analysis of algorithm GVNA, we now discuss the computational work performed per iteration. We have the following remark:

Remark 3.1 *Each iteration of algorithm GVNA requires at most*

$$T_C + O(mn)$$

operations, where T_C is the number of operations needed to solve an instance of (CSOP$_C$). The term $O(mn)$ derives from counting the matrix-vector and vector-vector multiplications. The number of operations required to perform these multiplications can be significantly reduced if M and g are sparse.

3.2 Algorithm HCI

In this section, we develop algorithm HCI (for Homogeneous Conic Inequalities) and analyze its complexity and the properties of solutions it generates. Algorithm HCI is designed

to obtain a solution of the problem

$$\text{(HCI)} \quad M^t s \in \text{int} C^*. \tag{3.26}$$

We will assume for the rest of this subsection that the system (HCI) of (3.26) is feasible. We denote the set of solutions of (HCI) by S_M, i.e.,

$$S_M \triangleq \{s : M^t s \in \text{int} C^*\}.$$

The solution s returned by algorithm HCI is "sufficiently interior" in the sense that the ratio $\frac{\|s\|}{\text{dist}(s, \partial S_M)}$ is not excessively large. (The notion of sufficiently interior solutions is very similar to the notion of reliable solutions. However, we wish to reserve the appellation "reliable" for solutions and certificates of infeasibility of the system (FP_d).)

Observe that the system (HCI) of (3.26) is of the form (3.3) (with $g = 0$). (HCI) is the "alternative" system (in the sense of Proposition 3.1) for the following problem:

$$\begin{aligned} \text{(PHCI)} \quad & Mx = 0 \\ & x \in C \\ & \bar{u}^t x = 1, \end{aligned} \tag{3.27}$$

which is a system of the form (3.2). Following (3.15) we define

$$r(M) \triangleq r(M, 0) = \inf\{\|h\| : h \in \partial \mathcal{H}_M\}, \tag{3.28}$$

where, as in (3.14), $\mathcal{H}_M \triangleq \{Mx : x \in C, \ \bar{u}^t x = 1\}$. Combining Proposition 3.3 and a separating hyperplane argument, we easily have the following result:

Proposition 3.5 *Suppose (HCI) of (3.26) is feasible. Then (PHCI) of (3.27) is infeasible and $r(M) = \min\{\|Mx\| : \ x \in C, \ \bar{u}^t x = 1\}$. Furthermore, $r(M) > 0$.*

Algorithm HCI, described below, consists of a single application of algorithm GVNA to the system (PHCI) and returns as output a sufficiently interior solution of the system (HCI).

Algorithm HCI

51

- *Data: M*

- Run algorithm GVNA with STOP = STOP2 on the data set $(M, 0, x^0)$ (where x^0 is an arbitrary admissible starting point). Let $\bar{v}$ be the residual at the last iteration of algorithm GVNA.

- Define $s \triangleq -\frac{\bar{v}}{\|\bar{v}\|}$. Return s.

The following theorem presents an analysis of the iteration complexity of algorithm HCI, and shows that the output s of HCI is a sufficiently interior solution of the system (HCI).

Theorem 3.1 *Suppose (HCI) is feasible. Algorithm HCI will terminate in at most*

$$\left\lfloor \frac{16\|M\|^2}{\beta_C^2 r(M)^2} \right\rfloor \tag{3.29}$$

iterations of algorithm GVNA.

Let s be the output of algorithm HCI. Then $s \in S_M$ and

$$\frac{\|s\|}{\operatorname{dist}(s, \partial S_M)} \leq \frac{2\|M\|}{\beta_C r(M)}. \tag{3.30}$$

Proof: Suppose that algorithm GVNA (called in algorithm HCI) has completed k iterations. From Lemma 3.3 we conclude that

$$\|v^k\| \leq \frac{4\|M\|}{\beta_C \sqrt{k}},$$

where $v^k = -Mx^k$ is the residual after k iterations. From Proposition 3.5, $r(M) \leq \|Mx\|$ for any admissible point x. Therefore,

$$r(M) \leq \|v^k\| \leq \frac{4\|M\|}{\beta_C \sqrt{k}}.$$

Rearranging yields

$$k \leq \frac{16\|M\|^2}{\beta_C^2 r(M)^2},$$

from which the first part of the theorem follows.

Next, observe that $\|s\| = 1$. Therefore, to establish the second part of the theorem, we need to show that $\mathrm{dist}(s, \partial S_M) \geq \frac{\beta_C r(M)}{2\|M\|}$. Equivalently, we need to show that for any $q \in Y^*$ such that $\|q\| \leq 1$, $M^t\left(s + \frac{\beta_C r(M)}{2\|M\|}q\right) \in C^*$. Let p be an arbitrary vector satisfying $p \in C$, $\bar{u}^t p = 1$. Then

$$\left(M^t\left(s + \frac{\beta_C r(M)}{2\|M\|}q\right)\right)^t p = s^t M p + \frac{\beta_C r(M)}{2\|M\|}q^t M p. \tag{3.31}$$

Observe that by definition of s

$$s^t M p = \frac{-\bar{v}^t M p}{\|\bar{v}\|} \geq \frac{\bar{v}^t w^{k-1}}{\|\bar{v}\|} > \frac{\|\bar{v}\|}{2},$$

where $\bar{v} = v^{k-1}$ is the residual at the last iteration of algorithm GVNA. (The first inequality follows since p is an admissible point, and the second inequality follows from the fact that the termination criterion of STOP2 is satisfied at the last iteration.) On the other hand,

$$\frac{\beta_C r(M)}{2\|M\|}q^t M p \geq -\frac{\beta_C r(M)}{2\|M\|}\|q\| \cdot \|M\| \cdot \|p\| \geq -\frac{r(M)}{2}.$$

Substituting the above two bounds into (3.31), we conclude that

$$\left(M^t\left(s + \frac{\beta_C r(M)}{2\|M\|}q\right)\right)^t p > \frac{\|\bar{v}\|}{2} - \frac{r(M)}{2} \geq 0.$$

Therefore, $M^t\left(s + \frac{\beta_C r(M)}{2\|M\|}q\right) \in C^*$ and hence $\mathrm{dist}(s, \partial S_M) \geq \frac{\beta_C r(M)}{2\|M\|}$, proving the last statement of the theorem. ∎

The following corollary presents a slight modification of Theorem 3.1 which will be useful in further analysis.

Corollary 3.1 *Suppose (HCI) is feasible. Let s be the output of algorithm HCI. Then $\|s\| = 1$ and $B\left(M^t s, \frac{\beta_C r(M)}{2}\right) \subset C^*$.*

3.3 Algorithm HCE

In this section, we develop algorithm HCE (for Homogeneous Conic Equalities) and analyze its complexity and the properties of solutions it generates. Algorithm HCE is designed

to obtain a solution of the problem

$$\text{(HCE)} \quad \begin{aligned} Mw &= 0 \\ w &\in C. \end{aligned} \tag{3.32}$$

We assume that M has full row rank. We denote the set of solutions of (HCE) by W_M, i.e.,

$$W_M \triangleq \{w : Mw = 0, \ w \in C\}.$$

The solution w returned by algorithm HCE is "sufficiently interior" in the sense that the ratio $\frac{\|w\|}{\text{dist}(w, \partial C)}$ is not excessively large. (The system (HCE) of (3.32) has a trivial solution $w = 0$. However this solution is not a sufficiently interior solution, since it is contained in the boundary of the cone C).

We define

$$\rho(M) \triangleq \quad \min_{v} \quad \max_{w} \quad \phi$$
$$\|v\| \leq 1 \quad \text{s.t.} \quad \begin{aligned} Mw - \phi v &= 0 \\ w &\in C \\ \|w\| &\leq 1. \end{aligned} \tag{3.33}$$

The following remark summarizes some important facts about $\rho(M)$:

Remark 3.2 *Suppose $\rho(M) > 0$. Then the set $\{w \in W_M : w \neq 0\}$ is non-empty, and M has full row rank. Moreover, $\rho(M) \leq \|M\|$ and*

$$\|(MM^t)^{-1}\| \leq \frac{1}{\rho(M)^2}. \tag{3.34}$$

This follows from the observation that $\rho(M)^2 \leq \lambda_1(MM^t)$, where $\lambda_1(MM^t)$ denotes the smallest eigenvalue of the matrix MM^t.

We will assume for the rest of this section that $\rho(M) > 0$. Then the second statement of Remark 3.2 implies that the earlier assumption that M has full row rank is satisfied. In order to obtain a sufficiently interior solution of (HCE) we will construct a transformation of the system (HCE) which has the form (3.2), and its solutions can be transformed into

sufficiently interior solutions of the system (HCE). The next subsection contains the analysis of the transformation, and its results are used to develop algorithm HCE in the following subsection.

3.3.1 Properties of a Parameterized Conic System of Equalities in Compact Form

In this subsection we work with a compact-form system

$$\begin{aligned} (\text{HCE}_0) \quad Mx &= 0 \\ x &\in C \\ \bar{u}^t x &= 1. \end{aligned} \tag{3.35}$$

The system (HCE_0) is derived from the system (HCE) by adding a compactifying constraint $\bar{u}^t x = 1$. Remark 3.2 implies that when $\rho(M) > 0$ the system (HCE_0) is feasible.

We will consider systems arising from parametric perturbations of the right-hand-side of (HCE_0). In particular, for a fixed vector $z \in Y$, we consider the perturbed compact-form system

$$\begin{aligned} (\text{HCE}_\delta) \quad Mx &= \delta z \\ x &\in C \\ \bar{u}^t x &= 1, \end{aligned} \tag{3.36}$$

where the scalar $\delta \geq 0$ is the perturbation parameter (observe that (HCE_0) can be viewed as an instance of (HCE_δ) with the parameter $\delta = 0$, justifying the notation). Since the case when $z = 0$ is trivial (i.e., (HCE_δ) is equivalent to (HCE_0) for all values of δ), we assume that $z \neq 0$. The following lemma establishes an estimate on the range of values of δ for which the resulting system is feasible, and establishes bounds on the parameters of the system (HCE_δ) in terms of δ.

Before stating the lemma, we will restate some facts about the geometric interpretation of (HCE_δ) and the parameter $r(M, \delta z)$ of (3.15). Recall that the system (HCE_δ) is feasible precisely when $\delta z \in \mathcal{H}_M \triangleq \{ Mx : x \in C, \ \bar{u}^t x = 1 \}$. Also, if the system (HCE_δ) is feasible, $r(M, \delta z)$ can be interpreted as the radius of the largest ball centered at δz and contained in

$\mathcal{H}_M$. Moreover, using the inequality $\beta_C \|x\| \leq \bar{u}^t x \leq \|x\|$ for all $x \in C$, it follows that

$$\beta_C r(M, 0) \leq \rho(M) \leq r(M, 0).$$

Lemma 3.4 *Suppose (HCE$_0$) of (3.35) is feasible, and $z \in Y$, $z \neq 0$. Define*

$$\bar{\delta} = \max\{\delta : (HCE_\delta) \text{ is feasible}\}. \tag{3.37}$$

Then $\frac{\rho(M)}{\|z\|} \leq \frac{r(M,0)}{\|z\|} \leq \bar{\delta} < +\infty$. Moreover, if $\rho(M) > 0$, then $\bar{\delta} > 0$, and for any $\delta \in [0, \bar{\delta}]$, the system (HCE$_\delta$) is feasible and $\|M - \delta z \bar{u}^t\| \leq \|M\| + \delta \|z\|$ and $r(M, \delta z) \geq \left(\frac{\bar{\delta} - \delta}{\bar{\delta}}\right) \rho(M)$.

Proof: Since $\mathcal{H}_M$ is a closed set, $\bar{\delta}$ is well defined. Note that the definition of $\bar{\delta}$ implies that $\bar{\delta} z \in \partial \mathcal{H}_M$. Also, since $z \neq 0$ and $\mathcal{H}_M$ is bounded, $\bar{\delta} < +\infty$. To establish the lower bound on $\bar{\delta}$, note that for any $y \in Y$ such that $\|y\| \leq 1$, $r(M, 0)y \in \mathcal{H}_M$. Therefore, if we take $y = \frac{z}{\|z\|}$, we have $\frac{r(M,0)}{\|z\|} z \in \mathcal{H}_M$, and so (HCE$_\delta$) is feasible for $\delta = \frac{r(M,0)}{\|z\|}$. Hence, $\bar{\delta} \geq \frac{r(M,0)}{\|z\|} \geq \frac{\rho(M)}{\|z\|}$.

The bound on $\|M - \delta z \bar{u}^t\|$ is a simple application of the triangle inequality for the operator norm, i.e., $\|M - \delta z \bar{u}^t\| \leq \|M\| + \delta \|z\| \cdot \|\bar{u}\|_* = \|M\| + \delta \|z\|$.

Finally, suppose that $\rho(M) > 0$. Then $\bar{\delta} > 0$. Let $\delta \in [0, \bar{\delta}]$ be some value of the perturbation parameter. Since $\delta \leq \bar{\delta}$, the system (HCE$_\delta$) is feasible. To establish the lower bound on $r(M, \delta z)$ stated in the lemma, we need to show that a ball of radius $\frac{\bar{\delta} - \delta}{\bar{\delta}} r(M, 0)$ centered at δz is contained in $\mathcal{H}_M$. Suppose $y \in Y$ is such that $\|y\| \leq 1$. As noted above, $\bar{\delta} z \in \mathcal{H}_M$ and $r(M, 0)y \in \mathcal{H}_M$. Therefore,

$$\delta z + \frac{\bar{\delta} - \delta}{\bar{\delta}} r(M, 0)y = \frac{\delta}{\bar{\delta}}(\bar{\delta} z) + \left(1 - \frac{\delta}{\bar{\delta}}\right)(r(M, 0)y) \in \mathcal{H}_M,$$

since the above is a convex combination of $\bar{\delta} z$ and $r(M, 0)y$. Therefore, $r(M, \delta z) \geq \frac{\bar{\delta} - \delta}{\bar{\delta}} r(M, 0) \geq \frac{\bar{\delta} - \delta}{\bar{\delta}} \rho(M)$, which concludes the proof. ∎

We now consider the system (HCE$_\delta$) of (3.36) with the vector $z \triangleq -Mu$, where u is as

specified in Assumption 3.1. The system (HCE_δ) becomes

$$(\text{HCE}_\delta) \quad Mx = -\delta Mu$$
$$x \in C \tag{3.38}$$
$$\bar{u}^t x = 1.$$

The following proposition indicates how approximate solutions of the system (HCE_δ) of (3.38) can be used to obtain sufficiently interior solutions of the system (HCE).

Proposition 3.6 *Suppose $\rho(M) > 0$ and $\delta > 0$. Suppose further that x is an admissible point for (HCE_δ), and in addition x satisfies*

$$\|Mx + \delta Mu\| \leq \frac{1}{2}\delta\tau_C \frac{\rho(M)^2}{\|M\|}.$$

Define

$$w \triangleq (I - M^t(MM^t)^{-1}M)(x + \delta u). \tag{3.39}$$

Then $Mw = 0$ and

$$\|w - (x + \delta u)\| \leq \frac{1}{2}\delta\tau_C \tag{3.40}$$

which implies that $w \in C$, $\text{dist}(w, \partial C) \geq \frac{1}{2}\delta\tau_C$, and $\|w\| \leq \frac{1}{2}\delta\tau_C + \frac{1}{\beta_C} + \delta$.

Proof: First, observe that w satisfies $Mw = 0$ by definition (3.39). To demonstrate (3.40) we apply the definition (3.39) of w to obtain

$$\|w - (x + \delta u)\| = \|M^t(MM^t)^{-1}M(x + \delta u)\| \leq \|M\| \cdot \|(MM^t)^{-1}\| \cdot \|M(x + \delta u)\|$$

$$\leq \frac{\delta\tau_C \rho(M)^2 \cdot \|M\| \cdot \|(MM^t)^{-1}\|}{2\|M\|} = \frac{\delta\tau_C \rho(M)^2 \cdot \|(MM^t)^{-1}\|}{2} \leq \frac{\delta\tau_C}{2},$$

since $\|(MM^t)^{-1}\| \leq \frac{1}{\rho(M)^2}$ from Remark 3.2.

The last three statements of the proposition are direct consequences of (3.40). Notice that $B(x + \delta u, \delta\tau_C) \subset C$ since $B(u, \tau_C) \subset C$ and $x \in C$. Combining this with (3.40) and

57

the triangle inequality for the norm we conclude that $w \in C$ and $\text{dist}(w, \partial C) \geq \frac{1}{2}\delta\tau_C$. Also,

$$\|w\| \leq \|w - (x + \delta u)\| + \|x + \delta u\| \leq \frac{1}{2}\delta\tau_C + \frac{1}{\beta_C} + \delta,$$

which completes the proof. ∎

Notice that w defined by (3.39) is the projection of $x + \delta u$ onto the set $\{w : Mw = 0\}$ with respect to the Euclidean norm on the space X. Although the norm on the space X may be different from the Euclidean norm, we will refer to the point w defined by (3.39) as the Euclidean projection of $x + \delta u$.

It is interesting to note that it is not necessary to have $\delta \leq \bar{\delta}$ for Proposition 3.6 to be applicable.

3.3.2 Algorithm HCE

The formal statement of algorithm HCE is as follows:

Algorithm HCE

- *Data: M*

- *Iteration k, $k \geq 1$*

Step 1 $\delta = \delta^k \triangleq 2^{1-k}$, compute $I(\delta)$:

$$I(\delta) \triangleq \left\lceil \frac{9}{2\beta_C^2\delta^2} \ln\left(\frac{1}{2\tau_C\delta^2}\left(1 + \frac{1}{\beta_C\delta}\right)\right) \right\rceil. \tag{3.41}$$

Step 2 Run GVNA with STOP = STOP3 with $I = I(\delta)$ on the data set $(M, -\delta Mu, x^0)$ (where x^0 is an arbitrary admissible starting point).

Step 3 Let x be the last iterate of GVNA in Step 2. Set

$$w = (I - M^t(MM^t)^{-1}M)(x + \delta u).$$

If $\|w - (x + \delta u)\| \leq \frac{1}{2}\tau_C\delta$, stop. Return w.

Else, set $k \leftarrow k + 1$ and repeat Step 1.

The following proposition states that when $\rho(M) > 0$ algorithm HCE will terminate and return as output a sufficiently interior solution of (HCE).

Theorem 3.2 *Suppose (HCE) satisfies $\rho(M) > 0$. Algorithm HCE will terminate in at most*

$$\left\lceil \log_2\left(\frac{\|M\|}{\rho(M)}\right) \right\rceil + 2 \tag{3.42}$$

iterations, performing at most

$$\frac{4}{3}\left\lceil \frac{216\|M\|^2}{\rho(M)^2\beta_C^2} \ln\left(\frac{40\|M\|}{\rho(M)\tau_C\beta_C}\right) \right\rceil + \left\lceil \log_2\left(\frac{\|M\|}{\rho(M)}\right) \right\rceil + 2 \tag{3.43}$$

iterations of algorithm GVNA.

Algorithm HCE will return a vector $w \in X$ with the following properties:

1. $w \in W_M$,

2. $\mathrm{dist}(w, \partial C) \geq \frac{\tau_C\rho(M)}{8\|M\|}$,

3. $\|w\| \leq \frac{5}{2\beta_C}$,

4. $\frac{\|w\|}{\mathrm{dist}(w,\partial C)} \leq \frac{11\|M\|}{\rho(M)\beta_C\tau_C}$.

Proof: We begin by establishing the maximum number of iterations algorithm HCE will perform. Suppose that x is an admissible point for the system (HCE$_\delta$) for some value $\delta > 0$. The residual at point x is defined in algorithm GVNA as $v = -\delta Mu - Mx = -M(x + \delta u)$. From Proposition 3.6, having a residual with a small norm will guarantee that the projection w of the point $x + \delta u$ will satisfy the property $\|w - (x + \delta u)\| \leq \frac{1}{2}\tau_C\delta$. In particular, it is sufficient to have $\|v\| \leq \epsilon$ with

$$\epsilon = \frac{1}{2}\delta\tau_C\frac{\rho(M)^2}{\|M\|}. \tag{3.44}$$

We now argue that if $\delta \leq \frac{1}{2}\frac{\rho(M)}{\|M\|}$, then Step 2 of algorithm HCE will terminate in $I(\delta)$ iterations and produce an iterate with the size of the residual no larger than ϵ given by (3.44).

Suppose $0 < \delta \leq \frac{1}{2}\frac{\rho(M)}{\|M\|}$. Let $\bar{\delta}$ be as defined in (3.37). Applying Lemma 3.4 for $z = -Mu$ we conclude that the system (HCE$_\delta$) is feasible for any $\delta \in [0, \bar{\delta}]$, and $\bar{\delta} \geq \frac{\rho(M)}{\|Mu\|} \geq \frac{\rho(M)}{\|M\|} > \delta$. Hence the system (HCE$_\delta$) is feasible, and furthermore

$$\|M + \delta Mu\bar{u}^t\| \leq (1 + \delta)\|M\| \leq \frac{3}{2}\|M\|$$

(since $\delta \leq \frac{1}{2}$), and

$$r(M, -\delta Mu) \geq \left(\frac{\bar{\delta} - \delta}{\bar{\delta}}\right)\rho(M) \geq \frac{1}{2}\rho(M).$$

Since the system (HCE$_\delta$) is feasible, from Proposition 3.2 it must be true that algorithm GVNA with STOP = STOP3 will perform $I = I(\delta)$ iterations, where

$$I(\delta) \triangleq \left\lceil \frac{9}{2\beta_C^2\delta^2} \ln\left(\frac{1}{2\tau_C\delta^2}\left(1 + \frac{1}{\beta_C\delta}\right)\right)\right\rceil \geq \frac{18\|M\|^2}{\rho(M)^2\beta_C^2} \ln\left(\frac{2\|M\|^2}{\rho(M)^2\tau_C}\left(1 + \frac{1}{\beta_C\delta}\right)\right), \quad (3.45)$$

since $\delta \leq \frac{1}{2}\frac{\rho(M)}{\|M\|}$. Applying Lemma 3.2 we conclude that after $I(\delta)$ iterations of GVNA the residual $v^{I(\delta)}$ satisfies:

$$\|v^{I(\delta)}\| \leq \|v^0\|e^{-\frac{I(\delta)}{2}\left(\frac{\beta_C r(M, -\delta Mu)}{\|M + \delta Mu\bar{u}^t\|}\right)^2} \leq \|Mx^0 + \delta Mu\|e^{-\frac{I(\delta)}{2}\left(\frac{\beta_C\rho(M)}{3\|M\|}\right)^2}$$

$$\leq \left(\frac{1}{\beta_C} + \delta\right)\|M\|e^{-\frac{9\|M\|^2}{\rho(M)^2\beta_C^2}\ln\left(\frac{2\|M\|^2}{\rho(M)^2\tau_C}\left(1 + \frac{1}{\beta_C\delta}\right)\right)\cdot\left(\frac{\beta_C\rho(M)}{3\|M\|}\right)^2} = \frac{\rho(M)^2\tau_C\delta}{2\|M\|} = \epsilon.$$

We conclude that if $0 < \delta \leq \frac{1}{2}\frac{\rho(M)}{\|M\|}$, then algorithm GVNA of Step 2 of HCE will perform $I(\delta)$ iterations and w defined in Step 3 will satisfy the termination criterion of HCE.

In principle, algorithm HCE might terminate with a solution after as little as one iteration, if the point w defined in Step 3 of that iteration happens to be sufficiently close to the point $x + \delta u$. However, in the worst case algorithm HCE will continue iterating until the value of δ becomes small enough to guarantee (by the analysis above) that the corresponding iteration will produce a point satisfying the termination criterion. To make this argument more precise, recall that during the kth iteration of the algorithm HCE, $\delta = \delta^k = 2^{1-k}$. Hence, HCE is guaranteed to stop at (or before) the iteration during which value of δ falls below $\frac{1}{2}\frac{\rho(M)}{\|M\|}$ for the first time. In other words, the number of iterations of HCE that are

60

performed is bounded above by

$$\min\left\{k : 2^{1-k} \leq \frac{1}{2}\frac{\rho(M)}{\|M\|}\right\}.$$

Therefore algorithm HCE will terminate in no more than

$$K = \left\lceil \log_2\left(\frac{\|M\|}{\rho(M)}\right)\right\rceil + 2 \tag{3.46}$$

iterations, which proves the first claim of the theorem. Also, notice that throughout the algorithm,

$$\delta^k > \frac{1}{4}\frac{\rho(M)}{\|M\|}. \tag{3.47}$$

To bound the total number of iterations of GVNA performed by HCE, we need to bound the sum of the corresponding $I(\delta)$'s:

$$\sum_{k=1}^{K} I(\delta^k) = \sum_{k=1}^{K}\left\lceil\frac{9\cdot 4^k}{8\beta_C^2}\ln\left(\frac{4^k}{8\tau_C}\left(1+\frac{2^{k-1}}{\beta_C}\right)\right)\right\rceil. \tag{3.48}$$

It can be shown by analyzing the geometric series $\sum_{k=1}^{K} 4^k$ that the sum in (3.48) satisfies $\sum_{k=1}^{K} I(\delta^k) \leq \frac{4}{3}I(\delta^K) + K$. Therefore

$$\sum_{k=1}^{K} I(\delta^k) \leq \frac{4}{3}\left\lceil\frac{9}{2\beta_C^2(\delta^K)^2}\ln\left(\frac{1}{2\tau_C(\delta^K)^2}\left(1+\frac{1}{\beta_C\delta^K}\right)\right)\right\rceil + K$$

$$\leq \frac{4}{3}\left\lceil\frac{72\|M\|^2}{\rho(M)^2\beta_C^2}\ln\left(\frac{8\|M\|^2}{\rho(M)^2\tau_C}\left(1+\frac{4\|M\|}{\rho(M)\beta_C}\right)\right)\right\rceil + \left\lceil\log_2\left(\frac{\|M\|}{\rho(M)}\right)\right\rceil + 2$$

$$\leq \frac{4}{3}\left\lceil\frac{72\|M\|^2}{\rho(M)^2\beta_C^2}\ln\left(\frac{40\|M\|^3}{\rho(M)^3\tau_C\beta_C}\right)\right\rceil + \left\lceil\log_2\left(\frac{\|M\|}{\rho(M)}\right)\right\rceil + 2$$

$$\leq \frac{4}{3}\left\lceil\frac{216\|M\|^2}{\rho(M)^2\beta_C^2}\ln\left(\frac{40\|M\|}{\rho(M)\tau_C\beta_C}\right)\right\rceil + \left\lceil\log_2\left(\frac{\|M\|}{\rho(M)}\right)\right\rceil + 2. \tag{3.49}$$

The first inequality in (3.49) follows from (3.47). We have thus established the second claim of the theorem.

It remains to show that the vector w returned by algorithm HCE satisfies conditions

1 through *4*. Let δ^K denote the value of δ during the last iteration of HCE. Applying Proposition 3.6 combined with (3.47) we conclude that conditions *1* and *2* are satisfied. Furthermore,

$$\|w\| \le \frac{1}{2}\delta^K\tau_C + \frac{1}{\beta_C} + \delta^K \le \frac{3}{2} + \frac{1}{\beta_C} \le \frac{5}{2\beta_C},$$

which establishes condition *3*, and

$$\frac{\|w\|}{\text{dist}(w, \partial C)} \le \frac{\frac{1}{2}\delta^K\tau_C + \frac{1}{\beta_C} + \delta^K}{\frac{1}{2}\tau_C\delta^K} = 2\left(\frac{1}{2} + \frac{1}{\beta_C\tau_C\delta^K} + \frac{1}{\tau_C}\right)$$

$$\le 2\left(\frac{1}{2} + \frac{4\|M\|}{\rho(M)\beta_C\tau_C} + \frac{1}{\tau_C}\right) \le \frac{11\|M\|}{\rho(M)\beta_C\tau_C},$$

which establishes condition *4* and completes the proof of the theorem. ∎

62

Chapter 4

Elementary Algorithm for General Conic Linear Systems

In this chapter we indicate how algorithms HCI and HCE of Chapter 3 can be used to obtain reliable solutions of a conic linear system in the most general form. A general conic linear system has the form

$$(\text{FP}_d) \quad b - Ax \in C_Y$$
$$x \in C_X$$

of (2.1), and the "strong alternative" system of (FP_d) is:

$$(\text{SA}_d) \quad A^t s \in C_X^*$$
$$s \in C_Y^*$$
$$b^t s < 0,$$

of (2.13).

We focus on three different cases of instances of (FP_d), namely:

Case 1: C_X is regular and $C_Y = \{0\}$,

Case 2: $C_X = X$ and C_Y is regular,

Case 3: C_X and C_Y are both regular.

The three cases above correspond to the following three forms of (FP_d):

$$\text{Case 1:} \quad Ax = b, \ x \in C_X,$$
$$\text{Case 2:} \quad b - Ax \in C_Y,$$
$$\text{Case 3:} \quad b - Ax \in C_Y, \ x \in C_X.$$

For each of the three cases above, we present an algorithm for solving an instance of the system (FP$_d$) and establish a complexity result for the algorithm. Due to the differences in the underlying geometry, a different algorithm has to be used in each of the three cases. We refer to these different algorithms as "CLS1" (for Conic Linear System, Case 1), "CLS2," and "CLS3." However, the general framework of all of the algorithms is the same. In every case i, the main algorithm CLSi is a combination of two other algorithms, namely algorithm FCLSi (Feasible Conic Linear System) which is used to find a reliable solution of (FP$_d$), and algorithm ICLSi (Infeasible Conic Linear System), which is used to find a reliable solution to the strong alternative system (SA$_d$). Each of the algorithms FCLSi and ICLSi consists of applying either algorithm HCE or algorithm HCI to a suitable transformation of the system (FP$_d$), and transforming the output of GVNA into a solution of the system (FP$_d$) or (SA$_d$).

We begin by presenting an analysis of Case 1. As it turns out, systems in this case yield themselves to the most natural transformations suitable for application of algorithms HCE and HCI, yet have enough richness to allow us to illustrate most of the aspects of the algorithm design and complexity analysis. We then discuss how the other two cases can be handled by similar algorithms.

For the remainder of this chapter we maintain the following notation:

Definition 4.1 *Whenever the cone C_X is regular, the coefficient of linearity of C_X is denoted by β, and the corresponding norm linearization vector is denoted by $\bar{f}$. The coefficient of linearity of C_X^* is denoted by τ, and the corresponding norm linearization vector is denoted by f. (Equivalently, the width of the cone C_X is τ and is attained for (f, τ).)*

Whenever the cone C_Y is regular, the coefficient of linearity of C_Y is denoted by $\check{\beta}$, and the corresponding norm linearization vector is denoted by $\bar{e}$. The coefficient of linearity of C_Y^* is denoted by $\check{\tau}$, and the corresponding norm linearization vector is denoted by e. (Equivalently, the width of the cone C_Y is $\check{\tau}$ and is attained for $(e, \check{\tau})$.)

4.1 Case 1: C_X is regular and $C_Y = \{0\}$

When the cone C_X is regular and $C_Y = \{0\}$, the systems (FP$_d$) and (SA$_d$) can be written as follows:

$$
\begin{array}{llll}
(\text{FP}_d) & Ax = b & (\text{SA}_d) & A^t s \in C_X^* \\
& x \in C_X & & b^t s < 0.
\end{array}
\tag{4.1}
$$

We develop algorithm CLS1, which is a combination of two other algorithms, namely algorithm FCLS1 (Feasible Conic Linear System, Case 1) which is used to find a reliable solution of (FP$_d$) of (4.1), and algorithm ICLS1 (Infeasible Conic Linear System, Case 1), which is used to find a reliable solution to the alternative system (SA$_d$) of (4.1). We first proceed by presenting algorithms FCLS1 and ICLS1, and studying their complexity. We then combine algorithms FCLS1 and ICLS1 to form algorithm CLS1 and study its complexity.

We will assume in this section that Y is an m-dimensional Euclidean space with Euclidean norm $\|y\| = \|y\|_2$ for $y \in Y$.

4.1.1 Algorithm FCLS1

Algorithm FCLS1 is designed to compute a reliable solution of (FP$_d$) of (4.1) when the system (FP$_d$) is feasible. Consider the following reformulation of the system (FP$_d$):

$$
\begin{aligned}
-b\theta + Ax &= 0 \\
\theta \geq 0, \; x &\in C_X.
\end{aligned}
\tag{4.2}
$$

System (4.2) is of the form (HCE) of (3.32) under the following assignments:

- $M = \begin{bmatrix} -b & A \end{bmatrix}$

- $C = \Re_+ \times C_X,$

with norms defined as follows:

- $\|(\theta, x)\| = |\theta| + \|x\|, \; (\theta, x) \in \Re \times X$

- $\|v\| = \|v\|_2, \; v \in Y.$

65

Applying Propositions 2.4 and 2.5, the norm linearization vector for C is easily seen to be $\bar{u} = (1, \bar{f})$ with $\beta_C = \beta$. Moreover, the width of the cone C is $\tau_C = \frac{\tau}{1+\tau} \geq \frac{1}{2}\tau$ and is attained at $u = \frac{1}{1+\tau}(\tau, f)$.

Proposition 4.1 *Suppose (FP$_d$) of (4.1) is feasible and $\rho(d) > 0$. Then the system (4.2) is feasible, M has full row rank, and we have*

$$\|M\| = \|d\|, \ \ and \ \rho(M) = \rho(d),$$

where $\rho(M)$ is defined in (3.33).

Proof: See Appendix B, Propostion B.1. ∎

We use algorithm HCE to find a sufficiently interior solution of the system (4.2) and transform its output into a reliable solution of (FP$_d$), as described below:

<div align="center">

Algorithm FCLS1

</div>

- *Data: $d = (A, b)$*

    Step 1   Apply algorithm HCE to the system (4.2). The algorithm will return a vector $\tilde{w} = (\tilde{\theta}, \tilde{x})$.

    Step 2   Define $\hat{x} = \frac{\tilde{x}}{\tilde{\theta}}$. Return $\hat{x}$ (a reliable solution of (FP$_d$)).

Lemma 4.1 *Suppose (FP$_d$) of (4.1) is feasible and $\rho(d) > 0$. Then algorithm FCLS1 will terminate in at most*

$$\frac{4}{3}\left\lceil \frac{216\mathcal{C}(d)^2}{\beta^2} \ln\left(\frac{80\mathcal{C}(d)}{\tau\beta}\right)\right\rceil + \lceil \log_2 \mathcal{C}(d)\rceil + 2 \tag{4.3}$$

iterations of algorithm GVNA. The output $\hat{x}$ will satisfy

1. $\hat{x} \in X_d$,

2. $\|\hat{x}\| \leq \frac{22\mathcal{C}(d)}{\beta\tau} - 1$,

3. $\mathrm{dist}(\hat{x}, \partial C_X) \geq \frac{\beta\tau}{22\mathcal{C}(d)}$,

4. $\dfrac{\|\hat{x}\|}{\operatorname{dist}(\hat{x}, \partial C_X)} \leq \dfrac{22\mathcal{C}(d)}{\beta\tau}$.

Proof: To simplify the expressions in this proof, define

$$\alpha \overset{\triangle}{=} \operatorname{dist}(\tilde{w}, \partial C) = \operatorname{dist}\left((\tilde{\theta}, \tilde{x}), \partial(\Re_+ \times C_X)\right).$$

From Theorem 3.2 we conclude that algorithm HCE in Step 1 will terminate in at most

$$\frac{4}{3}\left\lceil \frac{216\mathcal{C}(d)^2}{\beta^2}\ln\left(\frac{80\mathcal{C}(d)}{\tau\beta}\right)\right\rceil + \lceil\log_2 \mathcal{C}(d)\rceil + 2$$

iterations of algorithm GVNA, which establishes the first statement of the lemma.

Next, from Theorem 3.2 we conclude that the vector $\tilde{w} = (\tilde{\theta}, \tilde{x})$ returned by algorithm HCE in Step 1 satisfies:

$$-b\tilde{\theta} + A\tilde{x} = 0, \ (\tilde{\theta}, \tilde{x}) \in \Re_+ \times C_X, \ \alpha \geq \frac{\tau_C\rho(M)}{8\|M\|} \geq \frac{\tau}{16\mathcal{C}(d)}, \tag{4.4}$$

$$\|(\tilde{\theta}, \tilde{x})\| = |\tilde{\theta}| + \|\tilde{x}\| \leq \frac{5}{2\beta_C} = \frac{5}{2\beta}, \ \frac{\|(\tilde{\theta}, \tilde{x})\|}{\alpha} \leq \frac{11\|M\|}{\rho(M)\beta_C\tau_C} \leq \frac{22\mathcal{C}(d)}{\beta\tau}. \tag{4.5}$$

Note in particular that (4.4) implies that $\tilde{\theta} \geq \alpha > 0$, so that $\hat{x}$ is well-defined, and $A\hat{x} = b$, $\hat{x} \in C_X$, which establishes statement *1*.

Next,
$$\|\hat{x}\| = \frac{\|\tilde{x}\|}{\tilde{\theta}} = \frac{\|\tilde{w}\| - \tilde{\theta}}{\tilde{\theta}} \leq \frac{\|\tilde{w}\|}{\alpha} - 1 \leq \frac{22\mathcal{C}(d)}{\beta\tau} - 1,$$

which proves *2*.

To prove *3*, define $r \overset{\triangle}{=} \frac{\alpha}{\|\tilde{w}\|}(1 + \|\hat{x}\|)$. Then a simple application of (4.5) implies that $r \geq \frac{\beta\tau}{22\mathcal{C}(d)}$. Further, let $p \in X$ be an arbitrary vector satisfying $\|p\| \leq r$. Then

$$\|\tilde{\theta}p\| \leq \tilde{\theta} \cdot r = \tilde{\theta} \cdot \frac{\alpha}{\|\tilde{w}\|}(1 + \|\hat{x}\|) = \frac{\alpha}{\|\tilde{w}\|}(\tilde{\theta} + \|\tilde{x}\|) = \alpha,$$

and so $\tilde{x} + \tilde{\theta}p \in C_X$, and hence $\hat{x} + p = \frac{\tilde{x} + \tilde{\theta}p}{\tilde{\theta}} \in C_X$. Therefore, $\operatorname{dist}(\hat{x}, \partial C_X) \geq r \geq \frac{\beta\tau}{22\mathcal{C}(d)}$, establishing *3*.

Finally,

$$\frac{||\hat{x}||}{\mathrm{dist}(\hat{x}, \partial C_X)} \leq \frac{||\hat{x}||}{r} = \frac{||\hat{x}|| \cdot ||\tilde{w}||}{\alpha(1 + ||\hat{x}||)} \leq \frac{||\tilde{w}||}{\alpha} \leq \frac{22\mathcal{C}(d)}{\beta\tau},$$

which establishes 4. ■

4.1.2 Algorithm ICLS1

Algorithm ICLS1 is designed to compute a reliable solution of (SA_d) of (4.1) when the system (FP_d) is infeasible. Consider the following compact-form reformulation of the system (FP_d):

$$\begin{aligned} -b\theta + Ax &= 0 \\ \theta + \bar{f}^t x &= 1, \\ \theta &\geq 0, \; x \in C_X. \end{aligned} \tag{4.6}$$

The alternative system to (4.6) is given by

$$\begin{aligned} -b^t s &> 0 \\ A^t s &\in \mathrm{int}C_X^*. \end{aligned} \tag{4.7}$$

System (4.7) is of the form (HCI) under the following assignments:

- $M = \begin{bmatrix} -b & A \end{bmatrix}$

- $C = \Re_+ \times C_X,$

with norms defined as follows:

- $||(\theta, x)|| = |\theta| + ||x||, \; (\theta, x) \in \Re \times X$

- $||v|| = ||v||_2, \; v \in Y.$

From Proposition 2.4, the norm linearization vector for C is easily seen to be $\bar{u} = (1, \bar{f})$ with $\beta_C = \beta$.

Proposition 4.2 *Suppose (FP_d) of (4.1) is infeasible and $\rho(d) > 0$. Then the system (4.7) is feasible, and we have*

$$||M|| = ||d||,$$

$$\rho(d) \le r(M) \le \frac{\rho(d)}{\beta},$$

where $r(M)$ is defined in (3.28).

Proof: See Appendix B, Proposition B.2. ■

We use algorithm HCI to find a sufficiently interior solution of the system (4.7) and show that it is a reliable solution of (SA_d), as described below:

Algorithm ICLS1

- *Data: $d = (A, b)$*

  Step 1 Apply algorithm HCI to the system (4.7). The algorithm will return a vector s.

  Step 2 Return s (a reliable solution of (SA_d)).

Lemma 4.2 *Suppose (FP_d) of (4.1) is infeasible and $\rho(d) > 0$. Then algorithm ICLS1 will terminate in at most*

$$\left\lfloor \frac{16\mathcal{C}(d)^2}{\beta^2} \right\rfloor \tag{4.8}$$

iterations of GVNA. The output s will satisfy $s \in A_d$ and

$$\frac{\|s\|}{\mathrm{dist}(s, \partial A_d)} \le \frac{2\mathcal{C}(d)}{\beta}.$$

Proof: From Theorem 3.1 we conclude that algorithm HCI in Step 1 will terminate in at most

$$\left\lfloor \frac{16\|M\|^2}{\beta_C^2 r(M)^2} \right\rfloor \le \left\lfloor \frac{16\mathcal{C}(d)^2}{\beta^2} \right\rfloor$$

iterations of GVNA, which establishes the first statement of the lemma. Furthermore, the output s satisfies $s \in S_M$ and

$$\frac{\|s\|}{\mathrm{dist}(s, \partial S_M)} \le \frac{2\|M\|}{\beta_C r(M)} \le \frac{2\mathcal{C}(d)}{\beta}.$$

Since $S_M \subseteq A_d$, the second statement of the lemma follows. ■

4.1.3 Algorithm CLS1

Algorithm CLS1 described below is a combination of algorithms FCLS1 and ICLS1. Algorithm CLS1 is designed to solve the system (FP$_d$) of (4.1) by either finding a reliable solution of (FP$_d$) or demonstrating the infeasibility of (FP$_d$) by finding a reliable solution of (SA$_d$). Since it is not known in advance whether (FP$_d$) is feasible or not, algorithm CLS1 is designed to run both algorithms FCLS1 and ICLS1 in parallel, and will terminate when either one of the two algorithms terminates. The formal description of algorithm CLS1 is as follows:

Algorithm CLS1

- *Data:* $d = (A, b)$

  Step 1 Run algorithms FCLS1 and ICLS1 in parallel on the data set $d = (A, b)$, until one of them terminates.

  Step 2 If algorithm FCLS1 terminates first, return its output $\hat{x}$. If algorithm ICLS1 terminates first, return its output s.

Although Step 1 of algorithm CLS1 calls for algorithms FCLS1 and ICLS1 to be run in parallel, there is no necessity for parallel computation *per se*. Observe that both algorithms FCLS1 and ICLS1 consist of repetitively calling algorithm GVNA on a sequence of data instances. A sequential implementation of Step 1 is to run one iteration of algorithm GVNA called by algorithm FCLS1, followed by the next iteration of algorithm GVNA called by algorithm ICLS1, etc., until one of the iterations yields the termination of the algorithm.

Combining the complexity results for algorithms FCLS1 and ICLS1 from Lemmas 4.1 and 4.2 we obtain the following complexity analysis of algorithm CLS1:

Theorem 4.1 *Suppose that $\rho(d) > 0$. If the system (FP$_d$) of 4.1 is feasible, algorithm CLS1 will terminate in at most*

$$\frac{8}{3} \left\lceil \frac{216 \mathcal{C}(d)^2}{\beta^2} \ln \left(\frac{80 \mathcal{C}(d)}{\tau \beta} \right) \right\rceil + 2 \left\lceil \log_2 \mathcal{C}(d) \right\rceil + 4$$

iterations of GVNA, and will return a reliable solution $\hat{x}$ of (FP_d). That is, $\hat{x}$ will have the following properties:

- $\hat{x} \in X_d$,

- $\|\hat{x}\| \leq \frac{22\mathcal{C}(d)}{\beta\tau} - 1$,

- $\mathrm{dist}(\hat{x}, \partial C_X) \geq \frac{\beta\tau}{22\mathcal{C}(d)}$,

- $\frac{\|\hat{x}\|}{\mathrm{dist}(\hat{x}, \partial C_X)} \leq \frac{22\mathcal{C}(d)}{\beta\tau}$.

If the system (FP_d) is infeasible, algorithm CLS1 will terminate in at most

$$2 \left\lfloor \frac{16\mathcal{C}(d)^2}{\beta^2} \right\rfloor$$

iterations of GVNA, and will return a reliable solution s of (SA_d), thus demonstrating infeasibility of (FP_d). That is, s will satisfy the following properties:

- $s \in A_d$,

- $\frac{\|s\|}{\mathrm{dist}(s, \partial A_d)} \leq \frac{2\mathcal{C}(d)}{\beta}$.

Proof: The proof is an immediate consequence of Lemmas 4.1 and 4.2. The bounds on the number of iterations of algorithm GVNA in the theorem are precisely double the bounds in the lemmas, due to running algorithms FCLS1 and ICLS1 in parallel. ∎

4.2   Case 2: $C_X = X$ and C_Y is regular

When $C_X = X$ and the cone C_Y is regular, the systems (FP_d) and (SA_d) can be written as follows:

$$\begin{array}{lll} (\mathrm{FP}_d) & b - Ax \in C_Y & \qquad (\mathrm{SA}_d) \quad A^t s = 0 \\ & & \qquad\qquad\qquad\quad s \in C_Y^* \qquad\qquad (4.9) \\ & & \qquad\qquad\qquad\quad b^t s < 0. \end{array}$$

The basic idea behind the algorithm CLS2 is to convert the alternative system (SA_d) into the form similar to that of (FP_d) of (4.1). We will assume in this section that X is an

71

n-dimensional Euclidean space with Euclidean norm $\|x\| = \|x\|_2$ for $x \in X$ (and therefore the dual norm is also $\|q\|_* = \|q\|_2$ for $q \in X^*$).

4.2.1   Algorithm FCLS2.

Algorithm FCLS2 is designed to find a reliable solution of the system (FP$_d$) of (4.9) when (FP$_d$) is feasible. Consider the following compact-form reformulation of (SA$_d$):

$$
\begin{aligned}
b^t s \quad + \nu t &= 0 \\
-A^t s \quad\quad &= 0 \\
e^t s \quad + t &= 1 \\
s \in C_Y^*, \quad\quad t &\geq 0,
\end{aligned}
\tag{4.10}
$$

where $\nu > 0$ is a fixed constant. The alternative system to (4.10) is given by

$$
\begin{aligned}
b\theta \quad - Ax \quad &\in \mathrm{int} C_Y \\
\nu\theta \quad\quad &> 0.
\end{aligned}
\tag{4.11}
$$

System (4.11) is of the form (HCI) under the following assignments:

- $M = \begin{bmatrix} b^t & \nu \\ -A^t & 0 \end{bmatrix}$

- $C = C_Y^* \times \Re_+$,

with the norms defined as follows:

- $\|(s,t)\| = \|s\|_* + |t|$, $(s,t) \in Y^* \times \Re$,

- $\|(j,q)\| = \|(j,q)\|_2 = \sqrt{j^2 + \|q\|_2^2}$, $(j,q) \in \Re \times X^*$.

Then the norm linearization vector for C is easily seen to be $\bar{u} = (e,1)$ with $\beta_C = \breve{\tau}$.

Proposition 4.3 *Suppose (FP$_d$) of (4.9) is feasible and $\rho(d) > 0$. Then the system (4.11) is feasible, and we have*

$$
\max\{\|d\|, \nu\} \leq \|M\| \leq \sqrt{2}\max\{\|d\|, \nu\},
$$

$$\frac{\rho(d)}{3\max\{1,\|d\|/\nu\}} \le r(M) \le \frac{\sqrt{2}\rho(d)}{\breve{\tau}}.$$

Proof: See Appendix B, Proposition B.3. ∎

We use algorithm HCI to find a sufficiently interior solution of the system (4.11) and transform its output into a reliable solution of (FP_d), as described below:

<div align="center">

Algorithm FCLS2

</div>

- *Data:* $d = (A, b)$

  Step 1 Apply algorithm HCI to the system (4.11). The algorithm will return a vector $(\tilde{\theta}, \tilde{x})$.

  Step 2 Define $\hat{x} = \frac{\tilde{x}}{\tilde{\theta}}$. Return $\hat{x}$ (a reliable solution of (FP_d)).

Lemma 4.3 *Suppose (FP_d) of (4.9) is feasible and $\rho(d) > 0$. Then algorithm FCLS2 will terminate in at most*

$$\left\lfloor 288\mathcal{C}(d)^2 \frac{\max\{\|d\|/\nu, \nu/\|d\|\}^2}{\breve{\tau}^2} \right\rfloor$$

iterations of GVNA. The output $\hat{x}$ will satisfy:

1. $\hat{x} \in X_d$,

2. $\|\hat{x}\| \le 6\mathcal{C}(d)\frac{\max\{1,\nu/\|d\|\}}{\breve{\tau}}$,

3. $\mathrm{dist}(\hat{x}, \partial X_d) \ge \frac{1}{6\sqrt{2}\mathcal{C}(d)}\frac{\breve{\tau}}{\max\{\|d\|/\nu, \nu/\|d\|\}}$,

4. $\frac{\|\hat{x}\|}{\mathrm{dist}(\hat{x}, \partial X_d)} \le 6\sqrt{2}\mathcal{C}(d)\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\breve{\tau}}$.

Proof: From Theorem 3.1 we conclude that algorithm HCI in Step 1 will terminate in at most

$$\left\lfloor \frac{16\|M\|^2}{\beta_C^2 r(M)^2} \right\rfloor \le \left\lfloor \frac{16 \cdot 2\max\{\|d\|, \nu\}^2 \cdot 9\max\{1, \|d\|/\nu\}^2}{\breve{\tau}^2 \rho(d)^2} \right\rfloor$$

$$= \left\lfloor 288\mathcal{C}(d)^2 \frac{\max\{\|d\|/\nu, \nu/\|d\|\}^2}{\breve{\tau}^2} \right\rfloor$$

iterations of GVNA, which establishes the first statement of the lemma.

Define
$$\gamma \triangleq \mathrm{dist}\left((b\tilde{\theta} - A\tilde{x}, \nu\tilde{\theta}), \partial(C_Y \times \Re_+)\right).$$

Using Corollary 3.1 we conclude that the vector $(\tilde{\theta}, \tilde{x})$ returned by algorithm HCI in Step 1 satisfies:
$$\|(\tilde{\theta}, \tilde{x})\| = 1, \ \ B(b\tilde{\theta} - A\tilde{x}, \gamma) \subset C_Y, \ \ \nu\tilde{\theta} \geq \gamma,$$

$$\gamma \geq \frac{\beta_C r(M)}{2} \geq \frac{\check{\tau}\rho(d)}{6\max\{1, \|d\|/\nu\}}.$$

In particular, $\tilde{\theta} \geq \frac{\gamma}{\nu} > 0$, so that $\hat{x}$ is well-defined, and $b - A\hat{x} \in C_Y$, which establishes 1.
Also,
$$\|\hat{x}\| = \frac{\|\tilde{x}\|}{\tilde{\theta}} \leq \frac{1}{\tilde{\theta}} \leq \frac{\nu}{\gamma} \leq \frac{6\nu\max\{1, \|d\|/\nu\}}{\check{\tau}\rho(d)} = 6\mathcal{C}(d)\frac{\max\{1, \nu/\|d\|\}}{\check{\tau}},$$

proving 2.

To prove 3, let $r \triangleq \frac{\gamma}{\sqrt{2}\max\{\|d\|,\nu\}}(1 + \|\hat{x}\|)$, where γ is as above. Then it is easy to see that

$$r \geq \frac{1}{6\sqrt{2}\mathcal{C}(d)}\frac{\check{\tau}}{\max\{\|d\|/\nu, \nu/\|d\|\}}.$$

Further, let $p \in X$ be an arbitrary vector satisfying $\|p\| \leq r$. Then

$$\|\tilde{\theta}Ap\| \leq \tilde{\theta}\|A\|r = \tilde{\theta}\|A\|\frac{\gamma}{\sqrt{2}\max\{\|d\|,\nu\}}(1 + \|\hat{x}\|)$$

$$= \|A\|\frac{\gamma}{\sqrt{2}\max\{\|d\|,\nu\}}(\tilde{\theta} + \|\tilde{x}\|) \leq \gamma,$$

so that $b\tilde{\theta} - A\tilde{x} - \tilde{\theta}Ap \in C_Y$, and hence

$$b - A(\hat{x} + p) = \frac{b\tilde{\theta} - A\tilde{x} - \tilde{\theta}Ap}{\tilde{\theta}} \in C_Y,$$

and we conclude that $B(\hat{x}, r) \subset X_d$, which proves 3.

Finally,
$$\frac{\|\hat{x}\|}{r} = \frac{\|\hat{x}\|\sqrt{2}\max\{\|d\|,\nu\}}{\gamma(1 + \|\hat{x}\|)} \leq \frac{\sqrt{2}\max\{\|d\|,\nu\}}{\gamma}$$

$$\leq \frac{6\sqrt{2}\max\{\|d\|,\nu\}\max\{1, \|d\|/\nu\}}{\check{\tau}\rho(d)} = 6\sqrt{2}\mathcal{C}(d)\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\check{\tau}},$$

74

which proves *4*. ∎

4.2.2 Algorithm ICLS2.

Algorithm ICLS2 is designed to compute a reliable solution of (SA_d) of (4.9) when (FP_d) is infeasible. Consider the following reformulation of the system (SA_d):

$$
\begin{aligned}
b^t s \ \ +\nu t &= 0 \\
-A^t s \qquad\ &= 0 \\
s \in C_Y^*, \qquad t &\geq 0.
\end{aligned}
\tag{4.12}
$$

System (4.12) is of the form (HCE) under the following assignments:

- $M = \begin{bmatrix} b^t & \nu \\ -A^t & 0 \end{bmatrix}$

- $C = C_Y^* \times \Re_+,$

with norms defined as follows:

- $\|(s,t)\| = \|s\|_* + |t|, \ (s,t) \in Y^* \times \Re,$

- $\|(j,q)\| = \|(j,q)\|_2 = \sqrt{j^2 + \|q\|_2^2}, \ (j,q) \in \Re \times X^*.$

Then the norm linearization vector for C is easily seen to be $\bar{u} = (e, 1)$ with $\beta_C = \breve{\tau}$. Moreover, the width of C is $\tau_C = \frac{\breve{\beta}}{1+\breve{\beta}} \geq \frac{\breve{\beta}}{2}$ and is attained at $u = \frac{1}{1+\breve{\beta}}(\bar{e}, \breve{\beta})$.

Proposition 4.4 *Suppose (FP_d) of (4.9) is infeasible and $\rho(d) > 0$. Then the system (4.12) is feasible, M has full row rank, and we have*

$$
\max\{\|d\|, \nu\} \leq \|M\| \leq \sqrt{2}\max\{\|d\|, \nu\},
$$

$$
\frac{\rho(d)}{3\max\{1, \|d\|/\nu\}} \leq \rho(M) \leq \sqrt{2}\rho(d).
$$

Proof: See Appendix B, Proposition B.4. ∎

We use algorithm HCE to find a sufficiently interior solution of the system (4.12) and transform its output into a reliable solution of (SA_d), as described below:

75

Algorithm ICLS2

- *Data:* $d = (A, b)$

    Step 1 Apply algorithm HCE to the system (4.12). The algorithm will return a vector
    $$\tilde{w} = (\tilde{s}, \tilde{t})$$

    Step 2 Return $s = \frac{\tilde{s}}{\|\tilde{s}\|_*}$ (a reliable solution of (SA$_d$)).

Lemma 4.4 *Suppose (FP$_d$) of (4.9) is infeasible and $\rho(d) > 0$. Then algorithm ICLS2 will terminate in at most*

$$\frac{4}{3} \left\lceil \frac{3888 \mathcal{C}(d)^2 \max\{\|d\|/\nu, \nu/\|d\|\}^2}{\breve{\tau}^2} \ln\left(\frac{240\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\}}{\breve{\beta}\breve{\tau}} \right) \right\rceil$$

$$+ \left\lceil \log_2\left(3\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\} \right) \right\rceil + 2$$

iterations of GVNA. The output s will satisfy:

1. *$s \in A_d$,*

2. *$\|s\|_* = 1$,*

3. *$\mathrm{dist}(s, \partial(C_Y^* \cap \{s : b^t s \leq 0\})) \geq \frac{\breve{\beta}\breve{\tau}}{66\sqrt{2}\mathcal{C}(d)\max\{\|d\|/\nu,\nu/\|d\|\}} \cdot \min\left\{1, \frac{\nu}{\|b\|}\right\}.$*

Proof: From Theorem 3.2 we conclude that algorithm HCE in Step 1 will terminate in at most

$$\frac{4}{3} \left\lceil \frac{216\|M\|^2}{\rho(M)^2 \beta_C^2} \ln\left(\frac{40\|M\|}{\rho(M)\tau_C \beta_C} \right) \right\rceil + \left\lceil \log_2\left(\frac{\|M\|}{\rho(M)} \right) \right\rceil + 2$$

$$\leq \frac{4}{3} \left\lceil \frac{3888 \mathcal{C}(d)^2 \max\{\|d\|/\nu, \nu/\|d\|\}^2}{\breve{\tau}^2} \ln\left(\frac{240\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\}}{\breve{\beta}\breve{\tau}} \right) \right\rceil$$

$$+ \left\lceil \log_2\left(3\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\} \right) \right\rceil + 2$$

iterations of GVNA, which establishes the first statement of the lemma.

Define
$$\alpha \overset{\triangle}{=} \mathrm{dist}(\tilde{w}, \partial C) = \mathrm{dist}\left((\tilde{s}, \tilde{t}), \partial(C_Y^* \times \Re_+) \right)$$

76

From Theorem 3.2 we conclude that the vector $\tilde{w} = (\tilde{s}, \tilde{t})$ returned by algorithm HCE in Step 1 satisfies:

$$b^t \tilde{s} + \nu \tilde{t} = 0, \ -A^t \tilde{s} = 0, \ \alpha > 0, \tag{4.13}$$

$$\frac{\|(\tilde{s}, \tilde{t})\|}{\alpha} \leq \frac{11\|M\|}{\rho(M)\beta_C \tau_C} \leq \frac{66\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\}}{\breve{\beta}\breve{\tau}}. \tag{4.14}$$

Note in particular that (4.13) implies that $B(\tilde{s}, \alpha) \subset C_Y^*$ and $\tilde{t} \geq \alpha > 0$ so that the vector s returned by algorithm ICLS2 satisfies $s \in C_Y^*, A^t s = 0, \ b^t s = -\frac{\nu \tilde{t}}{\|\tilde{s}\|_*} < 0$, so that $s \in A_d$, establishing 1. Observe that 2 holds by the definition of s.

To prove 3, define $r \triangleq \frac{\alpha}{\|\tilde{s}\|_*} \cdot \min\left\{1, \frac{\nu}{\|b\|}\right\}$, where α is as above. Then a simple application of (4.14) implies that

$$r \geq \frac{\alpha}{\|(\tilde{s}, \tilde{t})\|} \cdot \min\left\{1, \frac{\nu}{\|b\|}\right\} \geq \frac{\breve{\beta}\breve{\tau}}{66\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\}} \cdot \min\left\{1, \frac{\nu}{\|b\|}\right\}.$$

Moreover,

$$B(s, r) \subseteq B\left(s, \frac{\alpha}{\|\tilde{s}\|_*}\right) \subset C_Y^*, \tag{4.15}$$

since $B(\tilde{s}, \alpha) \subset C_Y^*$. Further let $y \in Y^*$ be an arbitrary vector satisfying $\|y\|_* \leq r \leq \frac{\alpha \nu}{\|\tilde{s}\|_* \cdot \|b\|}$. Then

$$b^t(\tilde{s} + \|\tilde{s}\|_* y) \leq -\nu\tilde{t} + \|b\| \cdot \|\tilde{s}\|_* r \leq -\nu\alpha + \|b\| \cdot \|\tilde{s}\|_* \frac{\alpha\nu}{\|\tilde{s}\|_* \cdot \|b\|} = 0,$$

so that $b^t(s + y) \leq 0$, which, together with (4.15), proves 3. ∎

4.2.3 Algorithm CLS2

Similarly to algorithm CLS1, algorithm CLS2 described below is a combination of algorithms FCLS2 and ICLS2. Algorithm CLS2 is designed to solve the system (FP_d) of (4.9) by either finding a reliable solution of (FP_d) or demonstrating the infeasibility of (FP_d) by finding a reliable solution of (SA_d). The formal description of algorithm CLS2 is as follows:

Algorithm CLS2

- *Data:* $d = (A, b)$

Step 1 Run algorithms FCLS2 and ICLS2 in parallel on the data set $d = (A, b)$, until one of them terminates.

Step 2 If algorithm FCLS2 terminates first, return its output $\hat{x}$. If algorithm ICLS2 terminates first, return its output s.

Combining the complexity results for algorithms FCLS2 and ICLS2 from Lemmas 4.3 and 4.4 we obtain the following complexity analysis of algorithm CLS2:

Theorem 4.2 *Suppose that* $\rho(d) > 0$. *If the system (FP$_d$) of (4.9) is feasible, algorithm CLS2 will terminate in at most*

$$2 \left\lceil 288 \mathcal{C}(d)^2 \frac{\max\{\|d\|/\nu, \nu/\|d\|\}^2}{\check{\tau}^2} \right\rceil$$

iterations of GVNA, and will return a reliable solution $\hat{x}$ of (FP$_d$). That is, $\hat{x}$ will have the following properties:

- $\hat{x} \in X_d$,

- $\|\hat{x}\| \leq 6\mathcal{C}(d) \frac{\max\{1, \nu/\|d\|\}}{\check{\tau}}$,

- $\mathrm{dist}(\hat{x}, \partial X_d) \geq \frac{1}{6\sqrt{2}\mathcal{C}(d)} \frac{\check{\tau}}{\max\{\|d\|/\nu, \nu/\|d\|\}}$,

- $\frac{\|\hat{x}\|}{\mathrm{dist}(\hat{x}, \partial X_d)} \leq 6\sqrt{2}\mathcal{C}(d) \frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\check{\tau}}$.

If the system (FP$_d$) is infeasible, algorithm CLS2 will terminate in at most

$$\frac{8}{3} \left\lceil \frac{3888 \mathcal{C}(d)^2 \max\{\|d\|/\nu, \nu/\|d\|\}^2}{\check{\tau}^2} \ln \left(\frac{240\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\}}{\check{\beta}\check{\tau}} \right) \right\rceil$$

$$+2 \left\lceil \log_2 \left(3\sqrt{2}\mathcal{C}(d) \max\{\|d\|/\nu, \nu/\|d\|\} \right) \right\rceil + 4$$

iterations of GVNA, and will return a reliable solution s of (SA$_d$), thus demonstrating infeasibility of (FP$_d$). That is, s will satisfy the following properties:

- $s \in A_d$,

- $\|s\|_* = 1$,

78

- $\mathrm{dist}(s, \partial(C_Y^* \cap \{s : b^t s \leq 0\})) \geq \frac{\check{\beta}_{\check{r}}}{66\sqrt{2}\mathcal{C}(d)\max\{\|d\|/\nu, \nu/\|d\|\}} \cdot \min\left\{1, \frac{\nu}{\|b\|}\right\}.$

Proof: The proof is an immediate consequence of Lemmas 4.3 and 4.4. ∎

4.3 Case 3: C_X and C_Y are both regular

When both cones C_X and C_Y are regular, the systems (FP$_d$) and (SA$_d$) are written as follows:

$$\begin{array}{llll} (\mathrm{FP}_d) & b - Ax \in C_Y & (\mathrm{SA}_d) & A^t s \in C_X^* \\ & x \in C_X & & s \in C_Y^* \\ & & & b^t s < 0. \end{array} \qquad (4.16)$$

It is possible to convert the system (FP$_d$) of (4.16) into a system of the form (4.1) by using the transformation

$$b - Ax - \nu y = 0$$
$$(x, y) \in C_X \times C_Y,$$

where $\nu > 0$ is a fixed constant, and then apply algorithm CLS1 to the resulting system (see [36] and Appendix A for the analysis of the properties of the above system); this essentially is the approach we use in developing algorithm ICLS3. However, for the algorithm FCLS3, it is possible to use a different transformation, similar to the one used in algorithm FCLS2, which leads to a nicer complexity result. We will assume in this section that X is an n-dimensional Euclidean space with Euclidean norm $\|x\| = \|x\|_2$ for $x \in X$ (and therefore the dual norm is also $\|q\|_* = \|q\|_2$ for $q \in X^*$), and Y is an m-dimensional Euclidean space with Euclidean norm $\|y\| = \|y\|_2$ for $y \in Y$ (and therefore the dual norm is also $\|s\|_* = \|s\|_2$ for $s \in Y^*$).

4.3.1 Algorithm FCLS3.

Algorithm FCLS3 is designed to find a reliable solution of the system (FP$_d$) of (4.16)

when (FP$_d$) is feasible. Consider the following compact form reformulation of (SA$_d$):

$$
\begin{array}{rcll}
b^t s & +\nu t & = 0 \\
-A^t s & +\nu q & = 0 \\
e^t s & +f^t q & +t & = 1 \\
s \in C_Y^* & q \in C_X^* & t \geq 0,
\end{array}
\tag{4.17}
$$

where $\nu > 0$ is a fixed constant. The alternative system to (4.17) is given by

$$
\begin{array}{rcl}
b\theta & -Ax & \in \mathrm{int}C_Y \\
& \nu x & \in \mathrm{int}C_X \\
\nu\theta & & > 0.
\end{array}
\tag{4.18}
$$

System (4.18) is of the form (HCI) under the following assignments:

- $M = \begin{bmatrix} b^t & 0 & \nu \\ -A^t & \nu I & 0 \end{bmatrix}$

- $C = C_Y^* \times C_X^* \times \Re_+,$

with norms defined as follows:

- $\|(s, q, t)\| = \|s\|_* + \|q\|_* + |t|, \quad (s, q, t) \in Y^* \times X^* \times \Re,$

- $\|(j, q)\| = \|(j, q)\|_2 = \sqrt{j^2 + \|q\|_2^2}, \quad (j, q) \in \Re \times X^*.$

Then the norm linearization vector for C is easily seen to be $\bar{u} = (e, f, 1)$ with $\beta_C = \min\{\tau, \breve{\tau}\}$.

Proposition 4.5 *Suppose (FP$_d$) of (4.16) is feasible and $\rho(d) > 0$. Then the system (4.18) is feasible, and we have*

$$
\max\{\|d\|, \nu\} \leq \|M\| \leq \sqrt{2}\max\{\|d\|, \nu\},
$$

$$
\frac{\rho(d)}{5\max\{1, \|d\|/\nu\}} \leq r(M) \leq \frac{\sqrt{2}\rho(d)}{\breve{\tau}} \leq \frac{\sqrt{2}\rho(d)}{\min\{\tau, \breve{\tau}\}}.
$$

Proof: See Appendix B, Proposition B.5. ∎

80

We use algorithm (HCI) to find a sufficiently interior solution of the system (4.18) and transform its output into a reliable solution of (FP$_d$), as described below:

Algorithm FCLS3

- *Data: $d = (A, b)$*

    Step 1   Apply algorithm HCI to the system (4.18). The algorithm will return a vector $(\tilde{\theta}, \tilde{x})$

    Step 2   Define $\hat{x} = \frac{\tilde{x}}{\tilde{\theta}}$. Return $\hat{x}$ (a reliable solution of (FP$_d$)).

Lemma 4.5 *Suppose (FP$_d$) of (4.16) is feasible and $\rho(d) > 0$. Then algorithm FCLS3 will terminate in at most*

$$\left\lfloor 800\mathcal{C}(d)^2 \left(\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\tau, \check{\tau}\}} \right)^2 \right\rfloor \tag{4.19}$$

iterations of GVNA. The output $\hat{x}$ will satisfy:

1. *$\hat{x} \in X_d$,*

2. *$\|\hat{x}\| \leq 10\mathcal{C}(d)\frac{\max\{1, \nu/\|d\|\}}{\min\{\tau, \check{\tau}\}}$,*

3. *$\operatorname{dist}(\hat{x}, \partial X_d) \geq \frac{1}{10\sqrt{2}\mathcal{C}(d)} \frac{\min\{\tau, \check{\tau}\}}{\max\{\|d\|/\nu, \nu/\|d\|\}}$,*

4. *$\frac{\|\hat{x}\|}{\operatorname{dist}(\hat{x}, \partial X_d)} \leq 10\sqrt{2}\mathcal{C}(d)\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\tau, \check{\tau}\}}$.*

Proof: From Theorem 3.1 we conclude that algorithm HCI in Step 1 will terminate in at most

$$\left\lfloor \frac{16\|M\|^2}{\beta_\mathcal{C}^2 r(M)^2} \right\rfloor \leq \left\lfloor \frac{16 \cdot 2\max\{\|d\|, \nu\}^2 \cdot 25\max\{1, \|d\|/\nu\}^2}{\min\{\tau, \check{\tau}\}^2 \rho(d)^2} \right\rfloor$$

$$= \left\lfloor 800\mathcal{C}(d)^2 \frac{\max\{\|d\|/\nu, \nu/\|d\|\}^2}{\min\{\tau, \check{\tau}\}^2} \right\rfloor$$

iterations of GVNA, which establishes the first statement of the lemma.

Define

$$\gamma \triangleq \operatorname{dist}\left((b\tilde{\theta} - A\tilde{x}, \nu\tilde{x}, \nu\tilde{\theta}), \partial(C_Y \times C_X \times \Re_+) \right).$$

Using Corollary 3.1 we conclude that the vector $(\tilde{\theta}, \tilde{x})$ returned by algorithm HCI in Step 1 satisfies:

$$\|(\tilde{\theta}, \tilde{x})\| = 1, \ B(b\tilde{\theta} - A\tilde{x}, \gamma) \subset C_Y, \ B(\nu\tilde{x}, \gamma) \subset C_X, \ \nu\tilde{\theta} \geq \gamma,$$

$$\gamma \geq \frac{\beta_C r(M)}{2} \geq \frac{\min\{\tau, \breve{\tau}\}\rho(d)}{10 \max\{1, \|d\|/\nu\}}.$$

In particular, $\tilde{\theta} \geq \frac{\gamma}{\nu} > 0$, so that $\hat{x}$ is well-defined, and $b - A\hat{x} \in C_Y, \ \hat{x} \in C_X$, which establishes 1. Also,

$$\|\hat{x}\| = \frac{\|\tilde{x}\|}{\tilde{\theta}} \leq \frac{1}{\tilde{\theta}} \leq \frac{\nu}{\gamma} \leq \frac{10\nu \max\{1, \|d\|/\nu\}}{\min\{\tau, \breve{\tau}\}\rho(d)} = 10\mathcal{C}(d)\frac{\max\{1, \nu/\|d\|\}}{\min\{\tau, \breve{\tau}\}},$$

proving 2.

To prove 3, let $r \triangleq \frac{\gamma}{\sqrt{2}\max\{\|d\|,\nu\}}(1 + \|\hat{x}\|)$, where γ is as above. Then it is easy to see that

$$r \geq \frac{1}{10\sqrt{2}\mathcal{C}(d)} \cdot \frac{\min\{\tau, \breve{\tau}\}}{\max\{\|d\|/\nu, \nu/\|d\|\}}.$$

Further, let $p \in X$ be an arbitrary vector satisfying $\|p\| \leq r$. Then

$$\|\tilde{\theta}\nu p\| \leq \tilde{\theta}\nu r = \tilde{\theta}\nu\frac{\gamma}{\sqrt{2}\max\{\|d\|,\nu\}}(1 + \|\hat{x}\|) = \nu\frac{\gamma}{\sqrt{2}\max\{\|d\|,\nu\}}(\tilde{\theta} + \|\tilde{x}\|) \leq \gamma,$$

and we conclude that $\nu\tilde{x} + \nu\tilde{\theta}p \in C_X$. Thus, $\hat{x} = \frac{\tilde{x}}{\tilde{\theta}}$ satisfies $\hat{x} + p \in C_X$, and so $B(\hat{x}, r) \subset C_X$. By an almost identical argument, for any $p \in X$ such that $\|p\| \leq r$, $b - A(\hat{x} + p) \in C_Y$, and therefore, $B(\hat{x}, r) \subset X_d$, proving 3.

Finally,
$$\frac{\|\hat{x}\|}{r} = \frac{\|\hat{x}\|\sqrt{2}\max\{\|d\|,\nu\}}{\gamma(1 + \|\hat{x}\|)} \leq \frac{\sqrt{2}\max\{\|d\|,\nu\}}{\gamma}$$

$$\leq \frac{10\sqrt{2}\max\{\|d\|,\nu\}\max\{1, \|d\|/\nu\}}{\min\{\tau, \breve{\tau}\}\rho(d)} = 10\sqrt{2}\mathcal{C}(d)\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\tau, \breve{\tau}\}},$$

which proves 4. ∎

4.3.2 Algorithm ICLS3.

Algorithm ICLS3 is designed to compute a reliable solution of (SA_d) of (4.16) when (FP_d) is infeasible. Consider the following compact form reformulation of (FP_d):

$$
\begin{aligned}
-b\theta + Ax + \nu y &= 0 \\
\theta + \bar{f}^t x + \bar{e}^t y &= 1 \\
\theta \geq 0, \ x \in C_X, \ y &\in C_Y,
\end{aligned}
\tag{4.20}
$$

where $\nu > 0$ is a fixed constant. The alternative system to (4.20) is given by

$$
\begin{aligned}
-b^t s \ &> 0 \\
A^t s \ &\in \text{int} C_X^* \\
\nu s \ &\in \text{int} C_Y^*.
\end{aligned}
\tag{4.21}
$$

System (4.21) is of the form (HCI) under the following assignments:

- $M = \begin{bmatrix} -b & A & \nu I \end{bmatrix}$

- $C = \Re_+ \times C_X \times C_Y$,

with norms defined as follows:

- $\|(\theta, x, y)\| = |\theta| + \|x\| + \|y\|, \ (\theta, x, y) \in \Re \times X \times Y$

- $\|v\| = \|v\|_2, \ v \in Y.$

Then the norm linearization vector for C is easily seen to be $\bar{u} = (1, \bar{f}, \bar{e})$ with $\beta_C = \min\{\beta, \breve{\beta}\}$.

Proposition 4.6 *Suppose (FP_d) of (4.16) is infeasible and $\rho(d) > 0$. Then the system (4.21) is feasible and we have*

$$
\|M\| = \max\{\|d\|, \nu\},
$$

$$
\frac{\rho(d)}{3 \max\{1, \|d\|/\nu\}} \leq r(M) \leq \frac{\rho(d)}{\beta} \leq \frac{\rho(d)}{\min\{\beta, \breve{\beta}\}}.
$$

Proof: See Appendix B, Proposition B.6. ∎

We use algorithm (HCI) to find a sufficiently interior solution of the system (4.21) and show that it is a reliable solution of (SA$_d$), as described below:

<div align="center">

Algorithm ICLS3

</div>

- *Data: $d = (A, b)$*

  Step 1 Apply algorithm HCI to the system (4.21). The algorithm will return a vector s.

  Step 2 Return s (a reliable solution of (SA$_d$)).

Lemma 4.6 *Suppose (FP$_d$) of (4.16) is infeasible and $\rho(d) > 0$. Then algorithm ICLS3 will terminate in at most*

$$\left\lfloor 144\mathcal{C}(d)^2 \left(\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\beta, \check{\beta}\}} \right)^2 \right\rfloor \tag{4.22}$$

iterations of GVNA. The output s will satisfy $s \in A_d$ and

$$\frac{\|s\|}{\operatorname{dist}(s, \partial A_d)} \leq 6\mathcal{C}(d) \frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\beta, \check{\beta}\}}$$

Proof: From Theorem 3.1 we conclude that algorithm HCI in Step 1 will terminate in at most

$$\left\lfloor \frac{16\|M\|^2}{\beta_C^2 r(M)^2} \right\rfloor \leq \left\lfloor \frac{16\max\{\|d\|, \nu\}^2 \cdot 9\max\{1, \|d\|/\nu\}^2}{\min\{\beta, \check{\beta}\}^2 \rho(d)^2} \right\rfloor$$

$$= \left\lfloor 144\mathcal{C}(d)^2 \left(\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\beta, \check{\beta}\}} \right)^2 \right\rfloor$$

iterations of GVNA, which establishes the first statement of the lemma.

Next, from Theorem 3.1 we conclude that the output s satisfies $\|s\|_2 = 1$ and

$$B(s, \alpha) \subset S_M, \text{ where } \alpha \geq \frac{\beta_C r(M)}{2\|M\|} \geq \frac{1}{6\mathcal{C}(d)} \frac{\min\{\beta, \check{\beta}\}}{\max\{\|d\|/\nu, \nu/\|d\|\}}.$$

Since $S_M \subseteq A_d$, the second statement of the lemma follows. ∎

4.3.3 Algorithm CLS3

Similarly to algorithm CLS1, algorithm CLS3 described below is a combination of algorithms FCLS3 and ICLS3. Algorithm CLS3 is designed to solve the system (FP_d) of (4.16) by either finding a reliable solution of (FP_d) or demonstrating the infeasibility of (FP_d) by finding a reliable solution of (SA_d). The formal description of algorithm CLS3 is as follows:

<div align="center">Algorithm CLS3</div>

- *Data:* $d = (A, b)$

  Step 1 Run algorithms FCLS3 and ICLS3 in parallel on the data set $d = (A, b)$, until one of them terminates.

  Step 2 If algorithm FCLS3 terminates first, return its output $\hat{x}$. If algorithm ICLS3 terminates first, return its output s.

Combining the complexity results for algorithms FCLS3 and ICLS3 from Lemmas 4.5 and 4.6 we obtain the following complexity analysis of algorithm CLS3:

Theorem 4.3 *Suppose that $\rho(d) > 0$. If the system (FP_d) of (4.16) is feasible, algorithm CLS3 will terminate in at most*

$$2 \left\lceil 800 \mathcal{C}(d)^2 \left(\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\tau, \breve{\tau}\}} \right)^2 \right\rceil$$

iterations of GVNA, and will return a reliable solution $\hat{x}$ of (FP_d). That is, $\hat{x}$ will have the following properties:

- $\hat{x} \in X_d$,

- $\|\hat{x}\| \leq 10 \mathcal{C}(d) \frac{\max\{1, \nu/\|d\|\}}{\min\{\tau, \breve{\tau}\}}$,

- $\text{dist}(\hat{x}, \partial X_d) \geq \frac{1}{10\sqrt{2}\mathcal{C}(d)} \frac{\min\{\tau, \breve{\tau}\}}{\max\{\|d\|/\nu, \nu/\|d\|\}}$,

- $\frac{\|\hat{x}\|}{\text{dist}(\hat{x}, \partial X_d)} \leq 10\sqrt{2}\mathcal{C}(d) \frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\tau, \breve{\tau}\}}$.

If the system (FP_d) is infeasible, algorithm CLS2 will terminate in at most

$$2 \left\lceil 144 \mathcal{C}(d)^2 \left(\frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\beta, \check{\beta}\}} \right)^2 \right\rceil$$

iterations of GVNA, and will return a reliable solution s of (SA_d), thus demonstrating infeasibility of (FP_d). That is, s will satisfy the following properties:

- $s \in A_d$,

- $\frac{\|s\|}{\text{dist}(s, \partial A_d)} \leq 6\mathcal{C}(d) \frac{\max\{\|d\|/\nu, \nu/\|d\|\}}{\min\{\beta, \check{\beta}\}}.$

Proof: The proof is an immediate consequence of Lemmas 4.5 and 4.6. ∎

Chapter 5

Other Measures of Conditioning for Conic Linear Systems

5.1   Overview and Preliminaries

In this chapter we will motivate and define a new measure of conditioning, μ_d, for a conic linear system (FP$_d$) in special form. We will discuss the relationship of μ_d to the condition number $\mathcal{C}(d)$ as well as other measures of conditioning for linear programming that have recently received attention in the literature, and we will study the role of μ_d in the complexity of an interior-point method and a version of the ellipsoid algorithm for solving (FP$_d$).

It has been established that the condition number $\mathcal{C}(d)$ plays an important role in various aspects of analyzing conic linear systems. As discussed in Chapter 1, the condition number $\mathcal{C}(d)$ of a conic linear system (FP$_d$) is connected to such properties of (FP$_d$) as the size of solutions of the system and geometry of the feasible region (see, for example, the discussion of existence of reliable solutions). $\mathcal{C}(d)$ also plays an important role in the complexity analysis of various algorithms for solving (FP$_d$), such as interior-point methods (see, for example, Renegar [41]), the ellipsoid algorithm (see Freund and Vera, [14]), and elementary algorithms, as described in Chapter 4 of this thesis. Other important issues in

the analysis of the system (FP_d) that can be addressed through the condition number $\mathcal{C}(d)$ include sensitivity of solutions of (FP_d) to data perturbations (see Renegar, [39]), numerical properties of algorithms for solving (FP_d) (see Vera, [57]), etc.

Nevertheless, two different (but related) issues motivated the study of other measures of conditioning for conic linear systems, which is presented in the remaining chapters of this thesis.

In part, this study was prompted by the recent developments in interior-point algorithms for linear programming and the associated measures of conditioning. In particular, two measures, $\bar{\chi}_d$ and σ_d, were used in the analysis of interior-point algorithms for linear programming (see, for example, Vavasis and Ye [51, 52, 53]). These measures, defined in Section 5.3, provide a new perspective on the analysis of linear programming problems; for example, similarly to the condition number $\mathcal{C}(d)$, they do not require the data for the problem to be rational. Hence, the first motivating aspect of this research was to study the relationship between these measures and $\mathcal{C}(d)$, and to consider the possibility of extending these measures beyond the realm of linear problems to general conic linear systems. As a result, a new measure of conditioning μ_d was introduced for conic linear systems in special form; as demonstrated by Proposition 5.1, it can be viewed as a generalization of σ_d to conic linear systems. Furthermore, the results in Section 5.3 completely describe the relationships between all of the abovementioned measures for the case when the system (FP_d) is in fact a linear programming feasibility problem.

The second motivating aspect has to do with one of the potential drawbacks of studying the properties of the system (FP_d), especially the properties of the feasible region of (FP_d), through the condition number $\mathcal{C}(d)$. In many cases there can be several elements of the data space $\mathcal{D} = \{d = (A, b) : A \in L(X, Y), \ b \in Y\}$ that give rise to equivalent conic linear systems. Suppose, for example, that two data instances, $d \in \mathcal{D}$ and $\tilde{d} \in \mathcal{D}$ define systems (FP_d) and $(\text{FP}_{\tilde{d}})$, respectively, that are equivalent in the sense that $X_d = X_{\tilde{d}}$ (we can think of the systems (FP_d) and $(\text{FP}_{\tilde{d}})$ as different formulations of the same feasibility problem (FP)). Since the condition number $\mathcal{C}(d)$ is, in general, different from $\mathcal{C}(\tilde{d})$, analyzing the properties of solutions and the feasible region of the problem (FP) will lead to different results, depending on which formulation, (FP_d) or $(\text{FP}_{\tilde{d}})$, is being considered. This issue is

discussed in more detail in Chapter 6, where we show that the new measure μ_d is in fact invariant under changes in formulation of a conic linear system. Therefore, using the results of this chapter, many properties of (FP$_d$) can be analyzed independently of the formulation. (It is also interesting to note that the measures $\bar{\chi}_d$ and σ_d possess the same invariance property).

We will make some further assumptions throughout the rest of this thesis. We will be working with the convex feasibility problem of the form

$$(\text{FP}_d) \quad \begin{aligned} Ax &= b \\ x &\in C_X, \end{aligned} \tag{5.1}$$

where, as before, $A \in L(X, Y)$ is a linear operator between the n-dimensional normed linear vector space X and the m-dimensional normed linear vector space Y (with norms $\|x\|$ for $x \in X$ and $\|y\|$ for $y \in Y$, respectively), $C_X \subset X$ is a closed convex cone, $C_X \neq X$, and $b \in Y$. This form of system (FP$_d$) is equivalent to the system (2.1) with $C_Y = \{0\}$. Note that for the case when $C_Y \neq \{0\}$, the system (2.1) can be transformed into the form (5.1) by appropriately adding "slack variables" (see Appendix A for further details of this transformation).

We will assume throughout this and the following chapters that the system (FP$_d$) of (5.1) is feasible. At this point, we do not make any assumptions on the cone C_X and the norms on the spaces X and Y, unless stated otherwise. (We will make some additional assumptions later in this chapter and in Chapter 6.)

Recall from Theorem 2.1 that under the above assumptions the distance to ill-posedness $\rho(d)$ of the system (FP$_d$) can be expressed via the following characterization:

$$\begin{aligned} \rho(d) = \quad & \min_{\substack{v \in Y \\ \|v\| \leq 1}} \quad \max_{\theta, x, \phi} \quad \phi \\ & \text{s.t} \quad b\theta - Ax = v\phi \\ & \qquad \theta \geq 0, \ x \in C_X \\ & \qquad |\theta| + \|x\| \leq 1. \end{aligned} \tag{5.2}$$

It is important to note that the characterization of $\rho(d)$ presented by (5.2) has an elegant geometric interpretation. In particular, $\rho(d)$ can be viewed as the radius of the largest ball (defined with respect to the norm on the space Y) centered at 0 that is contained in the set

$$\mathcal{H}_d \triangleq \{y = b\theta - Ax : \theta \ge 0, \ x \in C_X, \ |\theta| + \|x\| \le 1\}$$

(as a consequence, observe that $\rho(d) > 0$ precisely when $0 \in \mathrm{int}\mathcal{H}_d$). This interpretation, noted by Renegar in [41] and Freund and Vera in [16], will serve as an important tool in developing further understanding of the properties of the system (FP_d).

5.2   The Symmetry Measure and Its Properties

We begin by defining the *symmetry* of a set with respect to a point.

Definition 5.1 *Let $D \subset Y$ be a bounded convex set. For $y \in \mathrm{int}D$ we define $\mathrm{sym}(D,y)$ to be the symmetry of D about y, i.e.,*

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

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

This definition of symmetry is similar to the definition in [41]. Observe that $\mathrm{sym}(D,y) \in [0,1]$ with $\mathrm{sym}(D,y) = 1$ if D is perfectly symmetric about y, and $\mathrm{sym}(D,y) = 0$ precisely when $y \in \partial D$. Moreover, the definition of $\mathrm{sym}(D,y)$ is independent of the norm on the space Y.

The following lemma will be used later on:

Lemma 5.1 *Suppose D is a compact convex set with a non-empty interior, and let $y \in \mathrm{int}D$. Then there exists an extreme point w of D such that*

$$\mathrm{sym}(D,y) = \mathrm{sym}_w(D,y) \triangleq \sup\{t \mid y - t(w - y) \in D\}.$$

Proof: Define $f(w) = \mathrm{sym}_w(D,y) = \sup\{t \mid y - t(w - y) \in D\}$. We will show that $f(w)$ is quasiconcave on D. This will imply (see [3], Section 3.5.3) that the minimum of $f(w)$ is

achieved at an extreme point of D.

Suppose $w = \lambda w_1 + (1 - \lambda)w_2$ for some $\lambda \in [0, 1]$ and some $w_1, w_2 \in D$. By definition of $f(w)$, $y - f(w_1)(w_1 - y) \in D$ and $y - f(w_2)(w_2 - y) \in D$. We would like to show that $f(w) \geq \min\{f(w_1), f(w_2)\}$. Without loss of generality, assume that $f(w_1) \leq f(w_2)$. Then $y - f(w_1)(w_2 - y) \in D$. It suffices to show that $y - f(w_1)(w - y) \in D$. Indeed,

$$y - f(w_1)(w - y) = y - f(w_1)(\lambda(w_1 - y) + (1 - \lambda)(w_2 - y))$$

$$= \lambda(y - f(w_1)(w_1 - y)) + (1 - \lambda)(y - f(w_1)(w_2 - y)) \in D. \ \blacksquare$$

To define the *symmetry measure* of the problem (FP$_d$) notice that if (FP$_d$) is feasible,

$$0 \in \mathcal{H}_d \triangleq \{b\theta - Ax : \theta \geq 0, \ x \in C_X, \ |\theta| + \|x\| \leq 1\}. \tag{5.3}$$

Hence, the following quantity is well-defined:

Definition 5.2 *Suppose the system (FP$_d$) is feasible. We define*

$$\mu_d \triangleq \frac{1}{\text{sym}(\mathcal{H}_d, 0)}. \tag{5.4}$$

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 (FP$_d$) is ill-posed.

We now establish two results that characterize geometric properties of the system (FP$_d$) in terms of μ_d. Recall that X_d denotes the set of solutions of (FP$_d$). Theorem 5.1 establishes a bound on the size of a solution of (FP$_d$) in terms of μ_d; this result is similar to the bound in terms of the condition number $\mathcal{C}(d)$ established by Renegar [39]. Theorem 5.2 demonstrates existence of a reliable solution of (FP$_d$). This result is similar to the one demonstrated by Freund and Vera in [16], 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 μ_d rather then $\mathcal{C}(d)$. Also, unlike for the condition number $\mathcal{C}(d)$, we can establish a converse result for μ_d, namely, if (FP$_d$) has a reliable solution, then μ_d is bounded by a function of

the size of the solution, its distance to the boundary of the cone C_X and the ratio of the

above quantities. This result is proven in Theorem 5.3.

Theorem 5.1 *Suppose $\mu_d < \infty$. Then there exists $x \in X_d$ such that $\|x\| \leq \mu_d$.*

Proof: Consider the following optimization problem:

$$
\begin{array}{rclcrl}
\min & \|x\| & = & \min & \|x\| \\
& x \in X_d & & & Ax = b & \qquad (5.5) \\
& & & & x \in C_X.
\end{array}
$$

Let z^* denote the optimal value of the problem (5.5). We need to show that $z^* \leq \mu_d$.

The Lagrange dual of this problem can be written as

$$
\begin{array}{rl}
\max & -b^t \lambda \\
& \|A^t \lambda - \pi\|_* \leq 1 \qquad (5.6) \\
& \pi \in C_X^*,
\end{array}
$$

and the optimal value of the problem (5.6) is equal to z^*, because strong duality between problems (5.5) and (5.6) can easily be shown to hold. If $z^* = 0$, it immediately follows that $z^* \leq \mu_d$. Assume that $z^* > 0$.

Applying the definition of symmetry to (5.4), we can characterize μ_d as the optimal value of the following optimization problem:

$$
\begin{array}{rl}
\frac{1}{\mu_d} = \mathrm{sym}(\mathcal{H}_d, 0) = & \min_{v \in \mathcal{H}_d} \quad \max \quad \psi \\
& \qquad\qquad \text{s.t.} \quad -\psi v \in \mathcal{H}_d
\end{array}
$$

$$
\begin{array}{rl}
= & \min_{v \in \mathcal{H}_d} \quad \max_{\delta, x, \psi} \quad \psi \\
& \qquad \text{s.t.} \quad b\theta - Ax + \psi v = 0 \qquad (5.7) \\
& \qquad\qquad\quad \theta \geq 0, \; x \in C_X \\
& \qquad\qquad\quad |\theta| + \|x\| \leq 1.
\end{array}
$$

Taking the Lagrange dual of the inner optimization problem in (5.7) we obtain another

characterization of μ_d:

$$
\begin{aligned}
\frac{1}{\mu_d} = \quad & \min_{v \in \mathcal{H}_d} \quad \min_{s,q,g} \quad \max\{\|A^t s - q\|_*, \ |b^t s + g|\} \\
& \text{s.t.} \quad q \in C_X^* \\
& \qquad\quad g \geq 0 \\
& \qquad\quad -v^t s \geq 1.
\end{aligned}
\tag{5.8}
$$

Let (λ, π) be an optimal solution of (5.6). Define $(s, q, g) = \frac{1}{z^*}(\lambda, \pi, z^*)$. Then $q \in C_X^*$, $g = 1 > 0$, $-b^t s = 1$, and so (q, g, s) is a feasible solution for the inner optimization problem of (5.8) for $v = b \in \mathcal{H}_d$. Therefore

$$
\frac{1}{\mu_d} \leq \max\{\|A^t s - q\|_*, |-b^t s + g|\} = \frac{\|A^t \lambda - \pi\|_*}{z^*} \leq \frac{1}{z^*}.
$$

Hence $z^* \leq \mu_d$, and so there exists a point $\hat{x} \in X_d$ such that $\|\hat{x}\| \leq \mu_d$. ∎

Theorem 5.2 *Suppose C_X is a regular cone and $\mu_d < \infty$. 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. *$\mathrm{dist}(\hat{x}, \partial C_X) \geq \frac{\tau}{2\mu_d + 1}$,*

4. *$\frac{\|\hat{x}\|}{r} \leq \frac{2\mu_d + 1}{\tau}$,*

where τ is the width of the cone C_X.

Proof: Let $(\tilde{\theta}, \tilde{x}) = \frac{1}{2}(1, u)$, where u is the norm linearization vector for the cone C_X^*. Then $(\tilde{\theta}, \tilde{x}) \in \Re_+ \times C_X$ and $|\tilde{\theta}| + \|\tilde{x}\| \leq 1$, so that $\frac{1}{2}b - \frac{1}{2}Au \in \mathcal{H}_d$. From the definition of μ_d we

conclude that $-\frac{1}{\mu_d}\left(\frac{1}{2}b - \frac{1}{2}Au\right) \in \mathcal{H}_d$, whereby there exists $(\bar{\theta}, \bar{x})$ satisfying

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

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 5.1 and 5.2 that, much like for the condition number $\mathcal{C}(d)$, if the symmetry measure μ_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 can also establish a converse result, that is, if the feasible region X_d possesses nice geometry, namely, contains a reliable solution, then μ_d can be nicely bounded by a function of the parameters associated with the reliable solution. This result is quite specific to μ_d; no result of such type is possible for the condition number $\mathcal{C}(d)$.

Theorem 5.3 *Suppose C_X is a regular cone and there exist $\hat{x} \in X_d$ and $r > 0$ such that* $\mathrm{dist}(\hat{x}, \partial C_X) \geq r$. *Let*

$$\gamma = \max\left\{\|\hat{x}\|, \frac{1}{r}, \frac{\|\hat{x}\|}{r}\right\}. \tag{5.9}$$

Then $\mu_d \leq 1 + 2\gamma$.

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

94

$\frac{\pi}{\delta+\pi}$, i.e., for an arbitrary vector $y \in \mathcal{H}_d$ we would like to show that $-\frac{\pi}{\delta+\pi}y \in \mathcal{H}_d$.

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

$$|\check{\theta}| + \|\check{x}\| \leq \frac{1}{\delta+\pi}\left(1 + \pi\|\bar{x}\| + \|\hat{x}\|\right) \leq 1,$$

and therefore $-\frac{\pi}{\delta+\pi}y \in \mathcal{H}_d$, establishing that $\frac{1}{\mu_d} \geq \frac{\pi}{\delta+\pi}$. Hence

$$\mu_d \leq 1 + \frac{\delta}{\pi} = 1 + \frac{\|\hat{x}\|+1}{\min\{r,1\}} = 1 + \frac{1}{\min\{r,1\}} + \frac{\|\hat{x}\|}{\min\{r,1\}} \leq 1 + \max\{\gamma,1\} + \gamma \leq 1 + 2\gamma.$$

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

We conclude this section by establishing a relationship between μ_d and $\mathcal{C}(d)$. As demonstrated in Theorem 5.4, if an instance of (FP_d) is "well-conditioned" in the sense that $\mathcal{C}(d)$ is small, then μ_d is also small. This relationship, however, is one-sided, since μ_d may carry no upper-bound information about $\mathcal{C}(d)$. In particular, in Remark 5.1 we exhibit a sequence of instances of (FP_d) with $\mathcal{C}(d)$ becoming arbitrarily large, while μ_d remains fixed.

Theorem 5.4

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

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

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

proving the theorem. ∎

Remark 5.1 μ_d *may carry no upper-bound information about* $\mathcal{C}(d)$.

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

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

Furthermore, let $C = \Re_+^4$ and $\|x\| = \|x\|_1$ for $x \in X$ and $\|y\| = \|y\|_2$ for $y \in Y$. Consider the values of parameter $\epsilon \in (0, 1]$. The set $\mathcal{H}_{d_\epsilon}$ is perfectly symmetric about 0, so $\mu_{d_\epsilon} = 1$ for any value of ϵ. 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 μ_d remains constant.

So far, we have made no assumptions on the norm on the space Y; in fact, it can be easily seen that μ_d is invariant under the changes in the norm on Y, which is not true for $\mathcal{C}(d)$. We conclude this section by providing another interpretation of the relationship between the measures μ_d and $\mathcal{C}(d)$. As Theorem 5.5 indicates, when the space Y is endowed with an appropriate norm, the two measures are within a constant factor of each other.

We will denote

$$\mathcal{T}_d \stackrel{\triangle}{=} -\mathcal{H}_d \cap \mathcal{H}_d. \tag{5.10}$$

Theorem 5.5 *Suppose* C_X *is regular and* $\rho(d) > 0$. *If the norm on the space* Y *is such that*

$$\mathcal{T}_d = B(0, 1) \subset Y,$$

then $\rho(d) = 1$ *and*

$$\mathcal{C}(d) \leq \frac{\mu_d}{\delta},$$

where δ *is the norm approximation coefficient of the cone* C_X.

Proof: First observe that since $\rho(d) > 0$, the set $\mathcal{T}_d$ is a symmetric bounded convex set such that $0 \in \mathrm{int}\mathcal{T}_d$, which implies that the norm on the space Y can be uniquely defined

so that the set $\mathcal{T}_d$ is the unit ball in the space Y (see, for example, [43, Theorem 15.2]). Furthermore, characterization (5.2) 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 2.1. To derive the second inequality, suppose that $y \in \mathcal{H}_d$. Then by the definition of μ_d, $-\frac{1}{\mu_d}y \in \mathcal{H}_d$. Moreover, $\frac{1}{\mu_d}y \in \mathcal{H}_d$ since $\mu_d \in (0,1]$. 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 $\max\{\|y\| : y \in \mathcal{H}_d\} = \mu_d$. $\blacksquare$

5.3   The Symmetry Measure and Other Measures of Conditioning for Linear Programming

In this section we will work with the cone $C_X = \Re^n_+$, in which case the problem (FP_d) becomes a linear feasibility problem, and can be written as follows:

$$(\mathrm{FP}_d) \quad \begin{aligned} Ax &= b \\ x &\geq 0, \end{aligned} \tag{5.11}$$

where $x \in \Re^n$, $b \in \Re^m$ and $A \in \Re^{m \times n}$.

In this section we consider the relationship between the symmetry measure μ_d and other measures of conditioning for linear programming.

We will assume in this subsection that the norms are as follows:

$$\|x\| = \|x\|_1, \ x \in \Re^n \text{ and } \|y\| = \|y\|_2, \ y \in \Re^m.$$

For simplicity of notation, we define an "expanded" matrix $\tilde{A} \triangleq [b; -A] \in \Re^{m \times (n+1)}$. Notice that

$$\|\tilde{A}\| \triangleq \max\{\|b\theta - Ax\| : \|(\theta, x)\|_1 \leq 1\} = \|d\|.$$

97

In addition to μ_d and $\mathcal{C}(d)$, we associate two other measures of conditioning with this problem. The first measure, σ_d, is closely related to the complexity measure introduced and used in the complexity analysis of an interior-point algorithm for solving (FP_d) by Vavasis and Ye (see, for example, [51]).

$$
\sigma_d \triangleq \quad \min_{j=1,\ldots,n+1} \quad \max_{x} \quad \begin{array}{l} e_j^t w \\ \tilde{A}w = 0 \\ e^t w = 1 \\ w \geq 0, \end{array} \tag{5.12}
$$

where e_j, $j = 1, \ldots, n+1$ denotes the jth unit vector and $e \in \Re^{n+1}$ is the vector of all ones.

The second measure of conditioning, $\bar{\chi}_d$, is closely related to the complexity measure used by Vavasis and Ye ([52, 53]) and Megiddo et. al. ([29]) 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

$$
\bar{\chi}_d = \max\{\|B^{-1}\tilde{A}\| : B \in \mathcal{B}(\tilde{A})\}, \tag{5.13}
$$

where $\mathcal{B}(\tilde{A})$ is the set of all bases (i.e., $m \times m$ non-singular sub-matrices) of $\tilde{A}$. See, for instance, [48] for the proof of the equivalence of these characterizations.

In this section we state both known and new results on the relationship between these measures of conditioning.

It has been established by Vavasis and Ye [51] that σ_d and $\bar{\chi}_d$ are related by the following inequality:

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

On the other hand, Tunçel in [49] established that in general σ_d may carry no upper-bound information about $\bar{\chi}_d$. Specifically, he provided a family of data instances d_ϵ 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.

We have established a relationship between μ_d and $\mathcal{C}(d)$ in Theorem 5.4 above. Below we establish the relationships between the other pairs of measures μ_d, $\mathcal{C}(d)$, $\bar{\chi}_d$, and σ_d, or provide counterexamples when no such relationship exists, in the spirit of [49].

Remark 5.2 *The condition number $\mathcal{C}(d)$ 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 constant or bounded.

First consider the family of matrices

$$\tilde{A}_\epsilon = \begin{bmatrix} \epsilon & 0 & -\epsilon \\ 0 & 1 & -1 \end{bmatrix}$$

For $\epsilon > 0$ sufficiently small, $\rho(d_\epsilon) = \frac{\epsilon}{\sqrt{\epsilon^2+4}}$. Furthermore, $\|d_\epsilon\| = \sqrt{1+\epsilon^2}$, and so

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

On the other hand, it is not hard to establish using (5.13) 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 ϵ as above. On the other hand it is not hard to establish using (5.13) that for any $\epsilon \in (0,1)$,

$$\bar{\chi}_{d_\epsilon} = \frac{1}{\epsilon} \to +\infty \text{ as } \epsilon \to 0.$$

In the following proposition we establish an exact relationship between the measures μ_d and σ_d.

Proposition 5.1 *Suppose the system (FP$_d$) of (5.11) is feasible. Then*

$$\sigma_d = \frac{1}{1+\mu_d}.$$

99

Proof: Observe that we can redefine σ_d as follows:

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

From Lemma 5.1, 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 $\{\tilde{A}_1, \ldots, \tilde{A}_{n+1}\}$, where $\tilde{A}_j \in \Re^m$ is the jth column of the matrix $\tilde{A}$, we can characterize μ_d as

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

We will now show that for any j,

$$\sigma_j = \frac{1}{1 + \mu_j}. \tag{5.14}$$

Without loss of generality we can consider $j = 1$. Consider the first column of $\tilde{A}$, $\tilde{A}_1$. By definition of μ_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 σ_1. Then $w_1 = \sigma_1$. Observe that, unless $\tilde{A}_1 = 0$, $\sigma_1 < 1$. If $\tilde{A}_1 = 0$, then $\sigma_1 = 1$ and $\frac{1}{\mu_1} = +\infty$, and (5.14) holds as a limiting relationship. If $\tilde{A}_1 \neq 0$, 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, $\frac{1}{\mu_1} \geq \frac{\sigma_1}{1-\sigma_1}$, or $\sigma_1 \leq \frac{1}{\mu_1+1}$. Combining this with the bound in the previous paragraph, we conclude that $\sigma_1 = \frac{1}{\mu_1+1}$, and by similar argument,

$$\sigma_j = \frac{1}{\mu_j + 1},\ j = 1,\ldots,n+1. \tag{5.15}$$

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, $\mu = \mu_j$ and hence

$$\sigma_d = \frac{1}{1 + \mu_d}. \quad \blacksquare$$

The following two remarks, which are easy consequences of Proposition 5.1, establish the remaining relationships between the measures of conditioning.

Remark 5.3 $\mu_d \leq \bar{\chi}_d$. *However, μ_d may carry no upper-bound information about $\bar{\chi}_d$.*

Remark 5.4 $\sigma_d \geq \frac{1}{\mathcal{C}(d)+1}$. *However, σ_d may carry no upper-bound information about $\mathcal{C}(d)$.*

In the light of Proposition 5.1, μ_d can in fact be viewed as a generalization of the Vavasis-Ye measure σ_d to conic linear systems. (Ho in [22] provides an argument indicating that extending $\bar{\chi}_d$ to general conic systems is not possible.)

5.4   The Symmetry Measure and the Complexity of Computing a Solution of (FP$_d$)

In this section we explore the role of the symmetry measure μ_d in the complexity of algorithms for finding a solution of the system (FP$_d$) of (5.1). We briefly indicate how a solution of (FP$_d$) can be found via an interior-point algorithm (when the cone C_X is the domain of a self-concordant barrier function) or via a version of the ellipsoid algorithm (when the cone C_X is represented via a separation oracle). We then study the complexity of these algorithms in terms of μ_d.

In this section we will 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 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 τ.

101

5.4.1 Complexity of an Interior-Point Algorithm for Computing a Solution of (FP$_d$)

In this subsection we will assume that the cone C_X is represented as the closure of the domain of a self-concordant barrier function. In this case a solution of (FP$_d$) can be found using the barrier method developed by Renegar, based on the theory of self-concordant functions of Nesterov and Nemirovskii [31]. Below we briefly describe the method as is it articulated in [42], 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

$$z^* = \inf\{c^t\omega : \omega \in S \cap L\}, \tag{5.16}$$

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 ϑ_f (see [31] and [42] 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

$$c^t\omega_k - \frac{6\vartheta_f}{5\eta_k} \leq z^* \leq c^t\omega_k, k = 0, 1, 2, \ldots \tag{5.17}$$

It follows from the analysis in [42] that if the barrier method is initialized at the point $\omega' \in \text{int}S \cap L$, then it will take at most

$$O\left(\sqrt{\vartheta_f}\ln\left(\frac{\vartheta_f(z_* - z^*)}{\text{sym}(S \cap L, \omega')} \cdot \bar{\eta}\right)\right) \tag{5.18}$$

iterations to bring the value of the barrier parameter η above the threshold of $\bar{\eta} \geq \eta_0$ while maintaining (5.17) (here, $z_* = \sup\{c^t\omega : \omega \in S \cap L\}$). This implies the main convergence

result for the barrier method:

Theorem 5.6 ([42], 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 ϑ_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 ω computed thereafter satisfy $\omega \in \text{int}S \cap L$ and

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

In order to find a solution of (FP$_d$) we will construct a closely related problem of the form (5.16) and apply the barrier method to this problem. This construction was carried out by Renegar in [41], where he analyzed the complexity of solving (FP$_d$) in terms of the condition number $\mathcal{C}(d)$. The optimization problem we consider is

$$
\begin{aligned}
z^* = \quad & \inf \quad t \\
& x, \theta, t \\
& \text{s.t.} \quad b\theta - Ax = t(\tfrac{1}{2}b - \tfrac{1}{2}Au) \\
& \qquad x \in \text{int}C_X \\
& \qquad \|x\| < 1 \\
& \qquad 0 < \theta < 1 \\
& \qquad -1 < t < 2,
\end{aligned}
\tag{5.19}
$$

where u is the norm linearization vector for C_X^*. We will use the barrier method to find a feasible solution $(\hat{\theta}, \hat{x}, \hat{t})$ of (5.19) such that $\hat{t} \le 0$, 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 (FP$_d$). (It is not hard to verify via algebraic manipulation that the point x above is indeed a solution of (FP$_d$).)

Let z_* be the optimal value of the problem obtained from (5.19) 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 $\vartheta_f \le \vartheta_{\tilde{f}} + 5$ (see, for example, [41] or [42] for details). If we define $L \triangleq \left\{(\theta, x, t) : b\theta - Ax = t\left(\frac{1}{2}b - \frac{1}{2}Au\right)\right\}$, then the problem (5.19) is of the form (5.16), and we can apply the barrier method initialized at the point $\omega' = (\theta', x', t') = \left(\frac{1}{2}, \frac{1}{2}u, 1\right)$.

The following proposition provides bounds on all of the parameters necessary in the analysis of the complexity of the barrier method via Theorem 5.6:

Proposition 5.2

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

Proof: The upper bound on z_* and the lower bound on z^* follow trivially from the last constraint of (5.19).

Let $y = \frac{1}{2}b - \frac{1}{2}Au \in \mathcal{H}_d$. From the definition of μ_d we conclude that $-\frac{y}{\mu_d} \in \mathcal{H}_d$, so there exists (θ, x) such that

$$\theta \ge 0, \ x \in C_X, \ |\theta| + \|x\| \le 1, \ b\theta - Ax = -\frac{1}{\mu_d}\left(\frac{1}{2}b - \frac{1}{2}Au\right).$$

Therefore $(\theta, x, -1/\mu_d)$ is in the closure of the feasible set of (5.19), and so $z^* \le -\frac{1}{\mu_d}$.

To establish the last statement of the proposition, we appeal to Proposition 3.3 of Renegar [41], where he established that ω' defined as above satisfies

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

Since $B\left(\frac{1}{2}u, \frac{1}{2}\tau\right) \subset C_X(1)$, it is not hard to verify that $\text{sym}\left(C_X(1), \frac{1}{2}u\right) \ge \frac{\tau}{3}$, establishing

104

the proposition. ∎

Theorem 5.7 *Suppose the barrier method applied to solving (5.19) is initialized at the point* $\left(\frac{1}{2}, \frac{1}{2}u, 1\right)$. *Then within*

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

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, when $\hat{t} \leq 0$, it is easy to check that 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 5.6 and substituting the bounds of Proposition 5.2 into the complexity bound, we conclude that after at most

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

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. ∎

5.4.2   Complexity of the Ellipsoid Algorithm for Computing a Solution of (FP$_d$)

In this subsection we will assume that the cone C_X is represented via a separation oracle. In this case a solution of (FP$_d$) can be found using a version of the ellipsoid algorithm (see, for example, [5] and [21]). Below is a generic theorem for analyzing the ellipsoid algorithm for finding a point ω in a convex set $S \subset \Re^k$ given by a separation oracle.

Theorem 5.8 ([14]) *Suppose that a convex set $S \subset \Re^k$ 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.

We approach solving (FP_d) by considering the following set:

$$S \triangleq \{(\theta, x) : \theta > 0, \ x \in C_X, \ b\theta - Ax = 0\}, \tag{5.20}$$

which is a convex set in the linear space $T \triangleq \{(\theta, x) : b\theta - Ax = 0\}$ of dimension $k = n+1-m$. Observe that it is not hard to construct a separation oracle for S provided a separation oracle for C_X is available. We will use the ellipsoid algorithm to find a point $(\hat{\theta}, \hat{x}) \in S$, and use the obvious transformation $x = \frac{\hat{x}}{\hat{\theta}}$ to transform the output of the algorithm into a solution of (FP_d).

Proposition 5.3 *Let S be as in (5.20). Then there exists a point $(\hat{\theta}, \hat{x}) \in S$ and $\hat{r} > 0$ such that*

$$B((\hat{\theta}, \hat{x}), \hat{r}) \cap \{(\theta, x) : b\theta - Ax = 0\} \subset S,$$

$$\|(\hat{\theta}, \hat{x})\| + \hat{r} \le 3, \ \ and \ \hat{r} \ge \frac{\tau}{2\mu_d}.$$

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

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

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 $\|\hat{\omega}\|_2 + \hat{r} = \sqrt{\left(\bar{\theta} + \frac{1}{2\mu_d}\right)^2 + \left\|\bar{x} + \frac{1}{2\mu_d}u\right\|^2} + \frac{\tau}{2\mu_d} \le 3.$ ∎

The following theorem is an immediate consequence of Theorem 5.8 and Proposition 5.3:

Theorem 5.9 *Suppose the ellipsoid algorithm applied to the set S in the linear space T is 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.

Chapter 6

Pre-Conditioners for Conic Linear Systems

In this chapter we continue the development of the properties of the measure of conditioning μ_d introduced in Chapter 5. We return to the issue of the importance of measures of conditioning of (FP_d) that are "data-independent," first mentioned in Chapter 1, and show that, similarly to $\bar{\chi}_d$ and σ_d, μ_d is in fact independent of the particular data d used to characterize the feasible region of the problem (FP_d). Therefore, the properties of (FP_d) and the complexity of algorithms studied in Chapter 5 can be characterized independently of the data d.

On the other hand, some properties of (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 (i.e., such that $(\mathrm{FP}_{\tilde{d}})$ has the same feasible region as (FP_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. We also construct an algorithm for computing an equivalent data instance whose condition number is within a known factor of this bound, and we analyze its complexity.

6.1 Motivation and Preliminaries

Recall that for a feasible instance of the system

$$(\text{FP}_d) \quad Ax = b$$
$$x \in C_X \tag{6.1}$$

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

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

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

$$(\text{FP}_{Bd}) \quad BAx = Bb$$
$$x \in C_X. \tag{6.2}$$

Notice that since B is non-singular, the systems (FP_d) of (6.1) and (FP_{Bd}) of (6.2) are equivalent in the sense that $X_d = X_{Bd}$. We can view the systems (FP_d) and (FP_{Bd}) as different formulations of the same feasibility problem (FP):

$$(\text{FP}) \text{ find } x \in \mathcal{A} \cap C_X, \tag{6.3}$$

where $\mathcal{A}$ is the affine subspace $\mathcal{A} \triangleq \{x : Ax = b\} = \{x : BAx = Bb\}$. 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 formulations, μ_d and μ_{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 $\mu_d = \mu_{Bd}$. We will show in Proposition 6.1 that, under certain reasonable assumptions, any data instance $\tilde{d}$ equivalent to d can be obtained as $\tilde{d} = Bd$ for an appropriate non-singular matrix B. Based on this observation, we can omit the dependence on the data instance from the notation for the symmetry measure, denoting it simply by μ.

In Chapter 5 we have analyzed some of the properties of the problem (FP$_d$) in terms of μ. In particular, in the analysis of reliable solutions of (FP$_d$), as well as in the analysis of the complexity of an interior-point algorithm and the ellipsoid algorithm for solving (FP$_d$), we have effectively replaced the dependence on the condition number $\mathcal{C}(d)$ by the data-independent measure μ. Therefore it may be beneficial to study the above properties via the symmetry measure μ rather then $\mathcal{C}(d)$. For example, suppose a feasibility problem can be represented via two equivalent systems (FP$_d$) and (FP$_{\tilde{d}}$), and suppose that $\mathcal{C}(d) << \mathcal{C}(\tilde{d})$. If one were to predict, for example, the performance of the interior-point algorithm for solving this problem 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 μ will give 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 the properties of the system which depend on the representation of the problem (FP$_{(\cdot)}$) by a specific data instance d. Such properties include sensitivity of the feasible region to perturbations of the data, numerical properties of algorithms for solving (FP$_d$), etc. Therefore, it is often beneficial to search for a representation of the problem (FP) through a data instance whose condition number is small, and which therefore will have better properties. In other words, if one is given a data instance d, it is often desirable to *pre-condition* the system (FP$_d$), i.e., find another data instance $\tilde{d}$ and the corresponding system (FP$_{\tilde{d}}$) which is equivalent to (FP$_d$), but has $\mathcal{C}(\tilde{d}) < \mathcal{C}(d)$. In this light, we can view the matrix B as above as a *pre-conditioner* for the system (FP$_d$), yielding the pre-conditioned system (FP$_{\tilde{d}}$) with $\tilde{d} = Bd$. Proposition 6.1 implies that, under some assumptions, any data instance $\tilde{d}$ which gives rise to a system equivalent to (FP$_d$) can be expressed as $\tilde{d} = Bd$ for an appropriate pre-conditioner B.

In this chapter we will maintain the assumption that the system (FP_d) is feasible. Further, we assume that the space Y is the m-dimensional Euclidean space $\Re^m$ with the norm $\|y\|$ for $y \in Y$ being the Euclidean norm $\|y\| = \|y\|_2 = \sqrt{y^t y}$ (and therefore, the dual norm $\|s\|_*$ for $s \in Y^*$ also being the Euclidean norm). We assume that $m \geq 2$ (in fact, the case $m = 1$ is trivial since in this case $\mu = \mathcal{C}(d)$, and the issue of pre-conditioning becomes irrelevant). We assume that the cone C_X is a regular cone with the norm approximation coefficient denoted by δ. The coefficient of linearity of the cone C_X is denoted by β and the corresponding norm linearization vector is denoted by $\bar{f}$; the width of the cone C_X is denoted by τ and is attained for (f, τ).

For any matrix $Q \in S_{++}^{m \times m}$ we define E_Q to be the ellipsoid $E_Q \triangleq \{y \in Y : y^t Q^{-1} y \leq 1\}$. An important notion for the results of this chapter is that of an α-rounding of a set.

Definition 6.1 *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 α-rounding of S if*

$$\alpha E_Q \subseteq S \subseteq E_Q,$$

that is, the set S is contained in the ellipsoid E_Q and contains another (smaller) ellipsoid concentric with E_Q. We refer to the parameter α as the tightness *of the rounding E_Q.*

If the set S above satisfies $S = -S$ (i.e., is centrally symmetric) then S possesses a $\frac{1}{\sqrt{m}}$-rounding, i.e., there exists an ellipsoid E_Q such that $\frac{1}{\sqrt{m}} E_Q \subseteq S \subseteq E_Q$. This important fact was established by John in [24]; see also [21] for a detailed discussion of geometric and algorithmic implications. In particular, the ellipsoid of minimum volume containing S is a $\frac{1}{\sqrt{m}}$-rounding of S; furthermore, this ellipsoid is unique and is often referred to as the Löwner-John ellipsoid of S.

In the following section we continue studying the notion of pre-conditioning of the system (FP_d); in particular, Theorem 6.1 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 μ.

6.2 Pre-Conditioners

In this section we will study the properties of pre-conditioners for the system (FP_d) of (6.1). First we establish that any well-posed data instance $\tilde{d} \in \mathcal{D}$ that gives rise to a system equivalent to (FP_d) can be obtained through pre-conditioning of the original data instance d. Therefore, in our study of equivalent re-formulations of (FP_d) we only need to consider pre-conditioning of the data to explore all possible representations.

Proposition 6.1 *Let $d = (A, b) \in \mathcal{D}$ and $\tilde{d} = (\tilde{A}, \tilde{b}) \in \mathcal{D}$ be such that $X_d = X_{\tilde{d}}$, $\rho(d) > 0$ and $\rho(\tilde{d}) > 0$. 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 $x_0 \in \text{int} C_X$. 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 $B \in \Re^{m \times m}$ such that $\tilde{A} = BA$. It is also easy to establish that $\tilde{b} = Bb$, and so $\tilde{d} = Bd$. ∎

In the following two lemmas we develop a geometric interpretation of a pre-conditioner B.

Lemma 6.1 *Let $B \in \Re^{m \times m}$ be a pre-conditioner for the system (FP_d) such that the condition number of the (pre-conditioned) system (FP_{Bd}) is $\mathcal{C}(Bd)$. 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, $\|B^t h\| \leq \rho(Bd)$. 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 6.1 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 $\mathcal{H}_d$.

Recall that the set $\mathcal{T}_d$ was defined in (5.10) as $\mathcal{T}_d \triangleq -\mathcal{H}_d \cap \mathcal{H}_d$.

Lemma 6.2 *Let $Q \in S_{++}^{m \times m}$ be such that E_Q is an α-rounding of $\mathcal{T}_d$. Let $B = Q^{-\frac{1}{2}}$. Then B is a pre-conditioner for the system (FP$_d$) such that*

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

Proof: We will 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 2.1 that

$$\begin{aligned}
\|Bd\| \leq \quad & \tfrac{1}{\delta} \max \quad \|v\| \\
& \text{s.t.} \quad v \in \mathcal{H}_{Bd}.
\end{aligned} \tag{6.4}$$

Let $v \in \mathcal{H}_{Bd}$. Then $y = B^{-1}v \in \mathcal{H}_d$. By the definition of μ, $-\frac{1}{\mu}y \in \mathcal{H}_d$. Moreover, $\frac{1}{\mu}y \in \mathcal{H}_d$ since $\mu \in (0,1]$. Therefore, $\frac{1}{\mu}y \in -\mathcal{H}_d \cap \mathcal{H}_d = \mathcal{T}_d \subseteq E_Q$, and hence $\|v\|^2 = y^t B^t B y = y^t Q^{-1} y \leq \mu^2$. Therefore, $\|Bd\| \leq \frac{\mu}{\delta}$, and so $\mathcal{C}(Bd) = \frac{\|Bd\|}{\rho(Bd)} \leq \frac{\mu}{\delta \alpha} \leq \frac{2\mu}{\tau \alpha}$. $\blacksquare$

We conclude this section with a theorem that demonstrates the existence of a pre-conditioner $\bar{B}$ such that the condition number of the pre-conditioned system (FP$_{Bd}$) is within the factor of $\frac{\sqrt{m}}{\delta}$ of the lower bound established in Theorem 5.4.

113

Theorem 6.1 *Suppose (FP$_d$) is feasible and $\mathcal{C}(d) < +\infty$. Then there exists a pre-conditioner $\bar{B}$ such that*

$$\mu \leq \mathcal{C}(\bar{B}d) \leq \frac{\sqrt{m}}{\delta} \cdot \mu. \tag{6.5}$$

Proof: By definition, $\mathcal{T}_d$ is a centrally-symmetric bounded convex set. Since $\mathcal{C}(d) < \infty$, $\mathcal{T}_d$ contains a ball of radius $\rho(d) > 0$ centered at 0, and so 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 6.2 with $\alpha = \frac{1}{\sqrt{m}}$ we obtain (6.5). ∎

Remark 6.1 *The upper bound in (6.5) is tight for any m.*

To show that the upper bound in (6.5) is attained for any m consider the system (FP$_d$) with $n = 2m$, $C_X = \Re_+^{2m}$, $\|x\| = \|x\|_1$ (so that $\delta = 1$) and the data $d = (A, b)$ as follows:

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

where e_i is the ith unit vector in $\Re^m$. Then

$$\mathcal{H}_d = \mathcal{T}_d = \mathrm{conv}\{\pm e_i, \ i = 1, \ldots, m\},$$

and it can be easily verified that $\mu = 1$, $\rho(d) = \frac{1}{\sqrt{m}}$, and $\|d\| = 1$, and therefore $\mathcal{C}(d) = \sqrt{m}$.

Suppose B is an arbitrary pre-conditioner. Using Lemma 6.1, we can construct a $\frac{1}{\mathcal{C}(Bd)}$-rounding of the set $\mathcal{T}_d$. However, it is well known (see, for example, [21]) that it is impossible to construct a rounding of the set $\mathrm{conv}\{\pm e_i, \ i = 1, \ldots, m\}$ with the tightness smaller than $\frac{1}{\sqrt{m}}$. Therefore, $\mathcal{C}(Bd) \leq \sqrt{m}$ for any pre-conditioner B.

6.3   On the Complexity of Computing a "Good" Pre-Conditioner

In this section we address the issue of computing a "good" pre-conditioner for the problem

(FP_d). We propose an algorithm that computes a pre-conditioner $\tilde{B}$ such that

$$\mathcal{C}(\tilde{B}d) \leq \frac{4m\mu}{\delta}. \tag{6.6}$$

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 $\|y\| = \|y\|_2$ for $y \in Y$). We also assume that the norm linearization vector $\bar{f} \in C_X^*$ for the cone C_X is known and given. Finally, we assume that an upper bound $\bar{d}$ on $\|d\|$ is known and given, or is easily computable. One could, for example, take

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

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

We have established in Lemma 6.2 that finding a "tight" rounding of the set $\mathcal{T}_d$ allows us to compute a good pre-conditioner for the system (FP_d). In Theorem 6.1 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

$$\mu \leq \mathcal{C}(\bar{B}d) \leq \frac{\sqrt{m}}{\delta} \cdot \mu,$$

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 (or approximate) a $\frac{1}{\sqrt{m}}$-rounding of the set $\mathcal{T}_d$ (see, for example, Grötschel, Lovász, and Schrijver [21]). We can, however, approximate a so-called weak Löwner-John ellipsoid, i.e., an ellipsoid which is an $\Omega\left(\frac{1}{m}\right)$-rounding. In particular, the algorithm we describe below will compute a matrix $\tilde{Q} \in S_{++}^{m \times m}$ such that

$$\frac{1}{4m}E_{\tilde{Q}} \subseteq \mathcal{T}_d \subseteq E_{\tilde{Q}}, \tag{6.7}$$

which can be used to obtain a pre-conditioner $\tilde{B}$ as in (6.6) via Lemma 6.2. We will therefore refer to the algorithm as *Algorithm WLJ* (for "Weak Löwner-John").

The algorithm WLJ is essentially a version of a parallel-cut ellipsoid algorithm. A generic

iteration of the algorithm WLJ is as follows. The current iterate of the algorithm is a matrix $Q \in S_{++}^{m \times m}$ such that $\mathcal{T}_d \subseteq E_Q$. We start by computing the spectral decomposition of the matrix Q. In particular, we compute the eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m$ of the matrix Q and the corresponding (orthonormal) eigenvectors $a_1, \ldots, a_m$. Then the axes of the ellipsoid E_Q are

$$v_i = \sqrt{\lambda_i} a_i, \ i = 1, \ldots, m.$$

We denote $V \triangleq [v_1, \ldots, v_m] \in \Re^{m \times m}$. It is trivial to verify that $Q = VV^t$. We will make repeated use of the following two simple facts. First, for all $i = 1, \ldots, m$, we have $\sqrt{\lambda_1} \leq \|v_i\| \leq \sqrt{\lambda_m}$. Also, for any vector $s \in Y^*$, $\|s\| \leq \frac{\|V^t s\|}{\sqrt{\lambda_1}}$.

Next, the algorithm makes a call to the following subroutine called *Weak Check*:

Subroutine *Weak Check*

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

(i) assert that $\pm \frac{1}{4\sqrt{m}} v_i \in \mathcal{T}_d$ for all $i = 1, \ldots, m$, or

(ii) find a vector s such that

$$s^t v_i = 1 \text{ for some } v_i, \text{ and } \mathcal{T}_d \subseteq E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\} \tag{6.8}$$

(i.e., find a parallel cut separating the two points $\pm \frac{1}{2\sqrt{m}} v_i$ from the set $\mathcal{T}_d$).

(We describe an implementation of the subroutine *Weak Check* in Subsection 6.3.1, along with a complexity analysis of the implementation).

If the subroutine terminates by asserting that $\pm \frac{1}{4\sqrt{m}} v_i \in \mathcal{T}_d$ for all $i = 1, \ldots, m$, we conclude that

$$\frac{1}{\sqrt{m}} \cdot \frac{1}{4\sqrt{m}} E_Q \subset \text{conv} \left\{ \pm \frac{1}{4\sqrt{m}} v_i, \ i = 1, \ldots, m \right\} \subseteq \mathcal{T}_d.$$

Therefore, algorithm WLJ terminates and returns the matrix $\tilde{Q} = Q$ as its output, which satisfies the condition (6.7), and hence $\tilde{B} \triangleq \tilde{Q}^{-\frac{1}{2}}$ is a pre-conditioner with the desired property (6.6).

On the other hand, if the subroutine terminates by finding a parallel cut for the set $\mathcal{T}_d$, the algorithm continues as follows. It computes $\hat{Q} \in S_{++}^{m \times m}$ such that the ellipsoid $E_{\hat{Q}}$ is the Löwner-John ellipsoid of the set

$$E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\}.$$

and proceeds to the next iteration with the matrix $\hat{Q}$ as the next iterate (observe that $\mathcal{T}_d \subseteq E_{\hat{Q}}$, and hence $\hat{Q}$ is a valid iterate). The formulas for computing $\hat{Q}$ are given in Subsection 6.3.2.

As is typical for the analysis of an ellipsoid-type algorithm, the finite termination and complexity analysis of algorithm WLJ follow from a volume-reduction argument. We provide the necessary tool for such analysis in Lemma 6.4 of Subsection 6.3.2, where we show that

$$\frac{\text{vol}(E_{\hat{Q}})}{\text{vol}(E_Q)} \leq \frac{1}{2} e^{\frac{3}{8}} \approx 0.73,$$

where $\text{vol}(S)$ denotes the volume of the set S. Finally, in Subsection 6.3.3, we provide the formal description of algorithm WLJ and its complexity analysis.

6.3.1   Implementing the Subroutine *Weak Check*

In this subsection we describe an implementation of the subroutine *Weak Check*.

Each iteration of algorithm WLJ makes use of the following subroutine *Weak Check*:

Subroutine *Weak Check*

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

(i) assert that $\pm \frac{1}{4\sqrt{m}} v_i \in \mathcal{T}_d$ for all $i = 1, \ldots, m$, or

(ii) find a vector s such that

$$s^t v_i = 1 \text{ for some } v_i, \text{ and } \mathcal{T}_d \subseteq E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\}. \qquad (6.9)$$

The purpose of the subroutine *Weak Check* is to determine whether the axes of the smaller ellipsoid, namely, $\frac{1}{4\sqrt{m}}E_Q$, are contained in $\mathcal{T}_d$, or produce a parallel cut otherwise. Consider the following $2m$ optimization problems (P_{ϕ_v}), where $v = \pm v_i$, $i = 1, \ldots, m$:

$$(P_{\phi_v}) \quad \phi_v = \max_{\phi v \in \mathcal{H}_d} \phi \quad = \quad \max_{\theta, x, \phi} \quad \phi \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad b\theta - Ax = v\phi \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad |\theta| + \|x\| \le 1 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad \theta \ge 0, \ x \in C_X. \tag{6.10}$$

It is not hard to verify that $\pm\frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$ for all $i = 1, \ldots, m$ precisely when

$$\phi_Q \triangleq \min_{\pm v_i} \phi_v \ge \frac{1}{4\sqrt{m}}. \tag{6.11}$$

(here $\min_{\pm v_i} \phi_v$ stands for $\min\{\phi_v : v = -v_i, v_i, \ i = 1, \ldots, m\}$ in order to shorten the notation). We will therefore implement the subroutine *Weak Check* by means of approximately solving the $2m$ optimization problems (6.10) and checking if condition (6.11) is satisfied.

The approach we will use to solve the optimization problems in the subroutine *Weak Check* uses the barrier method of Renegar described in Chapter 5. For every value of $v = \pm v_i$, $i = 1, \ldots, m$, we will solve (6.10) by considering its dual:

$$(P_{\gamma_v}) \quad \gamma_v = \min_{s, q, \gamma} \quad \gamma \\ \qquad\qquad\qquad\qquad \|A^t s - q\| \le \gamma \\ \qquad\qquad\qquad\qquad b^t s \le \gamma \\ \qquad\qquad\qquad\qquad q \in C_X^* \\ \qquad\qquad\qquad\qquad v^t s = 1. \tag{6.12}$$

It is not hard to verify that strong duality holds, 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 for the optimization problem at hand to have a bounded feasible region. To satisfy this condition, we consider the following

modification of (6.12):

$$(\mathrm{P}_{\tilde{\gamma}_v}) \quad \tilde{\gamma}_v = \min_{s,q,\gamma} \quad \gamma$$

$$\|A^t s - q\| \leq \gamma$$
$$b^t s \leq \gamma$$
$$\|V^t s\| \leq 2\sqrt{m} \tag{6.13}$$
$$\gamma \leq \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}$$
$$q \in C_X^*$$
$$v^t s = 1,$$

where $\bar{d}$ is the known upper bound on the norm of the data $\|d\|$. In the next proposition we argue that solving $(\mathrm{P}_{\tilde{\gamma}_v})$ instead of (P_{γ_v}) still yields a valid estimate of ϕ_Q.

Proposition 6.2 *For any v, $\gamma_v \leq \tilde{\gamma}_v$. Moreover,*

$$\phi_Q = \min_{\pm v_i} \gamma_v = \min_{\pm v_i} \tilde{\gamma}_v. \tag{6.14}$$

Proof: The first claim of the proposition is trivially true, since the feasible region of the program $(\mathrm{P}_{\tilde{\gamma}_v})$ is contained in the feasible region of the program (P_{γ_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_{γ_v}). Then we have

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

If the point $(\bar{s}, \bar{q}, \gamma_v)$ is feasible for the corresponding program $(\mathrm{P}_{\tilde{\gamma}_v})$, then $\gamma_v = \tilde{\gamma}_v$, and (6.14) 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 jth 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 (6.14) follows. $\blacksquare$

Let

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

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. Let $f^*(q)$ be a self-concordant barrier for the cone C_X^* with complexity parameter ϑ^*. 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 $\vartheta_f \leq \vartheta^* + 5$ (see, for example, [41] or [42] for details).

In order to use the barrier method to solve $(P_{\tilde{\gamma}_v})$, we need to have a point $(s', q', \gamma') \in$ int$S \cap L_v$ at which to initialize the method. The next proposition indicates that such point is readily available when the norm linearization vector $\bar{f} \in C_X^*$ for the cone C_X is known.

Proposition 6.3

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

Proof: First note that $v^t s' = 1$, so $(s', q', \gamma') \in L_v$. Also,

$$\|A^t s' - q'\| \le \frac{1}{\|v\|}(\|A\| + 2\bar{d}) \le \frac{3\bar{d}}{\sqrt{\lambda_1}} < \gamma',$$

$$b^t s' \le \frac{\|d\|}{\|v\|} < \gamma', \quad \|V^t s'\| = \sqrt{\sum_{i=1}^{m}(v_i^t s')^2} = 1 < 2\sqrt{m},$$

$$\gamma' < \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}, \quad \text{and } q' \in \text{int}C_X^*.$$

Therefore, $(s', q', \gamma') \in \text{int}S.$ ∎

The next proposition establishes a lower bound on $\text{sym}(S \cap L_v, (s', q', \gamma'))$ which is an important parameter for analyzing the complexity of the barrier method.

Proposition 6.4

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

Proof: Let (s, q, γ) be such that $(s' + s, q' + q, \gamma' + \gamma) \in S \cap L_v$. We want to find a scalar $t^* > 0$ such that for any $t \in [0, t^*]$, $(s' - ts, q' - tq, \gamma' - t\gamma) \in S \cap L_v$. The strategy is to find appropriate values $t_1, \ldots, t_5$ for each of the five constraints defining the region S, and set $t^* = \min\{t_1, \ldots, t_5\}$.

1. We have $\|(A^t s' - q') + (A^t s - q)\| \le \gamma' + \gamma$. We want to find a range of values of t such that $\|(A^t s' - q') - t(A^t s - q)\| \le \gamma' - t\gamma$.

It is sufficient to find a range of t such that $\|A^t s' - q'\| + t\|A^t s - q\| \le \gamma' - t\gamma$. Observe that

$$\|A^t s - q\| \le \|A^t s' - q'\| + \gamma + \gamma' \le \frac{\|d\|}{\|v\|} + \frac{2\bar{d}}{\|v\|} + \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} \le \frac{10\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}.$$

Also, $\gamma \le \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}$. Combining these inequalities, we see that for any $t \le t_1 \triangleq 1/13$,

$$(\|A^t s - q\| + \gamma)t \le \frac{1}{13}\left(\frac{10\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} + \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}\right) \le \frac{\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} \le \gamma' - \|A^t s' - q'\|.$$

2. We have $b^t s' + b^t s \le \gamma' + \gamma$. We want to find a range of values of t such that $b^t s' - t b^t s \le \gamma' - t\gamma$, that is, $t(\gamma - b^t s) \le \gamma' - b^t s'$.

First note that
$$\gamma' - b^t s' \geq \gamma' - \|b\| \cdot \|s'\| \geq \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}.$$

So, it is sufficient to have $t(\gamma - b^t s) \leq \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}$. Furthermore,

$$\|s\| \leq \|s + s'\| + \|s'\| \leq \frac{2\sqrt{m}}{\sqrt{\lambda_1}} + \frac{1}{\|v\|} \leq \frac{3\sqrt{m}}{\sqrt{\lambda_1}}.$$

We use this inequality to establish:

$$\gamma - b^t s \leq \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} + \bar{d}\|s\| \leq \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} + \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} = \frac{6\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}.$$

So, it is sufficient to have $t \leq t_2 \triangleq \frac{1}{2}$ to satisfy

$$t(\gamma - b^t s) \leq \frac{1}{2} \cdot \frac{6\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} \leq \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}.$$

3. We have $\|V^t(s + s')\| \leq 2\sqrt{m}$. We want to find a range of values of t such that $\|V^t(s' - ts)\| \leq 2\sqrt{m}$. Note that

$$\|V^t(s' - ts)\| \leq \|V^t s'\| + t\|V^t s\| = 1 + t\|V^t s\|$$

$$\leq 1 + t(\|V^t(s' + s)\| + \|V^t s'\|) \leq 1 + t(2\sqrt{m} + 1).$$

Hence, for any $t \leq t_3 \triangleq \frac{2\sqrt{m} - 1}{2\sqrt{m} + 1}$,

$$\|V^t(s' - ts)\| \leq 1 + \frac{2\sqrt{m} - 1}{2\sqrt{m} + 1}(2\sqrt{m} + 1) = 2\sqrt{m}.$$

4. We have $\gamma + \gamma' \leq \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}$, that is, $\gamma \leq \frac{3\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}$.

Since $\gamma + \gamma' \geq 0$, $\gamma \geq -\gamma'$. Therefore, for $t \leq t_4 \triangleq 3/4$, $\gamma' - t\gamma \leq \gamma' + \frac{3}{4}\gamma' = \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}$.

5. We have $q' + q \in C_X^*$. We want to find a range of values of t such that $q' - tq \in C_X^*$.

Since $B\left(q', \frac{2\bar{d}\beta}{\|v\|}\right) \subset C_X^*$, it is sufficient to have $t \leq \frac{2\bar{d}\beta}{\sqrt{\lambda_m}\|q\|}$. An upper bound on $\|q\|$ can be

obtained by noting that

$$\|q' + q\| \leq \gamma' + \gamma + \|A\| \cdot \|s + s'\| \leq \frac{7\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} + \frac{\bar{d} \cdot 2\sqrt{m}}{\sqrt{\lambda_1}} = \frac{9\sqrt{m}\bar{d}}{\sqrt{\lambda_1}}.$$

So,

$$\|q\| \leq \frac{9\sqrt{m}\bar{d}}{\sqrt{\lambda_1}} + \|q'\| \leq \frac{11\sqrt{m}\bar{d}}{\sqrt{\lambda_1}},$$

and taking $t \leq t_5 \triangleq \frac{2\beta}{11\sqrt{m}} \cdot \sqrt{\frac{\lambda_1}{\lambda_m}}$ is sufficient.

Finally,

$$\mathrm{sym}(S \cap L, (s', \gamma', q')) \geq \min\{t_1, \ldots, t_5\}$$

$$= \min\left\{\frac{1}{13}, \frac{1}{2}, \frac{2\sqrt{m}-1}{2\sqrt{m}+1}, \frac{3}{4}, \frac{2\beta}{11\sqrt{m}} \cdot \sqrt{\frac{\lambda_1}{\lambda_m}}\right\} \geq \frac{\beta}{13\sqrt{m}} \cdot \sqrt{\frac{\lambda_1}{\lambda_m}}. \blacksquare$$

As we can see, the square root of the ratio of the smallest and largest eigenvalues of Q, which can also be interpreted as the ratio of the lengths of the shortest and longest axes of E_Q, will play an important role in the complexity the barrier method. This ratio is often called the *skewness* of the ellipsoid E_Q.

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

<div align="center">

Subroutine *Weak Check*

</div>

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

- for $v = \pm v_i$, $i = 1, \ldots, m$,

    Step 1 Form the problem $(\mathrm{P}_{\tilde{\gamma}_v})$

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

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

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

Step 3 If $\gamma < \frac{1}{2\sqrt{m}}$, terminate. 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 $i = 1, \ldots, m$.

Lemma 6.3 *Subroutine* Weak Check *will terminate in at most*

$$O\left(m\sqrt{\vartheta^*}\ln\left(\frac{m\vartheta^*}{\beta} \cdot \frac{\bar{d}}{\sqrt{\lambda_1}} \cdot \sqrt{\frac{\lambda_m}{\lambda_1}}\right)\right) \tag{6.15}$$

iterations of the barrier method.

Upon termination, subroutine Weak Check *will either correctly assert that* $\pm\frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$ *for all* $i = 1, \ldots, m$, *or will return a vector s such that*

$$s^t v_i = 1 \text{ for some } v_i, \text{ and } \mathcal{T}_d \subseteq E_Q \cap \left\{y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}}\right\} \tag{6.16}$$

(i.e., will find a parallel cut separating the two points $\pm\frac{1}{2\sqrt{m}}v_i$ from the set $\mathcal{T}_d$).

Proof: We will use (5.18) 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 $(\mathrm{P}_{\tilde{\gamma}_v})$. 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 (5.18) and Proposition 6.4, we see that each of the (at most) $2m$ applications of the barrier method will terminate in at most

$$O\left(\sqrt{\vartheta_f}\ln\left(\frac{7\sqrt{m}\bar{d}\vartheta_f}{\sqrt{\lambda_1}} \cdot \frac{\bar{\eta}}{\mathrm{sym}(S\cap L_v,(s',q',\gamma'))}\right)\right)$$

$$\leq O\left(\sqrt{\vartheta^*}\ln\left(\frac{7\sqrt{m}\bar{d}\vartheta^*}{\sqrt{\lambda_1}} \cdot \frac{24\sqrt{m}\vartheta^*}{5} \cdot \frac{13\sqrt{m}}{\beta} \cdot \sqrt{\frac{\lambda_m}{\lambda_1}}\right)\right)$$

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

iterations of the barrier method. Multiplying the above bound by $2m$ we obtain (6.15).

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

$$\|A^t s - q\| \leq \gamma < \tfrac{1}{2\sqrt{m}}$$
$$b^t s \leq \gamma < \tfrac{1}{2\sqrt{m}}$$
$$\|V^t s\| \leq 2\sqrt{m}$$
$$q \in C_X^*, \ v_i^t s = 1.$$

The vector s above yields a parallel cut that separates $\pm \tfrac{v_i}{2\sqrt{m}}$ 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 $(\theta, x) \in \Re_+ \times C_X$ such that $|\theta| + \|x\| \leq 1$. Therefore

$$s^t h = s^t(b\theta - Ax) = \theta(b^t s) - x^t(A^t s) = \theta(b^t s) - x^t(A^t s - q) - x^t q$$

$$\leq (|\theta| + \|x\|)\gamma \leq \gamma < \frac{1}{2\sqrt{m}} = \frac{s^t v_i}{2\sqrt{m}}.$$

Applying the same argument for the point $-h \in \mathcal{H}_d$, we conclude that $s^t h > -\tfrac{s^t v_i}{2\sqrt{m}}$, and therefore the vector s returned by the subroutine *Weak Check* satisfies (6.16).

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 $\gamma \geq \tfrac{1}{2\sqrt{m}}$. Then, using (5.17),

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

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

6.3.2   Iterate Update

Suppose the subroutine *Weak Check* called by the algorithm WLJ has terminated and returned as output a vector s such that

$$s^t v_i = 1 \text{ for some } v_i, \text{ and } \mathcal{T}_d \subseteq E_Q \cap \left\{ y : -\frac{1}{2\sqrt{m}} \leq s^t y \leq \frac{1}{2\sqrt{m}} \right\}.$$

In this subsection we will show how the next iterate of the algorithm WLJ is computed and establish a volume-reduction argument typical for the analysis of an ellipsoid-type algorithm.

The next iterate of the algorithm WLJ is the matrix $\hat{Q}$ such that $E_{\hat{Q}}$ is the Löwner-John ellipsoid for the set

$$E_Q \cap \left\{ h : -\frac{1}{2\sqrt{m}} \leq s^t h \leq \frac{1}{2\sqrt{m}} \right\}.$$

Notice that $\sqrt{s^t Q s} = \|V^t s\| \geq s^t v_i = 1$, and therefore $-\frac{1}{\sqrt{m}} \leq -\frac{1}{2\sqrt{m}\sqrt{s^t Q s}}$. Hence we can use formula (3.1.20) of [21] to compute

$$\hat{Q} = \frac{m}{m-1} \left(1 - \frac{1}{4m\xi} \right) \left(Q - \frac{m(4\xi - 1)}{4m\xi - 1} \cdot \frac{Q s s^t Q}{\xi} \right), \tag{6.17}$$

where $\xi = s^t Q s$. We analyze the complexity of algorithm WLJ by looking at the decrease in the volumes of the ellipsoids from iteration to iteration. The main tool for such analysis is presented in the following lemma:

Lemma 6.4

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

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

$$\hat{Q} = \frac{m}{m-1} \left(1 - \frac{1}{4m\xi} \right) Q^{\frac{1}{2}} R^t \left(I - \frac{m(4\xi - 1)}{4m\xi - 1} e_1 e_1^t \right) R Q^{\frac{1}{2}}. \tag{6.18}$$

Therefore,

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

126

$$= \left(\frac{m}{m-1} \left(1 - \frac{1}{4m\xi} \right) \right)^m \left(1 - \frac{m(4\xi - 1)}{4m\xi - 1} \right) \det(Q).$$

We conclude that

$$\frac{\det(\hat{Q})}{\det(Q)} = \left(\frac{m}{m-1} \left(1 - \frac{1}{4m\xi} \right) \right)^m \left(1 - \frac{m(4\xi - 1)}{4m\xi - 1} \right)$$

$$= \frac{m^m (4m\xi - 1)^{m-1}}{(m-1)^{m-1} (4m\xi)^m} = \frac{1}{4\xi} \left(\frac{4m\xi - 1}{4m\xi - 4\xi} \right)^{m-1}$$

$$= \frac{1}{4\xi} \left(1 + \frac{4\xi - 1}{4\xi(m-1)} \right)^{m-1} \le \frac{1}{4\xi} e^{1 - \frac{1}{4\xi}} \le \frac{1}{4} e^{\frac{3}{4}}.$$

The last inequality follows since the function te^{1-t} is an increasing function for $t \in [0,1]$, and as we have established previously, $0 < \frac{1}{4\xi} \le \frac{1}{4}$.

Finally,

$$\frac{\mathrm{vol}(E_{\hat{Q}})}{\mathrm{vol}(E_Q)} = \frac{\sqrt{\det(\hat{Q})}\, \mathrm{vol}(B(0,1))}{\sqrt{\det(Q)}\, \mathrm{vol}(B(0,1))} \le \frac{1}{2} e^{\frac{3}{8}}. \quad \blacksquare$$

6.3.3   Complexity analysis of the algorithm WLJ

In this subsection we present a formal statement of algorithm WLJ and make use of the results of previous subsections to obtain a complexity analysis of the algorithm.

The statement of the algorithm is as follows (recall that an iterate of the algorithm is a matrix $Q \in S_{++}^{m \times m}$ such that $\mathcal{T}_d \subseteq E_Q \triangleq \{y : y^t Q^{-1} y \le 1\}$):

Algorithms WLJ (Weak Löwner-John)

- *Input:* $\bar{d}$ — an upper bound on $\|d\|$

- *Initialization:* The algorithm is initialized with the matrix $Q^0 = \bar{d}^2 I$.

- Iteration $k \ge 1$.

  Step 1 Let $Q = Q^k$. Compute the eigenvalues $0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_m$ of Q and the corresponding (orthonormal) eigenvectors $a_1, \ldots, a_m$. Define the axes of E_Q by

  $$v_i = \sqrt{\lambda_i} a_i, \ i = 1, \ldots, m.$$

127

Step 2 Call subroutine *Weak Check* with the input $(v_1, \ldots, v_m)$. If the subroutine terminates by asserting that $\pm \frac{1}{4\sqrt{m}} v_i \in \mathcal{T}_d$, $i = 1, \ldots, m$, stop, return $\tilde{B} = Q^{-\frac{1}{2}}$.

Step 3 Otherwise, the subroutine *Weak Check* returns a vector s. Define $\hat{Q}$ by (6.17).

Step 4 Let $Q^{k+1} = \hat{Q}$, $k \leftarrow k + 1$, go to Step 1.

Theorem 6.2 *Suppose $\mathcal{C}(d) < \infty$. Then algorithm WLJ will terminate in at most*

$$O\left(m^2 \sqrt{\vartheta^*} \ln^2\left(\frac{\bar{d}}{\rho(d)}\right) \ln\left(\frac{m\vartheta^*}{\beta}\right)\right) \tag{6.19}$$

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

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

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 $\mathcal{T}_d \subseteq E_{Q^0}$.

To establish a bound on the number of iterations of the barrier method required by algorithm WLJ, we proceed as follows. We first determine an upper bound on the number of iterations of algorithm WLJ via a volume-reduction argument. We then establish an upper bound on the skewness of the ellipsoids generated throughout the algorithm. We use this bound to analyze the running time of any of the calls to the subroutine *Weak Check* made by the algorithm, and combine all the above results to obtain the final complexity bound.

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 conclude that

$$\text{vol}(\mathcal{T}_d) \geq \text{vol}(B(0, \rho(d))) = \rho(d)^m \text{vol}(B(0, 1)).$$

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

$$K \leq m \ln \left(\frac{\bar{d}}{\rho(d)} \right) \cdot \frac{1}{\ln 2 - .375} \leq \frac{10}{3} m \ln \left(\frac{\bar{d}}{\rho(d)} \right) \tag{6.20}$$

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, $\lambda_1 \geq \rho(d)^2$.

Now let us estimate the change in the largest eigenvalue of the ellipsoid matrix from one iterations to the next. Suppose the matrix Q is the current iterate, and $\hat{Q}$ is the next iterate of the algorithm. Then from (6.18) we conclude that

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

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

Therefore, at iteration k, $\lambda_m^k \leq \lambda_m^0 e^{\frac{k}{m-1}} = \bar{d}^2 e^{\frac{k}{m-1}}$. Recall from (6.20) that the algorithm will perform at most

$$K = \frac{10}{3} m \ln \left(\frac{\bar{d}}{\rho(d)} \right) = \ln \left(\left(\frac{\bar{d}}{\rho(d)} \right)^{\frac{10}{3}m} \right)$$

iterations. Therefore, throughout the algorithm,

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

and the skewness of all ellipsoids generated by the algorithm is bounded above by

$$\sqrt{\frac{\lambda_m}{\lambda_1}} \leq \sqrt{\left(\frac{\bar{d}}{\rho(d)} \right)^{\frac{10m}{3(m-1)}+2}} \leq \left(\frac{\bar{d}}{\rho(d)} \right)^5. \tag{6.21}$$

Using (6.21) we conclude from Lemma 6.3 that any call to subroutine *Weak Check* will perform at most

$$O \left(m\sqrt{\vartheta^*} \ln \left(\frac{m\vartheta^*}{\beta} \cdot \frac{\bar{d}}{\rho(d)} \right) \right)$$

iterations of the barrier method. Combining this with (6.20), we can bound the total number of iterations of the barrier method performed by algorithm WLJ by

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

Finally, the properties of the pre-conditioner $\tilde{B}$ returned by algorithm WLJ follow from the discussion at the beginning of this section. $\blacksquare$

Remark 6.2 *Note that the skewness of the ellipsoids does not necessarily get worse at every iteration. In fact, the last ellipsoid of the algorithm (which satisfies $\frac{1}{4\sqrt{m}}v_i \in \mathcal{T}_d$ for all i) has the very nice property that $\sqrt{\frac{\lambda_m}{\lambda_1}} \leq 4\sqrt{m}\mathcal{C}(d)$.*

To further interpret the complexity result of Theorem 6.2, suppose for simplicity that $\bar{d} = \|d\|$, i.e., the size of the data $\|d\|$ is known. Then algorithm WLJ will perform at most

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

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., has a particularly large condition number, 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 the symmetry measure μ. This result, which may seem counterintuitive at first, is actually explained by the fact that, in order to obtain a pre-conditioner for the system (FP$_d$), algorithm WLJ has to work with the set $\mathcal{T}_d$, rather then $\mathcal{H}_d$, which is perfectly symmetric regardless of the geometry of the set $\mathcal{H}_d$.

Chapter 7

Conclusions and Future Work

In this thesis we have addressed several issues in the relatively new field of measures of conditioning for convex optimization and feasibility problems. We have explored the role that the condition number for a convex feasibility problem in conic linear form plays in studying the geometry of the problem and the complexity of certain algorithms for solving the problem. We have also discussed some potential drawbacks in using the condition number as the sole measure of conditioning of a conic linear system, motivating the study of "data-independent" measures. We have introduced one such measure, μ, for feasible conic linear systems in special form and studied many of its implications for problem geometry, conditioning, and algorithm complexity. The field of measures of conditioning for convex problems, however, is still in its infancy, and we believe that the work in this thesis has raised many issues and potential venues of research.

One such issue concerns the desire to find data-independent measures of conditioning for infeasible instances of the problem (FP_d) of (5.1), as well as for conic linear systems in the most general form (2.1) and problems of optimizing linear functions over the feasible region of (FP_d). Such measures might or might not be extensions of the measures discussed in this thesis, and in fact it seems possible that for some forms of the problem (FP_d) no such relevant measures can be defined. This would indicate that for certain classes of convex feasibility and optimization problems the properties of the problem and properties of algorithms for solving the problem cannot be addressed using only "geometric" tools of analysis, and require some

information about the specific representation of the problem. The pursuit and study of such measures of conditioning will hopefully lead to better understanding of the properties of the underlying problems, and potentially, lead to more efficient algorithms for solving the problems.

Another potential topic of research concerns incorporating pre-conditioning into algorithms for solving convex feasibility and optimization problems. The notion of having a pre-processing stage in an algorithm is not new. In fact, most the optimization software packages include some type of pre-processing options, such as variable and constraint elimination or data scaling. It would be interesting to study the potential benefits of including a pre-processing step that would attempt to improve the conditioning of the problem as measured by the condition number, or another relevant measure of conditioning.

Yet another potential research direction is the study of measures of conditioning of convex problems in forms other than the conic linear form used in this thesis and in the majority of the literature to date. Although any convex optimization (or feasibility) problem can be stated in conic linear form by appropriately defining the cones C_X and C_Y, such reformulations could be deceptively simple, since the conic notation "hides" non-linear aspects of the problem. It is often more efficient to work directly with the original or slightly modified problem formulation, or a slight modification, rather than with the cones resulting from the conic reformulation. Therefore, it would be useful to develop measures of conditioning for such problem formulations, providing additional tools for analyzing problem properties and algorithm performance.

Appendix A

Standard Form Reformulations

In this appendix we establish some properties of two "standard form" reformulations of the problem (FP_d) in its most general form. These reformulations are a direct extension of the standard approach used in linear programming applications to state the problem either in the form $Ax = b$, $x \geq 0$ (often referred to as the "standard primal form"), or in the form $Ax \leq b$, x unrestricted (the "standard dual form"). In linear programming such reformulations are often a helpful pre-processing step preceding, for example, the application of the primal (or dual) simplex algorithm for linear optimization. For a general conic linear system (FP_d) such reformulations can also be a useful pre-processing tool preceding an application of an algorithm that might be specially tailored to work on systems in a specific form. Also, the ability to reformulate a general system in a simpler form can motivate a helpful simplification of analysis of algorithms, see for example, [36].

The results of this appendix are not explicitly used in the thesis; nevertheless, we feel it is important to formally present these results at least as a reference tool. We also hope this analysis provides some helpful intuition into the analysis of perturbation-based condition numbers.

We will be working with a conic linear system (FP_d) of the form

$$(\text{FP}_d) \quad \begin{aligned} b - Ax &\in C_Y \\ x &\in C_X, \end{aligned} \tag{A.1}$$

where C_X and C_Y are closed convex cones and $d \triangleq (A, b)$ is the data for the system. We consider two "standard form" reformulations of the system (FP$_d$) and study their properties.

A.1 Reformulation I: Standard Primal Form

Consider the following reformulation of the system (FP$_d$):

$$(\text{FP}_{\tilde{d}}) \qquad \begin{aligned} Ax + \nu y &= b \\ (x, y) &\in C_X \times C_Y, \end{aligned} \qquad (\text{A.2})$$

where $\nu > 0$ is a fixed constant. Since $\nu > 0$, the system (FP$_{\tilde{d}}$) is feasible precisely when the system (FP$_d$) is feasible. It can also be viewed as a conic linear system under the following assignments:

- $\tilde{X} = X \times Y$, $\tilde{Y} = Y$

- $\tilde{d} = (\tilde{A}, \tilde{b})$, where $\tilde{A} = [A; \nu I] \in L(\tilde{X}, \tilde{Y})$, $\tilde{b} = b \in \tilde{Y}$

- $C_{\tilde{X}} = C_X \times C_Y$, $C_{\tilde{Y}} = \{0\}$,

with norms defined as follows:

- $\|\tilde{x}\| = \|(x, y)\| = \|x\| + \|y\|$, $\tilde{x} = (x, y) \in \tilde{X}$,

- $\|\tilde{y}\|$, $\tilde{y} \in \tilde{Y}$.

Note that, by the definition of the operator norm, $\|\tilde{A}\| = \max\{\|A\|, \|\nu I\|\} = \max\{\|A\|, \nu\}$, and therefore, $\|\tilde{d}\| = \max\{\|\tilde{A}\|, \|\tilde{b}\|\} = \max\{\|d\|, \nu\}$.

It is easy to see that $\rho(\tilde{d}) \leq \rho(d)$, whether (FP$_d$) is feasible or not, since any perturbation of the data $d = (A, b)$ that changes the feasibility status of the system (FP$_d$) corresponds to a perturbation of the data $\tilde{d}$ of the same size that changes the feasibility status of the system (FP$_{\tilde{d}}$). In the next two lemmas we derive a lower bound on $\rho(\tilde{d})$ in terms of $\rho(d)$.

Lemma A.1 was first established by Peña and Renegar in [36]; the proof presented here is almost identical to theirs. Lemma A.2 also replicates a result in the later version of [36], although it was established independently of [36].

Lemma A.1 ([36]) *If the system (FP$_d$) is feasible and $\nu > 0$, then*

$$\rho(\tilde{d}) \geq \frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}.$$

Proof: Denote $\eta \triangleq \max\{1, \|d\|/\nu\}$. We want to show that $\frac{\rho(d)}{3\eta} \leq \rho(\tilde{d})$. Let $v \in Y$ be an arbitrary vector satisfying $\|v\| \leq \frac{\rho(d)}{3\eta}$. By Theorem 2.1 there exists $(\bar{\theta}, \bar{x}, \bar{y})$ such that

$$b\bar{\theta} - A\bar{x} - \bar{y} = 3\eta v$$
$$\bar{\theta} \geq 0, \ \bar{x} \in C_X, \ \bar{y} \in C_Y$$
$$|\bar{\theta}| + \|\bar{x}\| \leq 1.$$

Define $(\tilde{\theta}, \tilde{x}, \tilde{y}) = \left(\frac{1}{3\eta}\bar{\theta}, \frac{1}{3\eta}\bar{x}, \frac{1}{3\eta\nu}\bar{y}\right)$. Then $\tilde{\theta} \geq 0$, $\tilde{x} \in C_X$, $\tilde{y} \in C_Y$, and $b\tilde{\theta} - A\tilde{x} - \nu\tilde{y} - v = 0 \in C_{\tilde{Y}}$. Taking into account the fact that

$$\|\bar{y}\| = \|b\bar{\theta} - A\bar{x} - 3\eta v\| \leq \|d\| + \rho(d) \leq 2\|d\|,$$

we derive:

$$\tilde{\theta} + \|\tilde{x}\| + \|\tilde{y}\| = \frac{1}{3\eta}\left(\bar{\theta} + \|\bar{x}\| + \frac{\|\bar{y}\|}{\nu}\right) \leq \frac{1}{3\eta}\left(1 + \frac{2\|d\|}{\nu}\right) \leq \frac{1}{3\eta} \cdot 3\max\left\{1, \frac{\|d\|}{\nu}\right\} = 1.$$

We conclude from Theorem 2.1 that $\rho(\tilde{d}) \geq \frac{\rho(d)}{3\eta}$, proving the lemma. ∎

Lemma A.2 *If the system (FP$_d$) is infeasible and $\nu > 0$, then*

$$\rho(\tilde{d}) \geq \frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}$$

Proof: Denote $\eta \triangleq \max\{1, \|d\|/\nu\}$. We want to show that $\rho(\tilde{d}) \geq \frac{\rho(d)}{3\eta}$. Let $v_1 \in X^*$ and $v_2 \in Y^*$ be arbitrary vectors satisfying $\max\{\|v_1\|_*, \|v_2\|_*\} \leq \frac{\rho(d)}{3\eta}$, and consider the vector $v = \frac{3}{2}v_1 - \frac{3}{2\nu}A^t v_2 \in X^*$. Then

$$\|v\|_* \leq \frac{3}{2}\left(\|v_1\|_* + \frac{\|d\|}{\nu}\|v_2\|_*\right) \leq \frac{3}{2}\left(1 + \frac{\|d\|}{\nu}\right)\frac{\rho(d)}{3\eta} \leq \rho(d).$$

By Theorem 2.3 there exists $\bar{s}$ such that

$$A^t \bar{s} - v \in C_X^*$$
$$-b^t \bar{s} \geq \rho(d)$$
$$\bar{s} \in C_Y^*$$
$$\|\bar{s}\|_* \leq 1.$$

Define $\tilde{s} = \frac{2}{3}\bar{s} + \frac{v_2}{\nu}$. Then

$$\|\tilde{s}\|_* \leq \frac{2}{3}\|\bar{s}\|_* + \frac{\|v_2\|_*}{\nu} \leq \frac{2}{3} + \frac{\rho(d)}{3\eta\nu} \leq 1,$$

$$A^t \tilde{s} - v_1 = \frac{2}{3}(A^t\bar{s} - v) \in C_X^*,$$

$$\nu\tilde{s} - v_2 = \frac{2\nu}{3}\bar{s} \in C_Y^*,$$

and
$$-b^t\tilde{s} = -\frac{2}{3}b^t\bar{s} - \frac{b^t v_2}{\nu} \geq \frac{2\rho(d)}{3} - \frac{\|d\|\rho(d)}{3\nu\eta} = \frac{\rho(d)}{3}\left(2 - \frac{\|d\|}{\nu\eta}\right) \geq \frac{\rho(d)}{3\eta}.$$

We conclude from Theorem 2.3 that $\rho(\tilde{d}) \geq \frac{\rho(d)}{3\eta}$, proving the lemma. ∎

Combining Lemmas A.1 and A.2 with the fact that $\|\tilde{d}\| = \max\{\|d\|, \nu\}$, we obtain

Theorem A.1
$$\mathcal{C}(d) \leq \mathcal{C}(\tilde{d}) \leq 3\mathcal{C}(d)\max\left\{\frac{\nu}{\|d\|}, \frac{\|d\|}{\nu}\right\}.$$

A.2   Reformulation II: Standard Dual Form

Consider another reformulation of the system (FP_d):

$$(\text{FP}_{\hat{d}}) \qquad \nu x \in C_X$$
$$b - Ax \in C_Y \qquad\qquad\qquad (\text{A.3})$$
$$x \text{ unrestricted,}$$

where $\nu > 0$ is a fixed constant. Since $\nu > 0$, the system $(\text{FP}_{\hat{d}})$ is feasible precisely when the

system (FP$_d$) is feasible. It can also be viewed as a conic linear system under the following assignments:

- $\hat{X} = X$, $\hat{Y} = X \times Y$

- $\hat{d} = (\hat{A}, \hat{b})$, where $\hat{A} = \begin{bmatrix} -\nu I \\ A \end{bmatrix} \in L(\hat{X}, \hat{Y})$, $\hat{b} = \begin{bmatrix} 0 \\ b \end{bmatrix} \in \hat{Y}$

- $C_{\hat{X}} \triangleq X$, $C_{\hat{Y}} = C_X \times C_Y$,

with norms defined as follows:

- $\|\hat{x}\|$, $\hat{x} \in X$,

- $\|\hat{y}\| = \|(x, y)\| = \max\{\|x\|, \|y\|\}$, $\hat{y} = (x, y) \in \hat{Y}$.

Note that, by the definition of the operator norm, $\|\hat{A}\| = \max\{\|A\|, \|\nu I\|\} = \max\{\|A\|, \nu\}$, and therefore, $\|\hat{d}\| = \max\{\|\hat{A}\|, \|\hat{b}\|\} = \max\{\|d\|, \nu\}$.

It is easy to see that $\rho(\hat{d}) \leq \rho(d)$, whether (FP$_d$) is feasible or not, since any perturbation of the data $d = (A, b)$ that changes the status of the system (FP$_d$) corresponds to a perturbation of the data $\hat{d}$ of the same size that changes the feasibility status of the system (FP$_{\hat{d}}$). In the next two lemmas we derive a lower bound on $\rho(\hat{d})$ in terms of $\rho(d)$.

Lemma A.3 *If the system (FP$_d$) is infeasible and $\nu > 0$, then*

$$\rho(\hat{d}) \geq \frac{\rho(d)}{3 \max\{1, \|d\|/\nu\}}.$$

Proof: Denote $\eta \triangleq \max\{1, \|d\|/\nu\}$. We want to show that $\rho(\hat{d}) \geq \frac{\rho(d)}{3\eta}$. Let $v \in X^*$ be an arbitrary vector satisfying $\|v\|_* \leq \frac{\rho(d)}{3\eta}$. By Theorem 2.3 there exits $(\bar{p}, \bar{s})$ such that

$$A^t \bar{s} - \bar{p} = 3\eta v$$
$$-b^t \bar{s} \geq \rho(d)$$
$$\bar{p} \in C_X^*, \ \bar{s} \in C_Y^*$$
$$\|\bar{s}\|_* \leq 1.$$

Define $(\hat{s}_1, \hat{s}_2) = \frac{1}{3\eta}\left(\frac{\bar{p}}{\nu}, \bar{s}\right)$. Then $\hat{s}_1 \in C_X^*$, $\hat{s}_2 \in C_Y^*$, $-\nu\hat{s}_1 + A^t\hat{s}_2 = v$, and $-b^t\hat{s}_2 \geq \frac{\rho(d)}{3\eta}$. Taking into account the fact that

$$\|\bar{p}\|_* = \|A^t\bar{s} - 3\eta v\|_* \leq \|A^t\bar{s}\|_* + 3\eta\|v\|_* \leq \|d\| + \rho(d) \leq 2\|d\|,$$

we derive:

$$\|(\hat{s}_1, \hat{s}_2)\|_* = \|\hat{s}_1\|_* + \|\hat{s}_2\|_* = \frac{1}{3\eta}\left(\frac{\|\bar{p}\|_*}{\nu} + \|\bar{s}\|_*\right)$$

$$\leq \frac{1}{3\eta}\left(\frac{2\|d\|}{\nu} + 1\right) \leq \frac{1}{3\eta} \cdot 3\max\left\{1, \frac{\|d\|}{\nu}\right\} = 1,$$

We conclude from Theorem 2.3 that $\rho(\hat{d}) \geq \frac{\rho(d)}{3\eta}$, proving the lemma. ∎

Lemma A.4 *If the system (FP_d) is feasible and $\nu > 0$, then*

$$\rho(\hat{d}) \geq \frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}.$$

Proof: Denote $\eta \overset{\triangle}{=} \max\{1, \|d\|/\nu\}$. We want to show that $\rho(\hat{d}) \geq \frac{\rho(d)}{3\eta}$. Let $v_1 \in X$ and $v_2 \in Y$ be arbitrary vectors satisfying $\max\{\|v_1\|, \|v_2\|\} \leq \frac{\rho(d)}{3\eta}$, and consider the vector $v = \frac{1}{\nu}Av_1 + v_2 \in Y$. Then

$$\|v\| = \left\|\frac{1}{\nu}Av_1 + v_2\right\| \leq \frac{\rho(d)}{3\eta}\left(\frac{\|d\|}{\nu} + 1\right) \leq \frac{2}{3}\rho(d).$$

By Theorem 2.1 there exists $(\bar{\theta}, \bar{x})$ such that

$$b\bar{\theta} - A\bar{x} - \tfrac{3}{2}v \in C_Y$$
$$\bar{\theta} \geq 0, \ \bar{x} \in C_X,$$
$$|\bar{\theta}| + \|\bar{x}\| \leq 1.$$

Define $(\hat{\theta}, \hat{x}) = \frac{2}{3}(\bar{\theta}, \bar{x} + \frac{3}{2\nu}v_1)$. Then

$$b\hat{\theta} - A\hat{x} - v_2 = \frac{2}{3}b\bar{\theta} - \frac{2}{3}A\bar{x} - \frac{Av_1}{\nu} - v_2 = \frac{2}{3}b\bar{\theta} - \frac{2}{3}A\bar{x} - v = \frac{2}{3}\left(b\bar{\theta} - A\bar{x} - \frac{3}{2}v\right) \in C_Y,$$

$$\nu \hat{x} - v_1 = \frac{2}{3} \nu \bar{x} \in C_X,$$

$$\hat{\theta} \geq 0, \ \hat{x} \in X,$$

and

$$\hat{\theta} + \|\hat{x}\| \leq \frac{2}{3} \left(\bar{\theta} + \|\bar{x}\| + \frac{\rho(d)}{2\eta\nu} \right) \leq \frac{2}{3} + \frac{1}{3} \frac{\|d\|}{\eta\nu} \leq 1.$$

We conclude from Theorem 2.1 that $\rho(\hat{d}) \geq \frac{\rho(d)}{3\eta}$, proving the lemma. ∎

Combining Lemmas A.3 and A.4 with the fact that $\|\hat{d}\| = \max\{\|d\|, \nu\}$, we obtain

Theorem A.2

$$\mathcal{C}(d) \leq \mathcal{C}(\hat{d}) \leq 3\mathcal{C}(d) \max \left\{ \frac{\nu}{\|d\|}, \frac{\|d\|}{\nu} \right\}.$$

Appendix B

Compact-form Reformulations

This appendix contains proofs of the six propositions of Chapter 4. We restate the six propositions as well as the notation of the corresponding sections of Chapter 4 in order to make the exposition easier to follow.

B.1   Case 1: C_X regular and $C_Y = \{0\}$.

$$\begin{array}{llll} (\text{FP}_d) & Ax = b & (\text{SA}_d) & A^t s \in C_X^* \\ & x \in C_X & & b^t s < 0. \end{array} \tag{B.1}$$

Here Y is an m-dimensional Euclidean space with Euclidean norm $\|y\| = \|y\|_2$ for $y \in Y$.

When the system (FP_d) of (B.1) is feasible, we consider the following reformulation:

$$\begin{aligned} -b\theta + Ax &= 0 \\ \theta \geq 0, \; x &\in C_X, \end{aligned} \tag{B.2}$$

which is of the form (HCE) of (3.32) under the following assignments:

$$M = \left[\begin{array}{cc} -b & A \end{array} \right], \; C = \Re_+ \times C_X$$

with norms defined as $\|(\theta, x)\| = |\theta| + \|x\|$, $(\theta, x) \in \Re \times X$ and $\|v\| = \|v\|_2$, $v \in Y$.

Proposition B.1 (Proposition 4.1) *Suppose the system (FP$_d$) of (B.1) is feasible and $\rho(d) > 0$. Then the system (B.2) is feasible, M has full row rank, and we have*

$$\|M\| = \|d\|, \ \text{and} \ \rho(M) = \rho(d),$$

where $\rho(M)$ is defined in (3.33).

Proof: Feasibility of the system (B.2) is trivially obvious. The expression for $\|M\| = \|d\|$ follows from the definition of the operator norm. Next we establish the bounds on $\rho(M)$. Using (3.33),

$$\rho(M) = \min_{\substack{v \in Y \\ \|v\| \le 1}} \max_{\theta, x, \phi} \phi \quad \text{s.t.} \quad \begin{array}{l} -b\theta + Ax - \phi v = 0 \\ \theta \ge 0, \ x \in C_X \\ \theta + \|x\| \le 1, \end{array}$$

which is a slightly altered version of program $\mathrm{P}_r(d)$ of [16] (for the case when $C_Y = \{0\}$). Therefore, $\rho(M) = \rho(d)$. Finally, since $\rho(M) = \rho(d) > 0$, Remark 3.2 implies that M has full row rank. ∎

When the system (FP$_d$) of (B.1) is infeasible, we consider the following reformulation:

$$\begin{aligned} -b\theta + Ax &= 0 \\ \theta + \bar{f}^t x &= 1 \\ \theta \ge 0, \ x &\in C_X. \end{aligned} \tag{B.3}$$

Its alternative system,

$$\begin{aligned} -b^t s &> 0 \\ A^t s &\in \mathrm{int} C_X^*, \end{aligned} \tag{B.4}$$

is of the form (HCI) of (3.26) under the following assignments:

$$M = \begin{bmatrix} -b & A \end{bmatrix}, \ C = \Re_+ \times C_X$$

with norms defined as $\|(\theta, x)\| = |\theta| + \|x\|$, $(\theta, x) \in \Re \times X$ and $\|v\| = \|v\|_2$, $v \in Y$.

Proposition B.2 (Proposition 4.2) *Suppose (FP$_d$) of (B.1) is infeasible and $\rho(d) > 0$.*
Then the system (B.4) is feasible, and we have

$$\|M\| = \|d\|, \ \text{ and } \rho(d) \le r(M) \le \frac{\rho(d)}{\beta},$$

where $r(M)$ is defined in (3.28).

Proof: Suppose the system (B.4) is infeasible. Then by Proposition 3.1 system (B.3) is feasible and, using Theorem 2.4, $\rho(d) = 0$, which is a contradiction. Hence, the system (B.4) is feasible.

The expression for $\|M\| = \|d\|$ follows from the definition of the operator norm. Next we establish the bounds on $r(M)$. Using Proposition 3.5,

$$
\begin{array}{llll}
r(M) = & \min & \|M(\theta, x)\| & = \min & \|b\theta - Ax\| \\
& & \theta + \bar{f}^t x = 1 & & \theta + \bar{f}^t x = 1 \\
& & \theta \ge 0, \ x \in C_X & & \theta \ge 0, \ x \in C_X,
\end{array}
\tag{B.5}
$$

which is exactly program P$_g(d)$ of [16] (for the case when $C_Y = \{0\}$). Therefore, applying Theorem 3.9 of [16] we conclude that $\beta r(M) \le \rho(d) \le r(M)$, that is, $\rho(d) \le r(M) \le \frac{\rho(d)}{\beta}$. $\blacksquare$

B.2 Case 2: $C_X = X$ and C_Y regular.

$$
\begin{array}{llll}
(\text{FP}_d) & b - Ax \in C_Y & (\text{SA}_d) & A^t s = 0 \\
& & & s \in C_Y^* \\
& & & b^t s < 0.
\end{array}
\tag{B.6}
$$

Here X is an n-dimensional Euclidean space with Euclidean norm $\|x\| = \|x\|_2$ for $x \in X$ (and therefore the dual norm is also $\|q\|_* = \|q\|_2$ for $q \in X^*$).

When the system (FP$_d$) of (B.6) is feasible, we consider the following reformulation:

$$
\begin{aligned}
b^t s \quad +\nu t &= 0 \\
-A^t s \quad\quad &= 0 \\
e^t s \quad +t &= 1 \\
s \in C_Y^*, \quad t &\geq 0,
\end{aligned}
\tag{B.7}
$$

where $\nu > 0$ is a fixed constant. Its alternative system,

$$
\begin{aligned}
b\theta \quad -Ax \quad &\in \mathrm{int} C_Y \\
\nu\theta \quad\quad\quad &> 0,
\end{aligned}
\tag{B.8}
$$

is of the form (HCI) of (3.26) under the following assignments:

$$
M = \begin{bmatrix} b^t & \nu \\ -A^t & 0 \end{bmatrix}, \quad C = C_Y^* \times \Re_+
$$

with norms defined as $\|(s,t)\| = \|s\|_* + |t|$, $(s,t) \in Y^* \times \Re$ and $\|(j,q)\| = \|(j,q)\|_2 = \sqrt{j^2 + \|q\|_2^2}$, $(j,q) \in \Re \times X^*$.

Proposition B.3 (Proposition 4.3) *Suppose (FP$_d$) of (B.6) is feasible and $\rho(d) > 0$. Then the system (B.8) is feasible, and we have*

$$
\max\{\|d\|, \nu\} \leq \|M\| \leq \sqrt{2} \max\{\|d\|, \nu\},
$$

$$
\frac{\rho(d)}{3\max\{1, \|d\|/\nu\}} \leq r(M) \leq \frac{\sqrt{2}\rho(d)}{\breve{\tau}}.
$$

Proof: Suppose the system (B.8) is infeasible. Then by Proposition 3.1 system (B.7) is feasible and, using Theorem 2.2, $\rho(d) = 0$, which is a contradiction. Hence, the system (B.8) is feasible.

The value of $\|M\|$ can be expressed as follows:

$$
\begin{aligned}
\|M\| = \max \quad & \sqrt{(b^t s + \nu t)^2 + \| - A^t s\|^2} \\
\text{s.t.} \quad & \|s\|_* + |t| = 1.
\end{aligned}
\tag{B.9}
$$

143

Let $\hat{a}$ and $\hat{b}$ be vectors satisfying

$$\hat{a} \in Y^*, \ \|\hat{a}\|_* = 1, \ \|A\| = \|A^t\| = \|A^t\hat{a}\|, \ \text{and } \hat{b} \in Y^*, \ \|\hat{b}\|_* = 1, \ \hat{b}^t b = \|b\|.$$

Then we obtain a lower bound on $\|M\|$ by combining the values of $\|M(s,t)\|$ for the following three vectors (s,t):

- letting $(s,t) = (0,1)$, we have $\|M\| \geq \nu$;

- letting $(s,t) = (\hat{a},0)$, we have $\|M\| \geq \|A\|$;

- letting $(s,t) = (\hat{b},0)$, we have $\|M\| \geq \|b\|$;

therefore, $\|M\| \geq \max\{\nu, \|A\|, \|b\|\} = \max\{\|d\|, \nu\}$.

To obtain an upper bound on $\|M\|$, notice that for (s,t) satisfying $\|s\|_* + |t| = 1$,

$$|b^t s + \nu t| \leq \max\{\|b\|, \nu\} \cdot (\|s\|_* + |t|) = \max\{\|b\|, \nu\} \leq \max\{\|d\|, \nu\},$$

and

$$\|A^t s\| \leq \|A\| \leq \max\{\|d\|, \nu\},$$

therefore, $\|M\| \leq \sqrt{\max\{\|d\|, \nu\}^2 + \max\{\|d\|, \nu\}^2} = \sqrt{2}\max\{\|d\|, \nu\}$, which completes the proof of the second claim of the proposition.

Next, we establish the bounds on $r(M)$. Using Proposition 3.5,

$$r(M) \overset{\triangle}{=} \min_{\substack{(s,t) \in C \\ e^t s + t = 1}} \|M(s,t)\| = \min_{\substack{s \in C_Y^*, \ t \geq 0 \\ e^t s + t = 1}} \sqrt{(b^t s + \nu t)^2 + \| - A^t s\|^2}$$

Taking the Lagrange dual of the above convex program, we can also express $r(M)$ as

$$r(M) = \max_{\theta, x, \delta} \quad \delta$$
$$\begin{aligned} & b\theta - Ax - \delta e \in C_Y \\ & \nu\theta - \delta \geq 0 \\ & \sqrt{\theta^2 + \|x\|^2} \leq 1. \end{aligned} \tag{B.10}$$

Let $(\bar{\theta}, \bar{x})$ be a solution of (B.10), i.e.,

$$b\bar{\theta} - A\bar{x} - r(M)e = \bar{y} \in C_Y, \ \nu\bar{\theta} \geq r(M), \ \sqrt{\bar{\theta}^2 + \|\bar{x}\|^2} = 1.$$

Let $v \in Y$ be an arbitrary vector satisfying $\|v\| \leq \frac{r(M)\check{r}}{\sqrt{2}}$. Define

$$\hat{\theta} = \frac{\bar{\theta}}{|\bar{\theta}| + \|\bar{x}\|}, \ \hat{x} = \frac{\bar{x}}{|\bar{\theta}| + \|\bar{x}\|}, \ \hat{y} = \frac{\bar{y} + r(M)e - (|\bar{\theta}| + \|\bar{x}\|)v}{|\bar{\theta}| + \|\bar{x}\|}.$$

Notice that $\hat{y} \in C_Y$, since for any $s \in C_Y^*$ such that $\|s\|_* = 1$ we have

$$s^t(\bar{y} + r(M)e - (|\bar{\theta}| + \|\bar{x}\|)v) \geq \check{r}r(M) - \sqrt{2}\sqrt{|\bar{\theta}|^2 + \|\bar{x}\|^2}\|v\| \geq \check{r}r(M) - \sqrt{2}\|v\| \geq 0.$$

Therefore $(\hat{\theta}, \hat{x})$ satisfy $\hat{\theta} \geq 0$, $\|\hat{x}\| + |\hat{\theta}| \leq 1$ and

$$b\hat{\theta} - A\hat{x} - v = \frac{b\bar{\theta} - A\bar{x} - (|\bar{\theta}| + \|\bar{x}\|)v}{|\bar{\theta}| + \|\bar{x}\|} = \frac{\bar{y} + r(M)e - (|\bar{\theta}| + \|\bar{x}\|)v}{|\bar{\theta}| + \|\bar{x}\|_2} = \hat{y} \in C_Y.$$

We conclude from Theorem 2.1 that $\rho(d) \geq \frac{r(M)\check{r}}{\sqrt{2}}$, that is, $r(M) \leq \frac{\sqrt{2}\rho(d)}{\check{r}}$, establishing the upper bound of $r(M)$ stated in the proposition.

Next, let $\eta = \max\{1, \|d\|/\nu\}$. Consider the vector $v = \frac{\rho(d)}{2\eta}\left(-\frac{b}{\nu} + e\right)$. Since $\|v\| \leq \frac{\rho(d)}{2\eta}\left(\frac{\|d\|}{\nu} + 1\right) \leq \rho(d)$, by Theorem 2.1, there exists $(\bar{\theta}, \bar{x})$ satisfying $b\bar{\theta} - A\bar{x} - v = \bar{y} \in C_Y$, $\bar{\theta} \geq 0$, and $|\bar{\theta}| + \|\bar{x}\| \leq 1$.

Define $(\tilde{\theta}, \tilde{x}) = \frac{2}{3}(\bar{\theta}, \bar{x}) + \frac{\rho(d)}{3\nu\eta}(1, 0)$. This vector is feasible for the program (B.10) with $\delta = \frac{\rho(d)}{3\eta}$ since $\sqrt{\tilde{\theta}^2 + \|\tilde{x}\|^2} \leq |\tilde{\theta}| + \|\tilde{x}\| \leq \frac{2}{3} + \frac{\rho(d)}{3\nu\eta} \leq 1$, $\nu\tilde{\theta} - \delta = \frac{2}{3}\nu\bar{\theta} + \frac{\rho(d)}{3\eta} - \frac{\rho(d)}{3\eta} = \frac{2}{3}\nu\bar{\theta} \geq 0$, and

$$b\tilde{\theta} - A\tilde{x} - \delta e = \frac{2}{3}(b\bar{\theta} - A\bar{x}) + \frac{\rho(d)}{3\nu\eta}b - \frac{\rho(d)}{3\eta}e = \frac{2}{3}(\bar{y} + v) + \frac{\rho(d)}{3\nu\eta}b - \frac{\rho(d)}{3\eta}e$$

$$= \frac{2}{3}\bar{y} + \frac{2}{3}\frac{\rho(d)}{2\eta}\left(-\frac{b}{\nu} + e\right) + \frac{\rho(d)}{3\nu\eta}b - \frac{\rho(d)}{3\eta}e = \frac{2}{3}\bar{y} \in C_Y.$$

Therefore, $r(M) \geq \delta = \frac{\rho(d)}{3\eta}$, completing the proof of the proposition. ∎

When the system (FP$_d$) of (B.6) is infeasible, we consider the following reformulation:

$$
\begin{aligned}
b^t s \quad + \nu t &= 0 \\
-A^t s \qquad\;\; &= 0 \\
s \in C_Y^*, \quad t &\geq 0,
\end{aligned}
\tag{B.11}
$$

where $\nu > 0$ is a fixed constant. System (B.11) is of the form (HCE) of (3.32) under the following assignments:

$$
M = \begin{bmatrix} b^t & \nu \\ -A^t & 0 \end{bmatrix}, \; C = C_Y^* \times \Re_+
$$

with norms defined as $\|(s,t)\| = \|s\|_* + |t|$, $(s,t) \in Y^* \times \Re$ and $\|(j,q)\| = \|(j,q)\|_2 = \sqrt{j^2 + \|q\|_2^2}$, $(j,q) \in \Re \times X^*$.

Proposition B.4 (Proposition 4.4) *Suppose (FP$_d$) of (B.6) is infeasible and $\rho(d) > 0$. Then the system (B.11) is feasible, M has full row rank, and we have*

$$
\max\{\|d\|, \nu\} \leq \|M\| \leq \sqrt{2}\max\{\|d\|, \nu\},
$$

$$
\frac{\rho(d)}{3\max\{1, \|d\|/\nu\}} \leq \rho(M) \leq \sqrt{2}\rho(d).
$$

Proof: Feasibility of the system (B.11) trivially follows from Proposition 2.2. The proof of the bounds on $\|M\|$ is identical to the proof of these bounds in Proposition B.3.

It remains to establish the bounds on $\rho(M)$ and to show that M has full row rank. Using (3.33),

$$
\begin{aligned}
\rho(M) = \quad &\min \qquad\quad \max \qquad\quad \phi \\
&(v_1, v_2) \in \Re \times X^* \quad s, t, \phi \\
&\|(v_1, v_2)\| \leq 1 \qquad \text{s.t.} \qquad b^t s \quad + \nu t \;\; -\phi v_1 = 0 \\
&\qquad\qquad\qquad\qquad\qquad\quad\; -A^t s \qquad\quad\;\; -\phi v_2 = 0 \\
&\qquad\qquad\qquad\qquad\qquad\quad\;\; \|s\|_* \quad + |t| \qquad\quad \leq 1 \\
&\qquad\qquad\qquad\qquad\qquad\quad\; s \in C_Y^*, \quad t \geq 0.
\end{aligned}
\tag{B.12}
$$

In order to show that $\rho(M) \geq \frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}$ it is sufficient to show that for any (γ, w) such

146

that $\sqrt{\gamma^2 + \|w\|^2} \le \frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}$, we have $(\gamma, w) \in \mathcal{H}$, where

$$\mathcal{H} \triangleq \{(b^t s + \nu t, -A^t s) : \|s\|_* + |t| \le 1, \ s \in C_Y^*, t \ge 0\}.$$

We will establish a somewhat stronger fact: for any $w \in X^*$ with $\|w\| \le \frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}$ there exist $\alpha_1 \ge \frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}$ and $\alpha_2 \le -\frac{\rho(d)}{3\max\{1, \|d\|/\nu\}}$ such that $(\alpha_1, w) \in \mathcal{H}$ and $(\alpha_2, w) \in \mathcal{H}$. Since $\{(\gamma, w) : \sqrt{\gamma^2 + \|w\|_2^2} \le \delta\} \subset \{(\gamma, w) : \|w\|_2 \le \delta, \ |\gamma| \le \delta\}$, the above fact will imply the lower bound on $\rho(M)$.

Denote $\eta \triangleq \max\{1, \|d\|/\nu\}$ and let $w \in X^*$ be an arbitrary vector satisfying $\|w\| \le \frac{\rho(d)}{3\eta} \le \rho(d)$. By Theorem 2.3 there exists $s \in C_Y^*$, $\|s\|_* \le 1$ such that $-b^t s \ge \rho(d)$, $A^t s = -w$. Let $t = 0$. We have: $s \in C_Y^*$, $t \ge 0$, $\|s\|_* + t = \|s\|_* \le 1$, and $A^t s = -w$. We define

$$\alpha_2 \triangleq b^t s + \nu t = b^t s \le -\rho(d) \le -\frac{\rho(d)}{3\eta}.$$

Therefore, $(\alpha_2, w) \in \mathcal{H}$, with $\alpha_2 \le -\frac{\rho(d)}{3\eta}$.

Next, let $v = -3\eta w$. Then $\|v\| \le \rho(d)$, therefore there exists $s \in C_Y^*$ such that $\|s\|_* \le 1$, $A^t s = -3\eta w$, $-b^t s \ge \rho(d)$. Let $\bar{s} = \frac{s}{3\eta} \in C_Y^*$, $\bar{t} = \frac{2\|d\|}{3\eta\nu} \ge 0$. Then $\|\bar{s}\|_* + \bar{t} \le \frac{1}{3\eta} + \frac{2}{3}\frac{\|d\|}{\eta\nu} \le 1$, $-A^t \bar{s} = w$, and
$$\alpha_1 \triangleq b^t \bar{s} + \nu \bar{t} \ge -\|b\| \cdot \frac{\|s\|_*}{3\eta} + \frac{2\|d\|}{3\eta} \ge \frac{\rho(d)}{3\eta}.$$

Therefore, $(\alpha_1, w) \in \mathcal{H}$, with $\alpha_1 \ge \frac{\rho(d)}{3\eta}$. This establishes the lower bound on $\rho(M)$. Also, from Remark 3.2, M has full row rank.

It remains to establish the upper bound on $\rho(M)$. Suppose $v \in X^*$ is an arbitrary vector satisfying $\|v\| \le \frac{\rho(M)}{\sqrt{2}}$. Let $w = -v$ and $\gamma = -\|w\|$. Then $\|(\gamma, w)\| = \sqrt{2}\|w\| \le \rho(M)$. Hence, there exist $s \in C_Y^*$ and $t \ge 0$ such that $\|s\|_* + |t| \le 1$, $A^t s = -w = v$ and $b^t s + \nu t = \gamma$. We conclude from Theorem 2.3 that $\rho(d) \ge \frac{\rho(M)}{\sqrt{2}}$, thus establishing the last statement of the proposition. $\blacksquare$

147

B.3 Case 3: C_X and C_Y are both regular

$$
\begin{array}{llll}
(\text{FP}_d) & b - Ax \in C_Y & (\text{SA}_d) & A^t s \in C_X^* \\
& x \in C_X & & s \in C_Y^* \\
& & & b^t s < 0.
\end{array}
\tag{B.13}
$$

Here, X is an n-dimensional Euclidean space with Euclidean norm $\|x\| = \|x\|_2$ for $x \in X$ (and therefore the dual norm is also $\|q\|_* = \|q\|_2$ for $q \in X^*$), and Y is an m-dimensional Euclidean space with Euclidean norm $\|y\| = \|y\|_2$ for $y \in Y$ (and therefore the dual norm is also $\|s\|_* = \|s\|_2$ for $s \in Y^*$).

When the system (FP_d) of (B.13) is feasible, we consider the following reformulation:

$$
\begin{array}{rcccll}
b^t s & & +\nu t & = 0 \\
-A^t s & +\nu q & & = 0 \\
e^t s & +f^t q & +t & = 1 \\
s \in C_Y^* & q \in C_X^* & t \geq 0
\end{array}
\tag{B.14}
$$

where $\nu > 0$ is a fixed constant. Its alternative system,

$$
\begin{array}{rll}
b\theta \ -Ax & \in \text{int} C_Y \\
\nu x & \in \text{int} C_X \\
\nu \theta & > 0,
\end{array}
\tag{B.15}
$$

is of the form (HCI) of (3.26) under the following assignments:

$$
M = \begin{bmatrix} b^t & 0 & \nu \\ -A^t & \nu I & 0 \end{bmatrix}, \quad C = C_Y^* \times C_X^* \times \Re_+
$$

with norms defined as $\|(s, q, t)\| = \|s\|_* + \|q\|_* + |t|$, $(s, q, t) \in Y^* \times X^* \times \Re$ and $\|(j, q)\| = \|(j, q)\|_2 = \sqrt{j^2 + \|q\|_2^2}$, $(j, q) \in \Re \times X^*$.

Proposition B.5 (Proposition 4.5) *Suppose (FP$_d$) of (B.13) is feasible and $\rho(d) > 0$.*

Then the system (B.15) is feasible, and we have

$$\max\{\|d\|, \nu\} \le \|M\| \le \sqrt{2}\max\{\|d\|, \nu\},$$

$$\frac{\rho(d)}{5\max\{1, \|d\|/\nu\}} \le r(M) \le \frac{\sqrt{2}\rho(d)}{\breve{\tau}} \le \frac{\sqrt{2}\rho(d)}{\min\{\tau, \breve{\tau}\}}.$$

Proof: Suppose the system (B.15) is infeasible. Then by Proposition 3.1 system (B.14) is feasible and, using Theorem 2.2, $\rho(d) = 0$, which is a contradiction. Hence, the system (B.15) is feasible.

By the definition of the operator norm,

$$\begin{aligned}\|M\| = \max \quad &\sqrt{(b^t s + \nu t)^2 + \| - A^t s + \nu q\|_2^2} \\ \text{s.t.} \quad &\|s\|_* + \|q\|_* + |t| = 1.\end{aligned} \tag{B.16}$$

Let $\hat{a}$ and $\hat{b}$ be vectors satisfying

$$\hat{a} \in Y^*, \ \|\hat{a}\|_* = 1, \ \|A\| = \|A^t\| = \|A^t \hat{a}\|, \ \text{and} \ \hat{b} \in Y^*, \ \|\hat{b}\|_* = 1, \ \hat{b}^t b = \|b\|.$$

Then we obtain a lower bound on $\|M\|$ by combining the values of $\|M(s, q, t)\|$ for the following three vectors (s, q, t):

- letting $(s, q, t) = (0, 0, 1)$, we have $\|M\| \ge \nu$;

- letting $(s, q, t) = (\hat{a}, 0, 0)$, we have $\|M\| \ge \|A\|$;

- letting $(s, q, t) = (\hat{b}, 0, 0)$, we have $\|M\| \ge \|b\|$;

therefore, $\|M\| \ge \max\{\nu, \|A\|, \|b\|\} = \max\{\|d\|, \nu\}$.

To obtain an upper bound on $\|M\|$, notice that for (s, q, t) satisfying $\|s\|_* + \|q\|_* + |t| = 1$,

$$|b^t s + \nu t| \le \max\{\|b\|, \nu\} \cdot (\|s\|_* + |t|) \le \max\{\|b\|, \nu\} \le \max\{\|d\|, \nu\},$$

and

$$\| - A^t s + \nu q\| \le \max\{\|A\|, \nu\} \cdot (\|s\|_* + \|q\|_*) \le \max\{\|A\|, \nu\} \le \max\{\|d\|, \nu\},$$

therefore, $||M|| \leq \sqrt{\max\{||d||, \nu\}^2 + \max\{||d||, \nu\}^2} = \sqrt{2}\max\{||d||, \nu\}$, which completes the proof of the second claim of the proposition.

Next, we establish bounds on $r(M)$. Using Proposition 3.5,

$$r(M) = \min \quad \sqrt{(b^t s + \nu t)^2 + ||-A^t s + \nu q||^2}$$
$$s \in C_Y^*, \ q \in C_X^*, \ t \geq 0$$
$$e^t s + f^t q + t = 1.$$

Taking the Lagrange dual of the above convex program, we can also express $r(M)$ as

$$r(M) = \max \quad \delta$$
$$b\theta - Ax - \delta e \in C_Y$$
$$\nu x - \delta f \in C_X \qquad\qquad\qquad \text{(B.17)}$$
$$\nu\theta - \delta \geq 0$$
$$\sqrt{\theta^2 + ||x||_2^2} \leq 1.$$

Let $(\bar{\theta}, \bar{x})$ be a solution of (B.17), i.e.,

$$b\bar{\theta} - A\bar{x} - r(M)e = \bar{y} \in C_Y, \ \nu\bar{x} - r(M)f \in C_X, \ \nu\bar{\theta} \geq r(M), \ \sqrt{\bar{\theta}^2 + ||\bar{x}||^2} = 1.$$

Let $v \in Y$ be an arbitrary vector satisfying $||v|| \leq \frac{r(M)\check{r}}{\sqrt{2}}$. Define

$$\hat{\theta} = \frac{\bar{\theta}}{|\bar{\theta}| + ||\bar{x}||}, \ \hat{x} = \frac{\bar{x}}{|\bar{\theta}| + ||\bar{x}||}, \ \hat{y} = \frac{\bar{y} + r(M)e - (|\bar{\theta}| + ||\bar{x}||)v}{|\bar{\theta}| + ||\bar{x}||}.$$

Notice that $\hat{y} \in C_Y$, since for any $s \in C_Y^*$ with $||s||_* = 1$ we have

$$s^t(\bar{y} + r(M)e - (|\bar{\theta}| + ||\bar{x}||)v) \geq \check{r}r(M) - \sqrt{2}\sqrt{|\bar{\theta}|^2 + ||\bar{x}||^2}\,||v|| \geq \check{r}r(M) - \sqrt{2}||v|| \geq 0.$$

Therefore, $(\hat{\theta}, \hat{x})$ satisfy $\hat{\theta} \geq 0$, $\hat{x} \in C_X$, $|\hat{\theta}| + ||\hat{x}|| = 1$, and

$$b\hat{\theta} - A\hat{x} - v = \frac{b\bar{\theta} - A\bar{x} - (|\bar{\theta}| + ||\bar{x}||)v}{|\bar{\theta}| + ||\bar{x}||} = \frac{\bar{y} + r(M)e - (|\bar{\theta}| + ||\bar{x}||)v}{|\bar{\theta}| + ||\bar{x}||} = \hat{y} \in C_Y.$$

We conclude from Theorem 2.1 that $\rho(d) \geq \frac{r(M)\check{\tau}}{\sqrt{2}}$, that is, $r(M) \leq \frac{\sqrt{2}\rho(d)}{\check{\tau}}$, establishing the upper bound of $r(M)$ stated in the proposition.

Next, let $\eta = \max\{1, \|d\|/\nu\}$. Consider the vector $v = \frac{\rho(d)}{3\eta}\left(-\frac{b}{\nu} + \frac{Af}{\nu} + e\right)$. Since $\|v\| \leq \frac{\rho(d)}{3\eta}\left(\frac{2\|d\|}{\nu} + 1\right) \leq \frac{\rho(d)}{3\eta}(3\eta) = \rho(d)$, therefore by Theorem 2.1 there exists $(\bar{\theta}, \bar{x})$ satisfying $b\bar{\theta} - A\bar{x} - v = \bar{y} \in C_Y$, $\bar{\theta} \geq 0$, $\bar{x} \in C_X$, and $|\bar{\theta}| + \|\bar{x}\| \leq 1$.

Define $(\tilde{\theta}, \tilde{x}) = \frac{3}{5}(\bar{\theta}, \bar{x}) + \frac{\rho(d)}{5\nu\eta}(1, f)$. This vector is feasible for the program (B.17) with $\delta = \frac{\rho(d)}{5\eta}$ since $\tilde{\theta}^2 + \|\tilde{x}\|^2 \leq |\tilde{\theta}| + \|\tilde{x}\| \leq \frac{3}{5} + \frac{2\rho(d)}{5\nu\eta} \leq 1$, $\nu\tilde{x} - \delta f = \frac{3}{5}\nu\bar{x} + \frac{\rho(d)}{5\eta}f - \frac{\rho(d)}{5\eta}f = \frac{3}{5}\nu\bar{x} \in C_X$, $\nu\tilde{\theta} - \delta = \frac{3}{5}\nu\bar{\theta} + \frac{\rho(d)}{5\eta} - \frac{\rho(d)}{5\eta} = \frac{3}{5}\nu\bar{\theta} \geq 0$, and

$$b\tilde{\theta} - A\tilde{x} - \delta e = \frac{3}{5}(b\bar{\theta} - A\bar{x}) + \frac{\rho(d)}{5\nu\eta}(b - Af) - \frac{\rho(d)}{5\eta}e = \frac{3}{5}(\bar{y} + v) + \frac{\rho(d)}{5\nu\eta}(b - Af) - \frac{\rho(d)}{5\eta}e$$

$$= \frac{3}{5}\bar{y} + \frac{3}{5}\frac{\rho(d)}{3\eta}\left(-\frac{b}{\nu} + \frac{Af}{\nu} + e\right) + \frac{\rho(d)}{5\nu\eta}(b - Af) - \frac{\rho(d)}{5\eta}e = \frac{3}{5}\bar{y} \in C_Y.$$

Therefore, $r(M) \geq \delta = \frac{\rho(d)}{5\eta}$, completing the proof of the proposition. ∎

When the system (FP_d) of (B.13) is infeasible, we consider the following reformulation:

$$\begin{aligned} -b\theta + Ax + \nu y &= 0 \\ \theta + \bar{f}^t x + \bar{e}^t y &= 1 \\ \theta \geq 0, \; x \in C_X, \; y \in C_Y, \end{aligned} \tag{B.18}$$

where $\nu > 0$ is a fixed constant. Its alternative system,

$$\begin{aligned} -b^t s &> 0 \\ A^t s &\in \mathrm{int}C_X^* \\ \nu s &\in \mathrm{int}C_Y^*, \end{aligned} \tag{B.19}$$

is of the form (HCI) of (3.26) under the following assignments:

$$M = \begin{bmatrix} -b & A & \nu I \end{bmatrix}, \; C = \Re_+ \times C_X \times C_Y$$

with norms defined as $\|(\theta, x, y)\| = |\theta| + \|x\| + \|y\|$, $(\theta, x, y) \in \Re \times X \times Y$ and $\|v\| = \|v\|_2$, $v \in$

Y.

Proposition B.6 (Proposition 4.6) *Suppose (FP$_d$) of (B.13) is infeasible and $\rho(d) > 0$.*
Then the system (B.19) is feasible and we have

$$\|M\| = \max\{\|d\|, \nu\},$$

$$\frac{\rho(d)}{3\max\{1, \|d\|/\tau\}} \le r(M) \le \frac{\rho(d)}{\beta} \le \frac{\rho(d)}{\min\{\beta, \breve{\beta}\}}.$$

Proof: Suppose the system (B.19) is infeasible. Then by Proposition 3.1 system (B.18) is feasible and, using Theorem 2.4, $\rho(d) = 0$, which is a contradiction. Hence, the system (B.19) is feasible.

By the definition of the operator norm,

$$\|M\| = \max \quad \|-br + Ax + \nu y\| \quad = \max\{\|b\|, \|A\|, \nu\} = \max\{\|d\|, \nu\}.$$
$$\text{s.t.} \quad |r| + \|x\| + \|y\| = 1$$

Next we compute the bounds on $r(M)$. Using Proposition 3.5,

$$
\begin{aligned}
r(M) = \quad \min \quad & \|b\theta - Ax - \nu y\| \\
& \theta + \bar{f}^t x + \bar{e}^t y = 1 \\
& \theta \ge 0, \; x \in C_X, \; y \in C_Y.
\end{aligned}
\tag{B.20}
$$

Taking the Lagrange dual of the above convex program, we can also express $r(M)$ as

$$
\begin{aligned}
r(M) = \max \quad & \delta \\
& -b^t s - \delta \ge 0 \\
& A^t s - \delta \bar{f} \in C_X^* \\
& \nu s - \delta \bar{e} \in C_Y^* \\
& \|s\|_* \le 1.
\end{aligned}
\tag{B.21}
$$

Suppose $(\bar{\theta}, \bar{x}, \bar{y})$ is a solution of (2.18). Let

$$(\theta, x, y) = \frac{(\bar{\theta}, \bar{x}, \bar{y}/\nu)}{\bar{\theta} + \bar{f}^t \bar{x} + \bar{e}^t \bar{y}/\nu}.$$

Then (θ, x, y) is feasible for (B.20), and so

$$r(M) \le \frac{\|b\bar{\theta} - A\bar{x} - \bar{y}\|}{\bar{\theta} + \bar{f}^t \bar{x} + \bar{e}^t \bar{y}/\nu} = \frac{\rho(d)}{\bar{\theta} + \bar{f}^t \bar{x} + \bar{e}^t \bar{y}/\nu} \le \frac{\rho(d)}{\beta}.$$

To establish the lower bound on $r(M)$, let $\eta = \max\{1, \|d\|/\nu\}$ and consider the vector $v = \frac{\rho(d)}{2\eta}\left(\bar{f} - \frac{A^t \bar{e}}{\nu}\right)$. Then

$$\|v\| \le \frac{\rho(d)}{2\eta}\left(1 + \frac{\|d\|}{\nu}\right) \le \rho(d),$$

and therefore by Theorem 2.3 there exists $\bar{s} \in C_Y^*$ such that $\|\bar{s}\|_* \le 1$, $A^t \bar{s} - v = \bar{q} \in C_X^*$, and $-b^t \bar{s} \ge \rho(d)$.

Define $\tilde{s} = \frac{2}{3}\bar{s} + \frac{\rho(d)\bar{e}}{3\nu\eta}$. This vector is feasible for (B.21) with $\delta = \frac{\rho(d)}{3\eta}$. Indeed,

$$\|\tilde{s}\|_* \le \frac{2}{3} + \frac{\rho(d)}{3\nu\eta} \le 1,$$

$$\nu\tilde{s} - \delta\bar{e} = \frac{2}{3}\nu\bar{s} + \frac{\rho(d)}{3\eta}\bar{e} - \frac{\rho(d)}{3\eta}\bar{e} = \frac{2}{3}\nu\bar{s} \in C_Y^*,$$

$$A^t\tilde{s} - \delta\bar{f} = \frac{2}{3}A^t\bar{s} + \frac{\rho(d)}{3\nu\eta}A^t\bar{e} - \frac{\rho(d)}{3\eta}\bar{f} = \frac{2}{3}\bar{q} \in C_X^*,$$

and

$$-b^t\tilde{s} = -\frac{2}{3}b^t\bar{s} - \frac{\rho(d)}{3\nu\eta}b^t\bar{e} \ge \frac{2\rho(d)}{3} - \frac{\rho(d)\|d\|}{3\nu\eta} \ge \frac{\rho(d)}{3} \ge \frac{\rho(d)}{3\eta} = \delta.$$

Thus $r(M) \ge \delta = \frac{\rho(d)}{3\eta}$, which establishes the proposition. ∎

Bibliography

[1] S. Agmon. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:382–392, 1954.

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

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

[4] Daniel Bienstock. Experiments with a network design algorithm using ϵ- approximate linear programs, December 1996.

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

[6] C. Chen and O.L. Mangasarian. Smoothing methods for convex inequalities and linear complementarity problems. *Mathematical Programming*, 71(1):51–70, 1995.

[7] G.B. Dantzig. Converting a converging algorithm into a polynomially bounded algorithm. Technical Report SOL 91-5, Stanford University, 1991.

[8] G.B. Dantzig. An ϵ-precise feasible solution to a linear program with a convexity constraint in $1/\epsilon^2$ iterations independent of problem size. Technical Report, Stanford University, 1992.

[9] B.C. Eaves. Piecewise linear retractions by reflection. *Linear Algebra and its Applications*, 7:93–98, 1973.

[10] Marina Epelman and Robert M. Freund. Condition number complexity of an elementary algorithm for resolving a conic linear system. Working Paper OR 319-97, Operations Research Center, Massachusetts Institute of Technology, February 1997.

[11] Marina Epelman and Robert M. Freund. Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system. Working Paper OR 330-98, Operations Research Center, Massachusetts Institute of Technology, December 1998.

[12] Sharon 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.

[13] Sharon 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.

[14] Robert M. Freund and Jorge R. Vera. Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm. *SIAM Journal on Optimization* (to appear), 1999.

[15] Robert M. Freund and Jorge R. Vera. On the complexity of estimating condition measures of conic linear systems. In preparation, 1999.

[16] Robert M. Freund and Jorge R. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Mathematical Programming* (to appear), 1999.

[17] J.L. Goffin. The relaxation method for solving systems of linear inequalities. *Mathematics of Operations Research*, 5(3):388–414, 1980.

[18] Andrew V. Goldberg, Jeffrey D. Oldham, Serge Plotkin, and Cliff Stein. An implementation of a combinatorial approximation algorithm for minimum-cost multicommodity flow, December 1997.

[19] Michael D. Grigoriadis and Leonid G. Khachiyan. Fast approximation schemes for convex programs with many blocks and coupling constraints. *SIAM Journal on Optimization*, 4(1):86–107, 1994.

[20] Michael D. Grigoriadis and Leonid G. Khachiyan. Coordination complexity of parallel price-directive decomposition. *Mathematics of Operations Research*, 21(2):321–340, 1996.

[21] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combiantorial Optimization*. Springer-Verlag, 1994.

[22] Jackie Cheuk Kin Ho. Structure of the central path and related complexity issues for linear programming. Master's thesis, University of Waterloo, Waterloo, Ontario, Canada, 1999.

[23] Richard A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.

[24] 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.

[25] David Karger and Serge Plotkin. Adding multiple cost constraints to combinatorial optimization problems, with applications to multicommodity flows. *Proceedings of the Twenty-Seventh Annual ACM Symposium on the Theory of Computing*, pages 18–25, 1995.

[26] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

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

[28] 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.

[29] 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.

[30] T.S. Motzkin and I.J. Schoenberg. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.

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

[32] Manuel A. Nunez. A characterization of ill-posed data instances for convex programming, September 1998.

[33] Manuel A. Nunez and Robert M. Freund. Condition measures and properties of the central trajectory of a linear program. *Mathematical Programming*, 83(1):1–28, 1998.

[34] Javier Peña. Understanding the geometry of infeasible perturbations of a conic linear system. Preprint.

[35] Javier Peña. Computing the distance to infeasibility: theoretical and practical issues. Technical report, Cornell University, October 1997.

[36] Javier Peña and James Renegar. Computing approximate solutions for conic systems of constraints, May 1998.

[37] Serge A. Plotkin, David B. Shmoys, and Éva Tardos. Fast approximation algorithms for fractional packing and covering problems, February 1995.

[38] J. Renegar. A polynomial-time algorithm, based on Newnton's method, for linear programming. *Mathematical Programming*, 40:59–93, 1988.

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

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

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

[42] James Renegar. A mathematical view of interior-point methods for convex optimization. Lecture notes, July 1998.

[43] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1997.

[44] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958.

[45] F. Rosenblatt. On the convergence of reinforcement procedures in simple perceptrons. Report, VG-1196-G-4, Cornell Aeronautical Laboratory, Buffalo, NY, 1960.

[46] F. Rosenblatt. Perceptron simulation experiments. *Proceedings of the Institute of Radio Engineers*, 48:301–309, 1960.

[47] F. Rosenblatt. *Principles of Neurodynamics*. Spartan Books, Washington, DC, 1962.

[48] M.J. Todd, Levent Tunçel, and Yinyu Ye. Probabilistic analysis of two complexity measures for linear programming problems. Research Report CORR 98-48, Faculty of Mathematics, University of Waterloo, October 1998.

[49] Levent Tunçel. On the condition numbers for polyhedra in Karmarkar's form. Research Report CORR 97-24, Faculty of Mathematics, University of Waterloo, November 1997.

[50] Levent 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.

[51] Stephen A. Vavasis and Yinyu Ye. Condition numbers for polyhedra with real number data. *Operations Research Letters*, 17:209–214, 1995.

[52] Stephen A. Vavasis and Yinyu Ye. A primal-dual interior point method whose running time depends only on the constraint matrix. *Mathematical Programming*, 74:79–120, 1996.

[53] Stephen A. Vavasis and Yinyu Ye. A simplification to "A primal-dual interior point method whose running time depends only on the constraint matrix", October 30 1996.

[54] Jorge R. Vera. Ill-posedness and the computation of solutions to linear programs with approximate data. Technical Report, Cornell University, May 1992.

[55] Jorge R. Vera. *Ill-Posedness in Mathematical Programming and Prob/lem Solving with Approximate Data*. PhD thesis, Cornell University, 1992.

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

[57] Jorge R. Vera. On the complexity of linear programming under finite precision arithmetic. *Mathematical Programming*, 80(1):91–123, 1998.

[58] Jorge Villavicencio and Michael D. Grigoriadis. Approximate Lagrangean decomposition with a modified Karmarkar logarithmic potential. Technical Report LCsR-TR 258, Rutgers University, June 1996.

[59] Stephen J. Wright. *Primal-dual Interior-point methods*. SIAM, 1997.