Condition Numbers and Properties of Central Trajectories in Convex Programming

Manuel A. Núñez Araya

BS in Mathematics, Universidad de Costa Rica (1986)
BS in Computer Science, Universidad de Costa Rica (1987)
MS in Operations Research, Stanford University (1989)
MS in Computer Science, Stanford University (1990)
Engineer's Degree in Operations Research, Stanford University (1990)
MS in Mathematics, Universidad de Costa Rica (1993)

Submitted to the Sloan School of Management in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY August 1997 © Massachusetts Institute of Technology 1997. All rights reserved.
Condition Numbers and Properties of Central Trajectories in Convex Programming

by

Manuel A. Núñez Araya

Submitted to the Sloan School of Management on August 15, 1997, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

Understanding the effect on optimal solutions to convex programs of changes in the input data associated with a given problem is of great concern in industrial applications as well as in optimization theory. A recently developed approach to address this concern is the use of perturbation theory, and in particular, the use of a measure of the stability of a convex program called the condition number. Roughly speaking, the larger the condition number of a convex program, the closer the program will be to being infeasible or unbounded, and so the harder it could be for a numerical method to find an optimal solution to the problem.

In this thesis we are interested in two applications of this theory: (i) continuous complexity, that is, determining new complexity bounds for algorithms applied to mathematical programs with real-number data; and (ii) interior-point methods, in particular, determining the relationship between interior-point methods and the conditioning of a problem.

We present exploratory research and show that condition numbers can be used to study and bound geometric objects of interest in convex programming such as analytic centers and central trajectories: objects that are intrinsically important in the analysis of complexity of interior-point methods and convex programming.

Thesis Supervisor: Robert M. Freund
Title: Seley Professor of Operations Research
Co-Director, Operations Research Center
Acknowledgments

More than being the natural consequence of my doctoral experience at MIT, this dissertation is the outcome of an ongoing learning process supported by the faith and trust of many people through all my life. In this process, I have come a long way in a journey that would not have been possible without the help of the people who saw in me more potential than I ever realized. I can only repay their confidence by doing my best effort to succeed. This document represents part of that effort.

First, I would like to express my infinite gratitude to Professor Robert M. Freund who, by being more a friend than an advisor, has helped me in innumerable ways to achieve my professional goals. My research experience with him has had a dramatic impact in my life. His passion for mathematics has been a constant motivation through my four years at MIT. This thesis is a product of his enthusiastic support and encouragement. Thanks a lot Rob.

I would also like to express my gratitude to my professors at Stanford University and MIT, who were able to instill in my mind the beauty of the optimization principles and to make me a fanatic convert of operations research. In particular, I would like to thank my former advisor Professor George B. Dantzig, for teaching me linear programming from a master craftsman point of view and for helping me to get in to MIT; and Professors Thomas L. Magnanti and Dimitris J. Bertsimas, for their help in writing and supervising this thesis and for their support in my professional endeavors.

Finally, in Costa Rica, I would like to thank my family (Cecilia, Gabriela, and Alvaro), for their love and caring through all my life; my elementary-school teachers Margarita Vargas and Elisa Esquivel, for giving me a very stimulating education at an early age; my teachers at the Marist High School, for teaching me important moral values and nurturing my love for mathematics; and my college professors, William Alvarado,
Edwin Castro, and Ricardo Estrada, for teaching me to think as a mathematician and for their encouragement in getting a college education outside my country.

**Dedication:**

To Victoria Kong.
Biographical Note

The author received his Bachelor's Degrees in Mathematics and Computer Science from the Universidad de Costa Rica in 1986 and 1987, respectively. From 1988 to 1990 he was a graduate student at Stanford University, from which he received two Master's Degrees in Operations Research and Computer Science, and an Engineer's Degree in Operations Research. His Engineer's Degree thesis advisor was Professor George B. Dantzig. In 1993 he received a Master's Degree in Mathematics from the Universidad de Costa Rica. He began his graduate studies at the Massachusetts Institute of Technology in the Fall of 1993. His teaching experience includes seven years as an academic instructor in a number of courses in operations research and computer science at the Universidad de Costa Rica and the Instituto Tecnológico de Costa Rica. He is interested in the broad spectrum of technical arenas that define operations research and their applications to operations management, including such topics as optimization models in production, linear and nonlinear programming, semi-definite programming, interior point methods, convex programming, and simulation modeling of production processes.
## Contents

1 **Introduction**  
1.1 Overview .............................................. 11  
1.2 Motivation to Ill-Posedness of Convex Programs ................. 14  
1.3 Main Issues Addressed in this Research .......................... 18  
1.4 Thesis Summary .......................................... 22  
1.5 Literature Review ........................................ 28  

2 **Notation, Definitions, and Preliminaries**  
2.1 Space of Data Instances .................................... 32  
2.2 Distances to Ill-Posedness .................................. 35  
2.3 Theorems of the Alternative ................................. 39  
2.4 Further Notation and Results ................................ 42  

3 **Characterization of Ill-Posed Data Instances**  
3.1 Overview ................................................ 46  
3.2 Characterizations of Ill-Posedness ............................. 51  
3.3 Further Properties ........................................ 68  

4 **Analytic Center Problems**  
4.1 Overview ................................................ 76  
4.2 Bounds on Analytic Centers ................................ 80
4.3 Changes in Optimal Solutions Under Data Perturbations ............ 89

5 Central Trajectory Problems for Linear Programming .................. 102
  5.1 Overview ................................................. 102
  5.2 Bounds on Solutions Along the Central Trajectory ................ 104
  5.3 Changes in Optimal Solutions Under Data Perturbations .......... 114

6 Conditioning of the Optimal Faces of a Linear Program ................. 133
  6.1 Overview ................................................. 133
  6.2 Properties of the Distance to Degeneracy ......................... 138
  6.3 Complexity Results ........................................ 144
  6.4 Supplementary Results ..................................... 151

7 Centering and Central Trajectory Problems for a Conic Linear System . 155
  7.1 Overview ................................................. 155
  7.2 The Central Trajectory of a Conic Linear System ................ 161
  7.3 Conic Analytic Center Problems ................................ 165
  7.4 Bounds on Solutions Along the Central Trajectory ................ 171
  7.5 Supplementary Results ..................................... 176

8 Semi-Definite Programming ........................................... 179
  8.1 Overview ................................................. 179
  8.2 Semi-Definite Analytic Center Problems .......................... 185
  8.3 Bounds on Solutions Along the Central Trajectory ................ 191
  8.4 Further Properties ......................................... 194

9 Conclusions and Future Work ........................................ 201

Bibliography ....................................................... 205
Chapter 1

Introduction

1.1 Overview

Ever since the initial development of mathematical programming during the late
1940s, understanding the effect on optimal solutions of changes in the input data
associated with a given problem has been of great concern in industrial applications,
as well as in economics. As early as 1955, in their classical paper [HMS75] Holt,
Modigliani, and Simon, in commenting on the stability of a quadratic programming
model for production scheduling in a factory, wrote: “Since estimates of the costs
parameters are subject to many sources of error, it is reassuring that the factory
(optimal) performance proves not to be critically dependent on the accuracy of the
cost function. Even if substantial errors are made in estimating the parameters of
the cost function, the factory (optimal) performance measured in cost terms will not
suffer seriously.”

Data perturbations occur frequently in optimization problems. For instance, part
of the data might be determined directly from physical measurements, and hence is
subject to observational errors. In other cases we cannot represent exactly the data
using a digital computer, even if the data is composed of rational numbers. Moreover, computing a solution of an optimization problem might involve calculations that could create rounding errors. Therefore, it is important to have an idea of how perturbations on the input data associated with a given optimization problem will affect the optimal solutions to such problem as well as the complexity of the algorithms employed to find such solutions.

In particular, when it comes to linear programming, this concern about stability of optimal solutions originated the well-known theory of sensitivity or parametric analysis, where the goal is to determine how data perturbations will affect the optimality of the basis corresponding to an optimal extreme point (see [BJS90, BHM77, BT97, Dan63, Mur83]). Since the appearance of the first paper on sensitivity analysis by Manne [Man53] in 1953, a large number of papers have been written on these subjects (for a survey on the history of sensitivity analysis see [Gal75, Gal84]). Nevertheless, it was not until recent years, and probably motivated by the development of modern interior-point methods for optimization, that researchers were able to provide a new theory of perturbation analysis in linear programming encompassing important issues such as continuous complexity and ill-conditioning of linear programs.

In the last five years, Renegar [Ren94, Ren95b, Ren95c, Ren96] introduced a new approach to the study of perturbations in mathematical programming. Instead of narrowing the analysis on the effect of data perturbations on some elements of a given data input (as occurs in classical sensitivity analysis), Renegar's approach is to allow simultaneous changes in all the elements of the data input until the feasibility region of the perturbed problem "breaks down" by becoming either empty or by having an unbounded objective. He calls the minimum size of a data perturbation that produces such a collapse the distance to ill-posedness. By using this new no-
1.1 Overview

tion, Renegar extended the concepts of *ill-posedness* and *condition numbers* from the context of matrix analysis in classical linear algebra to the context of mathematical programming. Furthermore, Renegar and several others [Fre94, FV95, NF97, Ver97] proved that this new perturbation theory is an adequate instrument for capturing the geometric and complexity properties of modern interior-point methods.

In this thesis we are concerned with two applications of this theory: (i) continuous complexity, that is, determining new complexity bounds for algorithms applied to mathematical programs with real-number data (not necessarily rational); and (ii) interior-point methods, in particular, determining the relationship between interior-point methods and the conditioning of a problem. We embrace Renegar's approach and extend the notions of ill-posedness and condition numbers to study and derive properties of key geometric objects such as analytic centers and central trajectories; objects that are intrinsically important in the analysis of complexity of interior-point methods and convex programming.

In the remainder of this chapter, we present an informal discussion of the main ideas that motivate this research. In particular, in Section 1.2, we develop some intuition concerning perturbation theory in convex programming, distances to ill-posedness, condition numbers, and interior-point methods. In Section 1.3, we formulate the main issues addressed and summarize our more relevant findings in this thesis. In Section 1.4, we present an outline of this thesis and describe the most important results established in each chapter. Finally, in Section 1.5 we present a literature review.
1.2 Motivation to Ill-Posedness of Convex Programs

We first provide some motivation and intuition underlying the concepts of distance to ill-posedness and condition number. Consider the following linear program written in standard form:

$$\min \left\{ c^T x : Ax = b, x \geq 0 \right\},$$

where $A$ is a matrix in $\mathbb{R}^{m \times n}$, $b$ is a vector in $\mathbb{R}^m$, and $c$ is a vector in $\mathbb{R}^n$. By considering all possible arrays of the form $(A, b, c)$ for fixed $m$ and $n$, we obtain a vector space, denoted by $\mathcal{D}$, whose elements are data instances associated with this linear program. Given a data instance $d = (A, b, c)$ in the data space $\mathcal{D}$, we are interested in studying the effects on $d$ of perturbing its elements, that is, determining what properties held by $d$ will remain intact for the data instance $d + \Delta d$, where $\Delta d = (\Delta A, \Delta b, \Delta c)$ denotes a data perturbation.

For example, assume that $m = 1$, $n = 2$, and consider the following data instances in the corresponding data space $\mathcal{D}$:

$$d_1 = \left( \begin{bmatrix} 1 & 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right),$$
$$d_2 = \left( \begin{bmatrix} 0 & 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right),$$
$$d_3 = \left( \begin{bmatrix} 1 & 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right).$$

Observe that the data instances $d_1$ and $d_3$ are both primal and dual feasible, and so they have optimal solutions. On the other hand, $d_2$ is neither primal nor dual feasible, and so it does not have an optimal solution. We are interested in determining whether or not the property of having or lacking an optimal solution of the data instances $d_1$,
$d_2$ and $d_3$ will remain intact after we apply a small perturbation to each of these data instances.

It is not difficult to see that after applying to $d_1$ any small perturbation of the form

$$\Delta \tilde{d} = \left( \begin{bmatrix} \epsilon_1 & \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix} \right),$$

where $|\epsilon_k| < 1$ for $k = 1, \ldots, 5$, we obtain a perturbed data instance $d_1 + \Delta \tilde{d}$ that is also both primal and dual feasible, and so the perturbed data instance $d_1 + \Delta \tilde{d}$ has an optimal solution. Nevertheless, if we apply a similar data perturbation to $d_2$, then we obtain a perturbed data instance $d_2 + \Delta \tilde{d}$ that is either primal or dual infeasible, and so the perturbed data instance $d_2 + \Delta \tilde{d}$ does not have an optimal solution. Hence, in both cases we have well-posed or well-conditioned programs in the sense that the property of existence or not of optimal solutions is not very sensitive to small changes in the data. In a sense, after applying the data perturbation $\Delta \tilde{d}$ to any of these data instances, we can tell without actually solving the corresponding linear program whether or not the resulting perturbed data instance will have an optimal solution.

On the other hand, no matter what kind of data perturbation $\Delta d$ we apply to $d_3$, the resulting perturbed data instance $d_3 + \Delta d$ might or might not have an optimal solution. In this inimical situation, where the existence of optimal solutions of the perturbed data instance $d_3 + \Delta d$ or its lack thereof cannot be established without solving the corresponding linear program for any perturbation $\Delta d$, we say that the data instance $d_3$ is ill-posed or ill-conditioned.

Returning to the general case, given a data instance $d$ in a fixed subset $F$ of data instances, the previous discussion naturally raises the question of what is the
1.2 Motivation to Ill-Posedness of Convex Programs

smallest change necessary to produce a perturbed data instance not in $\mathcal{F}$. In other words, if we endow the vector data space $\mathcal{D}$ with a norm $\|\cdot\|$, then we would like to find the smallest scalar $\rho(d)$ such that there is no data perturbation $\Delta d$ satisfying $d + \Delta d \notin \mathcal{F}$ and $\|\Delta d\| \leq \rho(d)$. Analogously, if $d \notin \mathcal{F}$, we would like to find the smallest scalar $\rho(d)$ such that there is no data perturbation $\Delta d$ satisfying $d + \Delta d \in \mathcal{F}$ and $\|\Delta d\| \leq \rho(d)$. When $\mathcal{F}$ is defined as the set of data instances that have an optimal solution, the scalar $\rho(d)$ corresponds to the distance to ill-posedness introduced by Renegar [Ren94, Ren95b, Ren95c, Ren96] for a linear program in standard form. In particular, if $d$ is a data instance on the boundary of $\mathcal{F}$, then $\rho(d) = 0$, and hence $d$ is an ill-posed data instance according to our previous definition. Conversely, given an ill-posed data instance $d$, we must have $\rho(d) = 0$ since otherwise all perturbed data instance $d + \Delta d$ with $\|\Delta d\| \leq \rho(d)$ will be in the same set to which $d$ belongs. Therefore, a data instance $d$ is ill-posed if and only if $d$ is on the boundary of $\mathcal{F}$, or equivalently, if and only if $\rho(d) = 0$.

For a given data instance $d$ with $\rho(d) > 0$, the ratio $C(d) := \|d\|/\rho(d)$ is called the condition number associated with the data instance $d$. The number $C(d)$ is a scale-invariant reciprocal of $\rho(d)$ that goes to infinity as the data instance $d$ approaches the boundary of $\mathcal{F}$, that is, as $d$ approaches ill-posedness. By using this condition number, many researchers (see [Fil94a, Fil94b, Fre94, FV95, Ren94, Ren95b, Ren95c, Ren96, Ver92, Ver96, Ver97] among many others) have proven several properties concerning optimal solutions to linear programs, sensitivity to data perturbations, and complexity of interior-point methods for solving such programs.

For instance, Renegar [Ren95c] proved that given a data instance $d = (A, b, c)$ of a linear programming in standard form, there exists an optimal primal solution $x$ such
1.2 Motivation to Ill-Posedness of Convex Programs

that

\[ \|x\|_1 \leq \mathcal{C}(d)^2, \]

so that if the data instance \( d \) is far from being ill-posed, then there is a small-sized optimal primal solution to the linear program associated with \( d \). In addition, Renegar [Ren95c] proved that if \( z(d) \) denotes the optimal objective value associated with the data instance \( d \), then

\[ |z(d) - z(d + \Delta d)| = O(\|\Delta d\|\mathcal{C}(d)^3), \]

where \( \|\Delta d\| < \rho(d) \), thereby establishing a bound on the sensitivity of the optimal objective to changes in the input data. Finally, Vera again [Ver97] proved that there exists an interior-point method that starting with an initial interior solution \( x_0 \) and initial duality gap of \( \epsilon_0 \) computes an \( \epsilon \)-optimal solution to the linear program

\[ \min \{c^T x : Ax \leq b\} \]

after at most

\[ O\left(\sqrt{m} \log \left(\frac{\epsilon_0}{\epsilon}\right)\right) \]

iterations, where each iteration involves a Newton step, and the algorithm requires no more than

\[ O\left(\log \left(\frac{nK\epsilon_0}{\epsilon} \mathcal{C}(d)\right)\right) \]

digits of precision. In this expression, \( K \) is a known upper bound on the optimal objective value of this linear program.

We conclude this section by remarking that we can analogously extend the notions of distance to ill-posedness and the corresponding condition number for any other convex optimization problem that takes its data input from the set \( \tilde{D} \). In other words, for a given optimization problem, we can define a set \( \mathcal{F} \) of data instances for
which this problem attains its optimal solution. By using this set $\mathcal{F}$, it is clear that we obtain a similar distance to ill-posedness in terms of the set $\mathcal{F}$. In particular, we study this type of extensions for optimization problems such as analytic center problems (Chapter 4), central trajectory problems (Chapter 5), general conic linear programs (Chapter 7), and semi-definite programming (Chapter 8).

1.3 Main Issues Addressed in this Research

From the discussion presented in the previous section, it is clear that condition numbers are relevant to the understanding of optimality and sensitivity of mathematical programs to data perturbations as well as the understanding of the behavior of interior-point methods used to solve such optimization problems. Because of the recent developments of this theory, there are many open issues concerning condition numbers in convex optimization, far more than we can address in this thesis. For this reason we concentrate our study on the following issues:

1. Characterizing the set of ill-posed data instances associated with a linear program in standard form. That is, provide necessary and sufficient conditions for a data instance to be ill-posed.

2. Determining how the optimal solution to an analytic center problem depends on the conditioning of the problem. Also, determining in terms of the conditioning of a problem how optimal solutions and optimal objective values of analytic center problems change when the input data is changed.

3. Determining how the optimal solution to a central trajectory problem depends on the conditioning of the problem. Determining in terms of the conditioning of a problem how optimal solutions and optimal objective values of central trajectory problems change when the input data is changed. Also, determining
how solutions along the central trajectory of a linear program depend on the conditioning of the program.

4. Determining complexity bounds for interior-point methods depending on the conditioning of the optimal face of a linear program in standard form.

5. Extending our study of the central trajectory of a linear program and the conditioning of a linear program to more general optimization problems such as semi-definite programming.

6. Extending our study of the central trajectory of a linear program and the conditioning of a linear program to more general optimization problems such as conic linear programs with a well-behaved (that is, self-concordant) barrier function.

We further detail these issues in the following paragraphs. With respect to the first issue, that is, characterizing the ill-posed data instances of a linear program in standard form, from the discussion in the previous section, this essentially comprises a description of the elements contained in the boundary $\partial \mathcal{F}$ of the set $\mathcal{F}$ of data instances that have optimal solutions. In turn, this entails describing the elements of four sets: the interior and closure of $\mathcal{F}$, and the interior and closure of the complement of $\mathcal{F}$. Although a characterization of the interior of $\mathcal{F}$ has been known for a long time (see [Ash81, Rob77]), no characterizations of the other three sets are available in the literature.

For example, consider the case when $m = n = 1$, that is, when the data space is simply $\mathcal{D} = \mathbb{R}^3$. In this case, we have that

$$\text{interior}(\mathcal{F}) = \{(a, b, c) : ab > 0\},$$

$$\text{closure}(\mathcal{F}) = \{(a, b, c) : ab \geq 0\},$$

$$\text{interior}(\mathcal{F}^C) = \{(a, b, c) : ab < 0\},$$
\[ \text{closure}(\mathcal{F}^C) = \{(a, b, c) : ab \leq 0\}, \]
\[ \partial \mathcal{F} = \{(a, b, c) : ab = 0\}, \]

where \( \mathcal{F}^C \) denotes the complement of the data instance set \( \mathcal{F} \). Our goal is to find similar characterizations for more general cases.

The second issue from the previous list has to do with the following two problems:

\[ AE(d) : \min \{ p(x) : Ax = b, x > 0 \}, \]
\[ AI(d) : \min \{ p(s) : A^Ty + s = c, s > 0 \}, \]

where \( d = (A, b, c) \) and for \( u > 0 \) in \( \mathbb{R}^n \), \( p(u) = -\sum_{j=1}^{n} \ln u_j \). The program \( AE(d) \) is called the analytic center problem in equality form, and the program \( AI(d) \) is called the analytic center problem in inequality form. Under certain conditions that are discussed in Chapter 4, these problems have unique optimal solutions, which we denote by \( x(d) \) and \( (y(d), s(d)) \), respectively. As discussed in the previous section, we can associate with these problems analogous condition numbers to the condition number for a linear program in standard form. We denote these condition numbers by \( C_E(d) \) and \( C_I(d) \), respectively. We would like to establish upper and lower bounds on \( \|x(d)\| \), and \( \|(y(d), s(d))\| \), in terms of the condition numbers \( C_E(d) \) and \( C_I(d) \), respectively, and so understand the relation of the optimal solutions to these two problems to the conditioning of the data \( d \). Furthermore, if possible, we would like to bound the change in optimal solutions \( \|x(d) - x(d + \Delta d)\| \) and \( \|(y(d), s(d)) - (y(d + \Delta d), s(d + \Delta d))\| \) in terms of the change \( \Delta d \) in the data input and the conditioning of the corresponding analytic center problems.
1.3 Main Issues Addressed in this Research

The third issue in the list has to do with the following optimization problems:

\[ P_\mu(d) : \min \{ c^T x + \mu p(x) : Ax = b, x > 0 \}, \]
\[ D_\mu(d) : \max \{ b^T y - \mu p(s) : A^T y + s = c, s > 0 \}, \]

where \( d = (A, b, c) \) and \( \mu \) is a positive scalar considered independent of the data instance \( d \). The programs \( P_\mu(d) \) and \( D_\mu(d) \) are a pair of Lagrangian dual programs. The program \( P_\mu(d) \) is called the central trajectory problem and is the key geometric notion behind path-following interior-point methods for solving linear programming. For this reason, the study of the central trajectory in terms of the conditioning of a data instance is very relevant to the study of the complexity and convergence of path-following interior-point methods for linear programming. As with analytic center problems, we are interested in finding lower and upper bounds on the size of the optimal solution \( x_\mu(d) \) to this problem, whenever it exists. Furthermore, if possible, we would like to bound the change in optimal solutions \( \| x_\mu(d) - x_\mu(d + \Delta d) \| \) in terms of the change \( \Delta d \) in the data input and the conditioning of the corresponding analytic center problems, and also to bound in terms of the conditioning of the problem the change in optimal solutions \( \| x_\mu(c) - x_{\bar{\mu}}(d) \| \) along the central trajectory associated with a data instance \( d \), where \( \mu, \bar{\mu} > 0 \).

The fourth issue concerns complexity bounds for interior-point methods depending on the conditioning of the optimal face of a linear program in standard form. Complexity bounds relating the general condition number \( C(d) \) to the performance of interior-point methods have already been obtained by Freund [Fre94] and Vera [Ver97] for algorithms that compute \( \epsilon \)-optimal solutions. Such complexity bounds are polynomial in terms of the dimensions \( m \) and \( n \) of the problem, \( \log(1/\epsilon), \log(C(d)) \), and the size of the data \( \|d\| \). Since the central trajectory associated with a data instance having multiple optima is very sensitive to data perturbations, it is natural to look
1.4 Thesis Summary

for new complexity bounds that solely depend on the conditioning of the optimal face.

The last two issues comprise extensions of the results on the central trajectory to more general problems. In particular, once we define a meaningful measure of ill-posedness in more general optimization contexts, we would like to extend the results concerning properties that relate the central trajectory of a semi-definite programming and, more generally, of a conic linear program, to the conditioning of such problems.

1.4 Thesis Summary

Chapter 2 contains the notation and description of the four optimization problems of interest in this research: linear programs in standard form, analytic center problems in equality and inequality form, and the central trajectory problem. The notation introduced therein is common for Chapters 3 through 6. (Chapters 7 and 8 can be read independently of the previous chapters.) In Section 2.2, we formally define the distances to ill-posedness and condition numbers corresponding to each of these optimization problems. Since many of the results derived in this research rely on theorems of the alternative, in Section 2.3 we list Farkas’ Lemma and many other equivalent versions that are used throughout this thesis. Finally, in Section 2.4, we state further results from classical matrix analysis theory concerning ill-conditioning of linear systems of equations.

In Chapter 3, we fully resolve the issue of characterizing the sets of ill-posed data instances of a linear program in standard form. By a characterization we mean the following: given a data instance $d = (A, b, c)$ in the space of data instances, for each optimization problem we provide a union of systems of linear equations and inequali-
ties in terms of the data \((A, b, c)\), such that \(d\) is an ill-posed data instance with respect to such optimization problem if and only if the corresponding union of linear systems has a solution. Furthermore, we fully characterize the sets of ill-posed data instances with respect to the analytic center problems in equality and inequality forms (see Chapter 2). All these characterizations are summarized in Theorem 3.1 for linear programs in standard form, in Theorem 3.2 for analytic center problems in equality form, and in Theorem 3.3 for analytic center problems in inequality form. In addition, these theorems also state characterizations of related data sets such as interiors and closures of the sets of feasible data instances (and their complements) corresponding to each of the three optimization problems. Since the set of ill-posed data instances for a linear program is the same set as the set of ill-posed data instances for the central trajectory problem (see Corollary 3.1), then there is no need to provide a separate characterization for the set of ill-posed data instances of the central trajectory problem.

Finally, in Section 3.3, we review the original distances to ill-posedness defined by Renegar in [Ren94, Ren95a, Ren95b, Ren95c, Ren96] and relate them to the extended versions used in this thesis.

In Chapter 4 we study the analytic center problems in equality and inequality form in terms of the conditions numbers \(C_E(\cdot)\) and \(C_I(\cdot)\), respectively (see previous section, or for further details see Chapter 2). In Section 4.2, we present some bounds on the norms of feasible solutions of the two analytic center problems \(AE(d)\) and \(AI(d)\). The most important results in this section are Lemma 4.1 and Lemma 4.2, which establish the boundeness of the feasible regions of \(AE(d)\) and \(AI(d)\), respectively, in terms of the conditions numbers \(C_E(d)\) and \(C_I(d)\), and \(\|d\|\). Proposition 4.3 and Proposition 4.4 state bounds on the optimal solutions to \(AED(d)\) and \(AID(d)\),
respectively, in terms of \( n \), \( C_E(d) \) and \( C_I(d) \), where \( AED(d) \) and \( AID(d) \) are the Lagrangian duals of \( AE(d) \) and \( AI(d) \), respectively. By using Lemmas 4.1 and 4.2 and Propositions 4.3 and 4.4, we prove in Lemma 4.3 and Lemma 4.4 lower bounds on the optimal solutions to \( AE(d) \) and \( AI(d) \), respectively, also in terms of \( n \), \( C_E(d) \) and \( C_I(d) \).

Furthermore, in Section 4.3 we use the bounds from Section 4.2 to study the effect of data perturbations on optimal solutions to the analytic centers \( AE(d) \) and \( AI(d) \), respectively. Our objective is to demonstrate upper bounds on the size of changes of optimal solutions to \( AE(d) \) and \( AI(d) \), respectively, in terms of polynomial expressions of the condition numbers \( C_E(d) \) and \( C_I(d) \); the size \( ||d|| \) of the data instance \( d \); the size \( ||\Delta d|| \) of the data perturbation \( \Delta d \); the dimensions \( m \) and \( n \); and some small constants. Among the most important results of Section 4.3 we have Theorem 4.1 and Theorem 4.2. Roughly speaking, Theorem 4.1 states that for a data perturbation \( \Delta d \) sufficiently small, the change in the optimal solutions \( x(d) \) and \( x(d + \Delta d) \) to \( AE(d) \) and \( AE(d + \Delta d) \), respectively, is of the order

\[
||x(d + \Delta d) - x(d)|| = O\left( mn^2 \frac{||\Delta d||}{||d||} C_E(d)^7 \right).
\]

It is important to notice the linear dependence on \( ||\Delta d|| \).

Analogously, Theorem 4.2 states that for a data perturbation \( \Delta d \) sufficiently small, the change in the optimal solutions \( (y(d), s(d)) \) and \( (y(d + \Delta d), s(d + \Delta d)) \) to \( AI(d) \) and \( AI(d + \Delta d) \), respectively, is of the order

\[
\begin{align*}
||y(d + \Delta d) - y(d)|| & = O\left( mn^2 \frac{||\Delta d||}{||d||} C_I(d)^7 \right), \\
||s(d + \Delta d) - s(d)|| & = O\left( mn^2 ||\Delta d|| C_I(d)^7 \right).
\end{align*}
\]
1.4 Thesis Summary

Again, it is important to notice the linear dependence on $\|\Delta d\|$.

Finally, in Theorem 4.3 and Theorem 4.4, we state upper bounds on the change of optimal objective values of the programs $AE(\cdot)$ and $AI(\cdot)$, respectively, under data perturbations. These bounds are linear in the size $\|\Delta d\|$ of the perturbation vector $\Delta d$, and also depend on the condition numbers $C_E(d)$ and $C_I(d)$; the dimension $n$; and some small constants.

In Chapter 5, we explore and demonstrate properties of solutions to the central trajectory problem that are inherently related to the condition number $C(d)$ of the data instance $d = (A, b, c)$. As discussed previously, in the context of the central trajectory problem, $\rho(d)$ essentially is the minimum change $\Delta d = (\Delta A, \Delta b, \Delta c)$ in the data $d = (A, b, c)$ necessary to create a data instance $d + \Delta d$ that is an infeasible instance of $P_\mu(\cdot)$ or $D_\mu(d)$. The main results in the chapter are stated in Sections 5.2 and 5.3.

In Section 5.2 we present upper and lower bounds on sizes of optimal solutions $x(\mu)$ and $(y(\mu), s(\mu))$ to the barrier problems $P_\mu(d)$ and $D_\mu(d)$ in terms of the conditioning of the data instance $d$. Theorems 5.1 and 5.2 state bounds on such solutions that are linear in $\mu$, where the constants in the bounds are polynomial functions of the condition number $C(d)$, the distance to ill-posedness $\rho(d)$, the dimension $n$, the norm of the data $\|d\|$, or their inverses. These theorems show in particular that as $\mu$ goes to zero, that $x_j(\mu)$ grows at least linearly in $\mu$; and as $\mu$ goes to $\infty$, $x_j(\mu)$ grows at most linearly in $\mu$. Moreover, in Theorem 5.3, we also show that when the feasible region of $P_\mu(d)$ is unbounded, then certain coordinates of $x(\mu)$ grow exactly linearly in $\mu$ as $\mu \to \infty$, all at rates bounded by polynomial functions of the condition number $C(d)$, the distance to ill-posedness $\rho(d)$, the dimension $n$, the norm of the data $\|d\|$.,
or their inverses.

In Section 5.3, we study the sensitivity of the optimal solutions to $P_\mu(d)$ and $D_\mu(d)$ as either the data $d = (A, b, c)$ changes or the barrier parameter $\mu$ changes. Theorems 5.4 and 5.7 state upper bounds on the sizes of the changes on optimal solutions as well as in the optimal objective values as the data $d = (A, b, c)$ is changed. Theorems 5.6 and 5.8 state upper bounds on the sizes of changes in optimal solutions and optimal objective values as the barrier parameter $\mu$ is changed. Along the way, we prove Theorem 5.5, which states bounds on the norm of the matrix $(AX^2(\mu)A^T)^{-1}$. This matrix is the main computational matrix in interior-point central trajectory methods. All of the bounds in this section are polynomial functions of the condition number $\mathcal{C}(d)$, the distance to ill-posedness $\rho(d)$, the dimension $n$, the norm of the data $\|d\|$, or their inverses.

In Chapter 6, the objective is to study the primal and dual optimal faces associated with a linear program in terms of a measure of the conditioning of these faces. For a given data instance $d = (A, b, c)$, we denoted this measure by $\eta(d)$. The measure $\eta(d)$ is called the distance to degeneracy and represents the minimum change $\Delta d = (\Delta A, \Delta b, \Delta c)$ in the data $d$ necessary to create a perturbed data instance $d + \Delta d$ whose optimal faces have a different combinatorial structure from the combinatorial structure of the optimal faces associated with $d$.

In Section 6.2 we state several properties of the distance to degeneracy $\eta(\cdot)$. These properties show relations with other measures such as the distance to ill-posedness $\rho(\cdot)$ (see Proposition 6.1), the distance to multiple optima $\zeta(\cdot)$ defined in Proposition 6.2, and the classical condition number $\kappa(\cdot)$ applied to a certain matrix associated with the optimal faces of a linear program (see Lemma 6.3). The most important results
in this section are Lemma 6.1 and Lemma 6.2. These lemmas establish that the distance to degeneracy $\eta(d)$ is positive if and only if $P(d)$ has unique primal and dual optimal solutions, or equivalently, if and only if $P(d)$ does not have primal nor dual degenerate optimal solutions.

In Section 6.3, we provide an algorithm that finds the exact solution of a linear program when the linear program exhibits unique primal and dual solutions. In particular, we prove Theorem 6.1, which states the following result: Let $d = (A, b, c)$ be a data instance in $\mathcal{D}$ such that $d \in \mathcal{F}$ and $\eta(d) > 0$. Then there exists a path-following interior-point algorithm that starting at $(x_0, y_0, s_0)$ sufficiently close to the central trajectory and with initial duality gap $\epsilon_0$ computes the optimal solution of $P(d)$ in no more than

$$O\left(\sqrt{n} \log \left(\frac{\epsilon_0 \|d\|}{\eta(d)^2}\right)^2\right)$$

iterations. Each iteration involves the solution of two $k \times k$ system of equations, with $k \leq n$.

Finally, in Section 6.4 we present further properties of the primal optimal face of a linear program in relation to the central trajectory associated with such program. The main result of this section is stated in Lemma 6.6. This lemma presents an upper bound on the error between the optimal solution to the central trajectory problem $P_\mu(d)$ and the analytic center of the primal optimal face for $\mu$ sufficiently small.

In Chapter 7 we extend some of the results concerning analytic center problems (from Chapter 4) and central trajectory problems (from Chapter 5) to conic linear systems with a self-concordant barrier function. In particular, by introducing a self-concordant logarithmically homogeneous barrier for the conic linear system, we prove bounds for analytic center problems and solutions along the central trajectory in this
1.5 Literature Review

generalized context.

In Section 7.2, we introduce the central trajectory problem associated with a conic linear system as well as the corresponding Karush-Kuhn-Tucker optimality conditions for this problem. Furthermore, we introduce Renegar's condition number in this context. Similarly, in Section 7.3, we introduce conic analytic center problems in equality and inequality form. In addition, we prove Lemma 7.1 and Lemma 7.2, which state bounds on the feasible solutions to these two conic analytic center problems in terms of their corresponding condition numbers, thereby generalizing Lemma 4.1 and Lemma 4.2, respectively, from Chapter 4. In Section 7.4, we generalize Theorem 5.1 and Theorem 5.2 from Chapter 5 to conic linear systems. In fact, Theorem 7.2 and Theorem 7.3, respectively, state analogous bounds to their counterparts from Chapter 5 regarding solutions along the central trajectory of a conic linear system.

Finally, in Chapter 8, we specialize the results from Chapter 7 to semi-definite programming. The most relevant results presented in this chapter are easy consequences of the results concerning conic linear systems. There are, however, a number of results that follow from specific properties of semi-definite programs. These results are not direct consequences of the theory developed in Chapter 7; rather, these results generalize some of the results for linear programs in standard form.

1.5 Literature Review

We may trace back the origin of Renegar's perturbation theory in mathematical programming to 1953, when Manne [Man53] introduced the concept of sensitivity or parametric analysis in linear programming. The main idea in sensitivity analysis is to determine how data perturbations might affect the optimality of the basis cor-
1.5 Literature Review

responding to an optimal extreme point. After this publication, many researchers have written papers on similar subjects exploring different kinds of data perturbations (see for instance [Gal75, Gal84, Mil56, Saa59, Fre85]). In particular, Wendell and Ravi [Wen84, Wen85, RW85, RW89] defined a similar measure to the distance to degeneracy studied in this thesis (see Chapter 6). They named their research the \textit{tolerance approach} to sensitivity analysis and their goal was to compute a maximum tolerance percentage within which the input data of a linear program may vary while still retaining the same set of basic variables in an optimal solution. Nevertheless, none of these approaches uses the distance to ill-posedness as defined by Renegar.

The study of perturbation theory and information complexity for convex programs in terms of the distance to ill-posedness $\rho(d)$ and the condition number $\mathcal{C}(d)$ of a given data instance $d$ was introduced by Renegar in [Ren94], who studied perturbations in the very general setting:

$$RLP: \quad z = \sup\{c^*x : Ax \leq b, x \geq 0, x \in \mathcal{X}\},$$

where $\mathcal{X}$ and $\mathcal{Y}$ denote real normed vector spaces, $A : \mathcal{X} \to \mathcal{Y}$ is a continuous linear operator, $c^* : \mathcal{X} \to \mathbb{R}$ is a continuous linear functional, and the inequalities $Ax \leq b$ and $x \geq 0$ are induced by any closed convex cones (linear or nonlinear) containing the origin in $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Previous to this paper of Renegar, many papers were written on perturbations of linear programs and systems of linear inequalities, but not in terms of the distance to ill-posedness (see [Man81, Rob75, Rob76, Rob77, Sch96]). There have also been many other different approaches to the study of stability of mathematical programs. For instance Karmanov [Kar77] uses the concept of \textit{well-posed problem in the weak sense}, that is, if $\Gamma$ denotes the set of solutions associated with a mathematical program
of the form $\min \{f(x) : x \in \mathcal{X}\}$, and for $\epsilon > 0$, $\Gamma_\epsilon$ denotes the set of solutions of an approximate mathematical program of the form $\min \{f_\epsilon(x) : x \in \mathcal{X}\}$, then the problem is well-posed in the weak sense with respect to the family of sets $\{\Gamma_\epsilon : \epsilon > 0\}$ if

$$\lim_{\epsilon \to 0} \sup_{\nu_e \in \Gamma, \nu \in \Gamma} \inf \|y - y_\epsilon\| = 0.$$ 

This definition of well-posedness has motivated a great number of publications; for a complete survey see [Lev94].

Going back to Renegar's work, in [Ren95b] and [Ren95c] Renegar introduced the concept of a fully efficient algorithm; and provided a fully-efficient algorithm that given any data instance $d$ answers whether the program $RLP$ associated with $d$ is consistent or not.

Vera in [Ver92] developed a fully-efficient algorithm for a certain form of linear programming that is a special case of $RLP$ in which the spaces are finite-dimensional. the linear inequalities are induced by the non-negative orthant, and non-negativity constraints $x \geq 0$ do not appear; that is, the problem $RLP$ is $\min \{c^T x : Ax \leq b, x \in \mathbb{R}^n\}$. In [Ver96], Vera established bounds similar to those in [Ren94] for norms of optimal primal and dual solutions and optimal objective function values. He then used these bounds to develop an algorithm for finding approximate optimal solutions of the original instance. In [Ver97] he provided a measure of the precision of a logarithmic barrier algorithm based upon the distance to ill-posedness of the instance. To do this, he followed the same arguments as Den Hertog, Roos, and Terlaky [DHRT92], making the appropriate changes when necessary to express their results in terms of the distance to ill-posedness.

Filipowski [Fil94a, Fil94b, Fil95] expanded upon Vera's results under the assump-
ation that it is known beforehand that the primal data instance is feasible. In addition, she developed several fully-efficient algorithms that approximate optimal solutions to the original instance under this assumption.

Todd and Ye [TY94] use a particular case of gauge duality theory [Fre87] to implement a criterion to determine whether a linear program or its dual is infeasible within the context of an infeasible-interior-point method depending on the well-posedness of certain inequality system.

Freund and Vera [FV95] addressed the issue of deciding feasibility of \( RLP \). The problem that they studied is defined as finding \( x \) that solves \( b - Ax \in C_Y \) and \( x \in C_X \), where \( C_X \) and \( C_Y \) are closed convex cones in the linear vector spaces \( X \) and \( Y \), respectively. They developed optimization problems that allow one to compute exactly or at least estimate the distance to ill-posedness. They also showed additional results relating the distance to ill-posedness to the existence of certain inscribed and circumscribed balls for the feasible region, which has implications for Khachiyan’s ellipsoid algorithm [Kha79].

Epelman and Freund [EF97] developed an algorithm for resolving a conic linear system. The algorithm resolves the system in that it either finds an \( \epsilon \)-solution for a pre-specified tolerance \( \epsilon > 0 \) or demonstrates that the system has no solution. The algorithm is based on a generalization of von Neumann’s algorithm for linear inequalities and runs in a number of iterations bounded by \( O(C(d)^2 \ln(C(d)) \ln(\|d\|/\epsilon)) \).
Chapter 2

Notation, Definitions, and Preliminaries

2.1 Space of Data Instances

We denote by \( \mathcal{D} \) the space of data instances, that is, \( \mathcal{D} = \{(A, b, c) : A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n\} \), where \( m \leq n \). Let \( d = (A, b, c) \) be a data instance in \( \mathcal{D} \), we are interested in studying the following mathematical programs:

\[
\begin{align*}
P(d) & : \min \left\{ c^T x : Ax = b, x \geq 0 \right\}, \\
AE(d) & : \min \left\{ p(x) : Ax = b, x > 0 \right\}, \\
AI(d) & : \min \left\{ p(s) : A^T y + s = c, s \geq 0 \right\}, \\
P_\mu(d) & : \min \left\{ c^T x + \mu p(x) : Ax = b, x > 0 \right\},
\end{align*}
\]

where \( p(u) = -\sum_{j=1}^n \ln(u_j) \) is the logarithmic barrier function, and the parameter \( \mu \) is a positive scalar considered independent of the data instance \( d = (A, b, c) \in \mathcal{D} \). The study of these programs and their conditioning is the main subject of this thesis. These programs arise in different optimization contexts: the program \( P(d) \) corre-
2.1 Space of Data Instances

Corresponds to a linear program in standard form, the programs $AE(d)$ and $AI(d)$ correspond to analytic center problems in equality and inequality form, respectively, and $P_{\mu}(d)$ corresponds to the central trajectory problem associated with a linear program in standard form.

It is also of great interest in this research to study the corresponding (Lagrangian) dual formulations of these problems. More precisely, we are also interested in the following mathematical programs, which are the duals of $P(d)$, $AE(d)$, $AI(d)$, and $P_{\mu}(d)$, respectively:

\[
D(d) : \max \left\{ b^T y : A^T y + s = c, s \geq 0 \right\},
\]

\[
AED(d) : \max \left\{ b^T u - p(t) : A^T u + t = 0, t > 0 \right\},
\]

\[
AID(d) : \max \left\{ -c^T v - p(v) : Av = 0, v > 0 \right\},
\]

\[
D_{\mu}(d) : \max \left\{ b^T y - \mu p(s) : A^T y + s = c, s \geq 0 \right\}.
\]

Observe that $AE(d)$ and $AED(d)$ are independent of the vector $c$. It is well-known that $AE(d)$ has a unique solution when its feasible region is bounded and non-empty. Similarly, $AI(d)$ and $AID(d)$ are independent of the vector $b$, and $AI(d)$ has a unique solution when its feasible region is bounded and non-empty.

In order to define a norm over the data set $\mathcal{D}$, we introduce the following norms on the space $\mathbb{R}^k$:

\[
\|v\|_{\alpha} = \left( \sum_{i=1}^{k} |v_i|^\alpha \right)^{1/\alpha},
\]

\[
\|v\|_{\infty} = \max_{1 \leq i \leq k} |v_i|.
\]
for each $v \in \mathbb{R}^k$, where $1 \leq \alpha < \infty$. When computing the norm of a given vector using these norms, we do not explicitly make the distinction between the spaces $\mathbb{R}^m$ and $\mathbb{R}^n$ because it is clear from the context or from the dimension of the vector what norm we are employing. Given an $m \times n$ matrix $A$, we define the norm of $A$ to be the operator norm:

$$
\|A\|_{\alpha, \beta} = \max \{\|Ax\|_\beta : x \in \mathbb{R}^n, \|x\|_\alpha \leq 1\},
$$

where $1 \leq \alpha, \beta \leq \infty$. In particular, when $\alpha = \beta$, we only use one subscript, that is.

$$
\|A\|_\alpha := \|A\|_{\alpha, \alpha}.
$$

Furthermore, we omit the subscripts when $\alpha = \beta = 1$, in other words,

$$
\|A\| := \|A\|_1.
$$

It follows that $\|A\| = \max_{1 \leq j \leq n} \sum_{i=1}^m |A_{ij}|$. Furthermore, it follows that $\|A^T\|_\infty = \|A\|_1 = \|A\|$. Observe that if $A = uv^T$, where $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, then $\|A\| = \|u\|_1 \|v\|_\infty$.

The following proposition establishes the well known equivalencies among the norms $\| \cdot \|_1, \| \cdot \|_2$, and $\| \cdot \|_\infty$.

**Proposition 2.1** The following inequalities hold:

(i) $\|v\|_2 \leq \|v\|_1 \leq \sqrt{k}\|v\|_2$ for any $v \in \mathbb{R}^k$.

(ii) $\|v\|_\infty \leq \|v\|_1 \leq k\|v\|_\infty$ for any $v \in \mathbb{R}^k$.

(iii) $(1/\sqrt{k})\|v\|_2 \leq \|v\|_\infty \leq \|v\|_2$ for any $v \in \mathbb{R}^k$.

(iv) $(1/\sqrt{n})\|A\|_2 \leq \|A\| \leq \sqrt{m}\|A\|_2$ for any $A \in \mathbb{R}^{m \times n}$. 

For \( d = (A, b, c) \in \tilde{D} \), we define the product norm on the Cartesian product \( \mathbb{R}^{m \times n} \times \mathbb{R}^m \times \mathbb{R}^n \) (and so the norm on \( \tilde{D} \)) as

\[
\|d\| = \max \{ \|A\|, \|b\|_1, \|c\|_{\infty} \}.
\]

We define the ball centered at \( d \in \tilde{D} \) with radius \( \delta \) as:

\[
B(d, \delta) = \{ d + \Delta d \in \tilde{D} : \|\Delta d\| \leq \delta \}.
\]

Given a subset of data instances \( S \subset \tilde{D} \), we denote by \( \text{cl}(S) \) the closure of \( S \), by \( \text{int}(S) \) the interior of \( S \), and by \( \partial S \) the boundary of \( S \), where in order to define these sets we are implicitly assuming that we are using the topology induced by the norm on \( \tilde{D} \) defined above.

### 2.2 Distances to Ill-Posedness

Consider the following subsets of the data space \( \tilde{D} \):

\[
\begin{align*}
\mathcal{F} &= \{ (A, b, c) \in \tilde{D} : \text{there exists } (x, y) \text{ with } Ax = b, x \geq 0, A^T y \leq c \}, \quad (2.1) \\
\mathcal{F}_E &= \{ (A, b, c) \in \tilde{D} : \text{there exists } (x, u) \text{ with } Ax = b, x > 0, A^T u < 0 \}, \quad (2.2) \\
\mathcal{F}_I &= \{ (A, b, c) \in \tilde{D} : \text{there exists } (y, v) \text{ with } A^T y < c, Av = 0, v > 0 \}, \quad (2.3) \\
\mathcal{F}_S &= \{ (A, b, c) \in \tilde{D} : \text{there exists } (x, y) \text{ with } Ax = b, x > 0, A^T y < c \}. \quad (2.4)
\end{align*}
\]

These subsets contain data instances for which the problems \( P(d) \), \( AE(d) \), \( AI(d) \), \( P_\mu(d) \), respectively, and their corresponding duals are feasible. For instance, the elements in \( \mathcal{F} \) correspond to those instances in \( \tilde{D} \) for which \( P(d) \) and \( D(d) \) are feasible. The complement of \( \mathcal{F} \), denoted by \( \mathcal{F}^C \), is the set of data instances \( d = (A, b, c) \) for which \( P(d) \) or \( D(d) \) is infeasible.
The following proposition establishes a few properties of the feasibility sets $\mathcal{F}$, $\mathcal{F}_E$, $\mathcal{F}_I$, and $\mathcal{F}_S$.

**Proposition 2.2** Let $\mathcal{F}$, $\mathcal{F}_E$, $\mathcal{F}_I$, and $\mathcal{F}_S$ be the subsets of data instances defined in (2.1), (2.2), (2.3), and (2.4), respectively. Let $\mathcal{H} = \{(A, b, c) \in \mathcal{D} : b \in \text{range}(A)\}$. Then

1. $\mathcal{F}_E \subset \mathcal{F}_S \subset \mathcal{F}$,

2. $\mathcal{F}_I \cap \mathcal{H} \subset \mathcal{F}_S$,

3. $\mathcal{F}_E \cap \mathcal{F}_I = \emptyset$,

where all the inclusions are proper.

The boundaries of the sets of data instances $\mathcal{F}$, $\mathcal{F}_E$, $\mathcal{F}_I$, and $\mathcal{F}_S$ are denoted as $\mathcal{B}$, $\mathcal{B}_E$, $\mathcal{B}_I$, and $\mathcal{B}_S$, respectively. The data instances in these boundary sets are called *ill-posed* with respect to the mathematical program associated with such boundary. For instance, the data instances $d = (A, b, c)$ in $\mathcal{B}$ are called the ill-posed data instances for $P(d)$, due to the fact that arbitrarily small changes in the data $d$ can yield data instances in $\mathcal{F}$ as well as data instances in its complement $\mathcal{F}^C$. Similar arguments apply to the data instances in the sets $\mathcal{B}_E$, $\mathcal{B}_I$, and $\mathcal{B}_S$. Note that none of these boundary sets is empty since the data instance $(0, 0, 0)$ belongs to $\mathcal{B} \cap \mathcal{B}_E \cap \mathcal{B}_I \cap \mathcal{B}_S$. In Chapter 3, we characterize these sets of ill-posed data instances and discuss their interrelationships.

**Definition 2.1 Distances to Ill-Posedness.** For a data instance $d \in \mathring{\mathcal{D}}$, the distances to ill-posedness with respect to the programs $P(d)$, $AE(d)$, $AI(d)$, and $P_\mu(d)$,
respectively, are defined as follows:

\[ \rho(d) = \inf\{ \|\Delta d\| : d + \Delta d \in B \}, \]
\[ \rho_E(d) = \inf\{ \|\Delta d\| : d + \Delta d \in B_E \}, \]
\[ \rho_I(d) = \inf\{ \|\Delta d\| : d + \Delta d \in B_I \}, \]
\[ \rho_S(d) = \inf\{ \|\Delta d\| : d + \Delta d \in B_S \}, \]

where \( B, B_E, B_I, \) and \( B_S \) are the boundaries of the sets \( \mathcal{F}, \mathcal{F}_E, \mathcal{F}_I, \) and \( \mathcal{F}_S, \) respectively.

The notion of a distance to ill-posedness in a mathematical programming context was originally conceived by Renegar in [Ren94, Ren95a, Ren95b, Ren95c, Ren96]. In Chapter 3, we show that for a data instance \( d \in \bar{D}, \rho_S(d) = \rho(d) \) (see Corollary 3.1). Therefore, from now on we also use the notation \( \rho(\cdot) \) to refer to \( \rho_S(\cdot) \). Observe that under a particular choice of norms used to define the norm on \( \bar{D} \), the distance to ill-posedness \( \rho(d) \) can be computed in polynomial time whenever \( d \in \bar{D} \) is rational (see Remark 3.1 of [FV95]). Also, notice that the definition of a distance to ill-posedness depends on the structure of the mathematical program associated with such distance. For instance, the distance to ill-posedness of a linear program expressed in standard form does not necessarily have the same meaning as the distance to ill-posedness of a linear program expressed in inequality form.

It is straightforward to show that

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

so that we could also define \( \rho(d) \) by employing (2.5). Analogous remarks apply to the distances \( \rho_E(d) \) and \( \rho_I(d) \).
From Proposition 2.2, it follows immediately the following corollary relating the distances \( \rho_E(d) \) and \( \rho_I(d) \) to the distance \( \rho(d) \).

**Corollary 2.1** If \( d = (A,b,c) \in \mathcal{F}_E \), then \( \rho_E(d) \leq \rho(d) \). If \( d = (A,b,c) \in \mathcal{F}_I \) and \( b \in \text{range}(A) \), then \( \rho_I(d) \leq \rho(d) \).

We are now ready to formally define conditions numbers associated with each of our optimization problems. As before, these definitions are based on the work of Renegar.

**Definition 2.2 Condition Numbers.** For a data instance \( d \in \mathcal{D} \), the condition numbers with respect to the programs \( P(d) \), \( AE(d) \), \( AI(d) \), and \( P_{\mu}(d) \), respectively, are defined as follows:

\[
C(d) = \begin{cases} 
\frac{\|d\|}{\rho(d)} & \text{if } \rho(d) > 0, \\
\infty & \text{if } \rho(d) = 0,
\end{cases}
\]

\[
C_E(d) = \begin{cases} 
\frac{\|d\|}{\rho_E(d)} & \text{if } \rho_E(d) > 0, \\
\infty & \text{if } \rho_E(d) = 0,
\end{cases}
\]

\[
C_I(d) = \begin{cases} 
\frac{\|d\|}{\rho_I(d)} & \text{if } \rho_I(d) > 0, \\
\infty & \text{if } \rho_I(d) = 0,
\end{cases}
\]

\[
C_S(d) = \begin{cases} 
\frac{\|d\|}{\rho_S(d)} & \text{if } \rho_S(d) > 0, \\
\infty & \text{if } \rho_S(d) = 0,
\end{cases}
\]

where \( \rho(d) \), \( \rho_E(d) \), \( \rho_I(d) \), and \( \rho_S(d) \) are the distances to ill-posedness defined in Definition 2.1, respectively.

As remarked before, since \( \rho_S(d) = \rho(d) \) for any data instance \( d \in \mathcal{D} \), we use \( C(\cdot) \) to refer to \( C_S(\cdot) \). For \( d \in \mathcal{D} \), the condition numbers \( C(d) \), \( C_E(d) \), and \( C_I(d) \) can be
viewed as a scale-invariant reciprocals of $\rho(d)$, $\rho_E(d)$, and $\rho_I(d)$, respectively, as it is
elementary to demonstrate that $C(d) = C(\alpha d)$, $C_E(d) = C_E(\alpha d)$, and $C_I(d) = C_I(\alpha d)$,
for any positive scalar $\alpha$. Moreover, for $d = (A, b, c)$, let $\Delta d = (-A, -b, -c)$ and
observe that since $d + \Delta d = (0, 0, 0) \in B \cap B_E \cap B_I$ and the boundaries $B$, $B_E$, and $B_I$
are closed sets, then for any $d \not\in B$ we have $\|d\| = \|\Delta d\| \geq \rho(d) > 0$, so that $C(d) \geq 1$.
Analogously, for $d \not\in B_E$ we have $C_E(d) \geq 1$ and for $d \not\in B_I$ we have $C_I(d) \geq 1$.

2.3 Theorems of the Alternative

We state elementary propositions that are well-known variants of classical "theorems
of the alternative" for linear inequality systems, see Gale [Gal60], and are stated in
the context of the mathematical programs studied here. The first two propositions
correspond to the well-known Farkas' Lemma and its corresponding inequality version.

Proposition 2.3 (Farkas' Lemma) Exactly one of the following two systems has a
solution:

- $Ax = b$ and $x \geq 0$.
- $A^T u \leq 0$ and $b^T u > 0$.

Proposition 2.4 Exactly one of the following two systems has a solution:

- $A^T y \leq c$.
- $Av = 0$, $v \geq 0$, and $c^T v < 0$.

The next two propositions correspond to the well-known Gordan's Theorem and
its corresponding homogeneous inequality version.

Proposition 2.5 (Gordan's Theorem) Exactly one of the following two systems has
a solution:
2.3 Theorems of the Alternative

- $Av = 0$, $v \geq 0$, and $v \neq 0$.

- $A^T u < 0$.

\textbf{Proposition 2.6} \textit{Exactly one of the following two systems has a solution:}

- $Av = 0$ and $v > 0$.

- $A^T u \leq 0$ and $A^T u \neq 0$.

\textit{Furthermore, exactly one of the following two statements is true:}

- \textit{The system $Av = 0$ and $v > 0$ has a solution and $A$ has full-row rank.}

- \textit{The system $A^T u \leq 0$ and $u \neq 0$ has a solution.}

The following two propositions are the analogous versions of Farkas' Lemma and its inequality form when considering strict inequalities, respectively.

\textbf{Proposition 2.7} \textit{Exactly one of the following two systems has a solution:}

- $Ax = b$ and $x > 0$.

- $A^T u \leq 0$, $b^T u \geq 0$, and $(A^T u, b^T u) \neq 0$.

\textit{Furthermore, exactly one of the following two statements is true:}

- \textit{The system $Ax = b$ and $x > 0$ has a solution and $A$ has full-row rank.}

- \textit{The system $A^T u \leq 0$, $b^T u \geq 0$, and $u \neq 0$ has a solution.}

\textbf{Proposition 2.8} \textit{Exactly one of the following two systems has a solution:}

- $A^T y < c$.

- $Av = 0$, $v \geq 0$, $c^T v \leq 0$, and $v \neq 0$. 
The following two corollaries are obvious consequences of Propositions 2.7 and 2.8. We state them in order to simplify the proofs of some results formulated in later chapters.

**Corollary 2.2** Exactly one of the following two systems has a solution:

- $Ax = -b$ and $x > 0$.
- $A^T u \leq 0$, $b^T u \leq 0$, and $(A^T u, b^T u) \neq 0$.

Furthermore, exactly one of the following two statements is true:

- The system $Ax = -b$ and $x > 0$ has a solution and $A$ has full-row rank.
- The system $A^T u \leq 0$, $b^T u \leq 0$, and $u \neq 0$ has a solution.

**Corollary 2.3** Exactly one of the following two systems has a solution:

- $A^T y < -c$.
- $Av = 0$, $v \geq 0$, $c^T v \geq 0$, and $v \neq 0$.

We now state two propositions concerning homogeneous systems of inequalities.

**Proposition 2.9** Exactly one of the following two systems has a solution:

- $Ax - br = 0$, $(x, r) \geq 0$, and $(x, r) \neq 0$.
- $A^T u < 0$ and $b^T u > 0$.

**Proposition 2.10** Exactly one of the following two systems has a solution:

- $A^T y - ct \leq 0$, $t \geq 0$, and $(t, A^T y - ct) \neq 0$.
- $Av = 0$, $c^T v < 0$, and $v > 0$.

Furthermore, exactly one of the following two statements is true:
2.4 Further Notation and Results

- The system $A^T y - ct \leq 0$, $t \geq 0$, and $(y, t) \neq 0$ has a solution.

- The system $Av = 0$, $c^T v < 0$, and $v > 0$ has a solution and $A$ has full-row rank.

Finally, the following proposition states a theorem of the alternative for data instances $d = (A, b, c)$ for which the corresponding linear programs $P(d)$ and $D(d)$ are both infeasible.

**Proposition 2.11** Exactly one of the following two systems has a solution:

- $Ax - br = 0$, $A^T y - ct \leq 0$, $(x, r, t) \geq 0$, and $r + t > 0$.

- $Av = 0$, $c^T v < 0$, $v \geq 0$, $A^T u \leq 0$, and $b^T u > 0$.

2.4 Further Notation and Results

We introduce the following notational convention which is standard in the field of interior point methods: if $x \in \mathbb{R}^n$, then $X = \text{diag}(x_1, \ldots, x_n)$. We also use a similar convention for the symbols $s$ and $v$ representing vectors in $\mathbb{R}^n$, that is, $S = \text{diag}(s_1, \ldots, s_n)$ and $V = \text{diag}(v_1, \ldots, v_n)$. We denote by $e$ a vector of ones whose dimension depends on the context of the expression where this vector appears, so that no confusion should arise when using this notation. Similarly, we denote by $e_j$ the $j$-th canonical vector in the vector space $\mathbb{R}^k$, where as before, the dimension $k$ of the vector depends on the context of the expression where this vector appears.

Given a subset $T \subset \{1, \ldots, l\}$ and a vector $v \in \mathbb{R}^l$, we denote by $v_T$ the vector in $\mathbb{R}^{|T|}$ whose components are $v_j$ for all $j \in T$. Furthermore, if $R$ denotes a matrix in $\mathbb{R}^{k \times l}$, then $R_T$ denotes the submatrix in $\mathbb{R}^{k \times |T|}$ whose columns are $R_{jT}$ for all $j \in T$.

If $H$ is a symmetric matrix in $\mathbb{R}^{k \times k}$, we denote by $\lambda(H) = (\lambda_1, \ldots, \lambda_k)^T$ the vector in $\mathbb{R}^k$ whose components are the ordered (real) eigenvalues of $H$, that is, each $\lambda_j$
is an eigenvalue of $H$, and $\lambda_1 \leq \ldots \leq \lambda_k$. Furthermore, we denote by $\lambda_j(H)$ the $j$-th eigenvalue of $H$ chosen according to the increasing order in $\lambda(H)$. In particular, $\lambda_1(H)$ and $\lambda_k(H)$ are the smallest and largest eigenvalues of $H$, respectively. The following result is a key proposition that relates the distance to ill-posedness of a data instance $d = (A, b, c)$ associated with a linear program in standard form to the smallest eigenvalue of the matrix $AA^T$.

**Proposition 2.12** Let $d = (A, b, c) \in \mathcal{F}$ and $\rho(d) > 0$. Then

(i) \( (1/m)\|(AA^T)^{-1}\|_2 \leq \|(AA^T)^{-1}\|_{1,\infty} \leq \|(AA^T)^{-1}\|_2, \)

and

(ii) \( \rho(d) \leq \sqrt{m\lambda_1(AA^T)}. \)

**Proof:** Observe that $A$ has rank $m$, otherwise it is easy to prove that $\rho(d) = 0$. Thus, $(AA^T)^{-1}$ exists. The proof of (i) follows directly from Proposition 2.1, inequalities (i) and (iii). For the proof of (ii), let $\lambda_1 = \lambda_1(AA^T)$. There exists $v \in \mathbb{R}^m$ with $\|v\|_2 = 1$ and $AA^Tv = \lambda_1v$, so that $\|ATv\|_2^2 = v^TAA^Tv = \lambda_1$. Let $\Delta A = -vv^TA$, $\Delta b = \epsilon v$ for any $\epsilon > 0$ and small. Then, $(A + \Delta A)^Tv = 0$ and $(b + \Delta b)^Tv = b^Tv + \epsilon \neq 0$, for all $\epsilon > 0$ small. Hence, $(A + \Delta A)x = b + \Delta b$ is an inconsistent system of equations for all $\epsilon > 0$ and small. Therefore, by Proposition 2.1, inequality (iv), $\rho(d) \leq \max\{\|\Delta A\|, \|\Delta b\|_1\} = \|\Delta A\| \leq \sqrt{m}\|\Delta A\|_2 = \sqrt{m}\|ATv\|_2 = \sqrt{m}\lambda_1$. thus proving (ii).

q.e.d.

The following are standard results from matrix analysis in linear algebra. For a proof of these results see [GVL97], [SB93], or [Wil65]. Given a matrix $H$ in $\mathbb{R}^{k\times k}$, we
introduce the following condition number:

\[ \kappa_\alpha(H) := \begin{cases} 
\|H\|_\alpha \|H^{-1}\|_\alpha & \text{if } H \text{ is non-singular,} \\
\infty & \text{if } H \text{ is singular.}
\end{cases} \]

From this definition, we have \( \kappa_\alpha(\mu H) = \kappa_\alpha(H) \) for all \( \mu \neq 0 \), and \( \kappa_\alpha(H) \geq 1 \). In particular, observe that if \( H \) is non-singular and symmetric, then for \( \alpha = 2 \) we have \( \|H^{-1}\|_2 = \lambda_1(H) \) and \( \|H\|_2 = \lambda_k(H) \). Thus,

\[ \kappa_2(H) = \frac{\lambda_k(H)}{\lambda_1(H)}. \]

More generally, let \( \delta_\alpha(H) \) denote the distance to singularity of the matrix \( H \) using the matrix operator norm \( \| \cdot \|_\alpha \), that is,

\[ \delta_\alpha(H) = \inf \{ \|\Delta H\|_\alpha : H + \Delta H \text{ is singular}\}. \quad (2.6) \]

If \( H \) is a non-singular matrix, then (see [GVL97])

\[ \|H^{-1}\|_\alpha^{-1} = \delta_\alpha(H), \quad (2.7) \]

so that \( \kappa_\alpha(H) \) is a scale-invariant reciprocal of the distance to singularity of the matrix \( H \). Moreover, the closer to singularity the matrix \( H \) is, the larger the value of \( \|H^{-1}\|_\alpha \) will be.

The following proposition states an upper bound on the size of a data perturbation \( \Delta H \) on a non-singular matrix \( H \) so that \( H + \Delta H \) is also non-singular. In particular, this proposition is important because it proves that the set of non-singular matrices in \( \mathbb{R}^{k \times k} \) is an open set, a result that we will use in Chapter 3.

**Proposition 2.13** Let \( H \) be a non-singular matrix in \( \mathbb{R}^{k \times k} \) and suppose that \( \Delta H \) is
a matrix in \( \mathbb{R}^{k \times k} \) such that \( \|\Delta H\|_\alpha < \|H^{-1}\|^{-1}_\alpha \). Then, \( H + \Delta H \) is non-singular and

\[
\|(H + \Delta H)^{-1} - H^{-1}\|_\alpha \leq \frac{\|H^{-1}\|^{-2}_\alpha \|\Delta H\|_\alpha}{1 - \|\Delta H\|_\alpha \|H^{-1}\|^{-1}_\alpha},
\]

where \( 1 \leq \alpha \leq \infty \). Moreover, given \( \epsilon > 0 \) we have that if

\[
\|\Delta H\|_\alpha < \frac{\epsilon \|H^{-1}\|^{-1}_\alpha}{\epsilon + \|H^{-1}\|^{-1}_\alpha},
\]

then \( \|(H + \Delta H)^{-1} - H^{-1}\|_\alpha < \epsilon \).

Finally, the following proposition establishes an upper bound on the change of solutions to linear systems of equations under data perturbations.

**Proposition 2.14** Let \( H \) be a non-singular matrix in \( \mathbb{R}^{k \times k} \) and \( \Delta H \) be a matrix in \( \mathbb{R}^{k \times k} \) such that \( \|\Delta H\|_\alpha < \|H^{-1}\|^{-1}_\alpha \). Let \( \hat{u} \) be the unique solution to the system \( Hu = b \) and \( \bar{u} \) be the unique solution to the system \( (H + \Delta H) u = b + \Delta b \). Then,

\[
\|\bar{u} - \hat{u}\|_\alpha \leq \|H^{-1}\|_\alpha \left( \frac{\|\Delta b\|_\alpha + \|\Delta H\|_\alpha \|\hat{u}\|_\alpha}{1 - \|\Delta H\|_\alpha \|H^{-1}\|^{-1}_\alpha} \right),
\]

where \( 1 \leq \alpha \leq \infty \). Moreover, given \( \epsilon > 0 \), if

\[
\|\Delta H\|_\alpha \leq \frac{\epsilon \|H^{-1}\|^{-1}_\alpha}{2(\|\hat{u}\|_\alpha + \epsilon)}, \quad \|\Delta b\|_\alpha \leq \frac{\epsilon}{2} \|H^{-1}\|^{-1}_\alpha,
\]

then \( \|\bar{u} - \hat{u}\|_\alpha < \epsilon \).
Chapter 3

Characterization of Ill-Posed Data Instances

3.1 Overview

In this chapter we characterize the sets of ill-posed data instances of the optimization problems of interest in this thesis. By a characterization we mean the following: given a data instance \( d = (A, b, c) \) in the space of data instances, for each optimization problem we provide a union of systems of linear equations and inequalities in terms of the data \((A, b, c)\), such that \( d \) is an ill-posed data instance with respect to such optimization problem if and only if the corresponding union of linear systems has a solution.

In Section 3.2, we characterize the sets of ill-posed data instances for the analytic center problems in equality and inequality form \((AE(d)\) and \(AI(d)\)), and a linear program in standard form \( P(d) \) (see Chapter 2). Since the set of ill-posed data instances for a linear program is the same set as the set of ill-posed data instances for the central trajectory problem (see Corollary 3.1 below), then we automatically ob-
tain a characterization for the set of ill-posed data instances of the central trajectory problem as well.

Throughout this chapter we use the following sets of data instances:

\[ \mathcal{P} \ := \ \{(A, b, c) : \text{there exists } x \text{ with } Ax = b, x \geq 0\}, \]
\[ \mathcal{P}_+ \ := \ \{(A, b, c) : \text{there exists } x \text{ with } Ax = b, x > 0, \text{ and } \text{rank}(A) = m\}, \]
\[ \mathcal{P}_- \ := \ \{(A, b, c) : \text{there exists } x \text{ with } Ax = -b, x > 0, \text{ and } \text{rank}(A) = m\}, \]
\[ \mathcal{P}_0 \ := \ \{(A, b, c) : \text{there exists } v \text{ with } Av = 0, v > 0, \text{ and } \text{rank}(A) = m\}. \]

Generally speaking, these sets correspond to data instances that are feasible for \(P(\cdot)\), strictly feasible for \(P(\cdot)\), strictly feasible for \(P(\cdot)\) after we change the sign of the right-hand-side vector \(b\), and data instances that have a strictly primal ray, respectively.

It follows from Proposition 2.3, Proposition 2.7, Corollary 2.2, and Proposition 2.6, respectively, that the complements of the sets above are given by:

\[ \mathcal{P}^C \ := \ \{(A, b, c) : \text{there exists } u \text{ with } A^T u \leq 0, b^T u > 0\}, \]
\[ \mathcal{P}_+^C \ := \ \{(A, b, c) : \text{there exists } u \text{ with } A^T u \leq 0, b^T u \geq 0, u \neq 0\}, \]
\[ \mathcal{P}_-^C \ := \ \{(A, b, c) : \text{there exists } u \text{ with } A^T u \leq 0, b^T u \leq 0, u \neq 0\}, \]
\[ \mathcal{P}_0^C \ := \ \{(A, b, c) : \text{there exists } u \text{ with } A^T u \leq 0, u \neq 0\}. \]

Analogously, we introduce the following sets of data instances associated with the problem \(D(\cdot)\):

\[ \mathcal{D} \ := \ \{(A, b, c) : \text{there exists } y \text{ with } A^T y \leq c\}, \]
\[ \mathcal{D}_+ \ := \ \{(A, b, c) : \text{there exists } y \text{ with } A^T y < c\}, \]
3.1 Overview

\[ \mathcal{D}_- \; := \; \{(A, b, c) : \text{there exists } y \text{ with } A^T y < -c\}, \]
\[ \mathcal{D}_0 \; := \; \{(A, b, c) : \text{there exists } u \text{ with } A^T u < 0\}. \]

Generally speaking, these sets correspond to data instances that are feasible for \( D(\cdot) \), strictly feasible for \( D(\cdot) \), strictly feasible for \( D(\cdot) \) after we change the sign of the right-hand-side vector \( c \), and data instances that have a strictly dual ray, respectively.

As before, it follows from Proposition 2.4, Proposition 2.8, Corollary 2.3, and Proposition 2.5, respectively, that the complements of the sets above are given by:

\[ \mathcal{D}^C \; = \; \{(A, b, c) : \text{there exists } v \text{ with } Av = 0, c^T v < 0, v \geq 0\}, \]
\[ \mathcal{D}_+^C \; = \; \{(A, b, c) : \text{there exists } v \text{ with } Av = 0, c^T v \leq 0, v \geq 0, v \neq 0\}, \]
\[ \mathcal{D}^- \; = \; \{(A, b, c) : \text{there exists } v \text{ with } Av = 0, c^T v \geq 0, v \geq 0, v \neq 0\}, \]
\[ \mathcal{D}_0^C \; = \; \{(A, b, c) : \text{there exists } v \text{ with } Av = 0, v \geq 0, v \neq 0\}. \]

We proceed to state the most important results in this chapter. The first theorem establishes a characterization in terms of the sets defined above of the interior, closure, and boundary of \( \mathcal{F} \), the set of feasible data instances associated with the linear program in standard form \( P(\cdot) \), and its complement \( \mathcal{F}^C \). To simplify the expressions presented in this theorem and the theorems below, we omit the symbol \( \cap \) denoting set intersection, so that "\( ST \)" denotes "\( S \cap T \)" for any pair of data sets \( S \) and \( T \). This notational convention is used in order to simplify the statements of the results, and is only used in this chapter.

**Theorem 3.1**

\[ \text{int}(\mathcal{F}) \; = \; \mathcal{P}_+ \mathcal{D}_+, \]
\[ \text{cl}(\mathcal{F}) \; = \; \mathcal{P}\mathcal{D} \cup \mathcal{P}\mathcal{D}_0^C \cup \mathcal{D}\mathcal{D}_0^C \cup \mathcal{P}^C \mathcal{D}^C \mathcal{D}^- \cup \mathcal{P}^C \mathcal{D}^C \mathcal{P}^- \]  

(3.1)  

(3.2)
3.1 Overview

\[
\text{int}(\mathcal{F}^C) = \mathcal{P}_+^D \mathcal{P}_-^D \cup \mathcal{P}_+^D \mathcal{P}_0 \cup \mathcal{P}_-^D \mathcal{P}_0, \tag{3.3}
\]
\[
\text{cl}(\mathcal{F}^C) = \mathcal{P}_+^C \cup \mathcal{D}_+^C, \tag{3.4}
\]
\[
\partial \mathcal{F} = (\mathcal{P}_+ \cup \mathcal{P}_0^C \cup \mathcal{D}_0^C) (\mathcal{P}_+^C \cup \mathcal{D}_+^C) \cup \mathcal{P}_+^D \mathcal{D}_-^C \cup \mathcal{P}_-^D \mathcal{P}_-^C, \tag{3.5}
\]

The next theorem states similar characterizations of the interior, closure, and boundary of \( \mathcal{F}_E \), the set of feasible data instances associated with the analytic center problem \( AE(\cdot) \), and its complement \( \mathcal{F}^C_E \).

**Theorem 3.2**

\[
\text{int}(\mathcal{F}_E) = \mathcal{P}_+ D_0, \tag{3.6}
\]
\[
\text{cl}(\mathcal{F}_E) = \mathcal{P}_+^C \cup \mathcal{D}_0^C \mathcal{P}_-^C, \tag{3.7}
\]
\[
\text{int}(\mathcal{F}^C_E) = \mathcal{P}_- \cup \mathcal{P}_+^C D_0, \tag{3.8}
\]
\[
\text{cl}(\mathcal{F}^C_E) = \mathcal{P}_+^C \cup \mathcal{D}_0^C, \tag{3.9}
\]
\[
\partial \mathcal{F}_E = \mathcal{P}_+^C \cup \mathcal{D}_0^C \mathcal{P}_-^C, \tag{3.10}
\]

Observe that from this theorem, identity (3.10), it follows that given a data instance \( d = (A, b, c) \), the program \( AE(d) \) is ill-posed if and only if at least one of the following statements holds:

- The program \( AE(d) \) is primal feasible, but does not admit a strictly feasible solution satisfying \( Ax = b \) and \( x > 0 \) (in other words, the feasible region associated with \( d \) is contained in one of the faces of the positive orthant \( \mathbb{R}_+^n \)).

- The program \( AE(d) \) is not dual feasible, and so there is a primal ray \( \bar{v} \) satisfying \( A\bar{v} = 0, \bar{v} \geq 0, \) and \( \bar{v} \neq 0 \), but there is no strictly positive primal ray, that is, a ray \( v \) satisfying \( Av = 0, v > 0 \).

- The matrix \( A \) is rank deficient.
3.1 Overview

The last theorem establishes characterizations of the interior, closure, and boundary of $\mathcal{F}_I$, the set of feasible data instances associated with the analytic center problem $AI(\cdot)$, and its complement $\mathcal{F}_I^C$.

**Theorem 3.3**

\[
\begin{align*}
\text{int}(\mathcal{F}_I) &= \mathcal{D}_+ \mathcal{P}_0, \\
\text{cl}(\mathcal{F}_I) &= \mathcal{D}^C_+ \cup \mathcal{P}^C_0 \mathcal{D}^C_-, \\
\text{int}(\mathcal{F}_I^C) &= \mathcal{D}^- \cup \mathcal{D}^C \mathcal{P}_0, \\
\text{cl}(\mathcal{F}_I^C) &= \mathcal{D}^C_+ \cup \mathcal{P}^C_0, \\
\partial \mathcal{F}_I &= \mathcal{D}^C_+ \cup \mathcal{P}^C_0 \mathcal{D}^C_-, 
\end{align*}
\]

As before, observe that from this theorem, identity (3.15), it follows that given a data instance $d = (A, b, c)$, the program $AI(d)$ is ill-posed if and only if at least one of the following statements holds:

- The program $AI(d)$ is primal feasible, but does not admit a strictly feasible solution satisfying $A^T y < c$.

- The program $AI(d)$ is not dual feasible, and so there is a primal ray $\bar{u}$ satisfying $A^T \bar{u} \leq 0$, and $\bar{u} \neq 0$, but there is no strict primal ray, that is, a ray $u$ satisfying $A^T u < 0$.

- The matrix $A$ is rank deficient.

Table 3.1 provides a guide to the proofs of the preceding theorems. The table shows where in Section 3.2 the reader can find the corresponding proof of the characterization of the interiors, closures, and boundaries of the sets of interest in this chapter.
3.2 Characterizations of Ill-Posedness

<table>
<thead>
<tr>
<th></th>
<th>$\mathcal{F}$</th>
<th>$\mathcal{F}_E$</th>
<th>$\mathcal{F}_I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{int}(\cdot)$</td>
<td>Lemma 3.1</td>
<td>Lemma 3.2</td>
<td>Lemma 3.2</td>
</tr>
<tr>
<td>$\text{cl}(\cdot)$</td>
<td>Lemma 3.4</td>
<td>Lemma 3.4</td>
<td>Lemma 3.4</td>
</tr>
<tr>
<td>$\text{int}((\cdot)^C)$</td>
<td>Lemma 3.5</td>
<td>Lemma 3.5</td>
<td>Lemma 3.5</td>
</tr>
<tr>
<td>$\text{cl}((\cdot)^C)$</td>
<td>Lemma 3.3</td>
<td>Lemma 3.3</td>
<td>Lemma 3.3</td>
</tr>
<tr>
<td>$\partial(\cdot)$</td>
<td>Lemma 3.6</td>
<td>Lemma 3.7</td>
<td>Lemma 3.7</td>
</tr>
</tbody>
</table>

Table 3.1: Guide to the proofs of Theorem 3.1, Theorem 3.2, and Theorem 3.3.

Finally, in Section 3.3, we review the original distances to ill-posedness defined by Renegar in [Ren94, Ren95a, Ren95b, Ren95c, Ren96] and we relate them to the extended versions used in this thesis.

3.2 Characterizations of Ill-Posedness

Our objective in this section is to characterize the boundary sets $\mathcal{B} = \partial \mathcal{F}$, $\mathcal{B}_E = \partial \mathcal{F}_E$, and $\mathcal{B}_I = \partial \mathcal{F}_I$ of ill-posed data instances with respect to a linear program in standard form $P(\cdot)$, and the analytic center problems $AE(\cdot)$ and $AI(\cdot)$, respectively. Recall from Chapter 2 that $AE(\cdot)$ corresponds to the analytic center problem in equality form and $AI(\cdot)$ to the analytic center problem in inequality form.
Throughout this section we denote by $\mathcal{R}$ the set of data instances $d = (A, b, c)$ for which the matrix $A$ has rank $m$ (recall from Chapter 2 that we assume that $m \leq n$):

$$\mathcal{R} := \{(A, b, c) \in \tilde{\mathcal{D}} : \text{rank}(A) = m\}. \quad (3.16)$$

In order to prove the characterizations associated with the sets $\mathcal{F}_E$ and $\mathcal{F}_I$, it is convenient to introduce the following relaxed versions of the sets of feasible data instances with respect to the analytic center problems in equality and inequality form, respectively:

$$\hat{\mathcal{F}}_E := \{(A, b, c) : \text{there exists } (x, u) \text{ with } Ax = b, x \geq 0, A^T u \leq 0, u \neq 0\},$$

$$\hat{\mathcal{F}}_I := \{(A, b, c) : \text{there exists } (v, y) \text{ with } Av = 0, v \geq 0, v \neq 0, A^T y \leq c\}.$$

Obviously, $\mathcal{F}_E \subset \hat{\mathcal{F}}_E$ and $\mathcal{F}_I \subset \hat{\mathcal{F}}_I$. Also observe that $\hat{\mathcal{F}}_E = \mathcal{P} \cap \mathcal{P}_0^C$ and $\hat{\mathcal{F}}_I = \mathcal{D} \cap \mathcal{D}_0^C$.

In addition, we introduce the following set $\mathcal{I}$ of data instances that are both primal and dual infeasible with respect to a linear program in standard form:

$$\mathcal{I} := \{(A, b, c) : \text{there exists } (u, v) \text{ with } Av = 0, c^T v < 0, v \geq 0, A^T u \leq 0, b^T u > 0\}.$$

In other words, $\mathcal{I} = \mathcal{P}^C \cap \mathcal{D}^C$.

We use the following well-known proposition that establishes the openness of four important sets of data instances. In particular, the proof of the openness of the first two sets of this proposition are a consequence of Proposition 2.13 and Proposition 2.14 on the perturbation of linear systems of equations.

**Proposition 3.1** The following subsets of data instances in the data space $\tilde{\mathcal{D}}$ are open sets: $\mathcal{P}_+$, $\mathcal{P}_-$, $\mathcal{P}_0$, $\mathcal{D}_+$, $\mathcal{D}_-$, and $\mathcal{D}_0$.

The first lemma characterizes the interior of the set $\mathcal{F}$ of feasible data instances for
3.2 Characterizations of Ill-Posedness

linear programs in standard form. This lemma is a well-known result in the literature of perturbation theory of mathematical programming (see [Ash81, Rob77]). Observe that since \( \mathcal{F}_S \cap \mathcal{R} = \mathcal{P}_+ \cap \mathcal{D}_+ \), we thus obtain the proof of identity (3.1) in Theorem 3.1.

**Lemma 3.1**

\[
\text{int}(\mathcal{F}) = \mathcal{F}_S \cap \mathcal{R}.
\]

**Proof:** The proof of this identity has been previously established by Robinson [Rob77] and Ashmanov [Ash81].

q.e.d.

The following proposition establishes useful identities with respect to the closures and interiors of the sets of feasible data instances of \( P(\cdot) \) and \( P_\mu(\cdot) \) (see Chapter 2). The identity for the closure of \( \text{cl}(\mathcal{F}) \) is particularly useful in proving the characterization of the set of ill-posed data instances for central trajectory problems.

**Proposition 3.2**

\[
\begin{align*}
\text{cl}(\mathcal{F}) &= \text{cl}(\mathcal{F}_S), \quad (3.17) \\
\text{int}(\mathcal{F}) &= \text{int}(\mathcal{F}_S). \quad (3.18)
\end{align*}
\]

**Proof:** Clearly, \( \mathcal{F}_S \subseteq \mathcal{F} \), and so \( \text{cl}(\mathcal{F}_S) \subseteq \text{cl}(\mathcal{F}) \) and \( \text{int}(\mathcal{F}_S) \subseteq \text{int}(\mathcal{F}) \). Given \( d = (A, b, c) \in \mathcal{F} \), there exist \( x \) and \( y \) such that \( Ax = b, \ x \geq 0, \ \text{and} \ \mathcal{A}^T y \leq c \). Thus, given any \( \epsilon > 0 \), we have \( A(x + \epsilon e) = b + \epsilon A e, \ x + \epsilon e > 0, \ \text{and} \ \mathcal{A}^T y < c + \epsilon e \).

Hence, \( d + \Delta d(\epsilon) \in \mathcal{F}_S \) for all \( \epsilon > 0 \), where \( \Delta d(\epsilon) = (0, \epsilon A e, \epsilon e) \). Therefore, \( d \in \text{cl}(\mathcal{F}_S) \), and so \( \mathcal{F} \subseteq \text{cl}(\mathcal{F}_S) \), which implies that \( \text{cl}(\mathcal{F}) \subseteq \text{cl}(\text{cl}(\mathcal{F}_S)) = \text{cl}(\mathcal{F}_S) \), thus proving (3.17). From Lemma 3.1 it follows that \( \text{int}(\mathcal{F}) \subseteq \mathcal{F}_S \), which implies \( \text{int}(\mathcal{F}) = \text{int}(\text{int}(\mathcal{F})) \subseteq \text{int}(\mathcal{F}_S) \), thus proving (3.18).
q.e.d.

In particular, the previous proposition implies that the boundary of $\mathcal{F}$ is the same as the boundary of $\mathcal{F}_S$. Therefore, we obtain the following corollary that asserts the equality between the distances to ill-posedness with respect to linear programming in standard form and the central trajectory problem (see the corresponding definitions in Chapter 2).

**Corollary 3.1** Let $d$ be a data instance in the space $\mathcal{D}$. Then $\partial \mathcal{F} = \partial \mathcal{F}_S$, and so

$$\rho(d) = \rho_S(d).$$

Next we state a proposition that describes the interior of the sets $\hat{\mathcal{F}}_E$ and $\hat{\mathcal{F}}_I$. This is an intermediate step towards characterizing the interiors of the sets $\mathcal{F}_E$ and $\mathcal{F}_I$.

**Proposition 3.3**

\[
\text{int}(\hat{\mathcal{F}}_E) = \mathcal{F}_E \cap \mathcal{R}, \tag{3.19}
\]

\[
\text{int}(\hat{\mathcal{F}}_I) = \mathcal{F}_I \cap \mathcal{R}. \tag{3.20}
\]

**Proof:** Proposition 3.1 implies that $\mathcal{F}_E \cap \mathcal{R}$ and $\mathcal{F}_I \cap \mathcal{R}$ are open sets, and so $\mathcal{F}_E \cap \mathcal{R} \subset \text{int}(\hat{\mathcal{F}}_E)$ and $\mathcal{F}_I \cap \mathcal{R} \subset \text{int}(\hat{\mathcal{F}}_I)$. Let $d = (A, b, c) \in \text{int}(\hat{\mathcal{F}}_E)$. If $d$ is such that $d \notin \mathcal{R}$ or $d$ does not satisfy the system $Ax = b$ and $x > 0$, then from Proposition 2.7 there exists a vector $u \in \mathbb{R}^m$ such that $A^Tu \leq 0$, $b^Tu \geq 0$ and $u \neq 0$. Let $\bar{u}$ be such that $\bar{u}^Tu > 0$. Given an arbitrary $\varepsilon > 0$, let $\Delta d(\varepsilon) = (0, \Delta b(\varepsilon), 0)$, where $\Delta b(\varepsilon) = \varepsilon \bar{u}$. Then, $A^Tu \leq 0$ and $(b + \Delta b(\varepsilon))^Tu > 0$, for all $\varepsilon > 0$. Therefore, by Proposition 2.3, the system $Ax = b + \Delta b(\varepsilon)$ and $x \geq 0$ does not have a solution, and so $d + \Delta d(\varepsilon) \notin \hat{\mathcal{F}}_E$ for all $\varepsilon > 0$, a contradiction. Thus, we have $d \in \mathcal{R}$ and the system $Ax = b$ and $x > 0$ is feasible. If the system $A^Tu < 0$ is not feasible, then
from Proposition 2.5 there exists a vector $v$ such that $Av = 0$, $v \geq 0$, and $v \neq 0$. For arbitrary $\epsilon > 0$, let $\Delta d(\epsilon) = (\Delta A(\epsilon), 0, 0)$, where $\Delta A(\epsilon) = -\epsilon Ae(v + \epsilon e)^T / \|v + \epsilon e\|_2^2$. Then, $(A + \Delta A(\epsilon))(v + \epsilon e) = 0$ and $v + \epsilon e > 0$, for all $\epsilon > 0$. Therefore, by Proposition 2.6, $d + \Delta d(\epsilon) \notin \hat{F}_E$ for all $\epsilon > 0$, a contradiction. Hence, $d \in \mathcal{F}_E \cap \mathcal{R}$, and so $\text{int}(\mathcal{F}_E) \subset \mathcal{F}_E \cap \mathcal{R}$, and identity (3.19) follows.

Now, let $d = (A, b, c) \in \text{int}(\hat{F}_I)$. If $d \notin \mathcal{R}$ or the system $Av = 0$ and $v > 0$ does not have a solution, then by Proposition 2.6 there exists a vector $u$ such that $A^T u \leq 0$ and $u \neq 0$ has a solution. Let $\bar{u}$ be such that $\bar{u}^T u = 1$. Given an arbitrary $\epsilon > 0$, let $\Delta d(\epsilon) = (\Delta A(\epsilon), 0, 0)$, where $\Delta A(\epsilon) = -\epsilon \bar{u} e^T$. Then, $(A + \Delta A(\epsilon))^T u < 0$, for all $\epsilon > 0$. Therefore, by Proposition 2.5, the system $(A + \Delta A(\epsilon))v = 0$, $v \geq 0$, and $v \neq 0$ does not have a solution, and so $d + \Delta d(\epsilon) \notin \hat{F}_I$ for all $\epsilon > 0$, a contradiction. Thus, we have $d \in \mathcal{R}$ and the system $Av = 0$ and $v > 0$ is feasible. If the system $A^T y < c$ is not feasible, then from Proposition 2.8 there exists a vector $v$ such that $Av = 0$, $v \geq 0$, $c^T v \leq 0$, and $v \neq 0$. For arbitrary $\epsilon > 0$, let $\Delta d(\epsilon) = (0, 0, \Delta c(\epsilon))$, where $\Delta c(\epsilon) = -\epsilon v$. Then, $Av = 0$, $(c + \Delta c(\epsilon))^T v < 0$, and $v \geq 0$, for all $\epsilon > 0$. Therefore, by Proposition 2.4, $d + \Delta d(\epsilon) \notin \hat{F}_I$ for all $\epsilon > 0$, a contradiction. Hence, $d \in \mathcal{F}_I \cap \mathcal{R}$, and so $\text{int}(\hat{F}_I) \subset \mathcal{F}_I \cap \mathcal{R}$, and identity (3.20) follows.

q.e.d.

We now prove an analogous result to Proposition 3.2 for the sets $\mathcal{F}_E$ and $\mathcal{F}_I$. This result will be used in the characterization of the closures of these two sets.

**Proposition 3.4**

\begin{align*}
\text{cl}(\mathcal{F}_E) &= \text{cl}(\hat{F}_E), \quad \text{(3.21)} \\
\text{int}(\mathcal{F}_E) &= \text{int}(\hat{F}_E), \quad \text{(3.22)} \\
\text{cl}(\mathcal{F}_I) &= \text{cl}(\hat{F}_I), \quad \text{(3.23)} \\
\text{int}(\mathcal{F}_I) &= \text{int}(\hat{F}_I). \quad \text{(3.24)}
\end{align*}
3.2 Characterizations of Ill-Posedness

Proof: Clearly, \( \text{cl}(\mathcal{F}_E) \subset \text{cl}(\hat{\mathcal{F}}_E) \) and \( \text{int}(\mathcal{F}_E) \subset \text{int}(\hat{\mathcal{F}}_E) \). Also, \( \text{cl}(\mathcal{F}_I) \subset \text{cl}(\hat{\mathcal{F}}_I) \) and \( \text{int}(\mathcal{F}_I) \subset \text{int}(\hat{\mathcal{F}}_I) \). Given \( d = (A, b, c) \in \hat{\mathcal{F}}_E \), then there exist \( x \) and \( u \) such that \( Ax = b, x \geq 0, A^Tu \leq 0 \), and \( u \neq 0 \). Thus, given any \( \epsilon > 0 \), let \( \bar{u} \) be such that \( \bar{u}^Tu = 1 \), and define \( \Delta A(\epsilon) = -\epsilon \bar{u}e^T, \Delta b(\epsilon) = \epsilon Ae - \epsilon \|x\|_1 \bar{u} - \epsilon^2 \bar{n} \bar{u}, \) and \( \Delta c(\epsilon) = 0 \). Then, it is easy to verify that for \( x(\epsilon) = x + \epsilon e \) we have \( (A + \Delta A(\epsilon))x(\epsilon) = b + \Delta b(\epsilon), x(\epsilon) > 0 \), and \( (A + \Delta A(\epsilon))^Tu < 0 \). Thus, for \( \Delta d(\epsilon) = (\Delta A(\epsilon), \Delta b(\epsilon), \Delta c(\epsilon)) \), we obtain that \( d + \Delta d(\epsilon) \in \mathcal{F}_E \) for all \( \epsilon > 0 \). Therefore, \( d \in \text{cl}(\mathcal{F}_E) \), and so \( \hat{\mathcal{F}}_E \subset \text{cl}(\mathcal{F}_E) \), which implies that \( \text{cl}(\hat{\mathcal{F}}_E) \subset \text{cl}(\text{cl}(\mathcal{F}_E)) = \text{cl}(\mathcal{F}_E) \), thus proving (3.21). From Proposition 3.3, identity (3.19), it follows that \( \text{int}(\hat{\mathcal{F}}_E) \subset \mathcal{F}_E \), which implies \( \text{int}(\hat{\mathcal{F}}_E) = \text{int}(\text{int}(\hat{\mathcal{F}}_E)) \subset \text{int}(\mathcal{F}_E) \), thus proving (3.22).

On the other hand, let \( d = (A, b, c) \in \hat{\mathcal{F}}_I \). Then there exist \( v \) and \( y \) such that \( Av = 0, v \geq 0, v \neq 0 \), and \( A^Ty \leq c \). Thus, given any \( \epsilon > 0 \), define \( \Delta A(\epsilon) = -\epsilon A(e^v + \epsilon e)^T v + \epsilon e \|v + \epsilon e\|^2_2, \Delta b(\epsilon) = \epsilon e - \epsilon e^T A^Ty(v + \epsilon e)/\|v + \epsilon e\|^2_2, \) and \( \Delta c(\epsilon) = 0 \). Then, it is easy to verify that for \( v(\epsilon) = v + \epsilon e \) we have \( (A + \Delta A(\epsilon))v(\epsilon) = 0, v(\epsilon) > 0 \), and \( (A + \Delta A(\epsilon))^Ty < c + \Delta c(\epsilon) \). Thus, for \( \Delta d(\epsilon) = (\Delta A(\epsilon), \Delta b(\epsilon), \Delta c(\epsilon)) \), we obtain that \( d + \Delta d(\epsilon) \in \mathcal{F}_I \) for all \( \epsilon > 0 \). Therefore, \( d \in \text{cl}(\mathcal{F}_I) \), and so \( \hat{\mathcal{F}}_I \subset \text{cl}(\mathcal{F}_I) \), which implies that \( \text{cl}(\hat{\mathcal{F}}_I) \subset \text{cl}(\text{cl}(\mathcal{F}_I)) = \text{cl}(\mathcal{F}_I) \), thus proving (3.23). From Proposition 3.3, identity (3.20), it follows that \( \text{int}(\hat{\mathcal{F}}_I) \subset \mathcal{F}_I \), which implies \( \text{int}(\hat{\mathcal{F}}_I) = \text{int}(\text{int}(\hat{\mathcal{F}}_I)) \subset \text{int}(\mathcal{F}_I) \), thus proving (3.24).

q.e.d.

By combining Proposition 3.3 and Proposition 3.4, we obtain the following lemma that characterizes the interior of the sets \( \mathcal{F}_E \) and \( \mathcal{F}_I \).

Lemma 3.2

\[
\text{int}(\mathcal{F}_E) = \mathcal{F}_E \cap \mathcal{R}, \quad (3.25)
\]

\[
\text{int}(\mathcal{F}_I) = \mathcal{F}_I \cap \mathcal{R}. \quad (3.26)
\]
3.2 Characterizations of Ill-Posedness

It follows from Lemma 3.2 that \( \text{int}(\mathcal{F}_E) = \mathcal{P}_+ \cap \mathcal{D}_0 \) and \( \text{int}(\mathcal{F}_I) = \mathcal{D}_+ \cap \mathcal{P}_0 \), and so we obtain identity (3.6) in Theorem 3.2 and identity (3.11) in Theorem 3.3.

By observing that \( \text{cl}(S^C) = \text{int}(S)^C \) for any set \( S \), we obtain the following lemma, which is a characterization of the closures of \( \mathcal{F}^C \), \( \mathcal{F}_E^C \), and \( \mathcal{F}_I^C \), respectively.

**Lemma 3.3**

\[
\begin{align*}
\text{cl}(\mathcal{F}^C) &= \mathcal{P}_+^C \cup \mathcal{D}_+^C, \\
\text{cl}(\mathcal{F}_E^C) &= \mathcal{P}_+^C \cup \mathcal{D}_0^C, \\
\text{cl}(\mathcal{F}_I^C) &= \mathcal{D}_+^C \cup \mathcal{P}_0^C.
\end{align*}
\] (3.27) (3.28) (3.29)

**Proof:** It follows from Lemma 3.1 that \( \text{cl}(\mathcal{F}^C) = \text{int}(\mathcal{F})^C = (\mathcal{F}_S \cap \mathcal{R})^C = (\mathcal{P}_+ \cap \mathcal{D}_+)^C = \mathcal{P}_+^C \cup \mathcal{D}_+^C \), and so identity (3.27) follows. Similarly, from Lemma 3.2, identity (3.25), we have \( \text{cl}(\mathcal{F}_E^C) = \text{int}(\mathcal{F}_E)^C = (\mathcal{F}_E \cap \mathcal{R})^C = (\mathcal{P}_+ \cap \mathcal{D}_0)^C = \mathcal{P}_+^C \cup \mathcal{D}_0^C \), and so identity (3.28) follows. Finally, again from Lemma 3.2, identity (3.26), we have \( \text{cl}(\mathcal{F}_I^C) = \text{int}(\mathcal{F}_I)^C = (\mathcal{F}_I \cap \mathcal{R})^C = (\mathcal{D}_+ \cap \mathcal{P}_0)^C = \mathcal{D}_+^C \cup \mathcal{P}_0^C \), and so identity (3.29) follows.

*Q.E.D.*

Observe that identities (3.27), (3.28), and (3.29) in Lemma 3.3 correspond to identity (3.4) in Theorem 3.1, identity (3.9) in Theorem 3.2, and identity (3.14) in Theorem 3.3.

The following proposition is instrumental in proving Lemma 3.4 below.

**Proposition 3.5** Consider the following subsets of \( \bar{\mathcal{D}} \):

\[
\mathcal{S} = \{ d = (A, b, c) : \text{there exists } (x, y, r, t) \text{ such that } Ax - br = 0, \}
\]
3.2 Characterizations of Ill-Posedness

\[ A^T y - ct \leq 0, \quad (x, r, t) \geq 0, \quad (x, r) \neq 0, \quad (y, t) \neq 0 \}, \]

\[ S_E = \{ d = (A, b, c) : \text{there exists } (x, u, r) \text{ such that } Ax - br = 0, \]
\[ A^T u \leq 0, \quad b^T u \leq 0, \quad (x, r) \geq 0, \quad (x, r) \neq 0, \quad u \neq 0 \}, \]

\[ S_I = \{ d = (A, b, c) : \text{there exists } (v, y, t) \text{ such that } Av = 0, \]
\[ c^T v \geq 0, \quad A^T y - ct \leq 0, \quad (v, t) \geq 0, \quad v \neq 0, \quad (y, t) \neq 0 \}. \]

Then

\[ \text{cl}(\mathcal{F}) \subset S, \quad (3.30) \]
\[ \text{cl}(\mathcal{F}_E) \subset S_E, \quad (3.31) \]
\[ \text{cl}(\mathcal{F}_I) \subset S_I. \quad (3.32) \]

**Proof:** Let \( d = (A, b, c) \) be a data instance in \( \text{cl}(\mathcal{F}) \), and let \( \{ d_k = (A_k, b_k, c_k) : k \geq 1 \} \) be a sequence of data instances in \( \mathcal{F} \) such that \( d_k \to d \) as \( k \to \infty \). For each \( k \geq 1 \), there exist \( x_k \) and \( y_k \) such that \( A_k x_k = b_k, \ x_k \geq 0, \) and \( A_k^T y_k \leq c_k \).

Let \( \bar{x}_k = x_k/\|x_k\|_1 + 1 \), \( \bar{y}_k = y_k/\|y_k\|_\infty + 1 \), \( \bar{r}_k = 1/(\|x_k\|_1 + 1) \), and \( \bar{t}_k = 1/(\|y_k\|_\infty + 1) \), for each \( k \geq 1 \). Then, for each \( k \geq 1 \), we have \( A_k \bar{x}_k - b_k \bar{r}_k = 0, \)
\[ A_k^T \bar{y}_k - c_k \bar{t}_k \leq 0, \quad (\bar{x}_k, \bar{r}_k, \bar{t}_k) \geq 0, \quad \|\bar{x}_k\|_1 + \bar{r}_k = 1, \quad \|\bar{y}_k\|_\infty + \bar{t}_k = 1. \]

Then, the sequence \( \{(\bar{x}_k, \bar{y}_k, \bar{r}_k, \bar{t}_k) : k \geq 1\} \) has an accumulation point \((\bar{x}, \bar{y}, \bar{r}, \bar{t})\) satisfying
\[ A\bar{x} - b\bar{r} = 0, \quad A^T \bar{y} - c\bar{t} \leq 0, \quad (\bar{x}, \bar{r}, \bar{t}) \geq 0, \quad \|\bar{x}\|_1 + \bar{r} = 1, \quad \|\bar{y}\|_\infty + \bar{t} = 1. \]
Therefore, \( d \in S \), and the inclusion (3.30) for \( \text{cl}(\mathcal{F}) \) follows.

Now, let \( d = (A, b, c) \) be a data instance in \( \text{cl}(\mathcal{F}_E) \), and let \( \{ d_k = (A_k, b_k, c_k) : k \geq 1 \} \) be a sequence of data instances in \( \mathcal{F}_E \) such that \( d_k \to d \) as \( k \to \infty \). For each \( k \geq 1 \), there exist \( x_k \) and \( u_k \) such that \( A_k x_k = b_k, \ x_k > 0, \) and \( A_k^T u_k < 0 \). Observe that we also have \( b_k^T u_k < 0 \) for all \( k \geq 1 \). Let \( \tilde{x}_k = x_k/\|x_k\|_1 + 1 \), \( \tilde{u}_k = u_k/\|u_k\|_\infty \), and \( \tilde{r}_k = 1/(\|x_k\|_1 + 1) \), for each \( k \geq 1 \). Then, for each \( k \geq 1 \), we have \( A_k \tilde{x}_k - b_k \tilde{r}_k = 0, \)
\[ A_k^T \tilde{u}_k < 0, \quad b_k^T \tilde{u}_k < 0, \quad (\tilde{x}_k, \tilde{r}_k) > 0, \quad \|\tilde{x}_k\|_1 + \tilde{r}_k = 1, \quad \|\tilde{u}_k\|_\infty = 1. \]
Then, the sequence
\{ (\bar{x}_k, \bar{u}_k, \bar{r}_k) : k \geq 1 \} has an accumulation point \((\bar{x}, \bar{u}, \bar{r})\) satisfying \(A\bar{x} - b\bar{r} = 0, A^T\bar{u} \leq 0, b^T\bar{u} \leq 0, (\bar{x}, \bar{r}) \geq 0, \|\bar{x}\|_1 + \bar{r} = 1,\) and \(\|\bar{u}\|_{\infty} = 1\). Therefore, \(d \in S_E, \) and the inclusion (3.31) for \(\text{cl}(F_E)\) follows.

Finally, let \(d = (A, b, c)\) be a data instance in \(\text{cl}(F_I)\), and let \(\{d_k = (A_k, b_k, c_k) : k \geq 1\}\) be a sequence of data instances in \(F_I\) such that \(d_k \to d\) as \(k \to \infty\). For each \(k \geq 1\), there exist \(v_k\) and \(y_k\) such that \(A_k v_k = 0, v_k > 0,\) and \(A_k^T y_k < c_k\). Observe that we also have \(c_k^T v_k > 0\) for all \(k \geq 1\). Let \(\bar{v}_k = v_k / \|v_k\|_1, \bar{y}_k = y_k / (\|y_k\|_{\infty} + 1),\) and \(\bar{c}_k = 1 / (\|y_k\|_{\infty} + 1)\), for each \(k \geq 1\). Then, for each \(k \geq 1\), we have \(A_k \bar{v}_k = 0, A_k^T \bar{y}_k - c_k \bar{c}_k < 0, c_k^T \bar{v}_k > 0, (\bar{v}_k, \bar{c}_k) > 0, \|\bar{y}_k\|_{\infty} + \bar{c}_k = 1,\) and \(\|\bar{v}_k\|_1 = 1\). Then, the sequence \(\{(\bar{v}_k, \bar{y}_k, \bar{c}_k) : k \geq 1\}\) has an accumulation point \((\bar{v}, \bar{y}, \bar{c})\) satisfying \(A\bar{v} = 0, A^T \bar{y} - \bar{c} \bar{T} \bar{v} \geq 0, (\bar{v}, \bar{c}) \geq 0, \|\bar{y}\|_{\infty} + \bar{c} = 1,\) and \(\|\bar{v}\|_1 = 1\). Therefore, \(d \in S_I,\) and the inclusion (3.32) for \(\text{cl}(F_I)\) follows.

q.e.d.

We now proceed to characterize the closures of \(F, F_E,\) and \(F_I,\) respectively.

**Lemma 3.4** Consider the following subsets of \(\bar{D}:\)

\[
G_1 = \{d = (A, b, c) : \text{there exists } (x, y, r, t) \text{ such that} \}
\]
\[
A x - b r = 0, A^T y - c t \leq 0, (x, r, t) \geq 0, (x, r) \neq 0, (y, t) \neq 0, r + t > 0 \}
\]

\[
G_2 = \{d = (A, b, c) : d \in I \text{ and there exists } w \text{ such that} \}
\]
\[
A w = 0, c^T w \geq 0, w \geq 0, w \neq 0 \}
\]

\[
G_3 = \{d = (A, b, c) : d \in I \text{ and there exists } \pi \text{ such that} \}
\]
\[
A^T \pi \leq 0, b^T \pi \leq 0, \pi \neq 0 \}
\]

Then

\[
\text{cl}(F) = G_1 \cup G_2 \cup G_3,
\] (3.33)
3.2 Characterizations of Ill-Posedness

\[ \text{cl}(\mathcal{F}_E) = S_E, \]  
\[ \text{cl}(\mathcal{F}_I) = S_I, \]  

where \( S_E \) and \( S_I \) are the sets defined in Proposition 3.5.

**Proof:** We first prove the identity (3.33) for \( \text{cl}(\mathcal{F}) \). Let \( d = (A, b, c) \) be a data instance in \( \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3 \). Thus, at least one of the following cases must hold:

(i) There exists \( (x, y) \) such that \( Ax = b, x \geq 0, \) and \( A^T y \leq c. \)

(ii) There exists \( (x, u) \) such that \( Ax = b, x \geq 0, A^T u \leq 0, u \neq 0. \)

(iii) There exists \( (v, y) \) such that \( Av = 0, v \geq 0, v \neq 0, \) and \( A^T y \leq c. \)

(iv) \( d \in \mathcal{I} \) and there exists \( w \) such that \( Aw = 0, c^T w \geq 0, w \geq 0, \) and \( w \neq 0. \)

(v) \( d \in \mathcal{I} \) and there exists \( \pi \) such that \( A^T \pi \leq 0, b^T \pi \leq 0, \) and \( \pi \neq 0. \)

For each case our goal is to find a data perturbation \( \Delta d(\epsilon) \) such that \( d + \Delta d(\epsilon) \in \mathcal{F} \) for all \( \epsilon > 0 \) small enough, and so \( d \in \text{cl}(\mathcal{F}) \). To do so, we list the corresponding data perturbations \( \Delta d(\epsilon) = (\Delta A(\epsilon), \Delta b(\epsilon), \Delta c(\epsilon)) \), and vectors \( x(\epsilon) \) and \( y(\epsilon) \) that make \( d + \Delta d(\epsilon) \) both primal and dual feasible with respect to the linear programming formulation in standard form. In this list, \( \bar{u}, \bar{v}, \bar{w}, \) and \( \bar{\pi} \) denote vectors such that \( \bar{u}^T u = 1, \bar{v}^T v = 1, \bar{w}^T w = 1, \) and \( \bar{\pi}^T \pi = 1, \) respectively, whenever \( u \neq 0, v \neq 0, \) \( w \neq 0, \) and \( \pi \neq 0. \) For cases (iv) and (v), \( u \) and \( v \) denote vectors such that \( Av = 0, c^T v < 0, v \geq 0, A^T u \leq 0, \) and \( b^T u > 0. \)

(i) \( \Delta d(\epsilon) = (0, 0, 0), x(\epsilon) = x, y(\epsilon) = y. \)

(ii) \( \Delta d(\epsilon) = (\epsilon \bar{u}^T c, \epsilon (c^T x) \bar{u}, 0), x(\epsilon) = x, y(\epsilon) = u/\epsilon. \)

(iii) \( \Delta d(\epsilon) = (\epsilon \bar{v}^T, 0, \epsilon (b^T y) \bar{v}), x(\epsilon) = v/\epsilon, y(\epsilon) = y. \)

(iv) \( \Delta d(\epsilon) = (\epsilon b (\epsilon + \epsilon \bar{w})^T, 0, \epsilon \bar{w}), x(\epsilon) = w/\epsilon (\epsilon + c^T w), y(\epsilon) = u/(\epsilon b^T u). \)
3.2 Characterizations of Ill-Posedness

\(\Delta d(\epsilon) = (-\epsilon(b - \epsilon \bar{\pi})c^T, -\epsilon \bar{\pi}, 0), x(\epsilon) = v/(-\epsilon c^Tv), y(\epsilon) = \pi/\epsilon(e - b^T\pi)\).

Therefore, \(G_1 \cup G_2 \cup G_3 \subset \text{cl}(\mathcal{F})\).

On the other hand, let \(d = (A, b, c) \in \text{cl}(\mathcal{F})\). From Proposition 3.5, identity (3.30), it follows that \(d \in \mathcal{S}\), and so there exists \((\bar{x}, \bar{y}, \bar{\pi}, \bar{t})\) such that \(A\bar{x} - b\bar{t} = 0, A^T\bar{y} - c\bar{t} \leq 0, (\bar{x}, \bar{\pi}, \bar{t}) \geq 0, (\bar{x}, \bar{\pi}) \neq 0, (\bar{y}, \bar{t}) \neq 0\). First consider the case when \(d \notin \mathcal{I}\), then from Proposition 2.11 there exists \((\hat{x}, \hat{y}, \hat{\pi}, \hat{t})\) such that \(A\hat{x} - b\hat{t} = 0, A^T\hat{y} - c\hat{t} \leq 0, (\hat{x}, \hat{\pi}, \hat{t}) \geq 0, \) and \(\hat{\pi} + \hat{t} > 0\). Therefore, by taking \((x, y, r, t) = (\bar{x}, \bar{y}, \bar{\pi}, \bar{t}) + \lambda(\hat{x}, \hat{y}, \hat{\pi}, \hat{t})\), there exists a \(\lambda > 0\) such that \(d\) satisfies the systems defined in \(G_1\), and so \(d \in G_1\).

Next consider the case when \(d \in \mathcal{I}\) and \(\text{rank}(A) = m\). Assume that the system \(Ax = -b, x > 0,\) and \(A^Ty < -c\) is feasible. We can choose \(\epsilon > 0\) small enough so that for all \(\Delta d = (\Delta A, \Delta b, \Delta c)\), \(||\Delta d|| < \epsilon\) implies that the system \((A + \Delta A)x = -(b + \Delta b), x > 0,\) and \((A + \Delta A)^Ty < -(c + \Delta c)\) remains feasible (see Propositions 2.13 and 2.14).

In fact, let \((\bar{x}, \bar{y})\) be a feasible solution to this system. If \(d + \Delta d\) is dual feasible, then there exists \(\hat{y}\) such that \((A + \Delta A)^T\hat{y} \leq c + \Delta c\). Let \(u = \bar{y} + \hat{y}\), then \((A + \Delta A)^Tu < 0\) and \((b + \Delta b)^Tu = -\bar{x}^T(A + \Delta A)^Tu > 0\), so that \(d + \Delta d\) is primal infeasible. Hence, \(d + \Delta d \in \mathcal{F}^C\) for all \(||\Delta d|| < \epsilon\). This implies that \(d \notin \text{cl}(\mathcal{F})\), a contradiction.

Therefore, the system \(Ax = -b, x > 0,\) and \(A^Ty < -c\) is not feasible, and so, by Corollary 2.2 and Corollary 2.3, \(d \in G_2 \cup G_3\).

To finish the proof of identity (3.33), consider the case when \(d \in \mathcal{I}\) and \(\text{rank}(A) < m\). Then, there exists a vector \(\pi\) such that \(A^T\pi = 0, b^T\pi \leq 0,\) and \(\pi \neq 0\). Therefore, \(d \in G_3\). By combining this case with the cases discussed in the previous paragraphs, we conclude that \(\text{cl}(\mathcal{F}) \subset G_1 \cup G_2 \cup G_3\) and the result follows.

In light of Proposition 3.5, identity (3.31), in order to prove identity (3.34), we only need to prove that \(S_E \subset \text{cl}(\mathcal{F}_E)\). Let \(d = (A, b, c)\) be a data instance in \(S_E\). At least one of the following cases must hold: there exists \((u, x)\) such that \(Ax = b, x \geq 0, A^Tu \leq 0, b^Tu \leq 0,\) and \(u \neq 0\); or there exists \((u, v)\) such that \(Au = 0, v \geq 0, v \neq 0, A^Tu \leq 0, b^Tu \leq 0,\) and \(u \neq 0\). In the first case, we readily obtain that \(d \in \mathcal{F}_E \subset \text{cl}(\mathcal{F}_E)\).
$\text{cl}(\tilde{\mathcal{F}}_E) = \text{cl}(\mathcal{F}_E)$, by Proposition 3.4, identity (3.21). In the second case, consider the data perturbation $\Delta d(\epsilon) = (\Delta A(\epsilon), \Delta b(\epsilon), \Delta c(\epsilon))$ given by $\Delta d(\epsilon) = (\epsilon b \epsilon^T, 0, 0)$, and the vectors $x(\epsilon) = v/(\epsilon||v||_1)$, $y(\epsilon) = u$. Then $(A + \Delta A(\epsilon))x(\epsilon) = b + \Delta b(\epsilon)$, $x(\epsilon) \geq 0$, $(A + \Delta A(\epsilon))^T y(\epsilon) \leq 0$, so that $d + \Delta d(\epsilon) \in \tilde{\mathcal{F}}_E$, for all $\epsilon > 0$. Therefore, $d \in \text{cl}(\tilde{\mathcal{F}}_E) = \text{cl}(\mathcal{F}_E)$, from Proposition 3.4, identity (3.21), and so $\mathcal{S}_E \subset \text{cl}(\mathcal{F}_E)$.

Finally, using Proposition 3.5, identity (3.32), we only need to show that $\mathcal{S}_I \subset \text{cl}(\mathcal{F}_I)$ to prove identity (3.35). Let $d = (A, b, c)$ be a data instance in $\mathcal{S}_I$. At least one of the following cases must hold: there exists $(y, v)$ such that $Av = 0$, $c^Tv \geq 0$, $v \geq 0$, $v \neq 0$, and $A^T y \leq c$; or there exists $(u, v)$ such that $Av = 0$, $c^Tv \geq 0$, $v \geq 0$, $v \neq 0$, $A^T u \leq 0$, and $u \neq 0$. In the first case, we readily obtain that $d \in \tilde{\mathcal{F}}_I \subset \text{cl}(\tilde{\mathcal{F}}_I) = \text{cl}(\mathcal{F}_I)$, by Proposition 3.4, identity (3.23). In the second case, without loss of generality we may also assume that the system $A^T y \leq c$ is not feasible (otherwise again $d \in \tilde{\mathcal{F}}_I \subset \text{cl}(\mathcal{F}_I)$ and the result follows). Hence, assume that there exists a vector $w$ such that $Aw = 0$, $c^Tw < 0$, $w \geq 0$. Let $\hat{u}$ be such that $\hat{u}^Tw = 1$ and consider the data perturbation $\Delta d(\epsilon) = (\Delta A(\epsilon), \Delta b(\epsilon), \Delta c(\epsilon))$ given by $\Delta d(\epsilon) = (\epsilon \hat{u}c^T, 0, 0)$, and the vectors $x(\epsilon) = v - (c^Tv/c^Tw)w$, $y(\epsilon) = u/\epsilon$. Then $(A + \Delta A(\epsilon))x(\epsilon) = 0$, $x(\epsilon) \geq 0$, $(A + \Delta A(\epsilon))^T y(\epsilon) \leq c + \Delta c(\epsilon)$, so that $d + \Delta d(\epsilon) \in \tilde{\mathcal{F}}_I$, for all $\epsilon > 0$. Therefore, $d \in \text{cl}(\tilde{\mathcal{F}}_I) = \text{cl}(\mathcal{F}_I)$, from Proposition 3.4, identity (3.23), and so $\mathcal{S}_I \subset \text{cl}(\mathcal{F}_I)$.

\textbf{q.e.d.}

Observe that $\mathcal{G}_1 = (\mathcal{P} \cap \mathcal{D}) \cup (\mathcal{P} \cap \mathcal{P}_0^c) \cup (\mathcal{D} \cap \mathcal{D}_0^c)$, $\mathcal{G}_2 = \mathcal{P}^c \cap \mathcal{D}^c \cap \mathcal{D}^c$, and $\mathcal{G}_3 = \mathcal{P}^c \cap \mathcal{D}^c \cap \mathcal{P}^c$, from which identity (3.2) in Theorem 3.1 follows. Similarly, notice that $\mathcal{S}_E = (\mathcal{P} \cap \mathcal{P}^c) \cup (\mathcal{D}_0^c \cap \mathcal{P}^c)$ and $\mathcal{S}_I = (\mathcal{D} \cap \mathcal{D}^c) \cup (\mathcal{P}_0^c \cap \mathcal{D}^c)$, from which identity (3.7) in Theorem 3.2 and identity (3.12) in Theorem 3.3 follow.

As a consequence of the previous lemma, we obtain a characterization of the
interiors of $\mathcal{F}^C$, $\mathcal{F}_E^C$ and $\mathcal{F}_I^C$, respectively, in the next lemma.

**Lemma 3.5** Consider the following subsets of $\bar{\mathcal{D}}$:

\[
\mathcal{H}_1 = \{ d = (A, b, c) : d \in \mathcal{I}, \text{rank}(A) = m, \text{ and there is } (x, y) \text{ such that } Ax = -b, x > 0, A^T y < -c \},
\]

\[
\mathcal{H}_2 = \{ d = (A, b, c) : \text{there is } u \text{ such that } A^T u < 0, b^T u > 0 \},
\]

\[
\mathcal{H}_3 = \{ d = (A, b, c) : \text{rank}(A) = m \text{ and there is } v \text{ such that } Av = 0, c^T v < 0, v > 0 \},
\]

Then

\[
\text{int}(\mathcal{F}^C) = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3, \quad (3.36)
\]

\[
\text{int}(\mathcal{F}_E^C) = \mathcal{P}_- \cup \mathcal{H}_2, \quad (3.37)
\]

\[
\text{int}(\mathcal{F}_I^C) = \mathcal{D}_- \cup \mathcal{H}_3. \quad (3.38)
\]

**Proof:** Observe that from Lemma 3.4, identity (3.33), we have $\text{int}(\mathcal{F}^C) = (\text{cl}(\mathcal{F}))^C = \mathcal{G}_1^C \cap \mathcal{G}_2^C \cap \mathcal{G}_3^C$. From Corollary 2.2, it follows that $\mathcal{G}_3^C = \mathcal{I}^C \cup \mathcal{P}_-$. Similarly, from Corollary 2.3, it follows that $\mathcal{G}_2^C = \mathcal{I}^C \cup \mathcal{D}_-$. Therefore, from elementary set algebra we have

\[
\mathcal{G}_2^C \cap \mathcal{G}_3^C = (\mathcal{I}^C \cup \mathcal{D}_-) \cap (\mathcal{I}^C \cup \mathcal{P}_-)
\]

\[
= \mathcal{I}^C \cup (\mathcal{P}_- \cap \mathcal{D}_-)
\]

\[
= \mathcal{I}^C \cup (\mathcal{P}_- \cap \mathcal{D}_- \cap \mathcal{I})
\]

\[
= \mathcal{I}^C \cup \mathcal{H}_1.
\]

Consider $d \in \mathcal{H}_2 \cup \mathcal{H}_3 \cup \mathcal{I}$. From Proposition 2.9, if $d \in \mathcal{H}_2$, then $d \in \mathcal{G}_1^C$. From Proposition 2.10, if $d \in \mathcal{H}_3$, then $d \in \mathcal{G}_1^C$. Moreover, if $d \in \mathcal{I}$, then $d$ is neither primal nor dual feasible, hence again $d \in \mathcal{G}_1^C$. Thus, $\mathcal{H}_2 \cup \mathcal{H}_3 \cup \mathcal{I} \subset \mathcal{G}_1^C$. On the other hand,
consider \( d \in \mathcal{G}_1^C \). If \( d \notin \mathcal{H}_2 \cup \mathcal{H}_3 \cup \mathcal{I} \), then, from Proposition 2.9, Proposition 2.10, and Proposition 2.11, there exist \((\bar{x}, \bar{y}, \bar{r}, \bar{t})\) and \((\check{x}, \check{y}, \check{r}, \check{t})\) such that \( A\bar{x} - b\bar{r} = 0, (\bar{x}, \bar{r}) \geq 0, (\bar{x}, \bar{r}) \neq 0, A^T\bar{y} - c\bar{t} \leq 0, \bar{t} \geq 0, (\bar{y}, \bar{t}) \neq 0, A\check{x} - b\check{r} = 0, A^T\check{y} - c\check{t} \leq 0, (\check{x}, \check{r}, \check{t}) \geq 0, \) and \( \check{r} + \check{t} > 0 \). Hence, by taking \((x, y, r, t) = (\bar{x}, \bar{y}, \bar{r}, \bar{t}) + \lambda(\check{x}, \check{y}, \check{r}, \check{t})\) we obtain that there exists \( \lambda > 0 \) such that \( d \) satisfies the systems defined in \( \mathcal{G}_1 \), and so \( d \in \mathcal{G}_1 \), a contradiction. Therefore, \( \mathcal{G}_1^C = \mathcal{H}_2 \cup \mathcal{H}_3 \cup \mathcal{I} \), and so

\[
\mathcal{G}_1^C = \mathcal{H}_2 \cup \mathcal{H}_3 \cup \mathcal{I}.
\]

Again by elementary set algebra we have

\[
\mathcal{G}_1^C \cap \mathcal{G}_2^C \cap \mathcal{G}_3^C = \left( \mathcal{H}_2 \cup \mathcal{H}_3 \cup \mathcal{I} \right) \cap \left( \mathcal{I}^C \cup \mathcal{H}_1 \right) = \left( \left( \mathcal{H}_2 \cup \mathcal{H}_3 \right) \cap \mathcal{I}^C \cup \mathcal{I} \right) \cap \left( \mathcal{I}^C \cup \mathcal{H}_1 \right) = \mathcal{H}_1 \cup \left( \left( \mathcal{H}_2 \cup \mathcal{H}_3 \right) \cap \mathcal{I}^C \right) = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3,
\]

where the last equality follows from observing that \( \mathcal{H}_2 \cup \mathcal{H}_3 \subset \mathcal{I}^C \). Therefore, identity (3.36) follows.

Identity (3.37) follows from Lemma 3.4, identity (3.34), since \( \text{int}(\mathcal{F}_E^C) = (\text{cl}(\mathcal{F}_E))^C = \mathcal{S}_E^C \). Hence, \( d = (A, b, c) \in \text{int}(\mathcal{F}_E^C) \) if and only if the system \( Ax - br = 0, (x, r) \geq 0, \) and \( (x, r) \neq 0 \) does not have a solution or the system \( A^Tu \leq 0, b^Tu \leq 0, \) and \( u \neq 0 \) does not have a solution. From this, Proposition 2.9 and Corollary 2.2, the result follows.

Finally, identity (3.38) follows from Lemma 3.4, identity (3.35), since \( \text{int}(\mathcal{F}_I^C) = (\text{cl}(\mathcal{F}_I))^C = \mathcal{S}_I^C \). Hence, \( d = (A, b, c) \in \text{int}(\mathcal{F}_I^C) \) if and only if the system \( A^Ty - ct = 0, y \geq 0, \) and \( (y, t) \neq 0 \) does not have a solution or the system \( Av = 0, c^Tv \geq 0, v \geq 0, \) and \( v \neq 0 \) does not have a solution. From this, Proposition 2.10 and Corollary 2.3, the result follows.
3.2 Characterizations of Ill-Posedness

q.e.d.

Notice that $\mathcal{H}_1 = \mathcal{P}^C \cap \mathcal{D}^C \cap \mathcal{P}_- \cap \mathcal{D}_-$, $\mathcal{H}_2 = \mathcal{P}^C \cap D_0$, and $\mathcal{H}_3 = \mathcal{D}^C \cap \mathcal{P}_0$, so that we readily obtain identity (3.3) in Theorem 3.1, identity (3.8) in Theorem 3.2, and identity (3.13) in Theorem 3.3.

In the following lemma we characterize the set of ill-posed data instances for a linear program in standard form, that is, $B = \partial \mathcal{F}$.

**Lemma 3.6** Consider the following subsets of $\hat{D}$:

\[
\begin{align*}
B_1 &= \{ d = (A, b, c) : \text{there exists } (x, y, r, t, u) \text{ such that} \} \\
&\quad Ax - br = 0, A^T y - ct \leq 0, (x, r, t) \geq 0, (x, r) \neq 0, (y, t) \neq 0, r + t > 0, \\
&\quad A^T u \leq 0, b^T u \geq 0, u \neq 0 \}, \\
B_2 &= \{ d = (A, b, c) : \text{there exists } (x, y, r, t, v) \text{ such that} \} \\
&\quad Ax - br = 0, A^T y - ct \leq 0, (x, r, t) \geq 0, (x, r) \neq 0, (y, t) \neq 0, r + t > 0, \\
&\quad Av = 0, c^T v \leq 0, v \geq 0, v \neq 0 \}.
\]

Then,

\[ B = \partial \mathcal{F} = B_1 \cup B_2 \cup \mathcal{G}_2 \cup \mathcal{G}_3, \]

where $\mathcal{G}_2$ and $\mathcal{G}_3$ are the data instance sets defined in Lemma 3.4.

**Proof:** Observe that $B_1 = \mathcal{G}_1 \cap \mathcal{P}_+^C$ and $B_2 = \mathcal{G}_1 \cap \mathcal{D}_+^C$. Since $\mathcal{G}_2 \cup \mathcal{G}_3 \subset \mathcal{I} \subset \mathcal{F}_- \subset cl(\mathcal{F}^C)$, then $(\mathcal{G}_2 \cup \mathcal{G}_3) \cap cl(\mathcal{F}^C) = \mathcal{G}_2 \cup \mathcal{G}_3$. The proof follows from elementary set algebra:

\[ B = cl(\mathcal{F}) \cap cl(\mathcal{F}^C) \]
\[ = (G_1 \cup G_2 \cup G_3) \cap cl(\mathcal{F}^C) \]
\[ = (G_1 \cap cl(\mathcal{F}^C)) \cup ((G_2 \cup G_3) \cap cl(\mathcal{F}^C)) \]
\[ = (G_1 \cap cl(\mathcal{F}^C)) \cup (G_2 \cup G_3) \]
\[ = (G_1 \cap (P_+^C \cup D_+^C)) \cup (G_2 \cup G_3) \]
\[ = (G_1 \cap P_+^C) \cup (G_1 \cap D_+^C) \cup (G_2 \cup G_3) \]
\[ = B_1 \cup B_2 \cup G_2 \cup G_3, \]

where we are using that \( cl(\mathcal{F}^C) = P_+^C \cup D_+^C \) from Lemma 3.3, identity (3.27).

q.e.d.

Observe that \( B_1 = G_1 \cap P_+^C \) and \( B_2 = G_1 \cap D_+^C \), so that \( B_1 \cup B_2 = G_1 \cap (P_+^C \cup D_+^C) \).

Therefore, we obtain identity (3.5) in Theorem 3.1.

Finally, we present a characterization of the sets of ill-posed data instances for the analytic center problems in equality and inequality form, respectively.

**Lemma 3.7** Consider the following subsets of \( \bar{D} \):

\[ J_1 = \{ d = (A, b, c) : \text{there exists } (x, u, \pi, r) \text{ such that } Ax - br = 0, \]
\[ (x, r) \geq 0, (x, r) \neq 0, A^T u \leq 0, b^T u \leq 0, u \neq 0, \]
\[ A^T \pi \leq 0, b^T \pi \geq 0, \pi \neq 0 \}, \]

\[ J_2 = \{ d = (A, b, c) : \text{there exists } (x, u, v, r) \text{ such that } Ax - br = 0, \]
\[ (x, r) \geq 0, (x, r) \neq 0, A^T u \leq 0, b^T u \leq 0, u \neq 0, \]
\[ Av = 0, v \geq 0, v \neq 0 \}, \]

\[ J_3 = \{ d = (A, b, c) : \text{there exists } (y, u, v, t) \text{ such that } A^T y - ct = 0, \]
\[ t \geq 0, (y, t) \neq 0, Av = 0, c^T v \geq 0, v \geq 0, v \neq 0, \]
\[ A^T u \leq 0, u \neq 0 \}, \]
\( \mathcal{J}_4 = \{d = (A, b, c) : \text{there exists } (y, v, w, t) \text{ such that } A^T y - ct = 0, \]
\[ t \geq 0, (y, t) \neq 0, Av = 0, c^T v \geq 0, v \geq 0, v \neq 0, \]
\[ Aw = 0, w \geq 0, c^T w \leq 0, w \neq 0 \} . \]

Then,

\[
\mathcal{B}_E = \mathcal{J}_1 \cup \mathcal{J}_2, \quad (3.39) \\
\mathcal{B}_I = \mathcal{J}_3 \cup \mathcal{J}_4. \quad (3.40)
\]

**Proof:** From Lemma 3.3, identity (3.28), and Lemma 3.4, identity (3.34), it follows that \( \mathcal{B}_E = \text{cl}(\mathcal{F}_E) \cap \text{cl}(\mathcal{F}_E^C) = S_E \cap (\mathcal{P}_+^C \cup \mathcal{D}_0^C) = (S_E \cap \mathcal{P}_+^C) \cup (S_E \cap \mathcal{D}_0^C) = \mathcal{J}_1 \cup \mathcal{J}_2, \) and so we obtain identity (3.39).

Analogously, from Lemma 3.3, identity (3.29), and Lemma 3.4, identity (3.35), it follows that \( \mathcal{B}_I = \text{cl}(\mathcal{F}_I) \cap \text{cl}(\mathcal{F}_I^C) = S_I \cap (\mathcal{D}_+^C \cup \mathcal{P}_0^C) = (S_I \cap \mathcal{D}_+^C) \cup (S_I \cap \mathcal{P}_0^C) = \mathcal{J}_4 \cup \mathcal{J}_3, \) and so we obtain identity (3.40).

q.e.d.

Lastly, since \( \mathcal{J}_1 = (\mathcal{P} \cup \mathcal{D}_0^C) \cap \mathcal{P}_+^C \cap \mathcal{P}_-^C, \mathcal{J}_2 = (\mathcal{P} \cup \mathcal{D}_0^C) \cap \mathcal{P}_-^C \cap \mathcal{D}_+^C, \) and \( \mathcal{P} \cap \mathcal{P}_-^C \subset \mathcal{P}_-^C, \) we obtain that

\[
\mathcal{B}_E = (\mathcal{P} \cup \mathcal{D}_0^C) \cap (\mathcal{P}_+^C \cup \mathcal{D}_0^C) \cap \mathcal{P}_-^C \\
= ((\mathcal{P} \cap \mathcal{P}_+^C) \cup \mathcal{D}_0^C) \cap \mathcal{P}_-^C \\
= (\mathcal{P} \cap \mathcal{P}_+^C \cap \mathcal{P}_-^C) \cup (\mathcal{D}_0^C \cap \mathcal{P}_-^C) \\
= (\mathcal{P} \cap \mathcal{P}_+^C) \cup (\mathcal{D}_0^C \cap \mathcal{P}_-^C).
\]

thus proving identity (3.10) in Theorem 3.2. Analogously, since \( \mathcal{J}_3 = (\mathcal{D} \cup \mathcal{P}_0^C) \cap \)
\[ \mathcal{D}^C \cap \mathcal{P}_0^C, \mathcal{J}_4 = (\mathcal{D} \cup \mathcal{P}_0^C) \cap \mathcal{D}^C \cap \mathcal{D}_4^C, \text{ and } \mathcal{D} \cap \mathcal{D}_4^C \subset \mathcal{D}^C, \text{ we obtain that} \]

\[
\begin{align*}
\mathcal{B}_i &= (\mathcal{D} \cup \mathcal{P}_0^C) \cap (\mathcal{D}_+^C \cup \mathcal{P}_0^C) \cap \mathcal{D}^C \\
&= ((\mathcal{D} \cap \mathcal{D}_+^C) \cup \mathcal{P}_0^C) \cap \mathcal{D}^C \\
&= (\mathcal{D} \cap \mathcal{D}_+^C \cap \mathcal{D}^C) \cup (\mathcal{P}_0^C \cap \mathcal{D}^C) \\
&= (\mathcal{D} \cap \mathcal{D}_+^C) \cup (\mathcal{P}_0^C \cap \mathcal{D}^C),
\end{align*}
\]

thus proving identity (3.15) in Theorem 3.3.

### 3.3 Further Properties

In this section we discuss how the distance to ill-posedness \( \rho(d) \) of a data instance \( d = (A, b, c) \) relates to the original distances to ill-posedness introduced by Renegar in [Ren94, Ren95a]. These distances are denoted by \( \rho_P(d) \) and \( \rho_D(d) \), respectively. When \( d \in \mathcal{F} \) the distance \( \rho_P(d) \) represents the minimum change \( \|\Delta d\| \) necessary to produce an infeasible data instance \( d + \Delta d \) with respect to the linear program \( F(\cdot) \). Similarly, when \( d \in \mathcal{F} \) the distance \( \rho_D(d) \) represents the minimum change \( \|\Delta d\| \) necessary to produce an infeasible data instance \( d + \Delta d \) with respect to the linear program \( D(\cdot) \).

We first introduce some notation that is used throughout this section. Recall that \( \mathcal{P} \) is the set of primal feasible data instances and \( \mathcal{D} \) is the set of dual feasible data instances, both with respect to a linear program in standard form. Observe that \( \mathcal{F} \), which is the set of instances for the linear programming problem in standard form \( P(d) \) (and its dual \( D(d) \)) that have optimal solutions, is characterized by \( \mathcal{F} = \mathcal{P} \cap \mathcal{D} \). Also, notice that \( \mathcal{F}^C = \mathcal{P}^C \cup \mathcal{D}^C \), an identity that we use in the proof of Lemma 3.10. It is also convenient to introduce the corresponding sets of ill-posed data instances:

\[ \mathcal{B}_P = \partial \mathcal{P} \text{ and } \mathcal{B}_D = \partial \mathcal{D}. \]
3.3 Further Properties

Similarly as we did in Chapter 2, we define the corresponding distances to ill-posedness of a data instance \( d = (A, b, c) \) with respect to the boundaries of the sets \( \mathcal{P} \) and \( \mathcal{D} \):

\[
\rho_P(d) = \inf \{ \| \Delta d \| : d + \Delta d \in \mathcal{B}_P \},
\]
\[
\rho_D(d) = \inf \{ \| \Delta d \| : d + \Delta d \in \mathcal{B}_D \}.
\]

In other words, \( \rho_P(d) \) and \( \rho_D(d) \) denote the distance to primal ill-posedness and the distance to dual ill-posedness of the data instance \( d \), respectively.

We also have alternative definitions of \( \rho_P(d) \) and \( \rho_D(d) \) analogous to the one given in definition (2.5):

\[
\rho_P(d) = \begin{cases} 
\sup \{ \delta : B(d, \delta) \subset \mathcal{P} \} & \text{if } d \in \mathcal{P}, \\
\sup \{ \delta : B(d, \delta) \subset \mathcal{P}^c \} & \text{if } d \in \mathcal{P}^c.
\end{cases} 
\]  

(3.41)

\[
\rho_D(d) = \begin{cases} 
\sup \{ \delta : B(d, \delta) \subset \mathcal{D} \} & \text{if } d \in \mathcal{D}, \\
\sup \{ \delta : B(d, \delta) \subset \mathcal{D}^c \} & \text{if } d \in \mathcal{D}^c.
\end{cases} 
\]  

(3.42)

Likewise, the corresponding condition measures for the primal problem and for the dual problem are

\[
C_P(d) = \| d \| / \rho_P(d) \quad \text{if } \rho_P(d) > 0 \quad \text{and} \quad C_P(d) = \infty, \text{ otherwise;}
\]

\[
C_D(d) = \| d \| / \rho_D(d) \quad \text{if } \rho_D(d) > 0 \quad \text{and} \quad C_D(d) = \infty, \text{ otherwise.}
\]

The following lemma describes the closure of various data instance sets.

**Lemma 3.8** The data instance sets \( \text{cl}(\mathcal{P}) \), \( \text{cl}(\mathcal{P}^c) \), \( \text{cl}(\mathcal{D}) \), and \( \text{cl}(\mathcal{D}^c) \) are character-
ized as follows:

\[ cl(\mathcal{P}) = \{(A, b, c) : \text{there exist } x \in \mathbb{R}^n \text{ and } r \in \mathbb{R} \text{ such that } Ax - br = 0, x \geq 0, r \geq 0, (x, r) \neq 0\}, \]

\[ cl(\mathcal{P}^C) = \{(A, b, c) : \text{there exists } u \in \mathbb{R}^m \text{ such that } A^Tu \leq 0, b^Tu \geq 0, u \neq 0\}, \]

\[ cl(\mathcal{D}) = \{(A, b, c) : \text{there exist } y \in \mathbb{R}^m \text{ and } t \in \mathbb{R} \text{ such that } A^Ty - ct \leq 0, t \geq 0, (y, t) \neq 0\}, \]

\[ cl(\mathcal{D}^C) = \{(A, b, c) : \text{there exists } v \in \mathbb{R}^n \text{ such that } Av = 0, c^Tv \leq 0, v \geq 0, v \neq 0\}. \]

**Proof:** Let \( d = (A, b, c) \in cl(\mathcal{P}) \), then there exists a sequence \( \{d_h = (A_h, b_h, c_h) : h \geq 1\} \) such that \( d_h \in \mathcal{P} \) for all \( h \geq 1 \), and \( d_h \to d \) as \( h \to \infty \). For each \( h \geq 1 \), there exists \( x_h \) such that \( A_h x_h = b_h \), and \( x_h \geq 0 \). Consider the sequence \( \{(\hat{x}_h, \hat{r}_h) : h \geq 1\} \), where \( \hat{x}_h = x_h/(\|x_h\|_1 + 1) \) and \( \hat{r}_h = 1/(\|x_h\|_1 + 1) \) for all \( h \geq 1 \). Observe that for each \( h \geq 1 \), \( A_h \hat{x}_h - b_h \hat{r}_h = 0, \|\hat{x}_h\|_1 + |\hat{r}_h| = 1 \), and \( (\hat{x}_h, \hat{r}_h) \geq 0 \). Hence, the sequence \( \{(\hat{x}_h, \hat{r}_h) : h \geq 1\} \) has an accumulation point \( (\hat{x}, \hat{r}) \in \mathbb{R}^{n+1} \) satisfying \( A\hat{x} - b\hat{r} = 0, (\hat{x}, \hat{r}) \geq 0, \) and \( (\hat{x}, \hat{r}) \neq 0 \). On the other hand, for a given data instance \( d = (A, b, c) \) assume that there exist \( x \) and \( r \) such that \( Ax - br = 0, (x, r) \geq 0, \) and \( (x, r) \neq 0 \). If \( r > 0 \), then we readily obtain that \( d \in \mathcal{P} \subset cl(\mathcal{P}) \). If \( r = 0 \), then consider \( \Delta d(\epsilon) = (\Delta A(\epsilon), 0, 0) \), where \( \Delta A(\epsilon) = \epsilon b e^T/\|x\|_1 \). Let \( x(\epsilon) = x/\epsilon \). Then we obtain that \( (A + \Delta A(\epsilon))x(\epsilon) = b \) and \( x(\epsilon) \geq 0 \), so that \( d + \Delta d(\epsilon) \in \mathcal{P} \) for all \( \epsilon > 0 \). Therefore, since \( \Delta d(\epsilon) \to 0 \) as \( \epsilon \to 0 \), we have \( d \in cl(\mathcal{P}) \). This concludes the proof of the characterization of \( cl(\mathcal{P}) \).

Similarly, for a given data instance \( d = (A, b, c) \in cl(\mathcal{P}^C) \), there exists a sequence
\{d_h = (A_h, b_h, c_h) : h \geq 1\}, such that \(d_h \in \mathcal{P}^C\) for all \(h \geq 1\), and \(d_h \to d\) as \(h \to \infty\).

For each \(h \geq 1\), we have from Proposition 2.3 that there exists \(u_h\) such that \(A_h^T u_h \leq 0\) and \(b_h^T u_h > 0\). Consider the sequence \(\{\hat{u}_h : h \geq 1\}\), where \(\hat{u}_h = u_h/\|u_h\|_\infty\) for all \(h \geq 1\). Observe that for each \(h \geq 1\), \(A_h^T \hat{u}_h \leq 0, b_h^T \hat{u}_h > 0,\) and \(\|\hat{u}_h\|_\infty = 1\). Hence, the sequence \(\{\hat{u}_h : h \geq 1\}\) has an accumulation point \(\hat{u} \in \mathbb{R}^m\) satisfying \(A^T \hat{u} \leq 0, b^T \hat{u} \geq 0,\) and \(\hat{u} \neq 0\). On the other hand, suppose that for a given data instance \(d = (A, b, c)\) there exists \(u\) such that \(A^T u \leq 0, b^T u \geq 0,\) and \(u \neq 0\). Let \(\tilde{u}\) be such that \(\tilde{u}^T u = 1\). For a given \(\epsilon > 0\) let \(\Delta b(\epsilon) = \epsilon \tilde{u}\), then since \(A^T u \leq 0\) and \((b + \Delta b(\epsilon))^T u > 0\), it follows from Proposition 2.3 that \(d + \Delta d(\epsilon) \in \mathcal{P}^C\) for all \(\epsilon > 0\), where \(\Delta d(\epsilon) = (0, \Delta b(\epsilon), 0)\). Since \(\Delta d(\epsilon) \to 0\) as \(\epsilon \to 0\), it follows that \(d \in cl(\mathcal{P}^C)\).

This concludes the proof of the characterization of \(cl(\mathcal{P}^C)\).

Next, let \(d = (A, b, c) \in cl(D)\), then there exists a sequence \(\{d_h = (A_h, b_h, c_h) : h \geq 1\}\) such that \(d_h \in D\) for all \(h \geq 1\), and \(d_h \to d\) as \(h \to \infty\). For each \(h \geq 1\), there exists \(y_h\) such that \(A_h^T y_h \leq c_h\). Consider the sequence \(\{\hat{y}_h, \hat{t}_h : h \geq 1\}\), where \(\hat{y}_h = y_h/(\|y_h\|_\infty + 1)\) and \(\hat{t}_h = 1/(\|y_h\|_\infty + 1)\) for all \(h \geq 1\). Observe that for each \(h \geq 1\), \(A_h^T \hat{y}_h - c_h \hat{t}_h \leq 0,\) \(\|\hat{y}_h\|_\infty + |\hat{t}_h| = 1,\) and \(\hat{t}_h \geq 0\). Hence, the sequence \(\{\hat{y}_h, \hat{t}_h : h \geq 1\}\) has an accumulation point \((\hat{y}, \hat{t}) \in \mathbb{R}^{m+1}\) satisfying \(A^T \hat{y} - c \hat{t} \leq 0,\) \(\hat{t} \geq 0,\) and \((\hat{y}, \hat{t}) \neq 0\). On the other hand, for a given data instance \(d = (A, b, c)\) assume that there exist \(y\) and \(t\) such that \(A^T y - ct \leq 0,\) \(t \geq 0,\) and \((y, t) \neq 0\). If \(t > 0,\) then we readily obtain that \(d \in D \subseteq cl(D)\). If \(t = 0,\) let \(\tilde{y}\) be such that \(\tilde{y}^T y = 1\) and consider \(\Delta d(\epsilon) = (\Delta A(\epsilon), 0, 0),\) where \(\Delta A(\epsilon) = \epsilon c \tilde{y}^T\). Let \(y(\epsilon) = y/\epsilon\). Then we obtain that \((A + \Delta A(\epsilon))^T y(\epsilon) \leq c,\) so that \(d + \Delta d(\epsilon) \in D\) for all \(\epsilon > 0\).

Therefore, since \(\Delta d(\epsilon) \to 0\) as \(\epsilon \to 0\), we have \(d \in cl(D)\). This concludes the proof of the characterization of \(cl(D)\).

Finally, for a given data instance \(d = (A, b, c) \in cl(D^C)\), there exists a sequence \(\{d_h = (A_h, b_h, c_h) : h \geq 1\}\), such that \(d_h \in D^C\) for all \(h \geq 1,\) and \(d_h \to d\) as \(h \to \infty\).

For each \(h \geq 1,\) we have from Proposition 2.4 that there exists \(v_h\) such that \(A_h v_h = 0,\)
3.3 Further Properties

$v_h \geq 0$, and $c^T v_h < 0$. Consider the sequence $\{\hat{v}_h : h \geq 1\}$, where $\hat{v}_h = v_h/\|v_h\|_1$ for all $h \geq 1$. Observe that for each $h \geq 1$, $A_h\hat{v}_h = 0$, $\hat{v}_h \geq 0$, $c^T \hat{v}_h < 0$, and $\|\hat{v}_h\|_1 = 1$. Hence, the sequence $\{\hat{v}_h : h \geq 1\}$ has an accumulation point $\hat{v} \in \mathbb{R}^n$ satisfying $A\hat{v} = 0$, $\hat{v} \geq 0$, $c^T \hat{v} \leq 0$, and $\hat{v} \neq 0$. On the other hand, suppose that for a given data instance $d = (A, b, c)$ there exists $v$ such that $Av = 0$, $v \geq 0$, $c^T v \leq 0$, and $v \neq 0$. For a given $\epsilon > 0$ let $\Delta c(\epsilon) = -\epsilon e$, then since $Av = 0$, $v \geq 0$, and $(c + \Delta c(\epsilon))^T v < 0$, it follows from Proposition 2.4 that $d + \Delta d(\epsilon) \in D^C$ for all $\epsilon > 0$, where $\Delta d(\epsilon) = (0, 0, \Delta c(\epsilon))$. Since $\Delta d(\epsilon) \to 0$ as $\epsilon \to 0$, it follows that $d \in cl(D^C)$. This concludes the proof of the characterization of $cl(D^C)$.

q.e.d.

As an immediate consequence of Lemma 3.8 we obtain:

**Corollary 3.2** The boundaries of $P$ and $D$ are given by

$$B_P = \{(A, b, c) : \text{there exist } x \in \mathbb{R}^n, r \in \mathbb{R}, \text{ and } u \in \mathbb{R}^m \text{ such that } Ax - br = 0, (x, r) \geq 0, (x, r) \neq 0, A^T u \leq 0, b^T u \geq 0, u \neq 0\},$$

$$B_D = \{(A, b, c) : \text{there exist } y \in \mathbb{R}^m, t \in \mathbb{R}, \text{ and } v \in \mathbb{R}^n \text{ such that } A^T y - ct \leq 0, t \geq 0, (y, t) \neq 0, Av = 0, c^T v \leq 0, v \geq 0, v \neq 0\},$$

respectively.

From this corollary it follows that given a data instance $d = (A, b, c)$, the system $Ax = b$ and $x \geq 0$ is ill-posed if and only if

- either the system is feasible, but does not have a strictly feasible solution satisfying $Ax = b$ and $x > 0$ (in other words, the feasible region associated with $d$ is contained in one of the faces of the positive orthant $\mathbb{R}_+^n$),
• or the system is not feasible, and so there is a dual ray $\bar{u}$ satisfying $A^T \bar{u} \leq 0$ and $b^T \bar{u} > 0$, but there is no dual ray satisfying the system $A^T u < 0$ and $b^T u > 0$.

Similarly, from the corollary we obtain that the system $A^T y \leq c$ is ill-posed if and only if

• either the system is feasible, but does not have a strictly feasible solution $y$ with $A^T y < c$,

• or the system is not feasible, and so there is a primal ray $\bar{v}$ satisfying $A\bar{v} = 0$, $c^T \bar{v} < 0$, and $\bar{v} \geq 0$, but there is no strictly positive primal ray, that is, a ray $v$ satisfying $Av = 0$, $v > 0$, and $c^T v < 0$.

The next lemma relates the three sets of ill-posed data instances.

**Lemma 3.9** $((B_P \cup B_D) \cap F) \subset B \subset (\bar{B}_P \cup B_D)$.

**Proof:** Suppose that $d \in B$. Then, given any $\epsilon > 0$, there exist $\bar{d}(\epsilon)$ and $\hat{d}(\epsilon)$ such that $\bar{d}(\epsilon) \in B(d, \epsilon) \cap F$ and $\hat{d}(\epsilon) \in B(d, \epsilon) \cap FC$. Since $\bar{d}(\epsilon) \in F$, it follows that $\bar{d}(\epsilon) \in B(d, \epsilon) \cap P$ and $\hat{d}(\epsilon) \in B(d, \epsilon) \cap D$. Therefore, by letting $\epsilon \to 0$ $d \in cl(P)$ and $d \in cl(D)$. On the other hand, $\hat{d}(\epsilon) \in FC$ implies that $\hat{d}(\epsilon) \in PC$ or $\hat{d}(\epsilon) \in DC$. Therefore, it also follows by letting $\epsilon \to 0$ that $d \in cl(PC)$ or $d \in cl(DC)$. In conclusion, $d \in B_P \cup B_D$.

Now, assume that $d \in (B_P \cup B_D) \cap F$. Since $d \in F$, then $d \in cl(F)$ and we only need to show that $d \in cl(FC)$. Given $\epsilon > 0$ and assuming that $d \in B_P$, it follows that $B(d, \epsilon) \cap PC \neq \emptyset$, so that $B(d, \epsilon) \cap FC \neq \emptyset$, and $d \in cl(FC)$. If $d \in B_D$, it follows that $B(d, \epsilon) \cap DC \neq \emptyset$, so that again $B(d, \epsilon) \cap FC \neq \emptyset$, and $d \in cl(FC)$, and the result follows.

q.e.d.
We now show by two examples that the two inclusions of Lemma 3.9 are proper
inclusions. First, consider the following data instance:

\[
\dd = \left( \begin{bmatrix} 0 & 0 \\ \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right).
\]

Observe that \( \dd \notin \mathcal{F} \), but \( \dd \in \mathcal{B} \), hence the first inclusion of Lemma 3.9 is proper.

Second, let \( \dd \) be the following data instance:

\[
\dd = \left( \begin{bmatrix} 0 & 1 \\ \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right).
\]

Then \( \dd \in \mathcal{P}^C \) and \( \dd \in \text{cl}(\mathcal{P}) \), so that \( \dd \in \mathcal{B}_P \). Also, \( \dd \in \mathcal{D}^C \) and \( \dd \in \text{cl}(\mathcal{D}) \), so that \( \dd \in \mathcal{B}_D \). Therefore, \( \dd \in \mathcal{B}_P \cap \mathcal{B}_D \). Nevertheless, \( \dd \notin \text{cl}(\mathcal{F}) \), hence \( \dd \notin \mathcal{B} \), and the second inclusion of Lemma 3.9 is also proper.

The final result of this chapter relates the three distances to ill-posed sets.

**Lemma 3.10** Let \( d \) be a data instance in \( \mathcal{D} \).

1. If \( d \in \mathcal{F} \), then \( \rho(d) = \min\{\rho_P(d), \rho_D(d)\} \).

2. If \( d \notin \mathcal{F} \), then \( \rho(d) \geq \max\{\rho_P(d), \rho_D(d)\} \).

**Proof:** In this lemma we use the alternative definitions (2.5), (3.41), and (3.42) of \( \rho(d) \), \( \rho_P(d) \), and \( \rho_D(d) \), respectively. First assume that \( d \in \mathcal{F} \). If \( \rho(d) = 0 \), then for each \( \epsilon > 0 \) there exists a data instance \( d_\epsilon \) such that \( d_\epsilon \in B(d, \epsilon) \cap \mathcal{F}^C \). Since \( \mathcal{F}^C = \mathcal{P}^C \cup \mathcal{D}^C \), without loss of generality we may assume that \( d_\epsilon \in \mathcal{P}^C \) for all \( \epsilon > 0 \). Hence, \( \rho_P(d) = 0 \), and so \( \rho(d) = 0 = \min\{\rho_P(d), \rho_D(d)\} \). Next, assume that \( \rho(d) > 0 \). Given any \( \delta \geq 0 \) such that \( \delta < \rho(d) \), then \( B(d, \delta) \subset \mathcal{F} = \mathcal{P} \cap \mathcal{D} \), and
it follows that $B(d, \delta) \subset \mathcal{P}$ and $B(d, \delta) \subset \mathcal{D}$. Hence, $\delta \leq \rho_F(d)$ and $\delta \leq \rho_D(d)$. That is, $\delta \leq \min\{\rho_F(d), \rho_D(d)\}$. Therefore, $\rho(d) \leq \min\{\rho_F(d), \rho_D(d)\}$. On the other hand, let $\epsilon$ be an arbitrary positive scalar, and without loss of generality assume that $\min\{\rho_F(d), \rho_D(d)\} = \rho_F(d)$. Since $\rho_F(d) = \sup\{\delta : B(d, \delta) \subset \mathcal{P}\}$, it follows that there exists $\delta$ such that $B(d, \delta) \subset \mathcal{P}$ and $\rho_F(d) - \epsilon \leq \delta < \rho_F(d) \leq \rho_D(d)$. Moreover, $\delta < \rho_D(d)$ implies $B(d, \delta) \subset \mathcal{D}$. Hence, $B(d, \delta) \subset \mathcal{P} \cap \mathcal{D} = \mathcal{F}$, so that $\rho_F(d) - \epsilon \leq \delta \leq \rho(d)$. Therefore, because $\epsilon$ is arbitrary, we have $\rho(d) \geq \rho_F(d) = \min\{\rho_F(d), \rho_D(d)\}$, concluding the proof for the case $d \in \mathcal{F}$.

Second, suppose that $d \notin \mathcal{F}$. Without loss of generality assume that $\rho_F(d) = \max\{\rho_F(d), \rho_D(d)\}$ and that $\rho_F(d) > 0$. Given any $\epsilon > 0$, there exists $\delta \geq 0$ such that $\rho_F(d) - \epsilon \leq \delta < \rho_F(d)$ and $B(d, \delta) \subset \mathcal{P}^C \subset \mathcal{F}^C$. Therefore, we have $\rho_F(d) - \epsilon \leq \delta \leq \rho(d)$. Since $\epsilon$ is arbitrary, we conclude that $\rho(d) \geq \rho_F(d)$.

q.e.d.
Chapter 4

Analytic Center Problems

4.1 Overview

As described in Chapter 2, given a data instance $d = (A, b, c)$ for a linear program, the equality form analytic center problem, denoted by $AE(d)$, is defined as:

$$AE(d) : \min \{ p(x) : Ax = b, x > 0 \},$$

where for $u > 0$ in $\mathbb{R}^n$, $p(u) = -\sum_{j=1}^{n} \ln u_j$. Its dual problem, denoted by $AED(d)$, is:

$$AED(d) : \max \{ b^T u - p(t) : A^T u + t = 0, t > 0 \}.$$  

Structurally, the program $AE(d)$ is closely related to the central trajectory problem $P_\mu(d)$ that we will study later, and was first extensively studied by Sonnevend, see [Son85a] and [Son85b]. In terms of data dependence, note that the program $AE(d)$ does not depend on the data $c$. It is well known that $AE(d)$ has a unique solution whenever its feasible region is bounded and non-empty.
Similarly, we define the inequality form analytic center problem, denoted by $AI(d)$, as:

$$AI(d) : \min \left\{ p(s) : A^T y + s = c, s > 0 \right\}.$$ 

Its dual problem, denoted by $AID(d)$, is:

$$AID(d) : \max \left\{ -c^T v - p(v) : Av = 0, v > 0 \right\}.$$ 

In terms of data dependence, the program $AI(d)$ does not depend on the data $b$. The program $AI(d)$ has a unique solution when its feasible region is bounded and non-empty. Note in particular that the two programs $AE(d)$ and $AI(d)$ are not duals of each other. (In fact, programs $AE(d)$ and $AI(d)$ cannot both be feasible at the same time.) As we will show later, the study of these problems is relevant to obtain certain results on the central trajectory problem.

As in Chapter 2, we use the data instance sets $\mathcal{F}_E$ and $\mathcal{F}_I$. Recall that $\mathcal{F}_E$ consists of data instances $d$ for which $AE(d)$ is feasible and attains its optimal solution, and $\mathcal{F}_I$ consists of data instances $d$ for which $AI(d)$ is feasible and attains its optimal solution. It is also appropriate to re-introduce the corresponding sets of ill-posed data instances $B_E := \partial \mathcal{F}_E$ for the program $AE(d)$, and $B_I := \partial \mathcal{F}_I$ for the program $AI(d)$. Furthermore, for the equality form analytic center problem $AE(d)$, the distance to ill-posedness of a data instance $d = (A, b, c)$ is denoted by $\rho_E(d)$, and for the inequality form analytic center problem $AI(d)$, the distance to ill-posedness of a data instance $d = (A, b, c)$ is denoted by $\rho_I(d)$. Likewise, the corresponding condition measures are denoted by $C_E(d)$ and $C_I(d)$ (see Chapter 2). Remember that $\rho_E(d)$ is a measure of how much the data instance $d$ can be perturbed before the feasible region of the corresponding equality form analytic center problem $AE(d)$ becomes either empty or
unbounded. Similarly, $\rho_I(d)$ can be interpreted as a measure of how much the data instance $d$ can be perturbed before the feasible region of the corresponding inequality form analytic center problem $AI(d)$ becomes either empty or unbounded.

In Section 4.2, we present some bounds on the norms of feasible solutions of the analytic center problems $AE(d)$ and $AI(d)$. The most important results in this section are Lemma 4.1 and Lemma 4.2, which establish the boundeness of the feasible regions of $AE(d)$ and $AI(d)$, respectively, in terms of the conditions numbers $C_E(d)$ and $C_I(d)$, and $\|d\|$. Proposition 4.3 and Proposition 4.4 state bounds on the optimal solutions to $AED(d)$ and $AID(d)$, respectively, in terms of $n$, $\rho_E(d)$, $\rho_I(d)$, $C_E(d)$ and $C_I(d)$. By using Lemmas 4.1 and 4.2 and Propositions 4.3 and 4.4, we prove in Lemma 4.3 and Lemma 4.4 lower bounds on the optimal solutions to $AE(d)$ and $AI(d)$, respectively, also in terms of $n$, $\rho_I(d)$, $C_E(d)$ and $C_I(d)$. As we show in Chapter 6, these kinds of bounds are useful to study the convergence of the central trajectory to the optimal face of a linear program in standard form. In particular, these bounds can be used to ascertain the optimal partition associated with the optimal face by using information at a current iterate of an interior-point method.

Furthermore, the bounds from Section 4.2 are used in Section 4.3 to study the effect of data perturbations on optimal solutions to the analytic centers $AE(d)$ and $AI(d)$, respectively. Our objective is to demonstrate upper bounds on the size of changes of optimal solutions to $AE(d)$ and $AI(d)$, respectively, in terms of polynomial expressions of the condition numbers $C_E(d)$ and $C_I(d)$; the size $\|d\|$ of the data instance $d$; the size $\|\Delta d\|$ of the data perturbation $\Delta d$; the dimensions $m$ and $n$; and some small constants. These results are also useful in Chapter 6 in the study of the convergence of the central trajectory to the optimal face of a linear program in standard form. In fact, we will prove the following theorem concerning the problem $AE(d)$:
Theorem 4.1 Let \( d = (A, b, c) \in \mathcal{F}_E \) and \( \rho_E(d) > 0 \). For a given \( \alpha \in [0,1) \), let \( \Delta d = (\Delta A, \Delta b, \Delta c) \in \bar{D} \) be such that \( \|\Delta d\| \leq \alpha \rho_E(d) \). Then

\[
\|\bar{x} - x\|_1 \leq \frac{\|\Delta d\|}{\|d\|} \frac{14mn^2}{(1 - \alpha)^3} C_E(d)^7, \quad (4.1)
\]

\[
\|\bar{u} - u\|_\infty \leq \frac{\|\Delta d\|}{\|d\|^2} \frac{10mn^3}{(1 - \alpha)^3} C_E(d)^7, \quad (4.2)
\]

\[
\|\bar{t} - t\|_\infty \leq \frac{\|\Delta d\|}{\|d\|} \frac{11mn^3}{(1 - \alpha)^3} C_E(d)^7, \quad (4.3)
\]

where \( x \) and \( \bar{x} \) are the optimal solutions to \( AE(d) \) and \( AE(d + \Delta d) \), respectively; and \( (u, t) \) and \( (\bar{u}, \bar{t}) \) are the optimal solutions to \( AED(d) \) and \( AED(d + \Delta d) \), respectively.

Roughly speaking, this theorem states that for a data perturbation \( \Delta d \) sufficiently small, the change in the optimal solutions \( x \) and \( \bar{x} \) to \( AE(d) \) and \( AE(d + \Delta d) \), respectively, is of the order

\[
\|\bar{x} - x\|_1 = O \left( \frac{mn^2}{\|d\|}, \frac{\|\Delta d\|}{\|d\|} C_E(d)^7 \right),
\]

where it is important to notice the linear dependence on \( \|\Delta d\| \).

Analogously, we prove the following theorem concerning the problem \( AI(d) \):

Theorem 4.2 Let \( d = (A, b, c) \in \mathcal{F}_I \) and \( \rho_I(d) > 0 \). For a given \( \alpha \in [0,1) \), let \( \Delta d = (\Delta A, \Delta b, \Delta c) \in \bar{D} \) be such that \( \|\Delta d\| \leq \alpha \rho_I(d) \). Then,

\[
\|\bar{y} - y\|_\infty \leq \frac{\|\Delta d\|}{\|d\|} \frac{64mn^2}{(1 - \alpha)^3} C_I(d)^7, \quad (4.4)
\]

\[
\|\bar{s} - s\|_\infty \leq \frac{\|\Delta d\|}{\|d\|} \frac{67mn^2}{(1 - \alpha)^3} C_I(d)^7, \quad (4.5)
\]

\[
\|\bar{v} - v\|_1 \leq \frac{\|\Delta d\|}{\|d\|^2} \frac{19mn^3}{(1 - \alpha)^3} C_I(d)^7, \quad (4.6)
\]

where \( v \) and \( \bar{v} \) are the optimal solutions to \( AID(d) \) and \( AID(d + \Delta d) \), respectively;
and \((y, s)\) and \((\tilde{y}, \tilde{s})\) are the optimal solutions to \(AI(d)\) and \(AI(d + \Delta d)\), respectively.

As before, roughly speaking this theorem states that for a data perturbation \(\Delta d\) sufficiently small, the change in the optimal solutions \((y, s)\) and \((\tilde{y}, \tilde{s})\) to \(AI(d)\) and \(AI(d + \Delta d)\), respectively, is of the order

\[
\|\tilde{y} - y\|_{\infty} = O\left( mn^2 \frac{\|\Delta d\|}{\|d\|} C_E(d)^7 \right),
\]

\[
\|\tilde{s} - s\|_{\infty} = O\left( mn^2 \|\Delta d\| C_E(d)^7 \right),
\]

where again it is important to notice the linear dependence on \(\|\Delta d\|\).

Finally, in Theorem 4.3 and Theorem 4.4, we state upper bounds on the change of optimal objective values of the programs \(AE(\cdot)\) and \(AI(\cdot)\), respectively, under data perturbations. These bounds are linear in the size \(\|\Delta d\|\) of the perturbation vector \(\Delta d\), and also depend on the condition numbers \(C_E(d)\) and \(C_I(d)\); the distances to ill-posedness \(\rho_E(d)\) and \(\rho_I(d)\); the dimension \(n\); and some small constants.

### 4.2 Bounds on Analytic Centers

We start by establishing the following two propositions that state the uniqueness of optimal solutions of the four analytic center considered in this chapter.

**Proposition 4.1** Let \(d = (A, b, c) \in \mathcal{F}_E\) and \(\rho_E(d) > 0\). Then \(AE(d)\) and \(AED(d)\) each have a unique optimal solution.

**Proof:** The strict convexity of \(p(x)\) implies that \(AE(d)\) has a unique optimal solution. Suppose that the program \(AED(d)\) has two optimal solutions \((u_1, t_1)\) and \((u_2, t_2)\). Because of the strict convexity of \(p(t)\), we must have \(t_1 = t_2\). Hence, \(A^T(u_1 - u_2) = 0\)
4.2 Bounds on Analytic Centers

and \( b^T(u_1 - u_2) = 0 \). Suppose that \( u_1 \neq u_2 \), then \( \text{rank}(A) < m \), which according to Theorem 3.2, identity (3.6), contradicts that \( \rho_E(d) > 0 \).

q.e.d.

**Proposition 4.2** Let \( d = (A, b, c) \in F_I \) and \( \rho_I(d) > 0 \). Then \( AI(d) \) and \( AID(d) \) each have a unique optimal solution.

**Proof:** The strict convexity of \( p(v) \) implies that \( AID(d) \) has a unique optimal solution. Suppose that the program \( AI(d) \) has two optimal solutions \((y_1, s_1)\) and \((y_2, s_2)\). Because of the strict convexity of \( p(s) \), we must have \( s_1 = s_2 \). Hence, \( A^T(y_1 - y_2) = 0 \). Suppose that \( y_1 \neq y_2 \), then \( \text{rank}(A) < m \), which according to Theorem 3.3, identity (3.11), contradicts that \( \rho_I(d) > 0 \).

q.e.d.

The following two lemmas present upper bounds on the norms of all feasible solutions for primal and dual analytic center problems, respectively.

**Lemma 4.1** Let \( d = (A, b, c) \in F_E \) and \( \rho_E(d) > 0 \). Then

\[
\|x\|_1 \leq C_E(d)
\]

for any feasible solution \( x \) of \( AE(d) \).

**Proof:** Let \( x \) be a feasible solution of \( AE(d) \). Define \( \Delta A = -be^T/\|x\|_1 \) and \( \Delta d = (\Delta A, 0, 0) \). Then, \((A + \Delta A)x = 0 \) and \( x > 0 \), and so, by Proposition 2.5, there cannot exist \( u \) for which \((A + \Delta A)^Tu < 0 \). Thus, \( d + \Delta d \in F_E^C \), whereby \( \rho_E(d) \leq \|\Delta d\| \). On the other hand, \( \|\Delta d\| \leq \|b\|_1/\|x\|_1 \leq \|d\|/\|x\|_1 \), so that \( \|x\|_1 \leq \|d\|/\rho_E(d) = C_E(d) \), and the result follows.

q.e.d.
Lemma 4.2 \( \text{Let } d = (A, b, c) \in \mathcal{F}_I \text{ and } \rho_I(d) > 0. \text{ Then} \)

\[
\|y\|_\infty \leq C_I(d), \\
\|s\|_\infty \leq 2\|d\|C_I(d),
\]

for any feasible solution \((y, s)\) of \(AI(d)\).

Proof: Let \((y, s)\) be a feasible solution of \(AI(d)\). If \(y = 0\), then \(s = c\) and the bounds are trivially true, so that we assume \(y \neq 0\). Let \(\tilde{y}\) be such that \(\|y\|_\infty = \tilde{y}^Ty\) and \(\|\tilde{y}\|_1 = 1\). Let \(\Delta A = -\tilde{y}c^T/\|y\|_\infty\) and \(\Delta d = (\Delta A, 0, 0)\). Hence, \((A + \Delta A)^Ty = A^Ty - c < 0\), and so, by Proposition 2.6, there cannot exist \(v\) for which \((A + \Delta A)v = 0\) and \(v > 0\). Thus, \(d + \Delta d \in \mathcal{F}_C\), whereby \(\rho_I(d) \leq \|\Delta d\|\). On the other hand, \(\|\Delta d\| = \|c\|_\infty/\|y\|_\infty \leq \|d\|/\|y\|_\infty\), so that \(\|y\|_\infty \leq \|d\|/\rho_I(d) = C_I(d)\). The bound for \(\|s\|_\infty\) can be easily derived using the fact that \(\|s\|_\infty \leq \|c\|_\infty + \|A^T\|_\infty \|y\|_\infty\).

\(\|A^T\|_\infty = \|A\|\), and \(C_I(d) \geq 1\).

q.e.d.

From the arithmetic-geometric mean inequality we have for \(x > 0\)

\[
\frac{\prod_{j=1}^n x_j}{(\sum_{j=1}^n x_j)^n} \leq \left(\frac{1}{n}\right)^n,
\]

from which we obtain

\[
\sum_{j=1}^n \ln(x_j) \leq n \ln\left(\sum_{j=1}^n x_j\right) - n \ln(n).
\]

Hence, for \(x > 0\),

\[-p(x) \leq n \ln(\|x\|_1) - n \ln(n).\]
Similarly, using that \(\|v\|_1 \leq n\|v\|_\infty\) for all \(v \in \mathbb{R}^n\), it follows that for any \(s > 0\),

\[-p(s) \leq n \ln(\|s\|_\infty).\]

Combining these results with the bounds from Lemmas 4.1 and 4.2, we obtain the following corollary:

**Corollary 4.1** Let \(d = (A, b, c) \in \tilde{D}\). If \(d \in \mathcal{F}_E\) and \(\rho_E(d) > 0\), then

\[p(x) \geq -n \ln(C_E(d)) + n \ln(n),\]

for all feasible solution \(x\) of \(AE(d)\). If \(d \in \mathcal{F}_I\) and \(\rho_I(d) > 0\), then

\[\mu(s) \geq -n \ln(2\|d\|C_I(d)),\]

for all feasible solution \((y, s)\) of \(AI(d)\).

The following propositions are useful results for finding lower bounds on optimal solutions to the equality and inequality forms of the analytic center problem.

**Proposition 4.3** Let \(d = (A, b, c) \in \mathcal{F}_E\) and \(\rho_E(d) > 0\). Then

\[\|u\|_\infty \leq \frac{n}{\rho_E(d)},\]

\[\|t\|_\infty \leq nC_E(d),\]

where \((u, t)\) is the optimal solution of the program \(AED(d)\).

**Proof:** We have \(b^Tu + n = p(x) + p(t) = 0\), where \(t = -A^Tu\) and \(x\) is the optimal solution to \(AE(d)\). If \(u = 0\), then \(t = 0\), and the inequalities are trivially true, thus we assume that \(u \neq 0\). Let \(\bar{u}\) be such that \(\|u\|_\infty = \bar{u}^Tu\) and \(\|\bar{u}\|_1 = 1\). Let \(\Delta b = n\bar{u}/\|u\|_\infty\), and \(\Delta d = (0, \Delta b, 0)\). Hence, \((b + \Delta b)^Tu = b^Tu + \Delta b^Tu = 0\).
Therefore, the instance $d + \Delta d$ corresponds to an unbounded program \( AED(d + \Delta d) \), and so it is also a primal infeasible instance. Thus, \( \|\Delta d\| \geq \rho_E(d) \). On the other hand, \( \|\Delta d\| = \|\Delta b\|_1 = n/\|u\|_\infty \), and the result for \( u \) follows. Finally, \( \|t\|_\infty \leq \|A^T\|_\infty \|u\|_\infty \leq \|d\| \|u\|_\infty \), and the result for \( t \) follows.

\[ \text{q.e.d.} \]

**Proposition 4.4** Let \( d = (A, b, c) \in \mathcal{F}_I \) and \( \rho_I(d) > 0 \). Then

\[ \|v\|_1 \leq \frac{n}{\rho_I(d)}, \]

where \( v \) is the optimal solution of the program \( AID(d) \).

**Proof:** We have \(-c^Tv + n = p(v) + p(s) = 0\), where \((y, s)\) is the optimal solution to \( AI(d) \). Let \( \Delta c = -ne/\|v\|_1 \) and \( \Delta d = (0, 0, \Delta c) \). Hence, \(- (c + \Delta c)^Tv = -c^Tv - \Delta c^Tv = 0\). Therefore, the instance \( d + \Delta d \) corresponds to an unbounded program \( AID(d + \Delta d) \), and so it is also a primal infeasible instance. Thus, \( \|\Delta d\| \geq \rho_I(d) \).

On the other hand, \( \|\Delta d\| = \|\Delta c\|_\infty = n/\|v\|_1 \), and the result follows.

\[ \text{q.e.d.} \]

From Lemmas 4.1 and 4.2 and Propositions 4.3 and 4.4, we obtain the following corollaries:

**Corollary 4.2** For a given \( \alpha \in [0, 1) \), let \( \delta \) be such that \( \delta \leq \alpha \rho_E(d) \), where \( d \in \mathcal{F}_E \) and \( \rho_E(d) > 0 \). If \( \Delta d \in \tilde{D} \) is such that \( \|\Delta d\| \leq \delta \), then

\[
\begin{align*}
\|x\|_1 & \leq \frac{1 + \alpha}{1 - \alpha} C_E(d), \\
\|u\|_\infty & \leq \frac{n}{1 - \alpha} \frac{1}{\rho_E(d)}, \\
\|t\|_\infty & \leq \frac{1 + \alpha}{1 - \alpha} C_E(d) n,
\end{align*}
\]
where \( x \) is a feasible solution of \( AE(d + \Delta d) \) and \((u,t)\) is the optimal solution to \( AED(d + \Delta d) \).

**Proof:** Observe that \( \rho_E(d + \Delta d) \geq (1 - \alpha) \rho_E(d) \) and \( \|d + \Delta d\| \leq \|d\| + \delta \leq \|d\| + \alpha \rho_E(d) \leq \|d\| + \alpha \|d\| = (1 + \alpha) \|d\| \). Therefore,

\[
C_E(d + \Delta d) \leq \frac{\|d\| + \delta}{(1 - \alpha) \rho_E(d)} \leq \left( \frac{1 + \alpha}{1 - \alpha} \right) \frac{\|d\|}{\rho_E(d)} = \left( \frac{1 + \alpha}{1 - \alpha} \right) C_E(d),
\]

so that the result follows.

q.e.d.

**Corollary 4.3** For a given \( \alpha \in [0, 1) \), let \( \delta \) be such that \( \delta \leq \alpha \rho_I(d) \), where \( d \in F_I \) and \( \rho_I(d) > 0 \). If \( \Delta d \in \bar{D} \) is such that \( \|\Delta d\| \leq \delta \), then

\[
\|y\|_\infty \leq \frac{1 + \alpha}{1 - \alpha} C_I(d),
\]

\[
\|s\|_\infty \leq \frac{2(1 + \alpha)^2}{1 - \alpha} C_I(d) \|d\|,
\]

\[
\|v\|_1 \leq \frac{1}{1 - \alpha} \frac{n}{\rho_I(d)},
\]

where \((y, s)\) is a feasible solution of \( AI(d + \Delta d) \) and \( v \) is the optimal solution to \( AID(d + \Delta d) \).

**Proof:** As in the previous corollary, \( \rho_I(d + \Delta d) \geq (1 - \alpha) \rho_I(d) \) and \( \|d + \Delta d\| \leq \|d\| + \delta \leq \|d\| + \alpha \rho_I(d) \leq \|d\|(1 + \alpha) \). Therefore,

\[
C_I(d + \Delta d) \leq \left( \frac{1 + \alpha}{1 - \alpha} \right) C_I(d),
\]

so that the result follows.

q.e.d.
4.2 Bounds on Analytic Centers

Next, from Propositions 4.3 and 4.4, we prove several results concerning lower bounds on solutions to $AE(d)$ and $AI(d)$.

**Lemma 4.3** If $d = (A, b, c) \in \mathcal{F}_E$ and $\rho_E(d) > 0$, then

$$\|x\|_1 \geq \frac{1}{C_E(d)},$$

for all feasible solution $x$ of $AE(d)$. Moreover, if $\hat{x}$ is the optimal solution to $AE(d)$ and $(\hat{u}, \hat{t})$ is the optimal solution to $AED(d)$ then

$$\hat{x}_j \geq \frac{1}{nC_E(d)},$$

$$\hat{t}_j \geq \frac{1}{C_E(d)},$$

for all $j = 1, \ldots, n$.

**Proof:** Let $(\hat{u}, \hat{t})$ be the optimal solution to $AED(d)$. Hence, for any feasible solution $x$ to $AE(d)$ we have $x^T \hat{t} = n$. Therefore, $\|x\|_1 \|\hat{t}\|_\infty \geq n$, that is, $\|x\|_1 \geq n/\|\hat{t}\|_\infty$. and by Proposition 4.3 the first inequality follows. Now, if $\hat{x}$ is the optimal solution to $AE(d)$, then from the Karush-Kuhn-Tucker optimality conditions we know that $\hat{x}_j \hat{t}_j = 1$, for all $j = 1, \ldots, n$. Therefore, since $\hat{t} > 0$, $\hat{x}_j = 1/\hat{t}_j \geq 1/\|\hat{t}\|_\infty$. On the other hand, since $\hat{x} > 0$, $\hat{t}_j = 1/\hat{x}_j \geq 1/\|\hat{x}\|_1$. Therefore, by using Proposition 4.3 and Lemma 4.1, we complete the proof.

q.e.d.

**Lemma 4.4** If $d = (A, b, c) \in \mathcal{F}_I$ and $\rho_I(d) > 0$, then

$$\|s\|_\infty \geq \rho_I(d),$$
for all feasible solution \((y, s)\) of \(AI(d)\). Moreover, if \((\hat{y}, \hat{s})\) is the optimal solution to \(AI(d)\) and \(\hat{v}\) the optimal solution to \(AID(d)\), then

\[
\hat{s}_j \geq \frac{\rho_I(d)}{n},
\]

\[
\hat{v}_j \geq \frac{1}{2\|d\|C_I(d)},
\]

for all \(j = 1, \ldots, n\).

**Proof:** Let \(\hat{v}\) be an optimal solution to \(AID(d)\). Hence, for any feasible solution \((y, s)\) to \(AI(d)\) we have \(s^T\hat{v} = n\). Therefore, \(\|s\|_{\infty} \|\hat{v}\|_1 \geq n\), that is, \(\|s\|_{\infty} \geq n/\|\hat{v}\|_1\), and by Proposition 4.4 the first inequality follows. Now, if \((\hat{y}, \hat{s})\) is the optimal solution to \(AI(d)\), then from the Karush-Kuhn-Tucker optimality conditions we know that \(\hat{s}_j\hat{v}_j = 1\), for all \(j = 1, \ldots, n\). Therefore, since \(\hat{v} > 0\), \(\hat{s}_j = 1/\hat{v}_j \geq 1/\|\hat{v}\|_1\). On the other hand, since \(\hat{s} > 0\), \(\hat{v}_j = 1/\hat{s}_j \geq 1/\|\hat{s}\|_{\infty}\). Therefore, by using Proposition 4.4 and Lemma 4.2, we complete the proof.

q.e.d.

The following corollary follows immediately from Lemmas 4.3 and 4.4.

**Corollary 4.4** If \(d \in F_E\) and \(\rho_E(d) > 0\), then

\[
p(\hat{x}) \leq n \ln (nC_E(d)),
\]

where \(\hat{x}\) is the optimal solution to \(AE(d)\). If \(d \in F_I\) and \(\rho_I(d) > 0\), then

\[
p(\hat{s}) \leq n \ln \left( \frac{n}{\rho_I(d)} \right),
\]

where \((\hat{y}, \hat{s})\) is the optimal solution to \(AI(d)\).

Finally, we present the following two corollaries that generalize the results of Lemmas 4.3 and 4.4. The proofs of the corollaries follow from the inequalities \(\rho_E(d +
\[ \Delta d \geq (1 - \alpha) \rho_E(d), \quad \rho_I(d + \Delta d) \geq (1 - \alpha) \rho_I(d), \quad \text{and} \quad \|d + \Delta d\| \leq \|d\| + \delta \leq \|d\|(1 + \alpha) \]

shown previously in the proofs of Corollary 4.2 and Corollary 4.3.

**Corollary 4.5** For a given \(\alpha \in [0, 1]\), let \(\delta\) be such that \(\delta \leq \alpha \rho_E(d)\), where \(d \in \mathcal{F}_E\) and \(\rho_E(d) > 0\). If \(\Delta d \in \tilde{D}\) is such that \(\|\Delta d\| \leq \delta\), then

\[
\|x\|_1 \geq \frac{1 - \alpha}{1 + \alpha} \frac{1}{C_E(d)},
\]

for all \(x\) feasible solution of \(AE(d + \Delta d)\). Moreover, if \(\hat{x}\) is the optimal solution to \(AE(d + \Delta d)\) and \((\hat{u}, \hat{t})\) is the optimal solution to \(AED(d + \Delta d)\), then

\[
\hat{x}_j \geq \frac{1 - \alpha}{1 + \alpha} \frac{1}{C_E(d)} \frac{1}{n}, \\
\hat{t}_j \geq \frac{1 - \alpha}{1 + \alpha} \frac{1}{C_E(d)},
\]

for all \(j = 1, \ldots, n\).

**Corollary 4.6** For a given \(\alpha \in [0, 1]\), let \(\delta\) be such that \(\delta \leq \alpha \rho_I(d)\), where \(d \in \mathcal{F}_I\) and \(\rho_I(d) > 0\). If \(\Delta d \in \tilde{D}\) is such that \(\|\Delta d\| \leq \delta\), then

\[
\|s\|_\infty \geq (1 - \alpha) \rho_I(d),
\]

for all \((y, s)\) feasible solution of \(AI(d + \Delta d)\). Moreover, if \((\hat{y}, \hat{s})\) is the optimal solution to \(AI(d + \Delta d)\) and \(\hat{v}\) is the optimal solution to \(AID(d + \Delta d)\), then

\[
\hat{s}_j \geq (1 - \alpha) \frac{\rho_I(d)}{n}, \\
\hat{v}_j \geq \frac{1 - \alpha}{2(1 + \alpha)^2} \frac{1}{C_I(d)\|d\|},
\]

for all \(j = 1, \ldots, n\).
4.3 Changes in Optimal Solutions Under Data Perturbations

In this section we study the effect of data perturbations on optimal solutions to our two versions of the analytic center problem. The main results in this section are Theorems 4.1, 4.2, 4.3, and 4.4, but before proving them, we establish a few preliminary lemmas.

**Lemma 4.5** For a given $\alpha \in [0, 1)$, let $\delta$ be such that $\delta \leq \alpha \rho_E(d)$, where $d \in F_E$ and $\rho_E(d) > 0$. Let $\Delta d \in \mathcal{D}$ be such that $\|\Delta d\| \leq \delta$. If $x$ is the optimal solution to $AE(d)$ and $\tilde{x}$ is the optimal solution to $AE(d + \Delta d)$, then for $j = 1, \ldots, n$,

$$\frac{1 - \alpha}{2n^2 C_E(d)^2} \leq x_j \tilde{x}_j \leq \frac{2C_E(d)^2}{1 - \alpha}.$$

**Proof:** This lemma follows directly by applying Lemmas 4.1, and 4.3, and Corollaries 4.2, and 4.5.

q.e.d.

**Lemma 4.6** For a given $\alpha \in [0, 1)$, let $\delta$ be such that $\delta \leq \alpha \rho_I(d)$, where $d \in F_I$ and $\rho_I(d) > 0$. Let $\Delta d \in \mathcal{D}$ be such that $\|\Delta d\| \leq \delta$. If $v$ is the optimal solution to $AID(d)$ and $\tilde{v}$ is the optimal solution to $AID(d + \Delta d)$, then for $j = 1, \ldots, n$,

$$\frac{1 - \alpha}{16\|d\|^2 C_I(d)^2} \leq v_j \tilde{v}_j \leq \frac{n^2}{(1 - \alpha)\rho_I(d)^2}.$$

**Proof:** This lemma follows directly by applying Proposition 4.4, Lemma 4.4, and Corollaries 4.3, and 4.6.

q.e.d.
4.3 Changes in Optimal Solutions Under Data Perturbations

The next lemma presents lower and upper bounds on the operator norm of the inverse of the linear operator $AX\bar{X}AT$.

**Lemma 4.7** For a given $\alpha \in [0, 1)$, let $\delta$ be such that $\delta \leq \alpha \rho_E(d)$, where $d \in \mathcal{F}_E$ and $\rho_E(d) > 0$. Let $\Delta d \in \hat{D}$ be such that $\|\Delta d\| \leq \delta$. If $x$ is the optimal solution to $AE(d)$ and $\bar{x}$ is the optimal solution to $AE(d + \Delta d)$,

$$\frac{1 - \alpha}{2mnC_E(d)^2\|d\|^2} \leq \|(AX\bar{X}AT)^{-1}\|_{1,\infty} \leq \frac{2mn^2C_E(d)^2}{(1 - \alpha)\rho_E(d)^2}.$$

**Proof:** Using identical logic to Proposition 2.12 part (i), we have

$$\|(AX\bar{X}AT)^{-1}\|_{1,\infty} \leq \|(AX\bar{X}AT)^{-1}\|_2 \leq \frac{\|(AA^T)^{-1}\|_2}{\min_j \{x_j\bar{x}_j\}}.$$

Now, by applying Proposition 2.12, part (ii), Proposition 2.2, Corollary 2.1, and Lemma 4.5, we obtain that

$$\|(AX\bar{X}AT)^{-1}\|_{1,\infty} \leq \frac{2n^2C_E(d)^2}{(1 - \alpha)\lambda_1(AA^T)} \leq \frac{2mn^2C_E(d)^2}{(1 - \alpha)\rho_E(d)^2}.$$

On the other hand, by identical logic to Proposition 2.12 part (i),

$$\|(AX\bar{X}AT)^{-1}\|_{1,\infty} \geq \frac{1}{m\|\|(AX\bar{X}AT)^{-1}\|_2} \geq \frac{\|(AA^T)^{-1}\|_2}{m \max_j \{x_j\bar{x}_j\}}.$$

Now, by applying Proposition 2.1, part (iv), and Lemma 4.5, we obtain that

$$\|(AX\bar{X}AT)^{-1}\|_{1,\infty} \geq \frac{1 - \alpha}{2mC_E(d)^2\lambda_1(AA^T)}.$$
\[
\frac{1 - \alpha}{2mC_E(d)^2\lambda_m(AA^T)} \\
\frac{1 - \alpha}{2mC_E(d)^2\|A\|^2} \\
\frac{1 - \alpha}{2mnC_E(d)^2\|A\|^2} \\
\frac{1 - \alpha}{2mnC_E(d)^2\|d\|^2}. 
\]

q.e.d.

The following result shows analogous bounds on \(||(AV\hat{V}A^T)^{-1}||_1\|\infty as in the preceding lemma.

**Lemma 4.8** For a given \(\alpha \in [0, 1]\), let \(\delta\) be such that \(\delta \leq \alpha \rho_I(d)\), where \(d \in F_I\) and \(\rho_I(d) > 0\). Let \(\Delta d \in \hat{D}\) be such that \(||\Delta d|| \leq \delta\). If \(\nu\) is the optimal solution to \(AID(d)\) and \(\bar{\nu}\) is the optimal solution to \(AID(d + \Delta d)\), then

\[
\frac{1 - \alpha}{mn^3C_I(d)^2} \leq ||(AV\hat{V}A^T)^{-1}||_1\|\infty \leq \frac{16mC_I(d)^2}{1 - \alpha}. 
\]

**Proof:** As in the proof of Lemma 4.7, using identical logic to Proposition 2.12 part (i) we have \(||(AV\hat{V}A^T)^{-1}||_1\|\infty \leq ||(AA^T)^{-1}||_2/\min_j \{v_j\bar{v}_j\}\). Now, by applying Proposition 2.12, part (ii), Proposition 2.2, Corollary 2.1 (\(b \in range(A)\) because \(A\) has full-row rank), and Lemma 4.6, we obtain that

\[
|| (AV\hat{V}A^T)^{-1} ||_1 \| \infty \leq \frac{16||d||^2C_I(d)^2}{(1 - \alpha)\lambda_1(AA^T)} \\
\leq \frac{16m||d||^2C_I(d)^2}{(1 - \alpha)\rho_I(d)^2},
\]

from which we obtain the upper bound on \(||(AV\hat{V}A^T)^{-1}||_1\|\infty\). On the other hand, by using identical logic to Proposition 2.12 part (i), we get the inequality \(||(AV\hat{V}A^T)^{-1}||_1\|\infty\)
\[ \geq (1/m)\|(AA^T)^{-1}\|_2 / \max_j \{v_j \bar{v}_j\} \]. Now, by applying Proposition 2.1, part (iv), and Lemma 4.6, we obtain that

\[
\|(AV\bar{V}A^T)^{-1}\|_{1,\infty} \geq \frac{(1 - \alpha)\rho_1(d)^2}{mn^2\lambda_1(\bar{A}A^T)} \\
\geq \frac{(1 - \alpha)\rho_1(d)^2}{mn^2\lambda_m(\bar{A}A^T)} \\
= \frac{(1 - \alpha)\rho_1(d)^2}{mn^2\|A\|_2^2} \\
\geq \frac{(1 - \alpha)\rho_1(d)^2}{mn^3\|d\|_2^2},
\]

from which we obtain the lower bound on \(\|(AV\bar{V}A^T)^{-1}\|_{1,\infty}\).

q.e.d.

We are now ready to prove Theorem 4.1, which states upper bounds on changes in optimal solutions to \(AE(\cdot)\) and \(AED(\cdot)\) as the data is changed.

**Proof of Theorem 4.1:** Let \(x\) and \(\bar{x}\) be the optimal solutions to \(AE(d)\) and \(AE(d+\Delta d)\), respectively; and let \((u, t)\) and \((\bar{u}, \bar{t})\) be the optimal solutions to \(AED(d)\) and \(AED(d+\Delta d)\), respectively. Then from the Karush-Kuhn-Tucker optimality conditions we have:

\[
Xt = e, \quad \bar{X}\bar{t} = e, \\
A^Tu + t = 0, \quad (A + \Delta A)^T\bar{u} + \bar{t} = 0, \\
Ax = b, \quad (A + \Delta A)\bar{x} = b + \Delta b, \\
x > 0, \quad \bar{x} > 0.
\]

Therefore,

\[
\bar{x} - x = X\bar{X}(t - \bar{t})
\]
4.3 Changes in Optimal Solutions Under Data Perturbations

\[ \begin{align*}
= & \quad X\tilde{X} \left((A + \Delta A)^T\tilde{u} - A^Tu\right) \\
= & \quad X\tilde{X}\Delta A^T\tilde{u} + X\tilde{X}A^T(\tilde{u} - u).
\end{align*} \quad (4.7) \]

On the other hand, \( A(\tilde{x} - x) = \Delta b - \Delta A\tilde{x} \). Since \( A \) has rank \( m \) (otherwise \( \rho_E(d) = 0 \)), then \( P = AX\tilde{X}A^T \) is a positive definite matrix. By combining both results together with (4.7), we obtain

\[ \Delta b - \Delta A\tilde{x} = AX\tilde{X}\Delta A^T\tilde{u} + P(\tilde{u} - u), \]

and so

\[ P^{-1}(\Delta b - \Delta A\tilde{x}) = P^{-1}AX\tilde{X}\Delta A^T\tilde{u} + \tilde{u} - u. \]

Therefore, we have the following identity:

\[ \tilde{u} - u = P^{-1}(\Delta b - \Delta A\tilde{x}) - P^{-1}AX\tilde{X}\Delta A^T\tilde{u}. \quad (4.8) \]

From this identity, it follows that

\[ \|\tilde{u} - u\|_\infty \leq \|P^{-1}\|_{1,\infty}\left(\|\Delta b - \Delta A\tilde{x}\|_1 + \|A\|\|X\tilde{X}\Delta A^T\tilde{u}\|_1\right). \quad (4.9) \]

Note that

\[ \|X\tilde{X}\Delta A^T\tilde{u}\|_1 \leq \|X\tilde{X}\|_{\infty,1}\|\Delta A^T\tilde{u}\|_\infty \leq \|x\|_1\|\tilde{x}\|_1\|\Delta A^T\tilde{u}\|_\infty. \quad (4.10) \]

From Corollary 4.2, we have

\[ \begin{align*}
\|\Delta b - \Delta A\tilde{x}\|_1 \leq & \quad \|\Delta d\|(1 + \|\tilde{x}\|_1) \\
\leq & \quad \|\Delta d\| \left(1 + \frac{2}{1 - \alpha}C_E(d)\right) \\
\leq & \quad \frac{3\|\Delta d\|}{1 - \alpha}C_E(d),
\end{align*} \quad (4.11) \]
and

\[ \| \Delta A^T \bar{u} \|_\infty \leq \| \Delta d \| \| \bar{u} \|_\infty \leq \frac{\| \Delta d \| n}{(1 - \alpha) \rho_E(d)}. \]  

(4.12)

Therefore, by combining (4.9), (4.10), (4.11), and (4.12), and by using Lemma 4.1, Corollary 4.2, and Lemma 4.7, we obtain the following bound on \( \| \bar{u} - u \|_\infty \):

\[ \| \bar{u} - u \|_\infty \leq \left( \frac{2mn^2C_E(d)^2(1 - \alpha)}{(1 - \alpha) \rho_E(d)^2} \right) \left( \frac{\| \Delta d \|}{1 - \alpha} \right) \left( 3C_E(d) + \| d \| \frac{2nC_E(d)^2}{(1 - \alpha) \rho_E(d)} \right) \]

\[ \leq 10mn^3\| \Delta d \| \frac{C_E(d)^5}{(1 - \alpha)^3 \rho_E(d)^5}, \]

thereby demonstrating the bound (4.2) on \( \| \bar{u} - u \|_\infty \). Now, by substituting identity (4.8) into equation (4.7), we obtain

\[ \bar{x} - x = X \bar{X} \left( I - A^T P^{-1} A X \bar{X} \right) \Delta A^T \bar{u} + X \bar{X} A^T P^{-1} (\Delta b - \Delta A \bar{x}) \]

\[ = D^{1/2} \left( I - D^{1/2} A^T P^{-1} A D^{1/2} \right) D^{1/2} \Delta A^T \bar{u} + DA^T P^{-1} (\Delta b - \Delta A \bar{x}), \]

where \( D = X \bar{X} \). Observe that the matrix \( Q = I - D^{1/2} A^T P^{-1} A D^{1/2} \) is a projection matrix, and so \( \| Qx \|_2 \leq \| x \|_2 \) for all \( x \in \mathbb{R}^n \). Hence, from Proposition 2.1 parts (i) and (iii), we obtain that

\[ \| \bar{x} - x \|_1 \leq \| D^{1/2} \left( I - D^{1/2} A^T P^{-1} A D^{1/2} \right) D^{1/2} \Delta A^T \bar{u} \|_1 + \| DA^T P^{-1} (\Delta b - \Delta A \bar{x}) \|_1 \]

\[ \leq \sqrt{n} \| D \|_2 \| D^{1/2} \|_2 \| \Delta A^T \bar{u} \|_2 + \| DA^T P^{-1} (\Delta b - \Delta A \bar{x}) \|_1 \]

\[ = \sqrt{n} \| D \|_2 \| \Delta A^T \bar{u} \|_2 + \| DA^T P^{-1} (\Delta b - \Delta A \bar{x}) \|_1 \]

\[ \leq n \max_j \{ x_j \bar{x}_j \} \| \Delta A^T \bar{u} \|_\infty + \| D \|_\infty,1 \| A^T \|_\infty \| P^{-1} \|_{1,\infty} \| \Delta b - \Delta A \bar{x} \|_1 \]

\[ = n \max_j \{ x_j \bar{x}_j \} \| \Delta A^T \bar{u} \|_\infty + \sum_{j=1}^n \{ x_j \bar{x}_j \} \| A \| \| P^{-1} \|_{1,\infty} \| \Delta b - \Delta A \bar{x} \|_1 \]
4.3 Changes in Optimal Solutions Under Data Perturbations

\[ \leq n \max_j \{x_j \bar{x}_j\} \|A^T \bar{u}\|_\infty + \|x\|_1 \|\bar{x}\|_1 \|d\|_1 \|P^{-1}\|_{1,\infty} \|\Delta b - \Delta A \bar{x}\|_1 \]

It follows from Lemma 4.5, Lemma 4.1, Corollary 4.2, Lemma 4.7, and inequalities (4.11) and (4.12) that

\[ \|\bar{x} - x\|_1 \leq n \left( \frac{2C_E(d)^2}{1 - \alpha} \right) \left( \frac{\|\Delta d\|n}{(1 - \alpha)\rho_E(d)} \right) + \|d\| \left( \frac{2C_E(d)^2}{1 - \alpha} \right) \left( \frac{2mn^2C_E(d)^2}{(1 - \alpha)\rho_E(d)^2} \right) \left( \frac{3\|\Delta d\|}{1 - \alpha} \right) C_E(d) \]

from which we obtain the following bound:

\[ \|\bar{x} - x\|_1 \leq 14mn^2 \|\Delta d\| \frac{C_E(d)^6}{(1 - \alpha)^3 \rho_E(d)}, \]

which thereby demonstrates the bound (4.1) on \( \|\bar{x} - x\|_1 \).

Finally, observe that \( \bar{t} - t = -\Delta A^T \bar{u} + A^T(u - \bar{u}) \), so that \( \|\bar{t} - t\|_\infty \leq \|\Delta A^T \bar{u}\|_\infty + \|A^T\|_\infty \|u - \bar{u}\|_\infty = \|\Delta A^T \bar{u}\|_\infty + \|A\| \|u - \bar{u}\|_\infty \). Using (4.12) and the bound on \( \|u - \bar{u}\|_\infty \), we obtain

\[ \|\bar{t} - t\|_\infty \leq \frac{\|\Delta d\|n}{(1 - \alpha)\rho_E(d)} + \|d\| \|\Delta d\| \frac{10mn^3C_E(d)^5}{(1 - \alpha)^3 \rho_E(d)^2} \]

\[ \leq 11mn^3 \|\Delta d\| \frac{C_E(d)^6}{(1 - \alpha)^3 \rho_E(d)}, \]

thus proving inequality (4.3).

q.e.d.

We next prove Theorem 4.2, which establishes analogous bounds to those in Theorem 4.1 on changes in optimal solutions to the analytic center problem in its inequality version \( AI(\cdot) \) and its dual \( AID(\cdot) \) as the data is perturbed.
4.3 Changes in Optimal Solutions Under Data Perturbations

**Proof of Theorem 4.2:** The proof of this theorem is very similar to the proof of Theorem 4.1, and so we omit some of the details and only include the most relevant parts. As in the proof of the previous theorem, let \((y, s)\) and \((\tilde{y}, \tilde{s})\) be the optimal solutions to \(AI(d)\) and \(AI(d + \Delta d)\), respectively; and \(v\) and \(\tilde{v}\) be the optimal solutions to \(AID(d)\) and \(AID(d + \Delta d)\), respectively. Then from the Karush-Kuhn-Tucker optimality conditions we have that:

\[
Vs = e, \quad \tilde{V}\tilde{s} = e,
\]

\[
ATy + s = c, \quad (A + \Delta A)^T\tilde{y} + \tilde{s} = c + \Delta c,
\]

\[
Av = 0, \quad (A + \Delta A)\tilde{v} = 0.
\]

\[
v > 0, \quad \tilde{v} > 0.
\]

Therefore, we obtain an analogous result to identity (4.7):

\[
\tilde{v} - v = V\tilde{V} \left( \Delta A^T\tilde{y} - \Delta c \right) + V\tilde{V}A^T(\tilde{y} - y). \quad (4.13)
\]

Observe that \(A(\tilde{v} - v) = -\Delta A\tilde{v}\), and as before, \(A\) has rank \(m\) (otherwise \(\rho_I(d) = 0\)). and so, \(P = AV\tilde{V}A^T\) is a positive definite matrix. By combining this result together with (4.13), we obtain

\[
\tilde{y} - y = -P^{-1}\Delta A\tilde{v} + P^{-1}AV\tilde{V} \left( \Delta c - \Delta A^T\tilde{y} \right). \quad (4.14)
\]

From this identity, it follows that

\[
\|\tilde{y} - y\|_\infty \leq \|P^{-1}\|_1,\infty \left( \|\Delta A\tilde{v}\|_1 + \|A\|\|v\|_1\|\tilde{v}\|_1\|\Delta c - \Delta A^T\tilde{y}\|_\infty \right). \quad (4.15)
\]

From Corollary 4.3, we have

\[
\|\Delta A\tilde{v}\|_1 \leq \|\Delta d\| \frac{n}{(1 - \alpha)\rho_I(d)}, \quad (4.16)
\]
and
\[ \| \Delta c - \Delta A^T \bar{y} \|_\infty \leq \| \Delta d \| \frac{3C_1(d)}{1 - \alpha}. \] (4.17)

Therefore, by combining (4.15), (4.16), and (4.17), and by using Proposition 4.4, Corollary 4.3, and Lemma 4.8, we obtain that
\[
\| \bar{y} - y \|_\infty \leq \left( \frac{16mC_1(d)^4}{1 - \alpha} \right) \frac{\| \Delta d \|}{1 - \alpha} \left( \frac{n}{\rho_1(d)} + \frac{3n^2 \| d \| C_1(d)}{(1 - \alpha)^2 \rho_1(d)^2} \right),
\]
\[
\leq \frac{64mn^2 \| \Delta d \| C_1(d)^6}{(1 - \alpha)^3 \rho_1(d)},
\]
thereby demonstrating the bound (4.4) on \( \| \bar{y} - y \|_\infty \). Now, by substituting equation (4.14) into equation (4.13), we obtain
\[
\bar{v} - v = D^{\frac{1}{2}} \left( I - D^{\frac{1}{2}} A^T P^{-1} A D^{\frac{1}{2}} \right) D^{\frac{1}{2}} \left( \Delta A^T \bar{y} - \Delta c \right) - DA^T P^{-1} \Delta A \bar{v},
\]
where \( D = V \bar{V} \). As before, the matrix \( Q = I - D^{\frac{1}{2}} A^T P^{-1} A D^{\frac{1}{2}} \) is a projection matrix, and so \( \| Q x \|_2 \leq \| x \|_2 \) for all \( x \in \mathbb{R}^n \). Hence, from Proposition 2.1 parts (i) and (iii), we obtain that
\[
\| \bar{v} - v \|_1 \leq n \max_j \{ \bar{v}_j \bar{c}_j \} \| \Delta A^T \bar{y} - \Delta c \|_\infty + \| v \|_1 \| \bar{v} \|_1 \| d \|_1 \| P^{-1} \|_{1, \infty} \| \Delta A \bar{v} \|_1.
\]
It follows from Proposition 4.4, Corollary 4.3, Lemma 4.6, Lemma 4.8, and inequalities (4.16) and (4.17) that
\[
\| \bar{v} - v \|_1 \leq n \left( \frac{n^2}{1 - \alpha \rho_1(d)^2} \right) \left( \frac{3\| \Delta d \| C_1(d)}{1 - \alpha} \right) + \| d \| \left( \frac{n^2}{1 - \alpha \rho_1(d)^2} \right) \left( \frac{16mC_1(d)^4}{1 - \alpha} \right) \left( \frac{n\| \Delta d \|}{(1 - \alpha)^3 \rho_1(d)} \right),
\]
4.3 Changes in Optimal Solutions Under Data Perturbations

from which we obtain the following bound:

$$\|\bar{v} - v\|_1 \leq 19mn^3 \|\Delta d\| \frac{C_I(d)^5}{(1 - \alpha)^3 \rho_I(d)^2},$$

which thereby demonstrates the bound (4.6) on $\|\bar{v} - v\|_1$.

Finally, observe that $s - \bar{s} = \Delta A^T \bar{y} - \Delta c + A^T (\bar{y} - y)$, so that $\|s - \bar{s}\|_\infty \leq \|\Delta A^T \bar{y} - \Delta c\|_\infty + \|A\|\|y - \bar{y}\|_\infty$. Using (4.17) and the bound on $\|y - \bar{y}\|_\infty$, we obtain

$$\|\bar{s} - s\|_\infty \leq \frac{3\|\Delta d\| C_I(d)}{1 - \alpha} + \|d\| \left( \frac{64mn^2\|\Delta d\| C_I(d)^6}{(1 - \alpha)^3 \rho_I(d)} \right) \leq 67mn^2 \|\Delta d\| \frac{C_I(d)^7}{(1 - \alpha)^3},$$

thus proving inequality (4.5).

q.e.d.

The following proposition establishes a relation between the distance to ill-posedness and the approximation error in the optimal objective when using approximate data. Before stating this theorem, we introduce the following notation. Let $d$ be a data instance in $F_E$, then we denote by $z_E(d)$ the corresponding optimal objective value associated with $AE(d)$, that is,

$$z_E(d) = \min \{ p(x) : Ax = b, x > 0 \}.$$

**Theorem 4.3** Let $d = (A, b, c)$ be a data instance in $F_E$ such that $\rho_E(d) > 0$. For a given $\alpha \in [0, 1)$, let $\Delta d = (\Delta A, \Delta b, \Delta c) \in \tilde{D}$ be such that $\|\Delta d\| \leq \alpha \rho_E(d)$. Then

$$|z_E(d + \Delta d) - z_E(d)| \leq \|\Delta d\| \frac{3nC_E(d)}{(1 - \alpha) \rho_E(d)}.$$
4.3 Changes in Optimal Solutions Under Data Perturbations

**Proof:** Let $\hat{x}$ and $\bar{x}$ be the optimal solutions to $AE(d)$ and $AE(d + \Delta d)$, respectively. Let $(\hat{u}, \hat{t})$ and $(\bar{u}, \bar{t})$ be the optimal solutions to $AED(d)$ and $AED(d + \Delta d)$, respectively. Observe that, $z_E(d) = \max_u \min_{x>0} \{p(x) + u^T(b - Ax)\} = \min_{x>0} \max_u \{p(x) + u^T(b - Ax)\}$. Similarly, $z_E(d + \Delta d) = \max_u \min_{x>0} \{p(x) + u^T(b + \Delta b - (A + \Delta A)x)\} = \min_{x>0} \max_u \{p(x) + u^T(b + \Delta b - (A + \Delta A)x)\}$. Hence, we have

$$z_E(d) = p(\hat{x}) + \hat{u}^T(b - A\hat{x})$$

$$= \max_u \{p(\hat{x}) + u^T(b - A\hat{x})\}$$

$$= \max_u \{p(\hat{x}) + u^T(b + \Delta b - (A + \Delta A)\hat{x}) + u^T(\Delta A\hat{x} - \Delta b)\}$$

$$\geq p(\hat{x}) + \hat{u}^T(b + \Delta b - (A + \Delta A)\hat{x}) + \hat{u}^T(\Delta A\hat{x} - \Delta b)$$

$$\geq z_E(d + \Delta d) + \hat{u}^T(\Delta A\hat{x} - \Delta b).$$

Thus $z_E(d) - z_E(d + \Delta d) \geq \hat{u}^T(\Delta A\hat{x} - \Delta b)$. Similarly, we can prove that $z_E(d) - z_E(d + \Delta d) \leq \hat{u}^T(\Delta A\bar{x} - \Delta b)$. Therefore, $|z_E(d) - z_E(d + \Delta d)| \leq |\hat{u}^T(\Delta A\hat{x} - \Delta b)|$, or $|z_E(d) - z_E(d + \Delta d)| \leq |\hat{u}^T(\Delta A\bar{x} - \Delta b)|$.

On the other hand, using Hölder’s inequality and the bounds from Lemma 4.1 and Corollary 4.2, we have

$$|\hat{u}^T(\Delta A\hat{x} - \Delta b)| \leq \|\hat{u}\|_\infty (\|\Delta b\|_1 + \|\Delta A\|\|\hat{x}\|_1)$$

$$\leq \frac{n}{(1 - \alpha)\rho_E(d)} (\|\Delta d\| + \|\Delta d\|C_E(d))$$

$$\leq 2n \|\Delta d\| \frac{C_E(d)}{(1 - \alpha)\rho_E(d)}.$$

Analogously, we have from Proposition 4.3 and Corollary 4.2 that

$$|\bar{u}^T(\Delta A\bar{x} - \Delta b)| \leq \|\bar{u}\|_\infty (\|\Delta b\|_1 + \|\Delta A\|\|\bar{x}\|_1)$$

$$\leq \frac{n}{\rho_E(d)} \left(\|\Delta d\| + \|\Delta d\| \frac{2}{1 - \alpha}C_E(d)\right).$$
4.3 Changes in Optimal Solutions Under Data Perturbations

\[ \leq 3n \|\Delta d\| \frac{C_E(d)}{(1-\alpha)\rho_E(d)}, \]

and the result follows.

q.e.d.

In order to state the last theorem of this section, we introduce the following notation. Let \( d \) be a data instance in \( \mathcal{F}_I \), then we denote by \( z_I(d) \) the corresponding optimal objective value associated with \( AI(d) \), that is,

\[ z_I(d) = \min \left\{ p(s) : A^T y + s = c, s > 0 \right\}. \]

**Theorem 4.4** Let \( d = (A, b, c) \) be a data instance in \( \mathcal{F}_I \) such that \( \rho_I(d) > 0 \). For a given \( \alpha \in [0, 1) \), let \( \Delta d = (\Delta A, \Delta b, \Delta c) \in \mathcal{D} \) be such that \( \|\Delta d\| \leq \alpha \rho_I(d) \). Then

\[ |z_I(d + \Delta d) - z_I(d)| \leq \|\Delta d\| \frac{3nC_I(d)}{(1-\alpha)\rho_I(d)}. \]

**Proof:** Let \( (\hat{y}, \hat{s}) \) and \( (\bar{y}, \bar{s}) \) be the optimal solutions to \( AI(d) \) and \( AI(d + \Delta d) \), respectively. Let \( \hat{v} \) and \( \bar{v} \) be the optimal solutions to \( AID(d) \) and \( AID(d + \Delta d) \), respectively. As in the previous theorem, we have

\[ z_I(d) = \max_{v > 0} \min_{(y, s), s > 0} \left\{ p(s) - v^T (c - A^T y - s) \right\} = \min_{(y, s), s > 0} \max_{v > 0} \left\{ p(s) - v^T (c - A^T y - s) \right\}. \]

Similarly,

\[ z_I(d + \Delta d) = \max_{v > 0} \min_{(y, s), s > 0} \left\{ p(s) - v^T (c + \Delta c - (A + \Delta A)^T y - s) \right\} = \min_{(y, s), s > 0} \max_{v > 0} \left\{ p(s) - v^T (c + \Delta c - (A + \Delta A)^T y - s) \right\}. \]

Hence, we have

\[
\begin{align*}
z_I(d) &= p(\hat{s}) - \hat{v}^T (c - A^T \hat{y} - \hat{s}) \\
&= \max_{v > 0} \left\{ p(\hat{s}) - v^T (c - A^T \hat{y} - \hat{s}) \right\} \\
&\geq p(\hat{s}) - \hat{v}^T \left( c + \Delta c - (A + \Delta A)^T \hat{y} - \hat{s} \right) - v^T \left( \Delta A^T \hat{y} - \Delta c \right) \\
&\geq z_I(d + \Delta d) - \hat{v}^T \left( \Delta A^T \hat{y} - \Delta c \right).
\end{align*}
\]
Thus $z_I(d) - z_I(d + \Delta d) \geq -\bar{\delta}^T (\Delta A^T \bar{y} - \Delta c)$. Similarly, we can prove that $z_I(d) - z_I(d + \Delta d) \leq -\bar{\delta}^T (\Delta A^T \bar{y} - \Delta c)$. Therefore, $|z_I(d) - z_I(d + \Delta d)| \leq \bar{\delta}^T (\Delta A^T \bar{y} - \Delta c)$, or $|z_I(d) - z_I(d + \Delta d)| \leq \bar{\delta}^T (\Delta A^T \bar{y} - \Delta c)$.

On the other hand, using Hölder's inequality and the bounds from Lemma 4.2 and Corollary 4.3, we have

$$||\bar{\delta}^T (\Delta A^T \bar{y} - \Delta c)|| \leq ||\bar{\delta}||_1 (||\Delta c||_\infty + ||\Delta A|| ||\bar{y}||_\infty)$$

$$\leq \frac{n}{(1 - \alpha) \rho_I(d)} (||\Delta d|| + ||\Delta d|| |C_I(d)|)$$

$$\leq 2n ||\Delta d|| \frac{C_I(d)}{(1 - \alpha) \rho_I(d)}.$$

Analogously, we have from Proposition 4.4 and Corollary 4.3 that

$$||\bar{\delta}^T (\Delta A^T \bar{y} - \Delta c)|| \leq ||\bar{\delta}||_1 (||\Delta c||_\infty + ||\Delta A|| ||\bar{y}||_\infty)$$

$$\leq \frac{n}{\rho_I(d)} (||\Delta d|| + ||\Delta d|| \frac{2}{1 - \alpha} C_I(d))$$

$$\leq 3n ||\Delta d|| \frac{C_I(d)}{(1 - \alpha) \rho_I(d)}.$$

and the result follows.

q.e.d.
Chapter 5

Central Trajectory Problems for Linear Programming

5.1 Overview

As described in Chapter 2, the central trajectory of a linear program consists of the set of optimal solutions \( x = x(\mu) \) and \((y, s) = (y(\mu), s(\mu))\) to the logarithmic barrier problems:

\[
\begin{align*}
P_\mu(d) & : \min \left\{ c^T x + \mu p(x) : Ax = b, x > 0 \right\}, \\
D_\mu(d) & : \max \left\{ b^T y - \mu p(s) : A^T y + s = c, s > 0 \right\},
\end{align*}
\]

where for \( u > 0 \) in \( \mathbb{R}^n \), \( p(u) = -\sum_{j=1}^{n} \ln(u_j) \) is the logarithmic barrier function, \( d = (A, b, c) \) is a data instance in the space of all data \( \bar{D} = \{(A, b, c) : A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n\} \), and the parameter \( \mu \) is a positive scalar considered independent of the data instance \( d = (A, b, c) \in \bar{D} \). The central trajectory is fundamental to the study of interior-point algorithms for linear programming, and has been the subject of an enormous volume of research, see among many others, the references cited in the sur-
5.1 Overview

veys by Gonzaga [Gon92] and Jansen et al. [JRT95], and the book by Wright [Wri97]. It is well known that programs $P_\mu(d)$ and $D_\mu(d)$ are related through Lagrangian duality; if each program is feasible, then both programs attain their optima, and optimal solutions $x = x(\mu)$ and $(y, s) = (y(\mu), s(\mu))$ satisfy $c^T x - b^T y = n\mu$, and hence exhibit a linear programming duality gap of $n\mu$ for the dual linear programming problems associated with $P_\mu(d)$ and $D_\mu(d)$.

From Theorem 3.1, identity (3.1), given a data instance $d \in F$ and $\rho(d) > 0$ (so that $d \in \text{int}(F)$), it follows from the strict convexity of the logarithmic barrier and the full rank of $A$ that the programs $P_\mu(d)$ and $D_\mu(d)$ will each have a unique optimal solution.

The purpose of this chapter is to explore and demonstrate properties of solutions to $P_\mu(d)$ and $D_\mu(d)$ that are inherently related to the condition number $C(d)$ of the data instance $d = (A, b, c)$. As discussed previously, in the context of the central trajectory problem, $\rho(d)$ essentially is the minimum change $\Delta d = (\Delta A, \Delta b, \Delta c)$ in the data $d = (A, b, c)$ necessary to create a data instance $d + \Delta d$ that is an infeasible instance of $P_\mu(\cdot)$ or $D_\mu(\cdot)$. The condition number of the data instance $d = (A, b, c)$, denoted $C(d)$, is defined to be $C(d) := \|d\|/\rho(d)$ and is a scale-invariant reciprocal of the distance to ill-posedness $\rho(d)$, so that $C(d)$ goes to $\infty$ as the data instance $d$ approaches infeasibility.

The main results in the chapter are stated in Sections 5.2 and 5.3. In Section 5.2 we present upper and lower bounds on sizes of optimal solutions to the barrier problems $P_\mu(d)$ and $D_\mu(d)$ in terms of the conditioning of the data instance $d$. Theorems 5.1 and 5.2 state bounds on such solutions that are linear in $\mu$, where the constants in the bounds are polynomial functions of the condition number $C(d)$, the distance
to ill-posedness $\rho(d)$, the dimension $n$, the norm of the data $\|d\|$, or their inverses. These theorems show in particular that as $\mu$ goes to zero, that $x_j(\mu)$ grows at least linearly in $\mu$; and as $\mu$ goes to $\infty$, $x_j(\mu)$ grows at most linearly in $\mu$. Moreover, in Theorem 5.3, we also show that when the feasible region of $P_\mu(d)$ is unbounded, then certain coordinates of $x(\mu)$ grow exactly linearly in $\mu$ as $\mu \to \infty$, all at rates bounded by polynomial functions of the condition number $C(d)$, the distance to ill-posedness $\rho(d)$, the dimension $n$, the norm of the data $\|d\|$, or their inverses.

In Section 5.3, we study the sensitivity of the optimal solutions to $P_\mu(d)$ and $D_\mu(d)$ as either the data $d = (A, b, c)$ changes or the barrier parameter $\mu$ changes. Theorems 5.4 and 5.7 state upper bounds on the sizes of the changes on optimal solutions as well as in the optimal objective values as the data $d = (A, b, c)$ is changed. Theorems 5.6 and 5.8 state upper bounds on the sizes of changes in optimal solutions and optimal objective values as the barrier parameter $\mu$ is changed. Along the way, we prove Theorem 5.5, which states bounds on the norm of the matrix $(AX^2(\mu)A^T)^{-1}$. This matrix is the main computational matrix in interior-point central trajectory methods. All of the bounds in this section are polynomial functions of the condition number $C(d)$, the distance to ill-posedness $\rho(d)$, the dimension $n$, the norm of the data $\|d\|$, or their inverses.

5.2 Bounds on Solutions Along the Central Trajectory

This section presents results on lower and upper bounds on sizes of optimal solutions along the central trajectory, for the pair of dual logarithmic barrier problems $P_\mu(d)$ and $D_\mu(d)$. As developed in the previous section, $d = (A, b, c)$ represents a data
5.2 Bounds on Solutions Along the Central Trajectory

instance. Before presenting the first bound, we define the following scalar quantity, denoted $\mathcal{K}(d, \mu)$, which appears in many of the results of this section as well as in Section 5.3:

$$\mathcal{K}(d, \mu) = C(d)^2 + \frac{\mu n}{\rho(d)}. \tag{5.1}$$

The first result concerns upper bounds on sizes of optimal solutions.

**Theorem 5.1** If $d = (A, b, c) \in \mathcal{F}$ and $\rho(d) > 0$, then

$$\|x(\mu)\|_1 \leq \mathcal{K}(d, \mu), \tag{5.2}$$

$$\|y(\mu)\|_{\infty} \leq \mathcal{K}(d, \mu), \tag{5.3}$$

$$\|s(\mu)\|_{\infty} \leq 2\|d\|\mathcal{K}(d, \mu), \tag{5.4}$$

for the optimal solution $x(\mu)$ to $P_\mu(d)$ and the optimal solution $(y(\mu), s(\mu))$ to the dual problem $D_\mu(d)$, where $\mathcal{K}(d, \mu)$ is the scalar defined in (5.1).

This theorem states that the norms of optimal solutions along the central trajectory are bounded above by quantities only involving the condition number $C(d)$ and the distance to ill-posedness $\rho(d)$ of the data $d$, as well as the dimension $n$ and the barrier parameter $\mu$. Furthermore, for example, the theorem shows that the norm of the optimal primal solution along the central trajectory grows at most linearly in the barrier parameter $\mu$, and at a rate no larger than $n/\rho(d)$, as $\mu$ goes to $\infty$.

**Proof of Theorem 5.1:** Let $\hat{x} = x(\mu)$ be the optimal solution to $P_\mu(d)$ and $(\hat{y}, \hat{s}) = (y(\mu), s(\mu))$ be the optimal solution to the corresponding dual problem $D_\mu(d)$. Note that the optimality conditions of $P_\mu(d)$ and $D_\mu(d)$ imply that $c^T\hat{x} = b^T\hat{y} + \mu n$.

Observe that since $\hat{s} = c - AT\hat{y}$, then $\|\hat{s}\|_{\infty} \leq \|c\|_{\infty} + \|AT\|_{\infty}\|\hat{y}\|_{\infty}$. Since $\|AT\|_{\infty} = \|A\|$, we have $\|\hat{s}\|_{\infty} \leq \|d\|(1 + \|\hat{y}\|_{\infty})$, and using the fact that $C(d) \geq 1$ the bound
(5.4) on $\|\hat{y}\|_\infty$ is a consequence of the bound (5.3) on $\|\hat{y}\|_\infty$. It therefore is sufficient to prove the bounds on $\|\hat{x}\|_1$ and on $\|\hat{y}\|_\infty$. In addition, the bound on $\|\hat{y}\|_\infty$ is trivial if $\hat{y} = 0$, so from now on we assume that $\hat{y} \neq 0$. Also, let $\bar{y}$ be a vector in $\mathbb{R}^m$ such that $\bar{y}^T \hat{y} = \|\hat{y}\|_\infty$ and $\|\bar{y}\|_1 = 1$.

The rest of the proof proceeds by examining three cases:

(i) $c^T \hat{x} \leq 0$,

(ii) $0 < c^T \hat{x} \leq \mu n$, and

(iii) $\mu n < c^T \hat{x}$.

In case (i), let $\Delta A = -be^T/\|\hat{x}\|_1$. Then $(A + \Delta A) \hat{x} = 0$, $\hat{x} > 0$, and $c^T \hat{x} \leq 0$. From Proposition 2.8, we have $D_\mu(d + \Delta d)$ is infeasible, and so $\rho(d) \leq \|\Delta d\| = \|\Delta A\| = \|b\|_1/\|\hat{x}\|_1 \leq \|d\|/\|\hat{x}\|_1$. Therefore, $\|\hat{x}\|_1 \leq \|d\|/\rho(d) = C(d) \leq K(d, \mu)$, since $C(d) \geq 1$ for any $d$. This proves (5.2) in this case.

Let $\theta = b^T \hat{y}$, $\Delta b = -\theta \bar{y}/\|\hat{y}\|_\infty$, $\Delta A = -\bar{y}c^T/\|\hat{y}\|_\infty$, and $d + \Delta d = (A + \Delta A, b + \Delta b, c)$. Observe that $(b + \Delta b)^T \hat{y} = 0$ and $(A + \Delta A)^T \hat{y} < 0$, so that $P_\mu(d + \Delta d)$ is infeasible from Proposition 2.7. Therefore, $\rho(d) \leq \|\Delta d\| = \max\{\|c\|_\infty, |\theta|\}/\|\hat{y}\|_\infty$. Hence, $\|\hat{y}\|_\infty \leq \max\{C(d), |\theta|/\rho(d)\}$. Furthermore, $|\theta| = |b^T \hat{y}| = |c^T \hat{x} - \mu n| \leq \|\hat{x}\|_1 \|c\|_\infty + \mu n \leq C(d)\|d\| + \mu n$. Therefore, again using the fact that $C(d) \geq 1$ for any $d$, we have (5.3).

In case (ii), let $d + \Delta d = (A + \Delta A, b, c + \Delta c)$, where $\Delta A = -be^T/\|\hat{x}\|_1$ and $\Delta c = -\mu n e/\|\hat{x}\|_1$. Observe that $(A + \Delta A) \hat{x} = 0$ and $(c + \Delta c)^T \hat{x} \leq 0$. From Proposition 2.8, $D_\mu(d + \Delta d)$ is infeasible, and so we conclude that $\rho(d) \leq \|\Delta d\| = \max\{\|\Delta A\|, \|\Delta c\|_\infty\} = \max\{\|b\|_1, \mu n\}/\|\hat{x}\|_1 \leq (\|d\| + \mu n)/\|\hat{x}\|_1$. Therefore, $\|\hat{x}\|_1 \leq C(d) + \mu n / \rho(d) \leq K(d, \mu)$. This proves (5.2) for this case.

Now, let $d + \Delta d = (A + \Delta A, b + \Delta b, c)$, where $\Delta A = -\bar{y}c^T/\|\hat{y}\|_\infty$ and $\Delta b = \mu n \bar{y}/\|\hat{y}\|_\infty$. Observe that $(b + \Delta b)^T \hat{y} = b^T \hat{y} + \mu n = c^T \hat{x} > 0$ and $(A + \Delta A)^T \hat{y} < 0$. Again, from Proposition 2.7, $P_\mu(d + \Delta d)$ is infeasible, and so we conclude that
5.2 Bounds on Solutions Along the Central Trajectory

\[ \rho(d) \leq \| \Delta d \| = \max \{ \| \Delta A \|, \| \Delta b \| \} = \max \{ \| c \|_\infty, \mu n \} / \| \hat{y} \|_\infty \leq (\| d \| + \mu n) / \| \hat{y} \|_\infty. \]

Therefore, \( \| \hat{y} \|_\infty \leq C(d) + \mu n / \rho(d) \leq K(d, \mu) \).

In case (iii), we first consider the bound on \( \| \hat{y} \|_\infty \). Let \( d + \Delta d = (A + \Delta A, b, c) \), where \( \Delta A = -\hat{y} c^T / \| \hat{y} \|_\infty \). Since \( (A + \Delta A)^T \hat{y} < 0 \) and \( b^T \hat{y} = c^T \hat{x} - \mu n > 0 \), it follows from Proposition 2.7 that \( P_\mu(d + \Delta d) \) is infeasible and so, \( \rho(d) \leq \| \Delta d \| = \| c \|_\infty / \| \hat{y} \|_\infty \).

Therefore, \( \| \hat{y} \|_\infty \leq C(d) \leq K(d, \mu) \).

Finally, let \( \Delta A = -be^T / \| \hat{x} \|_1 \) and \( \Delta c = -\theta e / \| \hat{x} \|_1 \), where \( \theta = c^T \hat{x} \). Observe that \( (A + \Delta A) \hat{x} = 0 \) and \( (c + \Delta c)^T \hat{x} = 0 \). Using Proposition 2.8, we conclude that \( D_\mu(d + \Delta d) \) is infeasible and so, \( \rho(d) \leq \| \Delta d \| = \max \{ \| \Delta A \|, \| \Delta c \|_\infty \} = \max \{ \| b \|_1, \theta \} / \| \hat{x} \|_1 \), so that \( \| \hat{x} \|_1 \leq \max \{ C(d), \theta / \rho(d) \} \). Furthermore, \( \theta = c^T \hat{x} = b^T \hat{y} + \mu n \leq \| b \|_1 \| \hat{y} \|_\infty + \mu n \leq \| d \| C(d) + \mu n \). Therefore, \( \| \hat{x} \|_1 \leq K(d, \mu) \).

q.e.d.

Note that the scalar quantity \( K(d, \mu) \) appearing in Theorem 5.1 is scale invariant in the sense that \( K(\lambda d, \lambda \mu) = K(d, \mu) \) for any \( \lambda > 0 \). From this it follows that the bounds in Theorem 5.1 on \( \| x(\mu) \|_1 \) and \( \| y(\mu) \|_\infty \) are also scale invariant. However, as one would expect, the bound on \( \| s(\mu) \|_\infty \) is not scale invariant, since \( \| s(\mu) \|_\infty \) is sensitive to a positive scaling of the data. Moreover, observe that as \( \mu \to 0 \) the bounds in Theorem 5.1 converge to the bounds presented by Vera in [Ver96] for optimal solutions to linear programs of the form \( \min \{ c^T x : Ax = b, x \geq 0 \} \).

Examining the proof of Theorem 5.1, it is clear that the bounds stated in Theorem 5.1 will not generally be achieved. Indeed, implicit in the proof is the fact that bounds tighter than those in the theorem can be proved, and will depend on which of the three cases in the proof are applicable. However, our goal lies mainly in establishing bounds that are polynomial in the condition number \( C(d) \), the parameter \( \mu \), the size of the data \( \| d \| \), and the dimensions \( m \) and \( n \), and not necessarily in establishing
the best achievable bounds.

We now present a simple example illustrating that the bounds in Theorem 5.1 are not necessarily tight. Let \( m = 1, n = 2 \), and

\[
d = (A, b, c) = \begin{pmatrix}
[1, 1], [1], \\
[1 - 1]
\end{pmatrix}.
\]

For this data instance, we have \( \|d\| = 1 \) and \( \rho(d) = 1 \), so that \( C(d) = 1 \) and \( \mathcal{K}(d, \mu) = 1 + n\mu \). Now observe that \( x(\mu) = (1/2, 1/2)^T \) for all \( \mu > 0 \), so that \( \|x(\mu)\|_1 = 1 < \mathcal{K}(d, \mu) = 1 + n\mu \) for all \( \mu > 0 \), which demonstrates that (5.2) is not tight in general. Furthermore, notice that in this example \( c^T x(\mu) < 0 \), and so case (i) of the proof implies that \( \|x(\mu)\|_1 \leq C(d) \) (in fact, \( \|x(\mu)\|_1 = C(d) = 1 \) in this example), which is a tighter bound than (5.2).

**Corollary 5.1** For a given \( \alpha \in [0, 1) \), let \( \delta \) be such that \( \delta \leq \alpha \rho(d) \), where \( d \in \mathcal{F} \) and \( \rho(d) > 0 \). If \( \Delta d \in \bar{D} \) is such that \( \|\Delta d\| \leq \delta \), then

\[
\begin{align*}
\|x(\mu)\|_1 & \leq \left(1 + \frac{\alpha}{1 - \alpha}\right)^2 \mathcal{K}(d, \mu), \\
\|y(\mu)\|_\infty & \leq \left(1 + \frac{\alpha}{1 - \alpha}\right)^2 \mathcal{K}(d, \mu), \\
\|s(\mu)\|_\infty & \leq 2(\|d\| + \delta) \left(1 + \frac{\alpha}{1 - \alpha}\right)^2 \mathcal{K}(d, \mu),
\end{align*}
\]

where \( x(\mu) \) is the optimal solution to \( P_\mu(d + \Delta d) \), \( (y(\mu), s(\mu)) \) is the optimal solution to \( D_\mu(d + \Delta d) \), and \( \mathcal{K}(d, \mu) \) is the scalar defined in (5.1).

**Proof:** The proof follows by observing that for \( \bar{d} \in B(d, \delta) \) we have \( \|\bar{d}\| \leq \|d\| + \delta \), and \( \rho(\bar{d}) \geq (1 - \alpha) \rho(d) \), so that

\[
C(\bar{d}) \leq \frac{\|d\| + \delta}{(1 - \alpha) \rho(d)} = \left(\frac{1 + \alpha}{1 - \alpha}\right)(C(d) + \delta/\rho(d)) \leq \left(\frac{1 + \alpha}{1 - \alpha}\right)(C(d) + \alpha) \leq C(d) \left(\frac{1 + \alpha}{1 - \alpha}\right),
\]
5.2 Bounds on Solutions Along the Central Trajectory

since $C(d) \geq 1$.

**q.e.d.**

Note that for a fixed value of $\alpha$ that Corollary 5.1 shows that the norms of solutions to any suitably perturbed problem are uniformly upper-bounded by a fixed constant times the upper bounds on the solutions to the original problem.

The next result presents a lower bound on the norm of the optimal solutions $x(\mu)$ and $s(\mu)$ to the central trajectory problems $P_\mu(d)$ and $D_\mu(d)$, respectively.

**Theorem 5.2** If the program $P_\mu(d)$ has an optimal solution and $\rho(d) > 0$, then

\[
\begin{align*}
\|x(\mu)\|_1 & \geq \frac{\mu n}{2\|d\|K(d, \mu)}, \\
\|s(\mu)\|_\infty & \geq \frac{\mu n}{K(d, \mu)}, \\
x_j(\mu) & \geq \frac{\mu}{2\|d\|K(d, \mu)}, \\
s_j(\mu) & \geq \frac{\mu}{K(d, \mu)},
\end{align*}
\]

for all $j = 1, \ldots, n$, where $x(\mu)$ is the optimal solution to $P_\mu(d)$, $(y(\mu), s(\mu))$ is the optimal solution to $D_\mu(d)$, and $K(d, \mu)$ is the scalar defined in (5.1).

This theorem shows that $\|x(\mu)\|_1$ and $x_j(\mu)$ are bounded from below by functions only involving the quantities $\|d\|$, $C(d)$, $\rho(d)$, $n$, and $\mu$. In addition, the theorem shows that for $\mu$ close to zero, that $x_j(\mu)$ grows at least linearly in $\mu$, and at a rate that is at least $1/(2\|d\|C(d)^2)$ (since $K(d, \mu) = C(d)^2 + \mu n/\rho(d) \approx C(d)^2$ near $\mu = 0$). Furthermore, the theorem also shows that for $\mu$ close to zero, that $s_j(\mu)$ grows at least linearly in $\mu$, and at a rate that is at least $1/C(d)^2$. 
5.2 Bounds on Solutions Along the Central Trajectory

The theorem offers less insight when \( \mu \to \infty \), since the lower bound on \( \|x(\mu)\|_1 \) presented in the theorem converges to \( (2C(d))^{-1} \) as \( \mu \to \infty \). When the feasible region is unbounded, it is well known (see also the results at the end of this section) that \( \|x(\mu)\| \to \infty \) as \( \mu \to \infty \), so that as \( \mu \to \infty \) the lower bound of Theorem 5.2 does not adequately capture the behavior of the sizes of optimal solutions to \( P_\mu(d) \) when the feasible region is unbounded. We will present a more relevant bound shortly, in Theorem 5.3. Similar remarks apply to the bound on \( \|s(\mu)\|_\infty \) as \( \mu \to \infty \).

Proof of Theorem 5.2: By the Karush-Kuhn-Tucker optimality conditions of the dual pair of problems \( P_\mu(d) \) and \( D_\mu(d) \), we have \( s(\mu)^T \! x(\mu) = \mu n \). Since \( s(\mu)^T x(\mu) \leq \|s(\mu)\|_\infty \|x(\mu)\|_1 \), it follows that \( \|x(\mu)\|_1 \geq \mu n / \|s(\mu)\|_\infty \) and \( \|s(\mu)\|_\infty \geq \mu n / \|x(\mu)\|_1 \). Therefore, the first two inequalities follow from Theorem 5.1.

For the remaining inequalities, observe that for each \( j = 1, \ldots, n \), \( \mu = s_j(\mu) x_j(\mu) \), \( x_j(\mu) \leq \|x(\mu)\|_1 \), and \( s_j(\mu) \leq \|s(\mu)\|_\infty \). Therefore, the result follows again from Theorem 5.1.

q.e.d.

The following corollary uses Theorem 5.2 to provide lower bounds for solutions to perturbed problems.

Corollary 5.2 For a given \( \alpha \in [0, 1) \), let \( \delta \) be such that \( \delta \leq \alpha \rho(d) \), where \( d \in \mathcal{F} \) and \( \rho(d) > 0 \). If \( \Delta d \in \tilde{\mathcal{D}} \) is such that \( \|\Delta d\| \leq \delta \), then

\[
\|x(\mu)\|_1 \geq \left( \frac{1 - \alpha}{1 + \alpha} \right)^2 \frac{\mu n}{2(\|d\| + \delta) K(d, \mu)},
\]

\[
\|s(\mu)\|_\infty \geq \left( \frac{1 - \alpha}{1 + \alpha} \right)^2 \frac{\mu n}{K(d, \mu)},
\]

\[
x_j(\mu) \geq \left( \frac{1 - \alpha}{1 + \alpha} \right)^2 \frac{\mu}{2(\|d\| + \delta) K(d, \mu)},
\]

\[
s_j(\mu) \geq \left( \frac{1 - \alpha}{1 + \alpha} \right)^2 \frac{\mu}{K(d, \mu)},
\]
5.2 Bounds on Solutions Along the Central Trajectory

for all \( j = 1, \ldots, n \), where \( x(\mu) \) is the optimal solution to \( P_\mu(d + \Delta d) \), \((y(\mu), \varepsilon(\mu))\) is the optimal solution to \( D_\mu(d + \Delta d) \), and \( K(d, \mu) \) is the scalar defined in (5.1).

**Proof:** The proof follows the same logic as that of Corollary 5.1.

q.e.d.

Note that for a fixed value of \( \alpha \) that Corollary 5.2 shows that the norms of solutions to any suitably perturbed problem are uniformly lower-bounded by a fixed constant times the lower bounds on the solutions to the original problem.

The last result of this section, Theorem 5.3, presents different lower bounds on components of \( x(\mu) \) along the central trajectory, that are relevant when \( \mu \to \infty \) and when the primal feasible region is unbounded. For a given data instance \( d = (A, b, c) \in \mathcal{D} \) and a set \( B \subset \{1, \ldots, n\} \), in this theorem we use a data instance \( d_B \in \mathbb{R}^{m \times |B| + m + |B|} \) defined as \( d_B := (A_B, b, c_B) \). Associated with the data instance \( d_B \) there is an analytic center problem in inequality form (see Chapter 4) given by

\[
AI(d_B) : \min \left\{ p(s_B) : A_B^T y_B + s_B = c_B, s_B > 0 \right\},
\]

and corresponding condition number \( C_I(d_B) = \|d_B\|/\rho_I(d_B) \). We assume that only perturbations of the form \( \Delta d_B = (\Delta A_B, \Delta b, \Delta c_B) \) are considered to compute \( \rho_I(d_B) \), and hence to compute the condition number \( C_I(d_B) \) associated with the problem \( AI(d_B) \). Also, notice that \( C_I(d_B) \) is independent of the scalar \( \mu \).

**Theorem 5.3** Let \( d \in \mathcal{F} \) and \( \rho(d) > 0 \). Let \( x(\mu) \) denote the optimal solution to \( P_\mu(d) \) and \((y(\mu), s(\mu))\) denote the optimal solution to \( D_\mu(d) \). Then there exists a unique partition of the indices \( \{1, \ldots, n\} \) into two subsets \( B \) and \( N \) such that

\[
x_j(\mu) \geq \frac{\mu}{2\|d\|C_I(d_B)},
\]
5.2 Bounds on Solutions Along the Central Trajectory

\[ s_j(\mu) \leq 2\|d\| C_1(d_B), \]

for all \( j \in B \), and \( x_j(\mu) \) is uniformly bounded for all \( \mu \geq 0 \) for all \( j \in N \), where \( d_B = (A_B, b, c_B) \) is a data instance in \( \mathbb{R}^{m \times |B| + |m + |B|} \) composed of those elements of \( d \) indexed by the set \( B \).

Note that the set \( B \) is the index set of components of \( x \) that are unbounded over the feasible region of \( P_\mu(d) \), and \( N \) is the index set of components of \( x \) that are bounded over the feasible region of \( P_\mu(d) \). Theorem 5.3 states that as \( \mu \to \infty \), that \( x_j(\mu) \) for \( j \in B \) will go to \( \infty \) at least linearly in \( \mu \) as \( \mu \to \infty \), and at a rate that is at least \( 1/(2\|d\| C_1(d_B)) \). Of course, from Theorem 5.3, it also follows that when the feasible region of \( P_\mu(d) \) is unbounded, that is, \( B \neq \emptyset \), that \( \lim_{\mu \to \infty} \|x(\mu)\|_1 = \infty \). Finally, note that Theorem 5.1 combined with Theorem 5.3 state that as \( \mu \to \infty \), that \( x_j(\mu) \) for \( j \in B \) will go to \( \infty \) exactly linearly in \( \mu \).

**Proof of Theorem 5.3:** From Tucker’s strict complementarity theorem (see Dantzig [Dan63], p. 139, and [Wil70]), there exists a unique partition \([B, N]\) of the set \( \{1, \ldots, n\} \) into subsets \( B \) and \( N \), \( B \cap N = \emptyset \) and \( B \cup N = \{1, \ldots, n\} \) satisfying the following two properties:

1. \( Au = 0, u \geq 0 \) implies \( u_N = 0 \) and there exists \( \hat{u} \) for which \( A\hat{u} = 0, \hat{u}_B > 0 \), and \( \hat{u}_N = 0 \),

2. \( A^T y = v, v \leq 0 \) implies \( v_B = 0 \) and there exists \((\hat{y}, \hat{v})\) for which \( A^T \hat{y} = \hat{v}, \hat{v}_B = 0 \), and \( \hat{v}_N < 0 \).

Consider the set \( S = \{ s_B \in \mathbb{R}^{|B|} : s_B = c_B - A_B^T y \text{ for some } y \in \mathbb{R}^m, s_B > 0 \} \). Because \( P_\mu(d) \) has an optimal solution, \( S \) is non-empty. Also, \( S \) is bounded. To see this, suppose instead that \( S \) is unbounded, in which case there exists \( \hat{y} \) such that \( A_B^T \hat{y} \leq 0 \) and \( A_B^T \hat{y} \neq 0 \). Then, using the vector \( \hat{y} \) from property 2 above, we obtain that
5.2 Bounds on Solutions Along the Central Trajectory

$A_T^N(\tilde{y} + \lambda \tilde{y}) = A_T^N\tilde{y} + \lambda \hat{v}_N \leq 0$ for $\lambda$ sufficiently large, and since $A_T^B\tilde{y} = \hat{v}_B = 0$, it follows that $A^T(\tilde{y} + \lambda \tilde{y}) \leq 0$ for $\lambda$ sufficiently large. By the definition of the partition $[B, N]$, we have $A_T^B(\tilde{y} + \lambda \tilde{y}) = 0$. This in turn implies that $A_T^B\tilde{y} = 0$, a contradiction.

Because $S$ is non-empty and bounded, $d_B = (A_B, b, c_B) \in F_I$, that is, $d_B$ belongs to the set of feasible data instances for the analytic center problem in inequality form (see Chapter 4). Moreover, we have $\rho_I(d_B) > 0$. If $\rho_I(d_B) = 0$, then for each $\epsilon > 0$ there exists a perturbation array $\Delta d_B = (\Delta A_B, \Delta b, \Delta c_B)$ such that $\|\Delta d_B\| \leq \epsilon$ and $AI(d_B + \Delta d_B)$ is not feasible. By letting $\Delta d = ([\Delta A_B, 0], \Delta b, [\Delta c_B, 0])$, it follows that $\|\Delta d\| \leq \epsilon$ and $D_\mu(d + \Delta d)$ is not feasible, for all $\epsilon > 0$, a contradiction with our assumption $\rho(d) > 0$.

Therefore, by Lemma 4.2, for any $s_B \in S$, $\|s_B\|_\infty \leq 2\|d_B\|C_I(d_B)$, in particular

$$\|s_B(\mu)\|_\infty \leq 2\|d_B\|C_I(d_B) \leq C_I(d_B).$$

Hence, for any $j \in B$, $s_j(\mu) = \|s_B(\mu)\|_\infty \leq 2\|d\|C_I(d_B)$. Moreover, since $x_j(\mu) s_j(\mu) = \mu$, then

$$x_j(\mu) \geq \frac{\mu}{2\|d\|C_I(d_B)},$$

for $j \in B$.

Finally, by definition of the partition of $\{1, \ldots, n\}$ into $B$ and $N$, $x_j(\mu)$ is bounded for all $j \in N$ and for all $\mu > 0$. This also ensures that $B$ is unique.

q.e.d.

We end this section with the following remark concerning the scalar quantity $\mathcal{K}(d, \mu)$ defined in (5.1). Rather than using the quantity $\mathcal{K}(d, \mu)$, the results in this section could alternatively have been expressed in terms of the following scalar quantity:

$$\mathcal{R}(d, \mu) = \left( \frac{\max \{\|d\|, n\mu\}}{\min \{\rho(d), \mu\}} \right)^2.$$ (5.5)
One can think of the quantity $\mathcal{R}(d, \mu)$ as the square of the condition number of the data instance $(A, b, c, \mu)$ associated with the problem $P_\mu(d)$, where now $\mu > 0$ is considered as part of the data. The use of $\mathcal{R}(d, \mu)$ makes more sense intuitively relative to other results obtained in similar contexts (see for instance [Ver96]). In this case, the norm on the data space would be defined as $\|(A, b, c, \mu)\| = \max\{\|A\|, \|b\|_1, \|c\|_\infty, n\mu\}$, and the corresponding distance to ill-posedness would be defined by $\rho(A, b, c, \mu) = \min\{\rho(d), \mu\}$. However, we prefer to use the scalar $\mathcal{K}(d, \mu)$ of (5.1), which arises more naturally in the proofs and conveniently leads to slightly tighter results, and also because it more accurately conveys the behavior of the optimal solutions to $P_\mu(d)$ as $\mu$ changes.

5.3 Changes in Optimal Solutions Under Data Perturbations

In this section, we present upper bounds on changes in optimal solutions to $P_\mu(d)$ and $D_\mu(d)$ as the data $d = (A, b, c)$ is changed or as the barrier parameter $\mu$ is changed. The major results of this section are contained in Theorems 5.4, 5.5, 5.6, 5.7, and 5.8. We first present all five theorems; the proofs of the theorems are deferred to the end of the section. As in the previous section, the bounds stated in these theorems are not necessarily the best achievable. Rather, it has been our goal to establish bounds that are polynomial in terms of the condition number $\mathcal{C}(d)$, the parameter $\mu$, the size of the data $\|d\|$, and the dimensions $m$ and $n$.

The first theorem, Theorem 5.4, presents upper bounds on the sizes of changes in optimal solutions to $P_\mu(d)$ and $D_\mu(d)$ as the data $d = (A, b, c)$ is changed to data $d + \Delta d = (A + \Delta A, b + \Delta b, c + \Delta c)$ in a suitably small neighborhood of the original data $d$. 
5.3 Changes in Optimal Solutions Under Data Perturbations

**Theorem 5.4** Let \( d = (A, b, c) \) be a data instance in \( F \) such that \( \rho(d) > 0 \), and let \( \mu > 0 \) be given and fixed. Given \( \alpha \in [0, 1) \) fixed, let \( \Delta d = (\Delta A, \Delta b, \Delta c) \in \tilde{D} \) be such that \( \|\Delta d\| \leq \alpha \rho(d) \). Then,

\[
\|\bar{x}(\mu) - x(\mu)\|_1 \leq \|\Delta d\| \frac{640n \ C(d)^2 K(d, \mu)^5 (\mu + \|d\|)}{\mu^2 (1 - \alpha)^6}, \tag{5.6}
\]

\[
\|\bar{y}(\mu) - y(\mu)\|_\infty \leq \|\Delta d\| \frac{640m \ C(d)^2 K(d, \mu)^5 (\mu + \|d\|)}{\mu^2 (1 - \alpha)^6}, \tag{5.7}
\]

\[
\|\bar{s}(\mu) - s(\mu)\|_\infty \leq \|\Delta d\| \frac{640m \ C(d)^2 K(d, \mu)^5 (\mu + \|d\|)^2}{\mu^2 (1 - \alpha)^6}, \tag{5.8}
\]

where \( x(\mu) \) and \( \bar{x}(\mu) \) are the optimal solutions to \( P_\mu(d) \) and \( P_\mu(d + \Delta d) \), respectively; \( (y(\mu), s(\mu)) \) and \( (\bar{y}(\mu), \bar{s}(\mu)) \) are the optimal solutions to \( D_\mu(d) \) and \( D_\mu(d + \Delta d) \), respectively; and \( K(d, \mu) \) is the scalar defined in (5.1).

Notice that the bounds are linear in \( \|\Delta d\| \) which indicates that the central trajectory associated with \( d \) changes at most linearly and in direct proportion to perturbations in \( d \) as long as the perturbations are smaller than \( \alpha \rho(d) \). Also, the bounds are polynomial in the condition number \( C(d) \) and the barrier parameter \( \mu \). Furthermore, notice that as \( \mu \to 0 \) these bounds diverge to \( \infty \). This is because small perturbations in \( d \) can produce extreme changes in the limit of the central trajectory associated with \( d \) as \( \mu \to 0 \).

The next theorem is important in that it establishes lower and upper bounds on the operator norm of the matrix \( (AX^2(\mu)A^T)^{-1} \), where \( x(\mu) \) is the optimal solution of \( P_\mu(d) \). This is of central importance in interior point algorithms for linear programming that use Newton’s method.

**Theorem 5.5** Let \( d = (A, b, c) \) be a data instance in \( F \) such that \( \rho(d) > 0 \). Let \( x(\mu) \)
be the optimal solution of $P_\mu(d)$, where $\mu > 0$. Then

$$\frac{1}{mn} \left( \frac{1}{\ell(d, \mu)\|d\|} \right)^2 \leq \|(AX^2(\mu)A^T)^{-1}\|_{1,\infty} \leq 4m \left( \frac{C(d)\ell(d, \mu)}{\mu} \right)^2,$$

where $\ell(d, \mu)$ is the scalar defined in (5.1).

Notice that the bounds in the theorem only depend on the condition number $C(d)$, the distance to ill-posedness $\rho(d)$, the size of the data instance $d = (A, b, c)$, the barrier parameter $\mu$, and the dimensions $m$ and $n$. Also note that as $\mu \to 0$, the upper bound on $\|(AX^2(\mu)A^T)^{-1}\|_{1,\infty}$ in the theorem goes to $\infty$ quadratically in $1/\mu$ in the limit. Incidentally, the matrix $(AX^2(\mu)A^T)^{-1}$ differs from the inverse of the Hessian of the dual objective function at its optimum by the scalar $-\mu^2$.

Theorem 5.6 presents upper bounds on the sizes of changes in optimal solutions to $P_\mu(d)$ and $D_\mu(d)$ as the barrier parameter $\mu$ is changed:

**Theorem 5.6** Let $d = (A, b, c)$ be a data instance in $F$ such that $\rho(d) > 0$. Given $\mu, \bar{\mu} > 0$, let $x(\mu)$ and $x(\bar{\mu})$ be the optimal solutions of $P_\mu(d)$ and $P_{\bar{\mu}}(d)$, respectively; and let $(y(\mu), s(\mu))$ and $(y(\bar{\mu}), s(\bar{\mu}))$ be the optimal solutions of $D_\mu(d)$ and $D_{\bar{\mu}}(d)$, respectively. Then

$$\|x(\bar{\mu}) - x(\mu)\|_1 \leq \frac{n}{\mu\bar{\mu}} |\bar{\mu} - \mu| \ell(d, \mu)\|\ell(d, \bar{\mu})\|d\|,$$  \hspace{1cm} (5.9)

$$\|y(\bar{\mu}) - y(\mu)\|_\infty \leq \frac{4m}{\mu\bar{\mu}} |\bar{\mu} - \mu| \ell(d, \mu)\|\ell(d, \bar{\mu})\|d\|C(d)^2,$$  \hspace{1cm} (5.10)

$$\|s(\bar{\mu}) - s(\mu)\|_\infty \leq \frac{4m}{\mu\bar{\mu}} |\bar{\mu} - \mu| \ell(d, \mu)\|\ell(d, \bar{\mu})\|d\|C(d)^2.$$  \hspace{1cm} (5.11)

where $\ell(d, \cdot)$ is the scalar defined in (5.1).

Notice that these bounds are linear in $|\bar{\mu} - \mu|$, which indicates that solutions along the central trajectory associated with $d$ change at most linearly and in direct proportion to changes in $\mu$. Also, the bounds are polynomial in the condition number $C(d)$.
and the barrier parameter \( \mu \).

The next result, Corollary 5.3, states upper bounds on the first derivatives of the optimal solutions \( x(\mu) \) and \( (y(\mu), s(\mu)) \) of \( P_\mu(d) \) and \( D_\mu(d) \), respectively, with respect to the barrier parameter \( \mu \). We first define the derivatives along the central trajectory as follows:

\[
\begin{aligned}
\dot{x}(\mu) &= \lim_{\bar{\mu} \to \mu} \frac{x(\bar{\mu}) - x(\mu)}{\bar{\mu} - \mu}, \\
\dot{y}(\mu) &= \lim_{\bar{\mu} \to \mu} \frac{y(\bar{\mu}) - y(\mu)}{\bar{\mu} - \mu}, \\
\dot{s}(\mu) &= \lim_{\bar{\mu} \to \mu} \frac{s(\bar{\mu}) - s(\mu)}{\bar{\mu} - \mu}.
\end{aligned}
\]

See Adler and Monteiro [AM91] for an application of these derivatives to the limiting behavior of central trajectories in linear programming.

**Corollary 5.3** Let \( d = (A, b, c) \) be a data instance in \( \mathcal{F} \) such that \( \rho(d) > 0 \), and let \( \mu > 0 \) be given and fixed. Let \( x(\mu) \) and \( (y(\mu), s(\mu)) \) be the optimal solutions of \( P_\mu(d) \) and \( D_\mu(d) \), respectively. Then

\[
\begin{aligned}
\|\dot{x}(\mu)\|_1 &\leq \frac{n}{\mu^2} \mathcal{K}(d, \mu)^2 \|d\|, \\
\|\dot{y}(\mu)\|_\infty &\leq \frac{4m}{\mu^2} \mathcal{K}(d, \mu)^2 \|d\|C(d)^2, \\
\|\dot{s}(\mu)\|_\infty &\leq \frac{4m}{\mu^2} \mathcal{K}(d, \mu)^2 \|d\|^2C(d)^2,
\end{aligned}
\]

where \( \mathcal{K}(d, \mu) \) is the scalar defined in (5.1).

The proof of this corollary follows immediately from Theorem 5.6.
5.3 Changes in Optimal Solutions Under Data Perturbations

Theorem 5.7 presents an upper bound on the size of the change in the optimal objective function value of \( P_\mu(d) \) as the data \( d \) is changed to data \( d + \Delta d \) in a specific neighborhood of the original data \( d \). Before stating this theorem, we introduce the following notation. Let \( d \) be a data instance in \( \mathcal{F} \), then we denote by \( z(d) \) the corresponding optimal objective value associated with \( P_\mu(d) \) by keeping the parameter \( \mu \) fixed, that is,

\[
z(d) = \min\{c^T x + \mu p(x) : Ax = b, x > 0\}.
\]

**Theorem 5.7** Let \( d = (A, b, c) \) be a data instance in \( \mathcal{F} \) such that \( \rho(d) > 0 \), and let \( \mu \geq 0 \) be given and fixed. Given \( \alpha \in [0, 1) \) fixed, let \( \Delta d = (\Delta A, \Delta b, \Delta c) \in \bar{D} \) be such that \( \|\Delta d\| \leq \alpha \rho(d) \). Then,

\[
|z(d + \Delta d) - z(d)| \leq 3 \|\Delta d\| \left(\frac{1 + \alpha}{1 - \alpha}\right)^4 K(d, \mu)^2,
\]

(5.12)

where \( K(d, \mu) \) is the scalar defined in (5.1).

Observe that, as in Theorem 5.4, the upper bound in the change in the objective function value is linear in \( \|\Delta d\| \) so long as \( \|\Delta d\| \) is no larger than \( \alpha \rho(d) \), which indicates that optimal objective values along the central trajectory will change at most linearly and in direct proportion to changes in \( d \) for small changes in \( d \). Note also that the bound is polynomial in the condition number \( C(d) \) and in the barrier parameter \( \mu \).

Last of all, Theorem 5.8 presents an upper bound on the size of the change in the optimal objective function value of \( P_\mu(d) \) as the barrier parameter \( \mu \) is changed. As before, it is convenient to introduce the following notation. Let \( d \) be a data instance in \( \mathcal{F} \), then we denote by \( z(\mu) \) the corresponding optimal objective value associated with \( P_\mu(d) \) by keeping the data instance \( d \) fixed, that is,

\[
z(\mu) = \min\{c^T x + \mu p(x) : Ax = b, x > 0\}.
\]
5.3 Changes in Optimal Solutions Under Data Perturbations

**Theorem 5.8** Let \( d = (A, b, c) \) be a data instance in \( \mathcal{F} \) such that \( \rho(d) > 0 \). Then,

\[
|z(\bar{\mu}) - z(\mu)| \leq |\bar{\mu} - \mu| n \left( \ln(2) + \ln(K(d, \mu)K(d, \bar{\mu})) + |\ln(||d||)| + \max\{|\ln(\mu)|, |\ln(\bar{\mu})|\} \right).
\]

for given \( \mu, \bar{\mu} > 0 \), where \( K(d, \cdot) \) is the scalar defined in (5.1).

As in Theorem 5.6, the upper bound given by this theorem is linear in \( |\bar{\mu} - \mu| \), which indicates that optimal objective function values along the central trajectory associated with \( d \) change at most linearly and in direct proportion to changes in \( \mu \). Also, the bounds are logarithmic in the condition number \( C(d) \) and in the barrier parameter \( \mu \).

**Remark 5.1** Since \( z(\mu) = c^T x(\mu) + p(x(\mu)) \), it follows from the smoothness of \( x(\mu) \) that \( z(\mu) \) is also a smooth function, and from Theorem 5.8 it then follows that

\[
|\dot{z}(\mu)| \leq n \left( \ln(2) + 2 \ln(K(d)) + |\ln(||d||)| + |\ln(\mu)| \right).
\]

Before proving the five theorems, we first prove a variety of intermediary results that will be used in the proofs of the five theorems. The following three results establish upper and lower bounds on certain quantities as the data \( d = (A, b, c) \) is changed to data \( d + \Delta d = (A + \Delta A, b + \Delta b, c + \Delta c) \) in a specific neighborhood of the original data \( d \); or as the parameter \( \mu \) is changed along the central trajectory. These results will also be used in the proofs of the theorems of this section.

**Lemma 5.1** Suppose that \( d = (A, b, c) \in \mathcal{F}, \rho(d) > 0 \). Let \( \alpha \in (0, 1) \) be given and fixed, and let \( \Delta d \) be such that \( ||\Delta d|| \leq \alpha \rho(d) \). If \( x(\mu) \) is the optimal solution to \( P_\mu(d) \), and \( \bar{x}(\mu) \) is the optimal solution to \( P_\mu(d + \Delta d) \), then for \( j = 1, \ldots, n \),

\[
\frac{1}{32} \left( \frac{\mu(1 - \alpha)}{||d||K(d, \mu)} \right)^2 \leq x_j(\mu) \bar{x}_j(\mu) \leq 4 \left( \frac{K(d, \mu)}{1 - \alpha} \right)^2,
\]

(5.13)

where \( \mu > 0 \) is given and fixed, and \( K(d, \mu) \) is the scalar defined in (5.1).
5.3 Changes in Optimal Solutions Under Data Perturbations

Proof: Let \( x = x(\mu) \) and \( \bar{x} = \bar{x}(\mu) \). From Theorem 5.1 we have \( \|x\|_1 \leq \mathcal{K}(d, \mu) \), and from Corollary 5.1 we also have \( \|\bar{x}\|_1 \leq (4/(1 - \alpha)^2)\mathcal{K}(d, \mu) \). Therefore, we obtain

\[
x_j \bar{x}_j \leq \|x\|_1 \|\bar{x}\|_1 \leq 4(\mathcal{K}(d, \mu)^2/(1 - \alpha)^2) \text{ for all } j = 1, \ldots, n.
\]

On the other hand, from Theorem 5.2 and Corollary 5.2, it follows that

\[
x_j \geq \frac{\mu}{2\|d\|\mathcal{K}(d, \mu)},
\]

\[
\bar{x}_j \geq \frac{(1 - \alpha)^2 \mu}{8\|d\| + \|\Delta d\|\mathcal{K}(d, \mu)} \geq \frac{(1 - \alpha)^2 \mu}{16\|d\|\mathcal{K}(d, \mu)}.
\]

for all \( j = 1, \ldots, n \). Therefore,

\[
x_j \bar{x}_j \geq \frac{1}{32} \left( \frac{\mu(1 - \alpha)}{\|d\|\mathcal{K}(d, \mu)} \right)^2,
\]

for all \( j = 1, \ldots, n \).

q.e.d.

Lemma 5.2 Suppose that \( d = (A, b, c) \in \mathcal{F} \) and \( \rho(d) > 0 \). For a given \( \alpha \in [0, 1) \), let \( \Delta d \) be such that \( \|\Delta d\| \leq \alpha \rho(d) \). If \( x = x(\mu) \) is the optimal solution to \( P_\mu(d) \), and \( \bar{x} = \bar{x}(\mu) \) is the optimal solution to \( P_\mu(d + \Delta d) \), then

\[
\frac{1}{4mn} \left( \frac{1 - \alpha}{\mathcal{K}(d, \mu)} \right)^2 \leq \|(AX\bar{X}AT)^{-1}\|_{1,\infty} \leq 32m \left( \frac{C(d)\mathcal{K}(d, \mu)}{\mu(1 - \alpha)} \right)^2,
\]

where \( \mu > 0 \) is given and fixed, and \( \mathcal{K}(d, \mu) \) is the scalar defined in (5.1).

Proof: Using identical logic to Proposition 2.12 part (i), we have

\[
\|(AX\bar{X}AT)^{-1}\|_{1,\infty} \leq \|(AX\bar{X}AT)^{-1}\|_2 \leq \frac{\|(AA^T)^{-1}\|_2}{\min_j\{x_j\bar{x}_j\}}.
\]
5.3 Changes in Optimal Solutions Under Data Perturbations

Now, by applying Proposition 2.12, part (ii), and Lemma 5.1, we obtain that

\[
\|(AXX^T)^{-1}\|_{1,\infty} \leq \frac{32\|d\|^2K(d, \mu)^2}{\mu^2(1 - \alpha)^2\lambda_1(AAT)} \\
\leq \frac{32m\|d\|^2K(d, \mu)^2}{\mu^2(1 - \alpha)^2\rho(d)^2} \\
= \frac{32mC(d)^2K(d, \mu)^2}{\mu^2(1 - \alpha)^2}.
\]

On the other hand, by identical logic to Proposition 2.12 part (i),

\[
\|(AXX^T)^{-1}\|_{1,\infty} \geq \frac{1}{m}\|(AXX^T)^{-1}\|_2 \\
\geq \frac{\|(AAT)^{-1}\|_2}{m \max_j \{x_j\bar{x}_j\}}.
\]

Now, by applying Proposition 2.1, part (iv), and Lemma 5.1, we obtain that

\[
\|(AXX^T)^{-1}\|_{1,\infty} \geq \frac{(1 - \alpha)^2}{4mK(d, \mu)^2\lambda_1(AAT)} \\
\geq \frac{(1 - \alpha)^2}{4mK(d, \mu)^2\lambda_m(AAT)} \\
= \frac{(1 - \alpha)^2}{4mK(d, \mu)^2\|A\|_2^2} \\
\geq \frac{(1 - \alpha)^2}{4mnK(d, \mu)^2\|d\|^2}.
\]

q.e.d.

**Lemma 5.3** Let \(d = (A, b, c)\) be a data instance in \(F\) such that \(\rho(d) > 0\). Let \(x = x(\mu)\) and \(\bar{x} = x(\bar{\mu})\) be the optimal solutions of \(P_\mu(d)\) and \(P_{\bar{\mu}}(d)\), respectively,
where $\mu, \bar{\mu} > 0$. Then

$$
\frac{1}{mnK(d, \mu)K(d, \bar{\mu})\|d\|^2} \leq \|(AX\bar{X}A^T)^{-1}\|_{1,\infty} \leq \frac{4mC(d)^2K(d, \mu)K(d, \bar{\mu})}{\mu\bar{\mu}},
$$

where $K(d, \cdot)$ is the scalar defined in (5.1).

**Proof:** Following the proof of Lemma 5.2, we have from Proposition 2.12 and Theorem 5.2 that

$$
\|(AX\bar{X}A^T)^{-1}\|_{1,\infty} \leq \frac{m}{\min_j \{x_j \bar{x}_j\} \rho(d)^2}
\leq \frac{4m\|d\|^2K(d, \mu)K(d, \bar{\mu})}{\mu\bar{\mu}\rho(d)^2}
= \frac{4mC(d)^2K(d, \mu)K(d, \bar{\mu})}{\mu\bar{\mu}}.
$$

On the other hand, we have again from Proposition 2.1, Theorem 5.1, and Proposition 2.12 that

$$
\|(AX\bar{X}A^T)^{-1}\|_{1,\infty} \geq \frac{\|(AA^T)^{-1}\|_2}{m \max_j \{x_j \bar{x}_j\}}
\geq \frac{1}{mK(d, \mu)K(d, \bar{\mu})\lambda_1(AA^T)}
\geq \frac{1}{mK(d, \mu)K(d, \bar{\mu})\lambda_m(AA^T)}
\geq \frac{1}{mK(d, \mu)K(d, \bar{\mu})\|A\|^2}
\geq \frac{1}{mnK(d, \mu)K(d, \bar{\mu})\|d\|^2}.
$$

q.e.d.
5.3 Changes in Optimal Solutions Under Data Perturbations

Note that the proof of Theorem 5.5 follows as an immediate application of Lemma 5.3, by setting \( \mu = \bar{\mu} \).

We are now ready to prove Theorem 5.4.

**Proof of Theorem 5.4:** Let \( x = x(\mu) \) and \( \bar{x} = \bar{x}(\mu) \) be the optimal solutions to \( P_\mu(d) \) and \( P_\mu(d + \Delta d) \), respectively; and let \( (y, s) = (y(\mu), s(\mu)) \) and \( (\bar{y}, \bar{s}) = (\bar{y}(\mu), \bar{s}(\mu)) \) be the optimal solutions to \( D_\mu(d) \) and \( D_\mu(d + \Delta d) \), respectively. Then from the Karush-Kuhn-Tucker optimality conditions we have:

\[
\begin{align*}
Xs &= \mu e, & \bar{X}\bar{s} &= \mu e, \\
A^Ty + s &= c, & (A + \Delta A)^T\bar{y} + \bar{s} &= c + \Delta c, \\
Ax &= b, & (A + \Delta A)\bar{x} &= b + \Delta b, \\
x &> 0, & \bar{x} &> 0.
\end{align*}
\]

Therefore,

\[
\bar{x} - x = \frac{1}{\mu}X\bar{X}(s - \bar{s})
\]

\[
= \frac{1}{\mu}X\bar{X} \left( (c - A^Ty) - (c + \Delta c - (A + \Delta A)^T\bar{y}) \right)
\]

\[
= \frac{1}{\mu}X\bar{X} \left( \Delta A^T\bar{y} - \Delta c \right) + \frac{1}{\mu}X\bar{X}A^T(\bar{y} - y). \tag{5.15}
\]

On the other hand, \( A(\bar{x} - x) = \Delta b - \Delta A\bar{x} \). Since \( A \) has rank \( m \) (otherwise \( \rho(d) = 0 \)), then \( P = AX\bar{X}A^T \) is a positive definite matrix. By combining these statements together with (5.15), we obtain

\[
\Delta b - \Delta A\bar{x} = \frac{1}{\mu}AX\bar{X} \left( \Delta A^T\bar{y} - \Delta c \right) + \frac{1}{\mu}P(\bar{y} - y),
\]
and so

\[ \mu P^{-1} (\Delta b - \Delta A \bar{x}) = P^{-1}AX\bar{X} (\Delta A^T \bar{y} - \Delta c) + \bar{y} - y. \]

Therefore, we have the following identity:

\[ \bar{y} - y = \mu P^{-1} (\Delta b - \Delta A \bar{x}) + P^{-1}AX\bar{X} (\Delta c - \Delta A^T \bar{y}). \quad (5.16) \]

From this identity, it follows that

\[ \| \bar{y} - y \|_\infty \leq \| P^{-1} \|_1,\infty (\mu \| \Delta b - \Delta A \bar{x} \|_1 + \| A \| \| X\bar{X} (\Delta c - \Delta A^T \bar{y}) \|_1). \quad (5.17) \]

Note that

\[ \| X\bar{X} (\Delta c - \Delta A^T \bar{y}) \|_1 \leq \| X\bar{X} \|_{\infty,1} \| \Delta c - \Delta A^T \bar{y} \|_\infty \leq \| x \|_1 \| \bar{x} \|_1 \| \Delta c - \Delta A^T \bar{y} \|_\infty. \quad (5.18) \]

From Corollary 5.1, we have

\[
\begin{align*}
\| \Delta b - \Delta A \bar{x} \|_1 &\leq \| \Delta d \| (1 + \| \bar{x} \|_1) \\
&\leq \| \Delta d \| \left(1 + \frac{4}{(1 - \alpha)^2} \mathcal{K}(d, \mu)\right) \\
&\leq \frac{5\| \Delta d \|}{(1 - \alpha)^2} \mathcal{K}(d, \mu),
\end{align*}
\]

\[
\begin{align*}
\| \Delta c - \Delta A^T \bar{y} \|_\infty &\leq \| \Delta d \| (1 + \| \bar{y} \|_\infty) \\
&\leq \| \Delta d \| \left(1 + \frac{4}{(1 - \alpha)^2} \mathcal{K}(d, \mu)\right) \\
&\leq \frac{5\| \Delta d \|}{(1 - \alpha)^2} \mathcal{K}(d, \mu).
\end{align*}
\]

Therefore, by combining (5.17), (5.18), (5.19), and (5.20), and by using Theo-
5.3 Changes in Optimal Solutions Under Data Perturbations

rem 5.1, Corollary 5.1, and Lemma 5.2, we obtain the following bound on $\|\tilde{y} - y\|_\infty$:

$$\|\tilde{y} - y\|_\infty \leq 32m \left( \frac{C(d)K(d, \mu)}{\mu(1 - \alpha)} \right)^2 \left( \frac{5\|\Delta d\|}{(1 - \alpha)^2} K(d, \mu) \right) \left( \mu + 4\|d\| \left( \frac{K(d, \mu)}{1 - \alpha} \right)^2 \right)$$

$$\leq 640m\|\Delta d\| \frac{C(d)K(d, \mu)^5(\mu + \|d\|)}{\mu^2(1 - \alpha)^6},$$

thereby demonstrating the bound (5.7) on $\|\tilde{y} - y\|_\infty$.

Now, by substituting identity (5.16) into equation (5.15), we obtain

$$\bar{x} - x = \frac{1}{\mu} \bar{X} \left(I - A^T P^{-1} A\bar{X}\right) \left(\Delta A^T \tilde{y} - \Delta c\right) + \bar{X} A^T P^{-1} \left(\Delta b - \Delta A\bar{x}\right)$$

$$= \frac{1}{\mu} D^{\frac{1}{2}} \left(I - D^{\frac{1}{2}} A^T P^{-1} A D^{\frac{1}{2}}\right) D^{\frac{1}{2}} \left(\Delta A^T \tilde{y} - \Delta c\right) + DA^T P^{-1} \left(\Delta b - \Delta A\bar{x}\right),$$

where $D = X\bar{X}$. Observe that the matrix $Q = I - D^{\frac{1}{2}} A^T P^{-1} A D^{\frac{1}{2}}$ is a projection matrix, and so $\|Qx\|_2 \leq \|x\|_2$ for all $x \in \mathbb{R}^n$. Hence, from Proposition 2.1 parts (i) and (iii), we obtain that

$$\|\bar{x} - x\|_1 \leq \frac{1}{\mu} \|D^{\frac{1}{2}} \left(I - D^{\frac{1}{2}} A^T P^{-1} A D^{\frac{1}{2}}\right) D^{\frac{1}{2}} \left(\Delta A^T \tilde{y} - \Delta c\right)\|_1$$

$$\quad + \|DA^T P^{-1} (\Delta b - \Delta A\bar{x})\|_1$$

$$\leq \frac{\sqrt{n}}{\mu} \|D^{\frac{1}{2}}\|_2 \|\Delta A^T \tilde{y} - \Delta c\|_2 + \|DA^T P^{-1} (\Delta b - \Delta A\bar{x})\|_1$$

$$= \frac{\sqrt{n}}{\mu} \|D\|_2 \|\Delta A^T \tilde{y} - \Delta c\|_2 + \|DA^T P^{-1} (\Delta b - \Delta A\bar{x})\|_1$$

$$\leq \frac{n}{\mu} \max_j \{x_j \bar{x}_j\} \|\Delta A^T \tilde{y} - \Delta c\|_\infty + \|D\|_\infty,1 \|A^T\|_\infty \|P^{-1}\|_{1,\infty} \|\Delta b - \Delta A\bar{x}\|_1$$

$$= \frac{n}{\mu} \max_j \{x_j \bar{x}_j\} \|\Delta A^T \tilde{y} - \Delta c\|_\infty + \sum_{j=1}^n \|x_j \bar{x}_j\| \|A\| \|P^{-1}\|_{1,\infty} \|\Delta b - \Delta A\bar{x}\|_1$$

$$\leq \frac{n}{\mu} \max_j \{x_j \bar{x}_j\} \|\Delta A^T \tilde{y} - \Delta c\|_\infty + \|x\|_1 \|\bar{x}\|_1 \|d\| \|P^{-1}\|_{1,\infty} \|\Delta b - \Delta A\bar{x}\|_1$$

It follows from Lemma 5.1, Theorem 5.1, Corollary 5.1, Lemma 5.2, and inequali-
ties (5.19) and (5.20) that
\[
\|\bar{x} - x\|_1 \leq \frac{4n}{\mu} \left( \frac{\mathcal{K}(d, \mu)}{1 - \alpha} \right)^2 \frac{5\|\Delta d\|}{(1 - \alpha)^2} \mathcal{K}(d, \mu) \\
+ 128m \left( \frac{\mathcal{K}(d, \mu)}{1 - \alpha} \right)^2 \|d\| \left( \frac{C(d)\mathcal{K}(d, \mu)}{\mu(1 - \alpha)} \right)^2 \frac{5\|\Delta d\|}{(1 - \alpha)^2} \mathcal{K}(d, \mu),
\]
from which we obtain the following bound (recall that \(n \geq m\)):
\[
\|\bar{x} - x\|_1 \leq 640n\|\Delta d\| \frac{C(d)^2\mathcal{K}(d, \mu)^5(\mu + \|d\|)}{\mu^2(1 - \alpha)^6},
\]
which thereby demonstrates the bound (5.6) on \(\|\bar{x} - x\|_1\).

Finally, observe that \(\bar{s} - s = \Delta c - \Delta A^T \bar{y} + A^T(y - \bar{y})\), so that \(\|\bar{s} - s\|_\infty \leq \|\Delta c - \Delta A^T \bar{y}\|_\infty + \|A^T\|\|y - \bar{y}\|_\infty = \|\Delta c - \Delta A^T \bar{y}\|_\infty + \|A\|\|y - \bar{y}\|_\infty\). Using our previous results, we obtain
\[
\|\bar{s} - s\|_\infty \leq \frac{5\|\Delta d\|}{(1 - \alpha)^2} \mathcal{K}(d, \mu) + 640m\|d\|\|\Delta d\| \frac{C(d)^2\mathcal{K}(d, \mu)^5(\mu + \|d\|)}{\mu^2(1 - \alpha)^6} \\
\leq 640m\|\Delta d\| \frac{C(d)^2\mathcal{K}(d, \mu)^5(\mu + \|d\|)^2}{\mu^2(1 - \alpha)^6},
\]
and this concludes the proof of this theorem.

q.e.d.

We now present the proof of Theorem 5.6.

**Proof of Theorem 5.6:** Let \(x = x(\mu)\) and \(\bar{x} = x(\bar{\mu})\) be the primal optimal solutions to \(P_\mu(d)\) and \(P_{\bar{\mu}}(d)\), respectively; and let \((y, s) = (y(\mu), s(\mu))\) and \((\bar{y}, \bar{s}) = (y(\bar{\mu}), s(\bar{\mu}))\) be the dual optimal solutions to \(D_\mu(d)\) and \(D_{\bar{\mu}}(d)\), respectively. From the Karush-
Kuhn-Tucker optimality conditions we have

\[ Xs = \mu e, \quad \bar{X}s = \bar\mu e, \]
\[ A^T y + s = c, \quad A^T \bar{y} + \bar{s} = c, \]
\[ Ax = b, \quad A\bar{x} = b, \]
\[ x > 0, \quad \bar{x} > 0. \]

Therefore,

\[ \bar{x} - x = \frac{1}{\mu\bar{\mu}} X\bar{X} (\bar{\mu}s - \mu\bar{s}) \]
\[ = \frac{1}{\mu\bar{\mu}} X\bar{X} (\bar{\mu}(c - A^T y) - \mu(c - A^T \bar{y})) \]
\[ = \frac{1}{\mu\bar{\mu}} X\bar{X} \left((\bar{\mu} - \mu)c - A^T (\bar{\mu}y - \mu\bar{y})\right). \quad (5.21) \]

On the other hand, \( A(\bar{x} - x) = b - b = 0. \) Since \( A \) has rank \( m \) (otherwise \( \rho(d) = 0 \)), then \( P = AX\bar{X}A^T \) is a positive definite matrix. By combining these statements together with (5.21), we obtain

\[ 0 = \frac{1}{\mu\bar{\mu}} AX\bar{X} \left((\bar{\mu} - \mu)c - A^T (\bar{\mu}y - \mu\bar{y})\right), \]

and so

\[ P (\bar{\mu}y - \mu\bar{y}) = (\bar{\mu} - \mu)AX\bar{X}c, \]

equivalently

\[ \bar{\mu}y - \mu\bar{y} = (\bar{\mu} - \mu)P^{-1}AX\bar{X}c. \quad (5.22) \]

By substituting identity (5.22) into equation (5.21) and by letting \( D = X\bar{X} \), we obtain:

\[ \bar{x} - x = \frac{\bar{\mu} - \mu}{\mu\bar{\mu}} X\bar{X} \left(c - A^T P^{-1}AX\bar{X}c\right) \]
$$= \frac{\bar{\mu} - \mu}{\mu \bar{\mu}} D \left( c - A^T P^{-1} A D c \right)$$

$$= \frac{\bar{\mu} - \mu}{\mu \bar{\mu}} D^{\frac{1}{2}} \left( I - D^{\frac{1}{2}} A^T P^{-1} A D^{\frac{1}{2}} \right) D^{\frac{1}{2}} c,$$

Observe that the matrix $Q = I - D^{\frac{1}{2}} A^T P^{-1} A D^{\frac{1}{2}}$ is a projection matrix, and so $\|Qx\|_2 \leq \|x\|_2$ for all $x \in \mathbb{R}^n$. Hence, from Proposition 2.1, parts (i) and (iii), and Theorem 5.1, we have

$$\|\bar{x} - x\|_1 \leq \sqrt{n} \|\bar{x} - x\|_2$$

$$\leq \frac{\sqrt{n}}{\mu \bar{\mu}} |\bar{\mu} - \mu| \|D^{\frac{1}{2}}\|_2 \|D^{\frac{1}{2}}\|_2 \|c\|_2$$

$$= \frac{\sqrt{n}}{\mu \bar{\mu}} |\bar{\mu} - \mu| \|D\|_2 \|c\|_2$$

$$\leq \frac{n}{\mu \bar{\mu}} |\bar{\mu} - \mu| \|x\|_1 \|\bar{x}\|_1 \|c\|_\infty$$

$$\leq \frac{n}{\mu \bar{\mu}} |\bar{\mu} - \mu| K(d, \mu) K(d, \bar{\mu}) \|d\|,$$

which demonstrates the bound (5.9) for $\|\bar{x} - x\|_1$.

Now, since $c = A^T y + s$ and $c = A^T \bar{y} + \bar{s}$, it follows that $A^T (\bar{y} - y) + \bar{s} - s = 0$, which yields the following equalities in logical sequence:

$$0 = A^T (\bar{y} - y) + X^{-1} \bar{X}^{-1} (\bar{\mu} x - \mu \bar{x}),$$

$$A^T (y - \bar{y}) = X^{-1} \bar{X}^{-1} (\bar{\mu} x - \mu \bar{x}),$$

$$X \bar{X} A^T (y - \bar{y}) = \bar{\mu} x - \mu \bar{x},$$

so that by pre-multiplying by $A$, we obtain

$$AX \bar{X} A^T (y - \bar{y}) = (\bar{\mu} - \mu) b,$$

$$P (y - \bar{y}) = (\bar{\mu} - \mu) b,$$

$$y - \bar{y} = (\bar{\mu} - \mu) P^{-1} b.$$
Therefore, from Corollary 5.3,

\[
\|\mathcal{g} - y\|_\infty \leq \|\bar{\mu} - \mu\| P^{-1} \|b\|_1 \\
\leq \frac{4m}{\mu\bar{\mu}} |\bar{\mu} - \mu| C(d)^2 \mathcal{K}(d, \mu) \mathcal{K}(d, \bar{\mu}) \|\mathcal{d}\|,
\]

which establishes the bound (5.10) for \(\|\mathcal{g} - y\|_\infty\).

Finally, using the fact that \(s - \bar{s} = A^T (\mathcal{g} - y)\), we obtain \(\|s - \bar{s}\|_\infty \leq \|A^T\|_\infty \|y - \mathcal{g}\|_\infty = \|A\| \|y - \mathcal{g}\|_\infty \leq \|d\| \|y - \mathcal{g}\|_\infty\), which establishes (5.11) from (5.10), and so this concludes the proof of this theorem.

q.e.d.

We now prove Theorem 5.7.

**Proof of Theorem 5.7:** Consider the Lagrangian functions associated with these problems,

\[
L(x, y) = c^T x + \mu p(x) + y^T (b - Ax), \\
\bar{L}(x, y) = (c + \Delta c)^T x + \mu p(x) + y^T (b + \Delta b - (A + \Delta A)x),
\]

and define \(\Phi(x, y) = L(x, y) - \bar{L}(x, y)\). Observe that

\[
z(d) = \max_y \min_{x>0} L(x, y) = \min_{x>0} \max_y L(x, y), \\
z(d + \Delta d) = \max_y \min_{x>0} \bar{L}(x, y) = \min_{x>0} \max_y \bar{L}(x, y).
\]

Hence, if \((x(\mu), y(\mu))\) is a pair of optimal solutions to the primal and dual programs corresponding to \(d\), and \((\bar{x}(\mu), \bar{y}(\mu))\) is a pair of optimal solutions to the primal and dual programs corresponding to \(d + \Delta d\), then

\[
z(d) = L(x(\mu), y(\mu))
\]
5.3 Changes in Optimal Solutions Under Data Perturbations

\[ = \max_y L(x(\mu), y) \]
\[ = \max_y \{ \bar{L}(x(\mu), y) + \Phi(x(\mu), y) \} \]
\[ \geq \bar{L}(x(\mu), \bar{y}(\mu)) + \Phi(x(\mu), \bar{y}(\mu)) \]
\[ \geq z(d + \Delta d) + \Phi(x(\mu), \bar{y}(\mu)). \]

Thus, \( z(d) - z(d + \Delta d) \geq \Phi(x(\mu), \bar{y}(\mu)) \). Similarly, we can prove that \( z(d) - z(d + \Delta d) \leq \Phi(\bar{x}(\mu), y(\mu)) \).

Therefore, we obtain the following bounds: either

\[ |z(d + \Delta d) - z(d)| \leq |\Phi(x(\mu), \bar{y}(\mu))|, \]

or

\[ |z(d + \Delta d) - z(d)| \leq |\Phi(\bar{x}(\mu), y(\mu))|. \]

On the other hand, using Hölder’s inequality and the bounds from Corollary 5.1 we have

\[ |\Phi(x(\mu), \bar{y}(\mu))| = |\Delta c^T x(\mu) + \bar{y}(\mu)^T \Delta b - \bar{y}(\mu)^T \Delta A x(\mu)| \]
\[ \leq \|\Delta c\|_\infty \|x(\mu)\|_1 + \|\bar{y}(\mu)\|_\infty \|\Delta b\|_1 + \|\bar{y}(\mu)\|_\infty \|\Delta A\| \|x(\mu)\|_1 \]
\[ \leq \|\Delta d\|_1 (\|x(\mu)\|_1 + \|\bar{y}(\mu)\|_\infty + \|\bar{y}(\mu)\|_\infty \|x(\mu)\|_1) \]
\[ \leq 3\|\Delta d\| \left( \frac{1 + \alpha}{1 - \alpha} \right)^4 \mathcal{K}(d, \mu)^2. \]

Similarly, we can show that

\[ |\Phi(\bar{x}(\mu), y(\mu))| \leq 3\|\Delta d\| \left( \frac{1 + \alpha}{1 - \alpha} \right)^4 \mathcal{K}(d, \mu)^2, \]

and the result follows.

q.e.d.
Finally, we prove Theorem 5.8.

**Proof of Theorem 5.8:** Let \( x(\mu) \) and \( x(\bar{\mu}) \) be the optimal solutions to \( P_\mu(d) \) and \( P_{\bar{\mu}}(d) \), respectively; and \( (y(\mu), s(\mu)) \) and \( (y(\bar{\mu}), s(\bar{\mu})) \) be the optimal solutions to \( D_\mu(d) \) and \( D_{\bar{\mu}}(d) \), respectively. As in Theorem 5.7, for given \( \mu, \bar{\mu} > 0 \), consider the following Lagrangian functions: 

\[
L(x, y) = c^T x + \mu p(x) + y^T (b - Ax) \quad \text{and} \quad \tilde{L}(x, y) = c^T x + \bar{\mu} p(x) + y^T (b - Ax).
\]

Define \( \Phi(x, y) = L(x, y) - \tilde{L}(x, y) = (\mu - \bar{\mu}) p(x) \).

By a similar argument as in the proof of Theorem 5.7, we have \( z(\mu) - z(\bar{\mu}) \geq \Phi(x(\mu), y(\bar{\mu})) \) and \( z(\mu) - z(\bar{\mu}) \leq \Phi(x(\bar{\mu}), y(\mu)) \). Therefore, we obtain the following bounds: either

\[
|z(\bar{\mu}) - z(\mu)| \leq | - \Phi(x(\mu), y(\bar{\mu}))| = |\bar{\mu} - \mu| |p(x(\mu))|,
\]

or

\[
|z(\bar{\mu}) - z(\mu)| \leq | - \Phi(x(\bar{\mu}), y(\mu))| = |\bar{\mu} - \mu| |p(x(\bar{\mu}))|.
\]

Therefore,

\[
|z(\bar{\mu}) - z(\mu)| \leq |\bar{\mu} - \mu| \max\{|p(x(\mu))|, |p(x(\bar{\mu}))|\}.
\]

On the other hand, from Theorem 5.1 and Theorem 5.2, we have

\[
\frac{\mu}{2\|d\|K(d, \mu)} \leq x_j(\mu) \leq K(d, \mu),
\]

for all \( j = 1, \ldots, n \). Hence,

\[
n \ln \left( \frac{\mu}{2\|d\|K(d, \mu)} \right) \leq -p(x(\mu)) \leq n \ln(K(d, \mu)),
\]

so that

\[
|p(x(\mu))| \leq n \max\left\{ \left| \ln \left( \frac{\mu}{2\|d\|K(d, \mu)} \right) \right|, \ln(K(d, \mu)) \right\}
\]
\[ \leq n \left( \ln(2) + \ln(K(d, \mu)K(d, \bar{\mu})) + |\ln(||d||)| + \max \{ |\ln(\mu)|, |\ln(\bar{\mu})| \} \right). \]

Similarly, using \( \bar{\mu} \) instead of \( \mu \) we also obtain

\[ |p(x(\bar{\mu}))| \leq n \left( \ln(2) + \ln(K(d, \mu)K(d, \bar{\mu})) + |\ln(||d||)| + \max \{ |\ln(\mu)|, |\ln(\bar{\mu})| \} \right), \]

and the result follows.

q.e.d.
Chapter 6

Conditioning of the Optimal Faces of a Linear Program

6.1 Overview

For a given data instance \( d = (A, b, c) \) of a linear program, the objective of this chapter is to study the pair of optimal faces corresponding to the linear programs \( P(d) \) and \( D(d) \), respectively, in terms of a measure of the conditioning of these faces. This measure is called the distance to degeneracy and represents the minimum change \( \Delta d = (\Delta A, \Delta b, \Delta c) \) in the data \( d \) necessary to create a perturbed data instance \( d + \Delta d \) whose optimal faces have a different combinatorial structure than the combinatorial structure of the optimal faces associated with \( d \). By using this measure, we aim to establish complexity results for interior-point methods in terms of the conditioning of the optimal faces associated with \( d \). In particular, we seek polynomial bounds on the number of iterations that it will take for an interior-point method to find an optimal solution to \( P(d) \) and \( D(d) \), or at least to find an \( \epsilon \)-approximation of an optimal solution. Furthermore, we seek bounds that are independent of the Turing-machine model of computation, so that it is not necessary to assume that our data is rational (as
happens in most of the literature of interior-point methods in linear programming). By a polynomial bound on the number of iterations we mean a polynomial expression involving the logarithm of the condition measure; the dimensions $n$ and $m$, $\log(1/\epsilon)$ (when using an algorithm that finds an $\epsilon$-approximate solution); and numerical constants independent of these previous parameters.

With respect to finding an $\epsilon$-approximate solutions to $P(d)$ and $D(d)$ there are many complexity results in the literature. For instance, Freund [Fre94] uses the distance to ill-posedness $\rho(d)$ associated with a linear program in standard form to establish a polynomial bound on the number of iterations that Khachiyan's ellipsoid algorithm [Kha79] takes to find an $\epsilon$-approximate solution to $P(d)$. Similarly, Renegar [Ren95c] developed a polynomial-time interior-point method for finding an $\epsilon$-approximate solution to the program $\min \{c^Tx : Ax \leq b, x \geq 0\}$ by means of an appropriate distance to ill-posedness $\rho(d)$ corresponding to this formulation of a linear program and the related condition number $C(d)$. In both cases the bounds on the number of iterations of the corresponding algorithms involve polynomial expressions in terms of $\log(C(d))$, $\log(1/\epsilon)$, and the dimensions $m$ and $n$, thereby proving the polynomiality of the bound on the number of iterations as defined in the previous paragraph.

In Section 6.2 we state several properties of the distance to degeneracy $\eta(\cdot)$ as defined in Definition 6.1. These properties show relations with other measures such as the distance to ill-posedness $\rho(\cdot)$ (see Proposition 6.1), the distance to multiple optima $\zeta(\cdot)$ defined in Proposition 6.2, and the classical condition number $\kappa(\cdot)$ applied to a certain matrix associated with the optimal faces of a linear program (see Lemma 6.3). For a given data instance $d = (A, b, c)$ with unique primal and dual optimal solutions, the distance to multiple optima represents the minimum change $\Delta d = (\Delta A, \Delta b, \Delta c)$
in the data $d$ necessary to create a perturbed data instance $d + \Delta d$ with either primal multiple optima or dual multiple optima. The most important results in this section are Lemma 6.1 and Lemma 6.2. These lemmas establish that the distance $\eta(d)$ is positive if and only if $P(d)$ has unique primal and dual optimal solutions, or equivalently, if and only if $P(d)$ does not have primal nor dual degenerate optimal solutions.

In Section 6.3, we provide an algorithm that finds the exact solution of a linear program when the program exhibits unique primal and dual solutions. Our results are based on the notion of the optimal partition of a linear program $P(d)$ and the distance to degeneracy $\eta(d)$, which we will define shortly. Given a data instance $d = (A, b, c)$ in $\mathcal{F}$, we denote by $S_P(d)$ the feasible region associated with the program $P(d)$ and by $S_D(d)$ the feasible region associated with the program $D(d)$, that is,

$$S_P(d) := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\},$$

$$S_D(d) := \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^Ty + s = c, s \geq 0\}.$$

Since $d \in \mathcal{F}$, the program $P(d)$ has an optimal objective function value that we denote by $z(d)$, that is,

$$z(d) := \min \{c^T x : x \in S_P(d)\} = \max \{b^T y : (y, s) \in S_D(d)\}.$$

Therefore, the primal and dual optimal faces associated with $d$ are given by

$$\text{OPT}_P(d) = \{x : c^T x = z(d) \text{ and } x \in S_P(d)\},$$

$$\text{OPT}_D(d) = \{(y, s) : b^T y = z(d) \text{ and } (y, s) \in S_D(d)\},$$

respectively. Consider the following two subsets of indexes $B(d)$ and $N(d)$ in $\{1, \ldots, n\}$
associated with the programs $P(d)$ and $D(d)$, respectively:

\[
B(d) := \{ j : (y, s) \in \text{OPT}_P(d) \text{ implies } s_j = 0 \},
\]
\[
N(d) := \{ j : x \in \text{OPT}_D(d) \text{ implies } x_j = 0 \}.
\]

It is well known that the sets $B(d)$ and $N(d)$ satisfy the following properties:

1. The sets $B(d)$ and $N(d)$ form a unique partition of $\{1, \ldots, n\}$, that is,

\[
B(d) \cup N(d) = \{1, \ldots, n\},
\]
\[
B(d) \cap N(d) = \emptyset.
\]

2. There exists an optimal solution $x$ of $P(d)$ such that $x_{B(d)} > 0$.

3. There exists an optimal solution $(y, s)$ of $D(d)$ such that $s_{N(d)} > 0$.

The pair $[B(d), N(d)]$ is called the optimal partition (see [GT56]). The following definition is key for establishing our complexity results:

**Definition 6.1** We define the distance to degeneracy $\eta(d)$ of a data instance $d = (A, b, c) \in \mathcal{F}$ as

\[
\eta(d) := \inf \{ \| \Delta d \| : d + \Delta d \notin \mathcal{F} \text{ or } B(d) \neq B(d + \Delta d) \},
\]  
(6.1)

or equivalently,

\[
\eta(d) := \sup \{ \delta : \| \Delta d \| \leq \delta \text{ implies } d + \Delta d \in \mathcal{F} \text{ and } B(d) = B(d + \Delta d) \}.
\]  
(6.2)

In other words, we can change a data instance $d$ using a data perturbation $\Delta d$ of size $\| \Delta d \| \leq \eta(d)$ and still obtain a data instance $d + \Delta d$ having the same combina-
torial structure as the optimal partition associated with $d$.

Using these concepts, we will prove the following theorem in Section 6.2:

**Theorem 6.1** Let $d = (A, b, c)$ be a data instance in $\mathcal{D}$ such that $d \in \mathcal{F}$ and $\eta(d) > 0$. Then there exists a path-following interior-point algorithm that starting at $(x_0, y_0, s_0)$ sufficiently close to the central trajectory and with initial duality gap $\epsilon_0$ computes the optimal solution of $\mathcal{P}(d)$ in no more than

$$O\left(\sqrt{n} \log \left(\frac{\epsilon_0 \|d\|}{\eta(d)^2}\right)\right)$$

iterations. Each iteration involves the solution of two $k \times k$ systems of equations, with $k \leq n$.

Finally, in Section 6.4 we present further properties of the primal optimal face of a linear program in relation to the central trajectory associated with such program. In order to prove these properties, we use the well-known result that the central trajectory of a linear program converges to the analytic center of the optimal face. By using this fact, we obtain a different conditioning of the primal optimal face, namely the condition number associated with the analytic center problem defining the primal optimal face. For a given data instance $d = (A, b, c)$, the primal conditioning is more general than using the distance to degeneracy $\eta(d)$ because it will remain positive, even in the presence of multiple primal optima, as long as the matrix $A_{B(d)}$ has full-row rank. The main result of this section is stated in Lemma 6.6. This lemma presents an upper bound on the error between the optimal solution to the central trajectory problem $P_{\mu}(d)$ and the analytic center of the optimal primal face for $\mu$ sufficiently small.
6.2 Properties of the Distance to Degeneracy

We start by establishing an important relation between the distance to degeneracy $\eta(d)$ of a data instance $d = (A, b, c)$ and the distance to ill-posedness $\rho(d)$ with respect to the linear program in standard form associated with such instance.

**Proposition 6.1** Let $d = (A, b, c)$ be a data instance in $\mathcal{F}$. Then

$$\eta(d) \leq \rho(d).$$

**Proof:** Without loss of generality we assume that $\eta(d) > 0$. Given a positive number $\epsilon < \eta(d)$, it follows from Definition 6.1, item (6.2), that $d + \Delta d \in \mathcal{F}$ and $B(d) = B(d + \Delta d)$ for all data perturbation $||\Delta d|| \leq \eta(d) - \epsilon$. Hence, by definition of $\rho(d)$, $\eta(d) - \epsilon \leq \rho(d)$. Since $\epsilon$ is arbitrary, we conclude that $\eta(d) \leq \rho(d)$.

q.e.d.

The following lemma establishes an equivalence between the positiveness of the distance to degeneracy of a data instance and the uniqueness of primal and dual solutions of the corresponding linear program.

**Lemma 6.1** Let $d = (A, b, c)$ be a data instance in $\mathcal{F}$, then $\eta(d) > 0$ if and only if both $P(d)$ and $D(d)$ have unique solutions.

**Proof:** First we assume that $\eta(d) > 0$. By Proposition 6.1, we have $\rho(d) > 0$, and so, by Theorem 3.1, identity (3.1), $d \in \mathcal{F}_S$ and $A$ has full-row rank. Hence, by Proposition 2.7 and Proposition 2.8, both the primal and dual optimal faces $\epsilon$ssociated with $d$ are bounded. Suppose that $P(d)$ has multiple optima. Then there exists a vector $v$ such that $Av = 0$, $c^Tv = 0$, and $v \neq 0$. Since the primal optimal face $\text{OPT}_P(d)$ is bounded, it follows that there are at least two non-zero components of $v$ of different sign. Let $x$ be an optimal solution to $P(d)$ such that $x_{B(d)} > 0$ and
6.2 Properties of the Distance to Degeneracy

$x_{N(d)} = 0$. By considering vectors of the form $x + \lambda v$ for any scalar $\lambda$, it is clear that there exists a non-empty set $\tilde{B}$ and corresponding optimal solution $\bar{x}$ to $P(d)$ such that $\tilde{B} \subset B(d)$, $\tilde{B} \neq B(d)$, $\bar{x}_{\tilde{B}} > 0$, and $\bar{x}_{\bar{N}} = 0$, where $\bar{N} = \{1, \ldots, n\} \setminus \tilde{B}$ is non-empty. Since $\bar{x}$ is an optimal solution to $P(d)$, if $(\bar{y}, \bar{s})$ is an optimal solution to $D(d)$, then $\bar{s} = c - A^T \bar{y}$, $\bar{s} \geq 0$ and $\bar{x}^T \bar{s} = 0$. Given any $\epsilon > 0$ and for fixed $j \in B(d) \setminus \tilde{B}$, let $\Delta c(\epsilon) = \epsilon e_j$ and $\Delta d(\epsilon) = (0, 0, \Delta c(\epsilon))$. Observe that if we define $s(\epsilon) = c + \Delta c(\epsilon) - A^T \bar{y}$, then $s(\epsilon) \geq 0$ and $\bar{x}^T s(\epsilon) = 0$. Hence, $\bar{x}$ is an optimal solution of $P(d + \Delta d(\epsilon))$ and $(\bar{y}, s(\epsilon))$ is an optimal solution to $D(d + \Delta d(\epsilon))$. Thus, since $s(\epsilon)_j > 0$, we have $B(d + \Delta d(\epsilon)) \neq B(d)$, so that $\eta(d) \leq \|\Delta d(\epsilon)\| = \epsilon \|e_j\|_\infty$. Since $\epsilon$ is arbitrary, it follows that $\eta(d) = 0$, a contradiction.

Now, suppose that $D(d)$ has multiple optima. Then there exists a vector $u$ such that $A^T u \neq 0$ and $b^T u = 0$. Since the dual optimal face $\text{OPT}_D(d)$ is bounded, it follows that there are at least two non-zero components of $A^T u$ of different sign. Let $(y, s)$ be an optimal solution to $D(d)$ such that $s_{N(d)} > 0$ and $s_{B(d)} = 0$. By considering vectors of the form $y + \lambda u$ for any scalar $\lambda$, it is clear that there exist a non-empty set $\tilde{B}$ and corresponding optimal solution to $D(d)$, $(\bar{y}, \bar{s})$, such that $B(d) \subset \tilde{B}$, $B(d) \neq \tilde{B}$, $\bar{s} = c - A^T \bar{y}$, $\bar{s}_{\tilde{B}} > 0$, and $\bar{s}_{\bar{N}} = 0$, where $\bar{N} = \{1, \ldots, n\} \setminus \tilde{B}$ is non-empty. Since $(\bar{y}, \bar{s})$ is an optimal solution to $D(d)$, if $\bar{x}$ is an optimal solution to $P(d)$, then $A \bar{x} = b$, $\bar{x} \geq 0$ and $\bar{x}^T \bar{s} = 0$. Given any $\epsilon > 0$ and for fixed $j \in \tilde{B} \setminus B(d)$, let $\Delta b(\epsilon) = \epsilon A e_j$ and $\Delta d(\epsilon) = (0, \Delta b(\epsilon), 0)$. Observe that if we define $x(\epsilon) = \bar{x} + \epsilon e_j$, then $x(\epsilon) \geq 0$ and $x(\epsilon)^T \bar{s} = 0$. Hence, $x(\epsilon)$ is an optimal solution of $P(d + \Delta d(\epsilon))$ and $(\bar{y}, \bar{s})$ is an optimal solution to $D(d + \Delta d(\epsilon))$. Thus, since $x(\epsilon)_j > 0$, we have $B(d + \Delta d(\epsilon)) \neq B(d)$, so that $\eta(d) \leq \|\Delta d(\epsilon)\| = \epsilon \|Ae_j\|_1$. Since $\epsilon$ is arbitrary, it follows that $\eta(d) = 0$, again a contradiction.

To prove the other direction, assume that $P(d)$ and $D(d)$ both have unique solutions. Let $B = B(d)$. The uniqueness of primal and dual solutions together with the strict complementarity condition at optimality imply that $|B| = m$, $A_B$ is a non-
singular matrix, \( A_B^{-1} b > 0 \), and \( c_N - A_N^T A_B^{-T} c_B > 0 \). In light of Proposition 2.13, Proposition 2.14, and Proposition 3.1, it follows that there exists \( \epsilon > 0 \) such that for all \( \Delta d = (\Delta A, \Delta b, \Delta c) \) satisfying \( d + \Delta d \in \mathcal{F} \) and \( \|\Delta d\| < \epsilon \), implies \((A + \Delta A)_B\) is a non-singular matrix, \((A + \Delta A)_B^{-1}(b + \Delta b) > 0\), and \((c + \Delta c)_N - (A + \Delta A)_N^T (A + \Delta A)_B^{-T} (c + \Delta c)_B > 0\). Therefore, \( B(d + \Delta d) = B \) for all \( \Delta d \) such that \( \|\Delta d\| < \epsilon \). Hence, \( \eta(d) \geq \epsilon > 0 \).

q.e.d.

Given a data instance \( d = (A, b, c) \in \mathcal{F} \), consider the following set:

\[
\text{Opt}(d) := \{(x, y, s) : x \text{ solves } P(d) \text{ and } (y, s) \text{ solves } D(d)\}.
\]

From the proof of Lemma 6.1 we obtain the following corollary:

**Corollary 6.1** Let \( d = (A, b, c) \) be a data instance in \( \mathcal{F} \) such that both \( P(d) \) and \( D(d) \) have unique solutions. Then there exists a positive number \( \epsilon \) such that \( \text{Opt}(d + \Delta d) \) is a singleton and \( B(d + \Delta d) = B(d) \) for all data perturbation \( \Delta d \) satisfying \( \|\Delta d\| \leq \epsilon \).

The following proposition states an equivalent characterization of \( \eta(d) \) in terms of the distance to multiple-optima.

**Proposition 6.2** Let \( d = (A, b, c) \) be a data instance in \( \mathcal{F} \). Let

\[
\zeta(d) := \inf \{\|\Delta d\| : d + \Delta d \notin \mathcal{F} \text{ or } \text{Opt}(d + \Delta d) \text{ is not a singleton}\}.
\]

Then

\[
\eta(d) = \zeta(d).
\]

**Proof:** Let \( B = B(d) \). From Lemma 6.1 if \( \eta(d) = 0 \), then \( \text{Opt}(d) \) is not a singleton and \( \zeta(d) = 0 \). Hence, without loss of generality we assume that \( \eta(d) > 0 \). Given a positive
number $\epsilon < \eta(d)$, we have $d + \Delta d \in \mathcal{F}$ and $B(d + \Delta d) = B$ for all data perturbation $\Delta d$ satisfying $\|\Delta d\| \leq \eta(d) - \epsilon$. Since $\|\Delta d\| < \eta(d)$ and $\eta(d) \leq \eta(d + \Delta d) + \|\Delta d\|$, we have $\eta(d + \Delta d) > 0$. It follows from Lemma 6.1 that $\text{Opt}(d + \Delta d)$ is a singleton for all data perturbation $\Delta d$ satisfying $\|\Delta d\| \leq \eta(d) - \epsilon$. Therefore, $\eta(d) - \epsilon \leq \zeta(d)$ for all $\epsilon > 0$ sufficiently small, and so $\zeta(d) \geq \eta(d) > 0$.

On the other hand, since $\zeta(d) > 0$, let $\epsilon$ be a positive number such that $\epsilon < \zeta(d)$. Hence, we have $d + \Delta d \in \mathcal{F}$ and $\text{Opt}(d + \Delta d)$ is a singleton for all data perturbation $\Delta d$ satisfying $\|\Delta d\| \leq \zeta(d) - \epsilon$. For a given data perturbation $\Delta d$ with $\|\Delta d\| \leq \zeta(d) - \epsilon$, let $\hat{\lambda} = \sup\{\lambda : B(d + \lambda \Delta d) = B\}$. Suppose that $B(d + \hat{\lambda} \Delta d) = B$. From Corollary 6.1, there exists a number $\delta > 0$ such that $B(d + \hat{\lambda} \Delta d + \Delta \bar{d}) = B$ for all data perturbation $\Delta \bar{d}$ satisfying $\|\Delta \bar{d}\| \leq \delta$, contradicting that $\hat{\lambda}$ is a supremum. Hence, $B(d + \hat{\lambda} \Delta d) \neq B$. If $\text{Opt}(d + \hat{\lambda} \Delta d)$ is a singleton, again from Corollary 6.1, then there exists a number $\delta > 0$ such that $B(d + \hat{\lambda} \Delta d + \Delta \bar{d}) = B(d + \hat{\lambda} \Delta d) \neq B$ for all data perturbation $\Delta \bar{d}$ satisfying $\|\Delta \bar{d}\| \leq \delta$, contradicting again that $\hat{\lambda}$ is a supremum. Therefore, since $\text{Opt}(d + \lambda \Delta d)$ is a singleton for all $\lambda \in [0, 1]$, it follows that $\hat{\lambda} > 1$ and, in particular, that $B(d + \Delta d) = B$. In conclusion, we obtain that $d + \Delta d \in \mathcal{F}$ and $B(d + \Delta d) = B$ for all data perturbation $\Delta d$ satisfying $\|\Delta d\| \leq \zeta(d) - \epsilon$. Thus, $\zeta(d) - \epsilon \leq \eta(d)$ for all $\epsilon > 0$ sufficiently small. Therefore, $\zeta(d) \leq \eta(d)$ and the result follows.

q.e.d.

Given a data instance $d = (A, b, c) \in \mathcal{F}$, we say that $d$ is a primal non-degenerate data instance if there exists an optimal basis $B \subset \{1, \ldots, n\}$ such that

$$A_B^{-1}b \geq 0,$$

$$c_N - c_B A_B^{-1} A_N \geq 0,$$
where $N = \{1, \ldots, n\} \setminus B$. Similarly, we say that $d$ is a dual non-degenerate data instance if there exists an optimal basis $B \subset \{1, \ldots, n\}$ such that

$$A_B^{-1}b \geq 0,$$

$$c_N - c_B A_B^{-1} A_N > 0,$$

where again $N = \{1, \ldots, n\} \setminus B$.

The following proposition states an important property of primal and dual non-degenerate data instances.

**Proposition 6.3** Let $d = (A, b, c)$ be a data instance in $F$.

- If $d$ is primal non-degenerate, then $D(d)$ has a unique solution.
- If $d$ is dual non-degenerate, then $P(d)$ has a unique solution.

**Proof:** First, suppose that $d$ is primal non-degenerate. Then there exists at least one optimal basis $R \subset \{1, \ldots, n\}$ such that $R \subset B(d)$. Hence, $\text{rank}(A_{B(d)}) = m$. Therefore, there cannot exist a vector $u \neq 0$ such that $A_{B(d)}^T u = 0$, and so $D(d)$ has a unique solution. Next, suppose that $d$ is dual non-degenerate. Then there exists at least one optimal basis $R \subset \{1, \ldots, n\}$ such that $\{1, \ldots, n\} \setminus R \subset N(d)$. Hence, $B(d) \subset R$ and $\text{rank}(A_{B(d)}) = |B(d)| \leq |R| = m$. Therefore, there cannot exist a vector $v \neq 0$ such that $A_{B(d)} v = 0$, and so $P(d)$ has a unique solution.

q.e.d.
Observe that in the absence of the non-degeneracy assumption, Proposition 6.3 is not necessarily true. For example, consider the following data instance:

\[
d = \begin{pmatrix}
1 & 1 & 0 & 1 \\
0 & 0 & 1 & -1
\end{pmatrix}
\begin{pmatrix}
1 \\
0 \\
0 \\
1
\end{pmatrix}
\begin{pmatrix}
0
\end{pmatrix}.
\]

Notice that \(d \in \mathcal{F}\) and \(\rho(d) > 0\). Moreover, \(B(d) = \{1, 2\}\) and \(\text{rank}(A_{B(d)}) = 1\). Furthermore, it is easy to see that \(d\) is neither primal non-degenerate, nor dual non-degenerate. Also, \(d\) has both primal and dual multiple-optima.

Using Lemma 6.1 and Proposition 6.3, we prove the following lemma, which justifies why we call \(\eta(\cdot)\) the distance to degeneracy.

**Lemma 6.2** Let \(d = (A, b, c)\) be a data instance in \(\mathcal{F}\), then \(\eta(d) > 0\) if and only if \(d\) is primal and dual non-degenerate, and there exists a unique basis \(B = B(d)\) for which \(A_B\) is a non-singular matrix, \(A_B^{-1}b > 0\), and \(c_N - A_N^T A_B^{-T} c_B > 0\).

**Proof:** If \(d\) is primal and dual non-degenerate, then from Proposition 6.3 \(P(d)\) and \(D(d)\) have unique solutions, respectively. From Lemma 6.1 it follows that \(\eta(d) > 0\). To prove the other direction, observe that if \(\eta(d) > 0\), then Lemma 6.1 implies that \(P(d)\) and \(D(d)\) both have unique solutions. Let \(B = B(d)\). The uniqueness of primal and dual solutions together with the strict complementarity condition at optimality imply that \(|B| = m\), \(A_B\) is a non-singular matrix, \(A_B^{-1}b > 0\), and \(c_N - A_N^T A_B^{-T} c_B > 0\). Therefore, \(d\) is primal and dual non-degenerate.

q.e.d.

Given a data instance \(d = (A, b, c)\) such that \(d \in \mathcal{F}\) and \(\eta(d) > 0\), we have \(A_{B(d)}\) is a non-singular matrix. The following lemma relates the distance to degeneracy \(\eta(d)\)
6.3 Complexity Results

to the "classical" condition number $\kappa_1(A_{B(d)})$ of the matrix $A_{B(d)}$.

**Lemma 6.3** Let $d = (A, b, c)$ be a data instance in $F$ such that $\eta(d) > 0$. Then,

$$\kappa_1(A_{B(d)}) \leq \frac{\|A_{B(d)}\|}{\eta(d)} \leq \frac{\|d\|}{\eta(d)}.$$ 

**Proof:** Let $B = B(d)$. Since $\eta(d) > 0$, observe that $A_B$ is a non-singular matrix from Lemma 6.2. Let $\epsilon$ denote an arbitrary positive scalar. By identities (2.6) and (2.7), it follows that there exists a matrix $\Delta A_B$ in $\mathbb{R}^{[B] \times [B]}$ such that $A_B + \Delta A_B$ is a singular matrix and $\|\Delta A_B\| \leq \|A_B^{-1}\|^{-1} + \epsilon$. Let $\Delta d = ([\Delta A_B][0], 0, 0)$. If $d + \Delta d \notin F$, then $\eta(d) \leq \|\Delta d\|$. If $d + \Delta d \in F$, then it is clear that since $A_B + \Delta A_B$ is a singular matrix, then $B(d + \Delta d) \neq B$, and again $\eta(d) \leq \|\Delta d\|$. On the other hand, $\|\Delta d\| = \|\Delta A_B\| \leq \|A_B^{-1}\|^{-1} + \epsilon$. Therefore, we have $\eta(d) \leq \|A_B^{-1}\|^{-1}$, and the result follows.

q.e.d.

6.3 Complexity Results

In this section we prove Theorem 6.2. This theorem states a polynomial bound on the number of iterations of certain path-following interior-point algorithm for computing the optimal solutions to the linear programs $P(d)$ and $D(d)$, respectively, associated with the data instance $d = (A, b, c)$, where $d$ is such that $\eta(d) > 0$. The next lemma is instrumental in proving this theorem.

**Lemma 6.4** Let $d = (A, b, c)$ be a data instance in $F$ such that $\eta(d) > 0$. Let $x$ denote the optimal solution to $P(d)$ and let $(y,s)$ denote the optimal solution to
6.3 Complexity Results

\( D(d). \) Then

\[ x_j \geq \frac{\eta(d)}{\|d\|} \text{ for } j \in B(d), \]  \hspace{1cm} (6.3)

\[ s_j \geq \eta(d) \text{ for } j \in N(d). \]  \hspace{1cm} (6.4)

**Proof:** Let \( B = B(d) \) and \( N = N(d) \). Given \( j \in B \) and \( \epsilon > 0 \), let \( \Delta c = \epsilon e_j \) and \( \Delta b = -x_j \epsilon e_j \). Consider the data perturbation \( \Delta d = (0, \Delta b, \Delta c) \) and the vectors \( \Delta x = -x_j e_j \) and \( \Delta s = \epsilon e_j \). It follows that \( A(x + \Delta x) = b + \Delta b, x + \Delta x \geq 0, \)
\( A^T y + s + \Delta s = c + \Delta c, \) and \( s + \Delta s \geq 0. \) Furthermore, if \( \tilde{B} = B \setminus \{j\} \) and \( \tilde{N} = \{1, \ldots, n\} \setminus \tilde{B}, \) then \( (x + \Delta x)_B > 0, (s + \Delta s)_N > 0, \) and \( (x + \Delta x)^T (s + \Delta s) = 0. \)

Hence, \( B(d + \Delta d) = \tilde{B} \) is a proper subset of \( B. \) Therefore, for \( \epsilon \) sufficiently small, we have \( \eta(d) \leq \|\Delta d\| = \max \{x_j\|Ae_j\|_1, \epsilon\|e_j\|_\infty\} \leq x_j\|d\|, \) from which the statement (6.3) follows.

On the other hand, given \( j \in N \) and \( \epsilon > 0, \) let \( \Delta c = -s_j e_j \) and \( \Delta b = \epsilon A e_j. \)

Consider the data perturbation \( \Delta d = (0, \Delta b, \Delta c) \) and the vectors \( \Delta x = \epsilon e_j \) and \( \Delta s = -s_j e_j. \) As before, it follows that \( A(x + \Delta x) = b + \Delta b, x + \Delta x \geq 0, \)
\( A^T y + s + \Delta s = c + \Delta c, \) and \( s + \Delta s \geq 0. \) Furthermore, if \( \tilde{B} = B \cup \{j\} \) and \( \tilde{N} = \{1, \ldots, n\} \setminus \tilde{B}, \) then \( (x + \Delta x)_B > 0, (s + \Delta s)_N > 0, \) and \( (x + \Delta x)^T (s + \Delta s) = 0. \)

Hence, \( B(d + \Delta d) = \tilde{B} \) is a proper superset of \( B. \) Therefore, for \( \epsilon \) sufficiently small, we have \( \eta(d) \leq \|\Delta d\| = \max \{\epsilon\|Ae_j\|_1, s_j\|e_j\|_\infty\} = s_j, \) from which the statement (6.4) follows.

q.e.d.

**Lemma 6.5** Let \( d = (A, b, c) \) be a data instance in \( \mathcal{F} \) such that \( \eta(d) > 0. \) Let \( x \) denote the optimal solution to \( P(d) \) and let \( (y, s) \) denote the optimal solution to \( D(d). \) Let \( x(\mu) \) denote the optimal solution to \( P_\mu(d) \) and let \( (y(\mu), s(\mu)) \) denote the
optimal solution to $D_\mu(d)$. Then

$$\|x_N(d)(\mu)\|_1 \leq \frac{n\mu}{\eta(d)},$$

(6.5)

$$\|s_B(d)(\mu)\|_1 \leq \frac{n\mu\|d\|}{\eta(d)}.$$  

(6.6)

**Proof:** Let $B = B(d)$ and $N = N(d)$. It is not difficult to prove the following identity:

$$(x(\mu) - x)^T(s(\mu) - s) = 0.$$  

Hence, we have

$$x_B^Ts_B(\mu) + s_N^Tx_N(\mu) = n\mu.$$  

From this we obtain the following inequalities:

$$\|x_N(\mu)\|_1 \leq \frac{n\mu}{\min_{j\in N}\{s_j\}},$$

$$\|s_B(\mu)\|_1 \leq \frac{n\mu}{\min_{j\in B}\{x_j\}}.$$  

By using Lemma 6.4, we complete the proof.

q.e.d.

In order to measure the degree of difficulty of solving $P(d)$ by using an interior-point method, Ye introduced the following "condition number" $\sigma(\cdot)$ (see [Ye94]):

$$\sigma_P(d) := \min_{j\in B(d)} \max\{x_j : (x, y, s) \in \text{Opt}(d)\},$$

(6.7)

$$\sigma_D(d) := \min_{j\in N(d)} \max\{s_j : (x, y, s) \in \text{Opt}(d)\},$$

(6.8)

$$\sigma(d) := \min\{\sigma_P(d), \sigma_D(d)\}.$$  

(6.9)
In particular, Ye proved in [Ye94] that there exists a path-following interior-point algorithm that starting at \((x_0, y_0, s_0)\) sufficiently close to the central trajectory and with initial duality gap \(\epsilon_0\) in no more than

\[
O\left(\sqrt{n} \log\left(\frac{n\epsilon_0}{\sigma(d)^2}\right)\right)
\]

iterations computes feasible solutions to \(P(d)\) and \(D(d)\) from which the optimal partition \([B(d), N(d)]\) associated with \(d\) can be identified, and an optimal solution to \(P(d)\) and \(D(d)\) can be computed in an additional \(O(n^3)\) operations.

The following corollary relates Ye's condition number \(\sigma(d)\) to the distance to degeneracy \(\eta(d)\). The corollary is an immediate consequence of Lemma 6.4.

**Corollary 6.2** Let \(d = (A, b, c)\) be a data instance in \(\bar{\mathcal{D}}\). Then

\[
\sigma_P(d) \geq \frac{\eta(d)}{\|d\|}, \\
\sigma_D(d) \geq \eta(d), \\
\sigma(d) \geq \frac{\eta(d)}{\max \{1, \|d\|\}}.
\]

The main result of this section is stated in the following theorem. From this result, it is not difficult to prove the complexity bound shown at the end of this section.

**Theorem 6.2** Let \(d = (A, b, c)\) be a data instance in \(\bar{\mathcal{D}}\) such that \(d \in \mathcal{F}\) and \(\eta(d) > 0\). Let \(x\) be any extreme point of \(S_P(d)\) and let \((y, s)\) be any extreme point of \(S_D(d)\). Then

\[
\|x_{N(d)}\|_1 \notin (0, \eta(d)/\|d\|), \tag{6.10}
\]
\[
c^T x - z(d) \notin (0, \eta(d)^2/\|d\|), \tag{6.11}
\]
\[
\|s_{B(d)}\|_\infty \notin (0, \eta(d)), \tag{6.12}
\]
\[
z(d) - b^T y \notin (0, \eta(d)^2/\|d\|). \tag{6.13}
\]
6.3 Complexity Results

**Proof:** Let $B = B(d)$ and $N = N(d)$. From Lemma 6.1, it follows that there is a unique optimal solution to $P(d)$, so that if $x$ is optimal for $P(d)$, then $x_N = 0$ and $c^T x - z(d) = 0$. Therefore, from here on we assume that $x$ is not the optimal extreme point. Hence, there exist sets $\bar{B} \subset \{1, \ldots, n\}$ and $\bar{N} = \{1, \ldots, n\} \setminus \bar{B}$ such that $\bar{B} \cap N \neq \emptyset$, $x_B > 0$, and $x_{\bar{N}} = 0$. Let $(\hat{y}, \hat{s})$ denote the optimal solution to $D(d)$ (notice that $\hat{s}_N > 0$). Given $\epsilon > 0$, consider the data perturbation $\Delta d = (0, \Delta b, \Delta c)$, where

\[
\Delta b = - \sum_{j \in \bar{B} \cap \bar{N}} x_j A e_j,
\]
\[
\Delta c = \epsilon \sum_{j \in \bar{B} \cap \bar{N}} e_j,
\]

and the vectors $\Delta x$ and $\Delta s$ in $\mathbb{R}^n$ given by

\[
\Delta x = - \sum_{j \in \bar{B} \cap \bar{N}} x_j e_j,
\]
\[
\Delta s = \epsilon \sum_{j \in \bar{B} \cap \bar{N}} e_j.
\]

Then we obtain that $A(x + \Delta x) = b + \Delta b$, $x + \Delta x \geq 0$, $A^T \hat{y} + \hat{s} + \Delta s = c + \Delta c$, and $\hat{s} + \Delta s \geq 0$. Furthermore, $(x + \Delta x)_{B \cap \bar{B}} > 0$, $(\hat{s} + \Delta s)_{(B \cap \bar{B})^c} > 0$, and $(x + \Delta x)^T (\hat{s} + \Delta s) = 0$. Thus, $B(d + \Delta d) = B \cap \bar{B}$ is a proper subset of $B$. Therefore, for $\epsilon$ sufficiently small,

\[
\eta(d) \leq \|\Delta d\| = \max \left\{\left\| \sum_{j \in \bar{B} \cap \bar{N}} x_j A e_j \right\|_1, \epsilon \left\| \sum_{j \in \bar{B} \cap \bar{N}} e_j \right\|_\infty \right\} \leq \|d\| \|x_N\|_1,
\]

and statement (6.10) follows.
Next, assume that \( \hat{x} \) is the optimal solution to \( P(d) \). Then \( c^T x - z(d) = c^T x - c^T \hat{x} = s^T x \). It follows from Lemma 6.4 that \( s^T x \geq \eta(d) \| x_N \|_1 \). Therefore, statement (6.11) follows from statement (6.10).

On the other hand, again from Lemma 6.1, it follows that there is a unique optimal solution to \( D(d) \), so that if \( (y, s) \) is optimal for \( D(d) \), then \( s_B = 0 \) and \( z(d) - b^T y = 0 \). Therefore, from here on we assume that \( (y, s) \) is not the optimal extreme point. Hence, there exist sets \( \hat{B} \subset \{1, \ldots, n\} \) and \( \hat{N} = \{1, \ldots, n\} \setminus \hat{B} \) such that \( B \cap \hat{N} \neq \emptyset \), \( s_{\hat{N}} = c_{\hat{N}} - A_{\hat{N}}^T y > 0 \), and \( s_{\hat{B}} = c_{\hat{B}} - A_{\hat{B}}^T y = 0 \). Let \( \hat{x} \) denote the optimal solution to \( P(d) \) (notice that \( \hat{x}_B > 0 \)). Given \( \epsilon > 0 \), consider the data perturbation \( \Delta d = (0, \Delta b, \Delta c) \), where

\[
\Delta b = \epsilon \sum_{j \in \hat{B} \cap \hat{N}} Ae_j,
\Delta c = -\sum_{j \in \hat{B} \cap \hat{N}} s_j e_j,
\]

and the vectors \( \Delta x \) and \( \Delta s \) in \( \mathbb{R}^n \) given by

\[
\Delta x = \epsilon \sum_{j \in \hat{B} \cap \hat{N}} e_j,
\Delta s = -\sum_{j \in \hat{B} \cap \hat{N}} s_j e_j.
\]

Then we obtain that \( A(\hat{x} + \Delta x) = b + \Delta b, \hat{x} + \Delta x \geq 0, A^T y + s + \Delta s = c + \Delta c \), and \( s + \Delta s \geq 0 \). Furthermore, \( (\hat{x} + \Delta x)_{(N \cap \hat{N})^c} > 0, (s + \Delta s)_{N \cap \hat{N}} > 0 \), and \( (\hat{x} + \Delta x)^T (s + \Delta s) = 0 \). Thus, \( N(d + \Delta d) = N \cap \hat{N} \) is a proper subset of \( N \). Therefore, for \( \epsilon \) sufficiently small,

\[
\eta(d) \leq \| \Delta d \|
= \max \left\{ \epsilon \| \sum_{j \in \hat{B} \cap \hat{N}} Ae_j \|_1, \| \sum_{j \in \hat{B} \cap \hat{N}} s_j e_j \|_\infty \right\}
\]
and statement (6.12) follows.

Finally, assume that \((\hat{y}, \hat{s})\) is the optimal solution to \(D(d)\). Then \(z(d) - b^T y = b^T \hat{y} - b^T y = \hat{x}^T s\). It follows from Lemma 6.4 that \(\hat{x}^T s \geq \eta(d) ||s_B||_\infty / ||d||\). Therefore, statement (6.13) follows from statement (6.12).

q.e.d.

The next corollary follows immediately from Theorem 6.2.

**Corollary 6.3** Let \(d = (A, b, c)\) be a data instance in \(\mathcal{D}\) such that \(d \in \mathcal{T}\) and \(\tau(d) > 0\). Let \(x\) be the optimal solution to \(P(d)\) and let \((y, s)\) be the optimal solution to \(D(d)\). Then \(x\) is the only extreme point of \(S_P(d)\) contained in the ball \(\{v : ||v - x||_1 \leq \eta(d) / ||d||\}\). Furthermore, \((y, s)\) is the only extreme point of \(S_D(d)\) such that \(s\) is contained in the ball \(\{v : ||v - s||_\infty \leq \eta(d)\}\).

**Proof of Theorem 6.1:** Given a data instance \(d = (A, b, c) \in \mathcal{T}\) and any vector \(x \in S_P(d)\), then we can compute an extreme point \(\hat{x}\) of \(S_P(d)\) such that \(c^T \hat{x} \leq c^T x\) in at most \(O(n^3)\) arithmetic operations (see [BJS90, Gon92]). Analogously, given any vector \((y, s) \in S_D(d)\), then we can compute an extreme point \((\hat{y}, \hat{s})\) of \(S_D(d)\) such that \(b^T \hat{y} \geq b^T y\) in at most \(O(n^3)\) arithmetic operations.

Hence, consider a path-following interior-point algorithm that at the \(k\)-th iteration uses iterate \((x^k, y^k, s^k)\) to compute extreme points \(\hat{x}^k\) in \(S_P(d)\) and \((\hat{y}^k, \hat{s}^k)\) in \(S_D(d)\) such that \(c^T \hat{x}^k \leq c^T x^k\) and \(b^T \hat{y}^k \geq b^T y^k\). Observe that if \((\hat{x}^k)^T \hat{s}^k = 0\), then we have found the optimal solutions to \(P(d)\) and \(D(d)\). If \((\hat{x}^k)^T \hat{s}^k \neq 0\), then the algorithm executes a new iteration following the central trajectory.

Furthermore, if this algorithm starts at \((x_0, y_0, s_0)\) sufficiently close to the central trajectory and with initial duality gap \(\epsilon_0\), updates the duality gap using the iteration \(\mu_{k+1} = (1 + 1/(6\sqrt{n}))^{-1} \mu_k\), then from Theorem 6.2 the algorithm must stop after
at most $K$ iterations, where $K$ is such that $n\mu_K < \eta(d)^2/\|d\|$. In fact, if $n\mu_K < \eta(d)^2/\|d\|$, the current iterates $x^K$ and $(y^K, s^K)$ are feasible solutions of $P(d)$ and $D(d)$, respectively, and must satisfy $(x^K)^T s^K < \eta(d)^2/\|d\|$, and so the corresponding extreme points $\hat{x}^K$ and $(\hat{y}^K, \hat{s}^K)$ derived from $x^K$ and $(y^K, s^K)$, respectively, must be optimal.

We complete the proof by observing that $K = O\left(\sqrt{n} \log (\epsilon_0 \|d\|/\eta(d)^2)\right)$.

q.e.d.

6.4 Supplementary Results

Given a data instance $d = (A, b, c) \in \bar{D}$ and a set $B \subset \{1, \ldots, n\}$, consider the data instance $d_B \in \mathbb{R}^{m \times |B| + m + |B|}$ defined as $d_B = (A_B, b, c_B)$. In particular, we are interested in studying the analytic center problem (in equality form) associated with $d_B$:

$$AE(d_B) : \min \{p(x_B) : A_B x_B = b, x_B > 0\},$$

where as usual, $p(\cdot)$ denotes the logarithmic barrier function defined in Chapter 2. Observe that in this context, the set $B$ is not considered as part of the data, and we are only considering data perturbations of the form $\Delta d_B = (\Delta A_B, \Delta b, \Delta c_B)$. Hence, we can extend the definitions of the distance to ill-posedness $\rho_E(d_B)$ and the corresponding condition number $C(d_B)$ as stated in Chapter 2 for the program $AE(d_B)$. Note that the dual program associated with $AE(d_B)$ is given by

$$AED(d_B) : \min \\left\{b^T y - p(s_B) : A_B^T y_B + s_B = 0, s_B > 0\right\}.$$

In the case when $B = B(d)$, the solution of the problem $AE(d_B)$ is the analytic center $\hat{x}_B$ of the optimal face of $P(d)$, see [AM91].
We readily obtain from Lemma 4.1 and Lemma 4.3 the following corollary:

**Corollary 6.4** Given \( B \subset \{1, \ldots, n\} \), let \( d = (A, b, c) \in \mathcal{F} \) be such that \( \rho_E(d_B) > 0 \). Then

\[
\|x_B\|_1 \leq C_E(d_B),
\]
\[
\|x_B\|_1 \geq C_E(d_B)^{-1},
\]

for any feasible solution \( x_B \) of \( AE(d_B) \). Moreover, if \( \hat{x}_B \) is the optimal solution to \( AE(d_B) \) then

\[
\hat{x}_j \geq (n C_E(d_B))^{-1},
\]

for all \( j \in B \).

Furthermore, from Corollary 4.2 and Corollary 4.5, we obtain the following proposition:

**Proposition 6.4** Given \( B \subset \{1, \ldots, n\} \), let \( d = (A, b, c) \in \mathcal{F} \) be such that \( \rho_E(d_B) > 0 \). For a given scalar \( \alpha \in [0, 1) \), let \( x \in S_P(d) \) be such that

\[
\|x_N\|_1 \leq \frac{\alpha \rho_E(d_B)}{\|d\|},
\]

where \( N = \{1, \ldots, n\} \setminus B \). Then

\[
\|x_B\|_1 \leq \left( \frac{1 + \alpha}{1 - \alpha} \right) C_E(d_B),
\]
\[
\|x_B\|_1 \geq \left( \frac{1 - \alpha}{1 + \alpha} \right) C_E(d_B)^{-1}.
\]

**Proof:** Let \( x \in S_P(d) \) and \( \Delta d_B = (\Delta A_B, \Delta b, \Delta c_B) = (0, -A_N x_N, 0) \). Observe that

\[
\|\Delta d_B\| = \|A_N x_N\|_1
\]
Furthermore, $x_B$ is a feasible solution of $AE(d_B + \Delta d_B)$. Therefore, the result follows from Corollary 4.2 and Corollary 4.5.

q.e.d.

The following lemma establishes a bound on the convergence of the central trajectory to the analytic center of the optimal face when we have unique primal and dual optimal solutions. The lemma follows directly from Theorem 4.1 and Lemma 6.5 by observing that given a data instance $d = (A, b, c) \in \mathcal{F}_S$, if $x(\mu)$ is the optimal solution to the central trajectory problem $P_\mu(d)$, $B = B(d)$, and $N = N(d)$, then as proven by Adler and Monteiro [AM91] $x_B(\mu)$ is the optimal solution to the analytic center problem $AE(d_B + \Delta d_B)$, where $\Delta d_B = (0, -A_Nx_N(\mu), 0)$.

**Lemma 6.6** Let $d = (A, b, c) \in \mathcal{F}$ be such that $\eta(d) > 0$ and $\rho_\mathcal{E}(d_B) > 0$. Let $B = B(d)$, $N = N(d)$. For a given scalar $\alpha \in [0, 1)$, let $x(\mu)$ be the optimal solution to the central trajectory problem $P_\mu(d)$, where $\mu$ is such that

$$0 < \mu \leq \alpha \frac{\rho_\mathcal{E}(d)\eta(d)}{n\|d\|}.$$  

(6.14)

Then

$$\|\hat{x}_B - x_B(\mu)\|_1 \leq \mu \frac{14mn^3}{(1 - \alpha^3)\eta(d)\|d\|} C_\mathcal{E}(d_B)^7,$$

(6.15)

and

$$\|x_N(\mu)\|_1 \leq \frac{\mu n}{\eta(d)},$$

(6.16)

where $\hat{x}_B$ is the optimal solution to $AE(d_B)$. 
6.4 Supplementary Results

Proof: We first observe that $x_B(\mu)$ is the optimal solution to the analytic center problem $AE(d_B + \Delta d_B)$, where $\Delta d_B = (0, -A_N x_N(\mu), 0)$. Furthermore, from Lemma 6.5, we have

$$\|x_N(\mu)\|_1 \leq \frac{n \mu}{\eta(d)} \leq \frac{\alpha \rho_E(d_B)}{\|d\|},$$

and so $\|\Delta d_B\| \leq \alpha \rho_E(d_B)$. It follows from Theorem 4.1 that

$$\|\hat{x}_B - x_B(\mu)\|_1 \leq \frac{\|\Delta d_B\|}{\|d_B\|} \frac{14 mn^2}{(1 - \alpha)^3} C_E(d_B)^7 \leq \frac{\|d\| \|x_N(\mu)\|_1}{\|d_B\|} \frac{14 mn^2}{(1 - \alpha)^3} C_E(d_B)^7 \leq \frac{\|d\| \mu}{\eta(d) \|d_B\|} \frac{14 mn^3}{(1 - \alpha)^3} C_E(d_B)^7,$$

and the result follows.

q.e.d.
Chapter 7

Centering and Central Trajectory Problems for a Conic Linear System

7.1 Overview

In this chapter we extend some of the results concerning analytic center problems (from Chapter 4) and central trajectory problems (from Chapter 5) to conic linear systems with a self-concordant barrier function. We start this section by introducing several concepts taken from Nesterov and Nemirovskii in [NN94] concerning self-concordant functionals, self-concordant barriers, and logarithmically homogeneous barriers. We present some of the well-known properties that these types of functionals satisfy. We end this section by introducing two important concepts: normal barriers and the Legendre transformation. These notions are key for our later discussion of conic analytic center problems and central trajectory problems.

In Section 7.2, we introduce the central trajectory problem associated with a conic
linear system as well as the corresponding Karush-Kuhn-Tucker optimality conditions for this problem. Furthermore, we introduce Renegar’s condition number in this context. Similarly, in Section 7.3, we introduce conic analytic center problems in equality and inequality form. In addition, we prove Lemma 7.1 and Lemma 7.2, which state bounds on the feasible solutions to these two conic analytic center problems in terms of their corresponding condition numbers, thereby generalizing Lemma 4.1 and Lemma 4.2, respectively, from Chapter 4.

In Section 7.4, we generalize Theorem 5.1 and Theorem 5.2 from Chapter 5 to conic linear systems. In fact, Theorem 7.2 and Theorem 7.3, respectively, state analogous bounds to those of their Chapter 5 counterparts for solutions along the central trajectory of a conic linear system. Finally, in Section 7.5, we include for completeness the proofs of some of the results for self-concordant functionals taken from [NN94].

Given a finite-dimensional normed vector space \((\mathcal{X}, \| \cdot \|)\), let \(C_{\mathcal{X}}\) denote a closed convex pointed cone in \(\mathcal{X}\) with non-empty interior. Let \((\mathcal{X}^*, \| \cdot \|_*)\) denote the dual space of \(\mathcal{X}\), where \(\| \cdot \|_*\) denotes the dual norm induced by \(\| \cdot \|\).

**Definition 7.1 [NN94] Self-Concordant Functional.** Let \(\phi\) be a real-valued functional defined on \(\text{int}(C_{\mathcal{X}})\) and let \(a\) be a positive scalar. The functional \(\phi\) is called an \(a\)-self-concordant functional on the cone \(C_{\mathcal{X}}\) if \(\phi\) is a convex three-times Fréchet differentiable functional on \(\text{int}(C_{\mathcal{X}})\) and the following inequality holds:

\[
\left| \nabla^3 \phi(x)[v, v, v] \right| \leq 2a^{-1/2} \left( v^T \nabla^2 \phi(x)v \right)^{3/2},
\]

(7.1)

for all \(x \in \text{int}(C_{\mathcal{X}})\) and \(v \in \mathcal{X}\).

Observe that by using the second-order differential of \(\phi\), an \(a\)-self-concordant
functional on the cone $C_X$, we obtain a semi-norm on the space $X$, namely

$$
\|v\|_x := \left( \frac{1}{a} v^T \nabla^2 \phi(x) v \right)^{1/2},
$$

for a given $x \in \text{int}(C_X)$ and for all $v \in X$. Hence, for any $v \in X$, inequality (7.1) can be rewritten as

$$
\left| \nabla^3 \phi(x)[v, v, v] \right| \leq 2a \|v\|_x^3,
$$

or equivalently, when $\|v\|_x > 0$, as

$$
\left| \nabla^3 \phi(x) \left[ \frac{v}{\|v\|_x}, \frac{v}{\|v\|_x}, \frac{v}{\|v\|_x} \right] \right| \leq 2a.
$$

**Definition 7.2 [NN94] Strongly Self-Concordant Functional.** Let $\phi$ be an a-self-concordant functional on $C_X$. The functional $\phi$ is called a strongly a-self-concordant functional on $C_X$ if

$$
\lim_{k \to \infty} \phi(x_k) = \infty,
$$

for all sequence $\{x_k : k \geq 1\}$ in $\text{int}(C_X)$ converging to a point $y$ on the boundary $\partial C_X$ of $C_X$. Equivalently, the functional $\phi$ is called a strongly a-self-concordant functional on $C_X$ if the level sets $\{x \in \text{int}(C_X) : \phi(x) \leq t\}$ are closed in $X$ for each scalar $t \in \mathbb{R}$.

The following set is called *Dikin's ellipsoid* of $\phi$ centered at $x \in \text{int}(C_X)$ and of radius $r$:

$$
\{v : \|v - x\|_x \leq r\}.
$$

The following theorem is one of the main technical tools used in this theory:

**Theorem 7.1 [NN94]** Let $\phi$ be an a-self-concordant functional on the cone $C_X$ and let $x$ be an element of $\text{int}(C_X)$. Then for each $y \in \text{int}(C_X)$ such that $\|y - x\|_x < 1$ we
have

\[(1 - r)\|v\|_z \leq \|v\|_y \leq \frac{1}{1 - r} \|v\|_z,\]

(7.2)

for all \(v \in \mathcal{X}\), where \(r := \|y - x\|_z\). Furthermore, if \(\phi\) is also a strongly a-self-concordant functional on \(C_X\), then every Dikin's ellipsoid of \(\phi\) centered at \(x\) of radius \(r < 1\) is contained in \(\text{int}(C_X)\).

**Definition 7.3 [NN94] Non-degenerate Self-Concordant Functional.** Let \(\phi\) be an a-self-concordant functional on \(C_X\). The functional \(\phi\) is called a non-degenerate a-self-concordant functional on \(C_X\) if \(\phi\) is strictly convex.

Observe that if \(\phi\) is a non-degenerate a-self-concordant functional on the cone \(C_X\) and \(x \in \text{int}(C_X)\), then the semi-norm \(\|\cdot\|_z\) becomes a norm on \(\mathcal{X}\).

**Definition 7.4 [NN94] Non-degenerate Self-Concordant Barrier.** Let \(\phi\) be a non-degenerate strongly 1-self-concordant functional on \(C_X\) and let \(\vartheta\) be a non-negative scalar. The functional \(\phi\) is called a non-degenerate \(\vartheta\)-self-concordant barrier for the cone \(C_X\) if

\[\|\nabla^2 \phi(x)^{-1} \nabla \phi(x)\|_z^2 \leq \vartheta,\]

(7.3)

for all \(x \in \text{int}(C_X)\).

**Definition 7.5 [NN94] Logarithmically Homogeneous Barrier.** Let \(\phi\) be a real-valued functional defined on \(\text{int}(C_X)\) and let \(\vartheta \geq 1\) be a scalar, where \(C_X\) is a convex cone. The functional \(\phi\) is called a \(\vartheta\)-logarithmically homogeneous barrier for the cone \(C_X\) if \(\phi\) is a convex twice continuously Fréchet differentiable functional on \(\text{int}(C_X)\) such that

\[\lim_{k \to \infty} \phi(x_k) = \infty,\]

for all sequence \(\{x_k : k \geq 1\}\) in \(\text{int}(C_X)\) converging to a point \(y\) on the boundary \(\partial C_X\)
of $C_\mathcal{X}$, and the following identity holds:

$$
\phi(tx) = \phi(x) - \vartheta \ln(t),
$$

(7.4)

for all $x \in \text{int}(C_\mathcal{X})$ and $t > 0$.

**Definition 7.6 [NN94] Normal Barrier.** A functional $\phi$ is called a $\vartheta$-normal barrier for the cone $C_\mathcal{X}$ if $\phi$ is a $\vartheta$-logarithmically homogeneous barrier for the cone $C_\mathcal{X}$ and $\phi$ is a 1-self-concordant functional on $C_\mathcal{X}$.

As proven in [NN94], if $\phi$ is a strictly convex $\vartheta$-normal barrier for the cone $C_\mathcal{X}$, then $\phi$ is a non-degenerate $\vartheta$-self-concordant barrier for the cone $C_\mathcal{X}$. From now on, we assume that all normal barriers are also strictly convex functionals, and hence non-degenerate self-concordant barriers. The following proposition taken from [NN94] states properties of logarithmically homogeneous barriers. We include its proof for completeness in Section 7.5 (at the end of the chapter).

**Proposition 7.1 [NN94]** Let $\phi$ be a $\vartheta$-logarithmically homogeneous barrier for $C_\mathcal{X}$ and $\vartheta \geq 0$. Then for each $x \in \text{int}(C_\mathcal{X})$ and $t > 0$,

$$
\nabla \phi(tx) = \frac{1}{t} \nabla \phi(x), \quad \nabla^2 \phi(tx) = \frac{1}{t^2} \nabla^2 \phi(x),
$$

(7.5)

$$
\nabla^2 \phi(x)x = -\nabla \phi(x),
$$

(7.6)

$$
\nabla \phi(x)^T x = -\vartheta,
$$

(7.7)

$$
x^T \nabla^2 \phi(x)x = \vartheta, \quad \nabla \phi(x)^T \left(\nabla^2 \phi(x)\right)^{-1} \nabla \phi(x) = \vartheta,
$$

(7.8)

If $\phi$ is a $\vartheta$-normal barrier for $C_\mathcal{X}$, then we denote by $\phi^*(s)$ its corresponding Legendre transformation, that is,

$$
\phi^*(s) = \sup\{-s^T x - \phi(x) : x \in \text{int}(C_\mathcal{X})\},
$$

(7.9)
for all $s \in \mathcal{X}^*$. In particular, $\phi^*(s) < \infty$ for all $s \in \text{int}(C_C^*)$, where $C_C^* = \{ s \in \mathcal{X}^* : s^T \mathbf{x} \ge 0 \text{ for all } \mathbf{x} \in C_C \}$ denotes the polar of the cone $C_C$. As proven in [NN94], $\phi^*$ is also a $\vartheta$-normal barrier for $C_C^*$. The following proposition contains properties relating the barrier $\phi$ to its Legendre transformation $\phi^*(s)$. A proof of this proposition can be found in [NN94] and [NT94], and is repeated in Section 7.5 (at the end of the chapter) for completeness.

**Proposition 7.2** [NN94, NT94] Let $\phi$ be a $\vartheta$-normal barrier for $C_C$ and $\vartheta \ge 0$. Then for each $\mathbf{x} \in \text{int}(C_C)$ and $s \in \text{int}(C_C^*)$,

\begin{align*}
-\nabla \phi(\mathbf{x}) & \in \text{int}(C_C^*), \quad (7.10) \\
-\nabla \phi^*(s) & \in \text{int}(C_C), \quad (7.11) \\
\phi^*(-\nabla \phi(\mathbf{x})) & = -\vartheta - \phi(\mathbf{x}), \quad (7.12) \\
\phi(-\nabla \phi^*(s)) & = -\vartheta - \phi^*(s), \quad (7.13) \\
\nabla \phi^*(-\nabla \phi(\mathbf{x})) & = -\mathbf{x}, \quad (7.14) \\
\nabla \phi(-\nabla \phi^*(s)) & = -s, \quad (7.15) \\
\nabla^2 \phi^*(-\nabla \phi(\mathbf{x})) & = (\nabla^2 \phi(\mathbf{x}))^{-1}, \quad (7.16) \\
\nabla^2 \phi(-\nabla \phi^*(s)) & = (\nabla^2 \phi^*(s))^{-1}, \quad (7.17) \\
\phi(\mathbf{x}) + \phi^*(s) & \ge -\vartheta + \vartheta \ln \vartheta - \vartheta \ln s^T \mathbf{x}, \quad (7.18)
\end{align*}

and the last inequality is satisfied as an equality if and only if $s = -\alpha \nabla \phi(\mathbf{x})$ for some $\alpha > 0$. 
7.2 The Central Trajectory of a Conic Linear System

Given two finite-dimensional normed vector spaces \( \mathcal{X} \) and \( \mathcal{Y} \), let \( C_\mathcal{X} \) denote a closed convex pointed cone in \( \mathcal{X} \) with non-empty interior. We denote by \( L(\mathcal{X}, \mathcal{Y}) \) the set of linear operators mapping \( \mathcal{X} \) into \( \mathcal{Y} \). Let us consider a data set \( \mathcal{D} \) defined as

\[
\mathcal{D} = \{ d = (A, b, c) : A \in L(\mathcal{X}, \mathcal{Y}), b \in \mathcal{Y}, c \in \mathcal{X}^* \},
\]

where \( \mathcal{X}^* \) denotes the dual space of \( \mathcal{X} \). We write \( x \in C_\mathcal{X} \) as \( x \geq 0 \) and \( x \in \text{int}(C_\mathcal{X}) \) as \( x > 0 \). Similarly, we write \( s \in C_\mathcal{X}^* \) as \( s \geq 0 \) and \( s \in \text{int}(C_\mathcal{X}^*) \) as \( s > 0 \). Although we are using the same notation \( \geq \) and \( > \) to denote inclusion in different sets, it will always be clear from the context which relationship we are employing.

Given \( \vartheta \geq 0 \), let \( \phi \) be a \( \vartheta \)-normal barrier for \( C_\mathcal{X} \) and \( \phi^*(s) \) be the corresponding Legendre transformation of \( \phi \). Given a data instance \( d = (A, b, c) \in \mathcal{D} \), we consider the following optimization problems:

\[
P_\mu(d) : \inf \{ c^T x + \mu \phi(x) : A x = b, x > 0 \},
\]

\[
D_\mu(d) : \sup \{ b^T y - \mu \phi^*(s) : s = c - A^T y > 0, y \in \mathcal{Y}^* \}.
\]

Observe that \( P_\mu(d) \) and \( D_\mu(d) \) are a pair of Lagrangian dual problems. To be precise, the dual problem corresponding to \( P_\mu(d) \) is \( \sup \{ b^T y - \mu \phi^*(s/\mu) : s = c - A^T y > 0, y \in \mathcal{Y}^* \} \), but since this dual and \( D_\mu(d) \) only differ in the constant \( -\mu \vartheta \ln \mu \), we refer to \( D_\mu(d) \) as the dual. We regard \( \mu \) as a positive real parameter independent of the data instance \( d \in \mathcal{D} \).

Associated with the problems \( P_\mu(d) \) and \( D_\mu(d) \), we have the following Karush-
Kuhn-Tucker optimality conditions:

\[-\mu \nabla \phi(x) = s, \text{ equivalently } -\mu \nabla \phi^*(s) = x, \quad (7.19)\]

\[Ax = b, A^T y + s = c, \quad (7.20)\]

\[x > 0, y \in Y^*, s > 0. \quad (7.21)\]

Observe that because of the convexity of $\phi(x)$, these conditions are necessary and sufficient. In particular, observe that if $x$, $y$, and $s$ satisfy these conditions, then

\[s^T x = \mu \theta. \quad (7.22)\]

Furthermore, using (7.19), (7.22), and Proposition 7.2, statement (7.18), we have

\[c^T x = b^T y - \mu \phi^*(s) - \mu \phi(x) - \mu \theta \ln \mu = b^T y + \mu \theta. \quad (7.23)\]

We proceed to define a norm over the data set $\mathcal{D}$. Although the norms defined on the vector spaces $\mathcal{X}$ and $\mathcal{Y}$ are not necessarily the same, in order to simplify our exposition we will use the same notation $\| \cdot \|$ to represent both norms. We allow this abuse of notation because it is always clear from the context which norm we are employing. Similarly, we will use the notation $\| \cdot \|_*$ to represent the dual norms on the dual spaces $\mathcal{X}^*$ and $\mathcal{Y}^*$, respectively. That is,

\[\|s\|_* = \max \{ |s^T x| : \|x\| \leq 1, x \in \mathcal{X} \},\]

\[\|y\|_* = \max \{ |y^T v| : \|v\| \leq 1, v \in \mathcal{Y} \},\]

for all $s \in \mathcal{X}^*$ and $y \in \mathcal{Y}^*$. For a given linear operator $A$ in $L(\mathcal{X}, \mathcal{Y})$ we define $\|A\|$
to be the operator norm, namely,

$$\|A\| = \max \{\|Ax\| : x \in \mathcal{X}, \|x\| \leq 1\}.$$ 

Given a data instance \(d = (A, b, c) \in \mathcal{D}\), we define the following norm on \(\mathcal{D}\):

$$\|d\| = \max\{\|A\|, \|b\|, \|c\|\}.$$  \hspace{1cm} (7.24)

Observe that if \(A^T \in L(\mathcal{Y}^*, \mathcal{X}^*)\) denotes the adjoint operator of \(A\), then the operator norm \(\|A^T\|_*\) of \(A^T\) satisfies \(\|A^T\|_* = \|A\|\). Finally, if \(v\) and \(w\) are vectors in \(\mathcal{V}\) and \(\mathcal{W}\), respectively, we can define the norm of the product vector \((v, w)\) as \(\|(v, w)\|_{\mathcal{V} \times \mathcal{W}} = \max\{\|v\|_{\mathcal{V}}, \|w\|_{\mathcal{W}}\}\), whose corresponding dual norm is \(\|(v^*, w^*)\|_{(\mathcal{V} \times \mathcal{W})^*} = \|v^*\|_{\mathcal{X}^*} + \|w^*\|_{\mathcal{W}^*}\), for all \(v^* \in \mathcal{V}^*\) and \(w^* \in \mathcal{W}^*\), respectively.

We are interested in studying the following subsets of data instances:

\[
\mathcal{F}_P = \{d = (A, b, c) \in \mathcal{D} : Ax = b, x > 0 \text{ is feasible}\},
\]

\[
\mathcal{F}_D = \{d = (A, b, c) \in \mathcal{D} : c - A^Ty > 0, y \in \mathcal{Y} \text{ is feasible}\},
\]

\[
\mathcal{F} = \mathcal{F}_P \cap \mathcal{F}_D.
\]

Observe that \(\mathcal{F}_P\) consists of data instances \(d\) for which \(P_\mu(d)\) is feasible. Similarly, \(\mathcal{F}_D\) consists of data instances \(d\) for which \(D_\mu(d)\) is feasible. Finally, \(\mathcal{F}\) consists of data instances \(d\) for which both \(P_\mu(d)\) and \(D_\mu(d)\) are feasible. It is also appropriate to introduce the so-called sets of "ill-posed" data instances: \(\mathcal{B}_P = \partial\mathcal{F}_P\), \(\mathcal{B}_D = \partial\mathcal{F}_D\), and \(\mathcal{B} = \partial\mathcal{F}\), where \(\partial\mathcal{V}\) denotes the boundary of the set \(\mathcal{V}\). For a data instance
7.2 The Central Trajectory of a Conic Linear System

\(d \in \tilde{D}\), we define the following distances to ill-posedness as:

\[
\rho_P(d) = \inf\{ \|\Delta d\| : d + \Delta d \in B_P \}, \\
\rho_D(d) = \inf\{ \|\Delta d\| : d + \Delta d \in B_D \}, \\
\rho(d) = \inf\{ \|\Delta d\| : d + \Delta d \in B \},
\]

see [Ren94, Ren95a, Ren95b]. For instance, if \(d \in \mathcal{F}\), then \(\rho(d)\) can be interpreted as a measure of how much the data instance \(d\) can be perturbed before one of the feasible regions of the problems \(P_\mu(d)\) or \(D_\mu(d)\) becomes empty. Corresponding to each distance to ill-posedness we have a condition number:

\[
C_P(d) = \begin{cases} 
\frac{\|d\|}{\rho_P(d)} & \text{if } \rho_P(d) > 0, \\
\infty & \text{if } \rho_P(d) = 0.
\end{cases}
\]

\[
C_D(d) = \begin{cases} 
\frac{\|d\|}{\rho_D(d)} & \text{if } \rho_D(d) > 0, \\
\infty & \text{if } \rho_D(d) = 0.
\end{cases}
\]

\[
C(d) = \begin{cases} 
\frac{\|d\|}{\rho(d)} & \text{if } \rho(d) > 0, \\
\infty & \text{if } \rho(d) = 0.
\end{cases}
\]

The condition number \(C(d)\) can be viewed as a scale-invariant reciprocal of \(\rho(d)\), as it is elementary to demonstrate that \(C(d) = C(\alpha d)\) for any positive scalar \(\alpha\). Observe that since \(\tilde{d} = (\tilde{A}, \tilde{b}, \tilde{c}) = (0, 0, 0) \in B\) and \(B\) is a closed set, then for any \(d \notin B\) we have \(\|d\| = \|d - \tilde{d}\| \geq \rho(d) > 0\), so that \(C(d) \geq 1\). The value of \(C(d)\) is a measure of the relative conditioning of the data instance \(d\). Similar arguments apply to \(C_P(d)\) and \(C_D(d)\).

The following proposition is a special case of the extension \(\text{inm}\) of the Hahn-Banach Theorem (see Corollary 2 in Luenberger [Lue68], p. 112). A simple and short proof for finite-dimensional spaces is presented in [FV95].

**Proposition 7.3** Given a finite dimensional normed space \(X\) and \(u \in X\), there exists
$\bar{u} \in X^*$ such that $\bar{u}^T u = \|u\|$ and $\|\bar{u}\|_\ast = 1$.

7.3 Conic Analytic Center Problems

In this section we study the feasibility regions of $P_\mu(d)$ and $D_\mu(d)$, for a given data instance $d \in \bar{D}$, assuming that either one of them is a bounded set. We will show that by using analytic center problems we can derive very interesting properties concerning these regions.

Given a data instance $d = (A, b, c) \in \bar{D}$ and a $\theta$-logarithmically homogeneous barrier $\phi$ for $C_X$ for some $\theta \geq 0$, the analytic center problem in equality form, denoted $AE(d)$, is defined as:

$$AE(d) : \inf\{\phi(x) : Ax = b, x > 0\}.$$

Structurally, the program $AE(d)$ is closely related to the central trajectory problem $P_\mu(d)$, and was first extensively studied by Sonnevend for the positive orthant of $\mathbb{R}^n$ using the barrier $\phi(x) = -\sum_{j=1}^n \ln x_j$, see [Son85a] and [Son85b]. In terms of data dependence, note that the program $AE(d)$ does not depend on the data $c$. It is well known that $AE(d)$ has a unique solution when its feasible region is bounded and non-empty. We call this unique solution the analytic center of the equality form.

Associated with the problem $AE(d)$, we have the following Karush-Kuhn-Tucker optimality conditions:

$$-\nabla \phi(x) = t, \text{equivalently } -\nabla \phi^*(t) = x,$$

$$Ax = b, A^T u + t = 0,$$  \hspace{1cm} (7.25) \hspace{1cm} (7.26)
\[ x > 0, u \in \mathcal{Y}^*, t > 0. \]  

(7.27)

Observe that because of the convexity of \( \phi(x) \), these conditions are necessary and sufficient. In particular, observe that if \( x, u, \) and \( t \) satisfy these conditions, then

\[ t^T x = \vartheta. \]  

(7.28)

Furthermore, using (7.25), (7.28), and Proposition 7.2, statement (7.18), we have

\[ b^T u + \vartheta = \phi(x) + \phi^*(t) + \vartheta = 0. \]  

(7.29)

Similarly, we define the analytic center problem in inequality form, denoted \( AI(d) \), as:

\[ AI(d) : \inf \{ \phi^*(s) : s = c - A^Ty > 0, y \in \mathcal{Y}^* \}. \]

In terms of data dependence, the program \( (AI(d)) \) does not depend on the data \( b \). The program \( AI(d) \) has a unique solution when its feasible region is bounded and non-empty, and we call this unique solution the analytic center of the inequality form. Note in particular that the two programs \( AE(d) \) and \( AI(d) \) are not Lagrangian duals of each other. As we will show later, the study of these problems is relevant to obtain certain results for the central trajectory problem.

As with the equality case, associated with the problem \( AI(d) \), we have the following Karush-Kuhn-Tucker optimality conditions:

\[ -\nabla \phi(v) = s, \text{equivalently } -\nabla \phi^*(s) = v, \]  

(7.30)

\[ Av = 0, A^Ty + s = c, \]  

(7.31)
\[ v > 0, y \in Y^*, s > 0. \] (7.32)

Again, observe that because of the convexity of \( \phi^*(s) \), these conditions are necessary and sufficient. In particular, observe that if \( v, y, \) and \( s \) satisfy these conditions, then
\[ v^T s = \vartheta. \] (7.33)

By combining (7.30), (7.33), and Proposition 7.2, statement (7.18), we have
\[ -c^T v + \vartheta = \phi(v) + \phi^*(s) + \vartheta = 0. \] (7.34)

Since we are interested in studying these analytic center problems when they have solutions, we introduce the following sets of feasible data instances to both \( AE(d) \) and \( AI(d) \) respectively:

\[ \mathcal{F}_E = \{ d = (A, b, c) \in \overline{D} : Ax = b, x > 0, A^T u > 0 \text{ is feasible} \}, \]
\[ \mathcal{F}_I = \{ d = (A, b, c) \in \overline{D} : c - A^T y > 0, Av = 0, v > 0, y \in Y^* \text{ is feasible} \}. \]

Observe that \( \mathcal{F}_E \) consists of data instances \( d \) for which \( AE(d) \) is feasible and attains its optimal solution. Similarly, \( \mathcal{F}_I \) consists of data instances \( d \) for which \( AI(d) \) is feasible and attains its optimal solution. As before, we introduce sets of “ill-posed” data instances corresponding to these analytic center problems: \( B_E = \partial \mathcal{F}_E \) and \( B_I = \partial \mathcal{F}_I \).

For a data instance \( d \in \overline{D} \), we define the following distances to ill-posedness as:
\[ \rho_E(d) = \inf\{ \| \Delta d \| : d + \Delta d \in B_E \}, \]
\[ \rho_I(d) = \inf\{ \| \Delta d \| : d + \Delta d \in B_I \}. \]
7.3 Conic Analytic Center Problems

For instance, if \( d \in \mathcal{F}_E \), then \( \rho_E(d) \) can be interpreted as a measure of how much the data instance \( d \) can be perturbed before the feasible region of the problem \( AE(d) \) becomes either empty or unbounded. Corresponding to each distance to ill-posedness we have a condition number:

\[
C_E(d) = \begin{cases} \frac{\|d\|}{\rho_E(d)} & \text{if } \rho_E(d) > 0, \\ \infty & \text{if } \rho_E(d) = 0. \end{cases}
\]

\[
C_I(d) = \begin{cases} \frac{\|d\|}{\rho_I(d)} & \text{if } \rho_I(d) > 0, \\ \infty & \text{if } \rho_I(d) = 0. \end{cases}
\]

The value of \( C_E(d) \) is a measure of the relative conditioning of the data instance \( d \). A similar argument applies to \( C_I(d) \).

We will now present some particular upper bounds on the norms of feasible solutions of the analytic center problems \( AE(d) \) and \( AI(d) \).

**Lemma 7.1** Let \( d = (A, b, c) \in \mathcal{F}_E \) and \( \rho_E(d) > 0 \). Then

\[
\|x\| \leq C_E(d)
\]

for any feasible \( x \) of \( AE(d) \).

**Proof:** Let \( x \) be a feasible solution of \( AE(d) \). By Proposition 7.3, there is a vector \( \bar{x} \in X^* \) such that \( \bar{x}^T x = \|x\| \) and \( \|\bar{x}\|_* = 1 \). Define \( \Delta A = -\frac{1}{\|x\|} b \bar{x}^T \), and \( \Delta d = (\Delta A, 0, 0) \). Then, \((A + \Delta A)x = 0 \) and \( x > 0 \). Now, consider the program \((AE(d + \Delta d))\) defined as \( \inf \{ \phi(x) : (A + \Delta A)x = b, x > 0 \} \). Because the system \((A + \Delta A)^Tu = 0 \) and \( u > 0 \) has a solution, there cannot exist \( u \in Y^* \) for which \((A + \Delta A)^Tu > 0 \), and so \( d + \Delta d \in \mathcal{F}_E^C \), whereby \( \rho_E(d) \leq \|\Delta d\| \). On the other hand, \( \|\Delta d\| = \|\Delta b\| = (\|b\|\|\bar{x}\|_*)/\|x\| \leq \|d\|/\|x\| \), so that \( \|x\| \leq \|d\|/\rho_E(d) = C_E(d) \).
Lemma 7.2 Let $d = (A, b, c) \in \mathcal{F}_I$ and $\rho_I(d) > 0$. Then

$$\|y\|_\bullet \leq C_I(d),$$

$$\|s\|_\bullet \leq 2\|d\|C_I(d),$$

for any feasible $(y, s)$ of $AI(d)$.

Proof: Let $(y, s)$ be a feasible solution of $AI(d)$. If $y = 0$, then $s = c$ and the bounds are trivially true, so that we assume $y \neq 0$. By Proposition 7.3, there is a vector $\tilde{y} \in \mathcal{Y}$ such that $\|y\|_\bullet = y^T \tilde{y}$ and $\|\tilde{y}\| = 1$. Let $\Delta A = -\frac{1}{\|w\|} \tilde{y} c^T$, and $\Delta d = (\Delta A, 0, 0)$. Hence, $-(A + \Delta A)^T y = c - A^T y > 0$. Because $(A + \Delta A)^T u > 0$ has a solution, there cannot exist $v$ for which $(A + \Delta A)v = 0$ and $v > 0$, and so $d + \Delta d \in \mathcal{F}_I^C$, whereby $\rho_I(d) \leq \|\Delta d\|$. On the other hand, $\|\Delta d\| = \|c\|_\bullet /\|y\|_\bullet \leq \|d\| /\|y\|_\bullet$, so that $\|y\|_\bullet \leq \|d\| /\rho_I(d) = C_I(d)$. The bound for $\|s\|_\bullet$ can be easily derived using the fact that $\|s\|_\bullet \leq \|c\|_\bullet + \|A^T\|_\bullet \|y\|_\bullet$ and $C_I(d) \geq 1$.

q.e.d.

We now introduce the dual programs $AED(\cdot)$ and $AID(\cdot)$ of the conic analytic center problems $AE(\cdot)$ and $AI(\cdot)$, respectively:

\[
AED(d) \quad : \quad \max \left\{ b^T u - \phi^*(t) : A^T u + t = 0, t > 0, u \in \mathcal{Y}^* \right\},
\]

\[
AID(d) \quad : \quad \max \left\{ -c^T v - \phi(v) : Av = 0, v > 0 \right\},
\]

for $d = (A, b, c) \in \mathcal{D}$. 
The following two propositions show upper bounds on the optimal solutions to the programs \( AED(\cdot) \) and \( AID(\cdot) \), respectively.

**Proposition 7.4** Let \( d = (A, b, c) \in \mathcal{F}_E \) and \( \rho_E(d) > 0 \). Then

\[
\|u\|_* \leq \frac{\vartheta}{\rho_E(d)},
\]

\[
\|t\|_* \leq \vartheta C_E(d),
\]

where \((u, t)\) is the optimal solution of the program \( AED(d) \).

**Proof:** From (7.29), we have \( b^T u + \vartheta = 0 \), where \( t = -A^T u \) and \( x \) is the optimal solution to \( AE(d) \). If \( u = 0 \), then \( t = 0 \), which cannot be true since \( \phi^*(t) = \infty \) at \( t = 0 \), thus we assume that \( u \neq 0 \). By Proposition 7.3, there exists \( \tilde{u} \) such that \( \|u\|_* = \tilde{u}^T u \) and \( \|\tilde{u}\| = 1 \). Let \( \Delta b = \vartheta \tilde{u}/\|u\|_* \), and \( \Delta d = (0, \Delta b, 0) \). Hence, \( (b + \Delta b)^T u = b^T u + \Delta b^T u = 0 \). Therefore, the instance \( d + \Delta d \) corresponds to an unbounded program \( AED(d + \Delta d) \), and so it is also a primal infeasible instance. Thus, \( \|\Delta d\| \geq \rho_E(d) \). On the other hand, \( \|\Delta d\| = \|\Delta b\| = \vartheta/\|u\|_* \), and the result for \( u \) follows. Finally, \( \|t\|_* \leq \|A^T\|_* \|u\|_* \leq \|d\| \|u\|_* \), and the result for \( t \) follows.

q.e.d.

**Proposition 7.5** Let \( d = (A, b, c) \in \mathcal{F}_I \) and \( \rho_I(d) > 0 \). Then

\[
\|v\| \leq \frac{\vartheta}{\rho_I(d)},
\]

where \( v \) is the optimal solution of the program \( AID(d) \).

**Proof:** From (7.34), we have \( -c^T v + \vartheta = 0 \), where \((y, s)\) is the optimal solution to \( AI(d) \). By Proposition 7.3, there exists \( \tilde{v} \) such that \( \|v\| = \tilde{v}^T v \) and \( \|\tilde{v}\|_* = 1 \). Let \( \Delta c = -\vartheta \tilde{v}/\|v\| \) and \( \Delta d = (0, 0, \Delta c) \). Hence, \( -(c + \Delta c)^T v = -c^T v - \Delta c^T v = 0 \).
Therefore, the instance $d + \Delta d$ corresponds to an unbounded program $AID(d + \Delta d)$, and so it is also a primal infeasible instance. Thus, $||\Delta d|| \geq \rho_I(d)$. On the other hand, $||\Delta d|| = ||\Delta c||_\star = \vartheta/\|v\|$, and the result follows.

q.e.d.

7.4 Bounds on Solutions Along the Central Trajectory

This section presents results on lower and upper bounds on sizes of optimal solutions along the central trajectory, for the pair of dual barrier problems $P_\mu(d)$ and $D_\mu(d)$, as well as upper bounds on the sizes of changes in optimal solutions as the data is changed. As in the previous section, we assume that $d = (A, b, c) \in \tilde{\mathcal{D}}$ represents a data instance. Before presenting the first bound, we define the following constant, denoted $\mathcal{K}(d, \mu)$, which arises in many of the results to come:

$$
\mathcal{K}(d, \mu) = \mathcal{C}(d)^2 + \frac{\mu \vartheta}{\rho(d)}.
$$

Throughout this section we assume that $\phi$ is a $\vartheta$-logarithmically homogeneous barrier for $\mathcal{C}_\mathcal{X}$ for some $\vartheta \geq 0$.

**Theorem 7.2** If $d = (A, b, c) \in \mathcal{F}$ and $\rho(d) > 0$, then

$$
\|x(\mu)\| \leq \mathcal{K}(d, \mu),
$$
$$
\|y(\mu)\|_\star \leq \mathcal{K}(d, \mu),
$$
$$
\|s(\mu)\|_\star \leq 2\|d\|\mathcal{K}(d, \mu),
$$
7.4 Bounds on Solutions Along the Central Trajectory

for the optimal solution $x(\mu)$ to $P_\mu(d)$ and the optimal solution $(y(\mu), s(\mu))$ to the dual problem $D_\mu(d)$.

The proof of this theorem follows exactly the same logic as the proof of Theorem 3.1 in [NF97]. The only difference between these two theorems is the use of the notation for the more general context defined in this section.

**Proof of Theorem 7.2:** Let $\hat{x} = x(\mu)$ be the optimal solution to $P_\mu(d)$ and $(y(\mu), s(\mu)) = (\hat{y}, \hat{s})$ the optimal solution to the corresponding dual problem $D_\mu(d)$. Note that the optimality conditions (7.19)-(7.21) of $P_\mu(d)$ and $D_\mu(d)$, together with (7.23), imply that $c^T\hat{x} = b^T\hat{y} + \mu\vartheta$. Note also that by Proposition 7.3, there exists a vector $\bar{x} \in \mathcal{X}^*$ such that $\bar{x}^T\hat{x} = \|\hat{x}\|$ and $\|\bar{x}\|_* = 1$. Similarly, by Proposition 7.3, there exists a vector $\bar{y} \in \mathcal{Y}$ such that $\bar{y}^T\hat{y} = \|\hat{y}\|_*$ and $\|\bar{y}\| = 1$.

Observe that since $\hat{s} = c - A^T\hat{y}$, then $\|\hat{s}\|_* \leq \|c\|_* + \|A^T\|_*\|\hat{y}\|_*$. Moreover, $\|A^T\|_* = \|A\|$. Thus, $\|\hat{s}\|_* \leq \|d\|(1 + \|\hat{y}\|_*)$, and using the fact that $C(d) \geq 1$ the bound on $\|\hat{s}\|_*$ is a consequence of the bound on $\|\hat{y}\|_*$. Therefore, it is sufficient to prove the bounds on $\|\bar{x}\|$ and on $\|\hat{y}\|_*$.

The rest of the proof proceeds by examining three cases:

1. $c^T\hat{x} \leq 0$,

2. $0 < c^T\hat{x} \leq \mu\vartheta$, and

3. $\mu\vartheta < c^T\hat{x}$.

In case (1), let $\Delta A = -\frac{1}{\|\hat{x}\|}b\bar{x}^T$. Then $(A + \Delta A)\hat{x} = 0$, $\hat{x} > 0$, and $c^T\hat{x} \leq 0$. Let $\Delta d = (\Delta A, 0, 0)$. If $d + \Delta d$ is primal infeasible we have $\|\Delta d\| \geq \rho(d) > 0$. If $d + \Delta d$ is primal feasible, then $P_\mu(d + \Delta d)$ is unbounded ($\hat{x}$ is a ray of $P_\mu(d + \Delta d)$) and so its dual $D_\mu(d + \Delta d)$ is infeasible, so that again $\|\Delta d\| \geq \rho(d) > 0$. In either instance, $\rho(d) \leq \|\Delta d\| = \|\Delta A\| = (\|\bar{x}\|_*\|b\|)/\|\hat{x}\| = \|b\|/\|\bar{x}\| \leq \|d\|/\|\hat{x}\|$. Therefore,
\[ \| \dot{x} \| \leq C(d) \leq C(d)^2 + \mu \vartheta / \rho(d), \text{ since } C(d) \geq 1 \text{ for any } d. \] This proves the bound on \( \| \dot{x} \| \) for this case.

The bound on \( \| \dot{y} \|_\ast \) is trivial if \( \dot{y} = 0 \), so we assume that \( \dot{y} \neq 0 \). Let \( \theta = b^T \dot{y} \), \( \Delta b = -\theta \frac{\dot{y}}{\| \dot{y} \|} x \), \( \Delta A = -\frac{1}{\| \dot{y} \|} yc^T \), and \( \Delta d = (\Delta A, \Delta b, 0) \). Observe that \( (b + \Delta b)^T \dot{y} = 0 \) and \( -(A + \Delta A)^T \dot{y} > 0 \), so that \( \rho(d) \leq \| \Delta d \| = \max \{ \| c \|_\ast, \| \theta \| / \| \dot{y} \| \} \). Hence, \( \| \dot{y} \|_\ast \leq \max \{ C(d), |\theta| / \rho(d) \} \). Furthermore, \( |\theta| = |c^T \dot{x} - \mu \vartheta| \leq \| \dot{x} \| \| c \|_\ast + \mu \vartheta \leq C(d) \| d \| + \mu \vartheta \). Therefore, again using the fact that \( C(d) \geq 1 \) for any \( d \), we have \( \| \dot{y} \|_\ast \leq C(d)^2 + \mu \vartheta / \rho(d) \).

In case (2), let \( \Delta d = (\Delta A, 0, \Delta c) \), where \( \Delta A = -\frac{1}{\| \dot{x} \|} b \dot{x}^T \), and \( \Delta c = -\mu \vartheta \frac{\dot{y}}{\| \dot{y} \|} \). Observe that \( (A + \Delta A) \dot{x} = 0 \) and \( (c + \Delta c)^T \dot{x} \leq 0 \). Using a similar logic to that in the first part of case (1), we conclude that \( \rho(d) \leq \| \Delta d \| = \max \{ \| \Delta A \|, \| \Delta c \|_\ast \} = \max \{ \| b \|, \mu \vartheta \} / \| \dot{x} \| \leq (\| d \| + \mu \vartheta) / \| \dot{x} \| \). Therefore, \( \| \dot{x} \| \leq C(d) + \mu \vartheta / \rho(d) \leq C(d)^2 + \mu \vartheta / \rho(d) \), since \( C(d) \geq 1 \) for any \( d \). This proves the bound on \( \| \dot{x} \| \) for this case.

As before, the bound on \( \| \dot{y} \|_\ast \) is trivial if \( \dot{y} = 0 \), so we assume that \( \dot{y} \neq 0 \). Let \( \Delta d = (\Delta A, \Delta b, 0) \), where \( \Delta A = -\frac{1}{\| \dot{y} \|} yc^T \), and \( \Delta b = \mu \vartheta \frac{\dot{y}}{\| \dot{y} \|} \). Observe that \( (b + \Delta b)^T \dot{y} = b^T \dot{y} + \mu \vartheta = c^T \dot{x} > 0 \) and \( -(A + \Delta A)^T \dot{y} > 0 \). Using a similar logic to that in the first part of case (1), we conclude that \( \rho(d) \leq \| \Delta d \| = \max \{ \| \Delta A \|, \| \Delta b \| \} = \max \{ \| c \|_\ast, \mu \vartheta \} / \| \dot{y} \|_\ast \leq (\| d \| + \mu \vartheta) / \| \dot{y} \|_\ast \). Therefore, \( \| \dot{y} \|_\ast \leq C(d) + \mu \vartheta / \rho(d) \leq C(d)^2 + \mu \vartheta / \rho(d) \).

In case (3), we first consider the bound on \( \| \dot{y} \|_\ast \). Again noting that this bound on \( \| \dot{y} \|_\ast \) is trivial if \( \dot{y} = 0 \), we assume that \( \dot{y} \neq 0 \). Then let \( \Delta d = (\Delta A, 0, 0) \), where \( \Delta A = -\frac{1}{\| \dot{y} \|} yc^T \). Since \( -(A + \Delta A)^T \dot{y} > 0 \) and \( b^T \dot{y} = c^T \dot{x} - \mu \vartheta > 0 \), it follows from the same logic as in the first part of case (1) that \( \rho(d) \leq \| \Delta d \| = \| c \|_\ast / \| \dot{y} \|_\ast \). Therefore, \( \| \dot{y} \|_\ast \leq \| c \|_\ast / \rho(d) \leq C(d) \leq C(d)^2 + \mu \vartheta / \rho(d) \).

Finally, let \( \Delta A = -\frac{1}{\| \dot{x} \|} b \dot{x}^T \), \( \Delta c = -\theta \frac{\dot{x}}{\| \dot{x} \|} \), and \( \Delta d = (\Delta A, 0, \Delta c) \), where \( \theta = c^T \dot{x} \). Observe that \( (A + \Delta A) \dot{x} = 0 \), and \( (c + \Delta c)^T \dot{x} = 0 \). Using the same argument as in the previous cases, we conclude that \( \rho(d) \leq \| \Delta d \| = \max \{ \| \Delta A \|, \| \Delta c \|_\ast \} = \max \{ \| c \|_\ast, \mu \vartheta \} / \| \dot{x} \| \leq (\| d \| + \mu \vartheta) / \| \dot{x} \| \). Therefore, \( \| \dot{x} \| \leq C(d) + \mu \vartheta / \rho(d) \leq C(d)^2 + \mu \vartheta / \rho(d) \).
m.x{||b||, \theta}/||\hat{x}||$, so that $||\hat{x}|| \leq \max\{C(d), \theta/\rho(d)\}$. Furthermore, $\theta = b^T \hat{y} + \mu \theta \leq ||b|| \|\hat{y}\^*\| + \mu \theta \leq ||d||C(d) + \mu \theta$. Therefore, $||\hat{x}|| \leq C(d)^2 + \mu \theta/\rho(d)$, because $C(d) \geq 1$.

q.e.d.

We next consider upper bounds on solutions of $P_\mu(d + \Delta d)$ and $D_\mu(d + \Delta d)$, where $d + \Delta d$ is a data instance representing a small perturbation of the data instance $d$.

**Corollary 7.1** Let $d \in \mathcal{F}$ be such that $\rho(d) > 0$. If $||\Delta d|| \leq \rho(d)/3$, then

$$
\begin{align*}
||x|| & \leq 4K(d, \mu), \\
||y||^* & \leq 4K(d, \mu), \\
||s||^* & \leq 16||d||K(d, \mu),
\end{align*}
$$

where $x$ is the optimal solution to $P_\mu(d + \Delta d)$ and $(y, s)$ is the optimal solution to $D_\mu(d + \Delta d)$.

**Proof:** The proof follows from Theorem 7.2 by observing that if $||\Delta d|| \leq \rho(d)/3$, then we have $||d + \Delta d|| \leq ||d|| + \rho(d)/3$, and $\rho(d + \Delta d) \geq \frac{2}{3} \rho(d)$, so that

$$
C(d + \Delta d) \leq \frac{||d|| + \frac{1}{3} \rho(d)}{\frac{2}{3} \rho(d)} = \frac{3}{2} \left( C(d) + \frac{1}{3} \right) \leq 2C(d),
$$

since $C(d) \geq 1$.

q.e.d.

The next result presents lower bounds on the norm of any primal and dual optimal solution to the central trajectory problems $P_\mu(d)$ and $D_\mu(d)$, respectively.

**Theorem 7.3** If $d \in \mathcal{F}$ and $\rho(d) > 0$, then

$$
||x(\mu)|| \geq \frac{\mu \theta}{2||d||K(d, \mu)},
$$


\[ \|s(\mu)\| \geq \frac{\mu \vartheta}{\mathcal{K}(d, \mu)}, \]

where \( x(\mu) \) is the optimal solution to \( P_\mu(d) \) and \((y(\mu), s(\mu)) \) is the optimal solution to \( D_\mu(d) \).

This theorem shows that \( \|x(\mu)\| \) is bounded from below by a function only involving the quantities \( \|d\|, \mathcal{C}(d), \rho(d), \vartheta, \) and \( \mu \). Furthermore, the theorem shows that for \( \mu \) close to zero, that \( \|x(\mu)\| \) grows at least linearly in \( \mu \), and at a rate that is at least \( \vartheta/(2\|d\|\mathcal{C}(d)^2) \). The theorem offers less insight when \( \mu \to \infty \), since the lower bound on \( \|x(\mu)\| \) presented in the theorem converges to \( (2\mathcal{C}(d))^{-1} \) as \( \mu \to \infty \). When the feasible region is unbounded, it is well known that \( \|x(\mu)\| \to \infty \) as \( \mu \to \infty \), so that as \( \mu \to \infty \) the lower bound of Theorem 7.3 does not adequately capture the behavior of the sizes of optimal solutions to \( P_\mu(d) \) when the feasible region is unbounded.

**Proof of Theorem 7.3:** By the Karush-Kuhn-Tucker optimality conditions (7.19)-(7.21) of the dual pair of problems \( P_\mu(d) \) and \( D_\mu(d) \), together with (7.22), we have

\[ s(\mu)^T x(\mu) = \mu \vartheta. \]

Since \( s(\mu)^T x(\mu) \leq \|s(\mu)\| \|x(\mu)\| \), it follows that \( \|x(\mu)\| \geq \mu \vartheta / \|s(\mu)\| \), and the first inequality follows from Theorem 7.2. The second inequality can be proved analogously.

q.e.d.

The following corollary uses Theorem 7.3 to provide lower bounds for solutions to perturbed problems.

**Corollary 7.2** Let \( d \in \mathcal{F} \) be such that \( \rho(d) > 0 \). If \( \|\Delta d\| \leq \rho(d)/3 \), then

\[ \|x\| \geq \frac{\mu \vartheta}{8(\|d\| + \rho(d))\mathcal{K}(d, \mu)}, \]

\[ \|s\| \geq \frac{\mu \vartheta}{4\mathcal{K}(d, \mu)}, \]
where $x$ is the optimal solution to $P_\mu(d + \Delta d)$ and $(y, s)$ is the optimal solution to $D_\mu(d + \Delta d)$.

**Proof:** The proof follows the same logic as that of Corollary 7.1.

**q.e.d.**

### 7.5 Supplementary Results

**Proof of Proposition 7.1:** Because $\phi$ is a $\vartheta$-logarithmically homogeneous barrier for $C_X$ and $\vartheta \geq 0$, then for each $x \in \text{int}(C_X)$ and $t > 0$, by definition, we have above:

$$
\phi(tx) = \phi(x) - \vartheta \ln t.
$$

By taking derivatives with respect to $t$ on both sides, we obtain

$$
\nabla \phi(tx)^T x = -\frac{\vartheta}{t},
$$

and so, by choosing $t = 1$, we get (7.7). By taking derivatives again with respect to $t$, we obtain

$$
x^T \nabla^2 \phi(tx)x = \frac{\vartheta}{t^2},
$$

from which the first identity of (7.8) follows.

Going back to identity (7.7), if we take derivatives with respect to $x$ on both sides of the identity, we obtain

$$
\nabla^2 \phi(x)x + \nabla \phi(x) = 0,
$$

and (7.6) follows. Using (7.6) and the first identity of (7.8), we obtain the second identity of (7.8).

Finally, by using again the identity (7.4) and by taking the first and second deriva-
tives with respect to \( x \), we obtain (7.5).

\textbf{q.e.d.}

\textbf{Proof of Proposition 7.2:} Given any \( s \in \text{int}(C^*_X) \) and because of the Karush-Kuhn-Tucker optimality conditions, it is necessary and sufficient for \( \hat{x} \in \text{int}(C_X) \) to satisfy the following equation

\begin{equation}
    s = -\nabla \phi(\hat{x}),
\end{equation}

(7.35)

in order to solve the optimization problem \( \max \{-s^T x - \phi(x) : x \in \text{int}(C_X)\} \). From (7.35) it follows that \( -\nabla \phi(\hat{x}) \in \text{int}(C^*_X) \) and hence, we have the first statements (7.10) and (7.11). Moreover, observe that \( \phi(\hat{x}) \geq -s^T \hat{x} - \phi^*(s) \) for all \( s \in \text{int}(C^*_X) \), and for \( \hat{s} = -\nabla \phi(\hat{x}) \), \( \phi(\hat{x}) = -\hat{s}^T \hat{x} - \phi^*(\hat{s}) \). Therefore, \( \phi(x) = \phi^{**}(x) \) for all \( x \in \text{int}(C_X) \).

Since \( \phi^* \) is also a \( \vartheta \)-normal barrier for \( C^*_X \), then again from the condition (7.35) applied to \( \phi^* \) we have \( -\nabla \phi^*(s) \in \text{int}(C_X) \) for all \( s \in \text{int}(C^*_X) \).

Again using (7.35) and from (7.7) in Proposition 7.1, we have (7.12). Similarly, using that \( \phi = \phi^{**} \), (7.35), and (7.7) applied to \( \phi^* \), we obtain (7.13). From identity (7.12), if we take derivative with respect to \( x \) on both sides, then we obtain that

\begin{equation}
    -\nabla^2 \phi(x) \nabla \phi^*(\nabla \phi(x)) = -\nabla \phi(x),
\end{equation}

and so,

\begin{equation}
    \nabla \phi^*(\nabla \phi(x)) = -(\nabla^2 \phi(x))^{-1}(\nabla \phi(x)) = -x,
\end{equation}

where the last equality follows from (7.6) in Proposition 7.1. A similar argument applied to \( \phi^* \) proves that \( \nabla \phi(-\nabla \phi^*(s)) = -s \), and hence, we have (7.14) and (7.15). Identities (7.16) and (7.17) immediately follow from taking derivatives on both sides of equalities (7.14) and (7.15).

Finally, by minimizing the convex functional \( \phi(x) + \phi^*(s) - \vartheta \ln s^T x \) for all \( x \in \text{int}(C_X) \) and \( s \in \text{int}(C^*_X) \), we obtain (7.18).
q.e.d.
Chapter 8

Semi-Definite Programming

8.1 Overview

The main objective of this chapter is to particularize the results from Chapter 7 to semi-definite programming. The most relevant results presented herein are easy consequences of the results concerning conic linear systems. There are, however, a number of results that follow from specific properties of semi-definite programs. These results generalize some of the results for linear programs in standard form; however they are not direct consequences of the theory developed in Chapter 7.

In Section 8.2 we discuss semi-definite analytic center problems. Lemma 8.1 and Lemma 8.2 state upper bounds on solutions to two kinds of analytic center problems. The bounds depend on the corresponding conditioning of the data as well as on the size of the data. On the other hand, Lemma 8.3 and Lemma 8.4 present lower bounds on optimal solutions to semi-definite analytic center problems. These lemmas are also important because they establish lower bounds on the eigenvalues of the optimal solutions in terms of the conditioning of the data.
In Section 8.3 we discuss the semi-definite central trajectory problem. As in the analytic center case, in Theorem 8.1 and Theorem 8.2 we derive upper and lower bounds on optimal solutions along the central trajectory in terms of the conditioning of the data and other relevant parameters. Furthermore, in Theorem 8.2, we also derive similar lower bounds on the eigenvalues of optimal solutions along the central trajectory.

Finally, in Section 8.4 we characterize the interior of the set of feasible data instances of a semi-definite program (see Proposition 8.3). This result generalizes the classical characterization proven by Robinson [Rob77] and Ashmanov [Ash81] for linear programs in standard form. We also prove that the set of ill-posed data instances with respect to semi-definite programming is the same as the set of ill-posed data instances with respect to semi-definite central trajectory problems. At the end of this section, we conjecture about the behavior of the central trajectory associated with problems having an unbounded feasible region and present some partial results concerning this subject.

We now formally start our discussion by introducing the notation used throughout this chapter. This notation concerns semi-definite programs, and the related barrier programs such as the analytic center problem and the central trajectory problem in the semi-definite context.

Let $S_n$ denote the subspace of $\mathbb{R}^{n \times n}$ consisting of symmetric matrices. Given a matrix $U \in S_n$, let $U \succeq 0$ denote that $U$ is a positive semi-definite matrix, and let $U \succ 0$ denote that $U$ is a positive definite matrix. We denote by $S_n^+$ the set of positive semi-definite matrices in $S_n$, that is, $S_n^+ = \{ U \in S_n : U \succeq 0 \}$. Observe that $S_n^+$ is a closed convex pointed cone in $S_n$ with non-empty interior given by $\{ U \in S_n : U \succ 0 \}$. 
8.1 Overview

Given matrices $U$ and $V$ in $\mathcal{S}_n$, we define the following inner-product in the space $\mathcal{S}_n$:

$$U \cdot V := \text{trace}(UV),$$

where $\text{trace}(U) = \sum_{j=1}^{n} U_{jj}$ for all $U \in \mathcal{S}_n$. This inner-product induces a norm on the space $\mathcal{S}_n$, that is,

$$\|U\|_2 := \sqrt{\text{trace}(UU)} = \left(\sum_{j=1}^{n} \lambda_j(U)^2\right)^{1/2},$$

(8.1)

for all $U$ in $\mathcal{S}_n$, where $\lambda_j(U)$ denotes the $j$-th eigenvalue of $U$ chosen in increasing order (for more details see Chapter 2). In addition, we define the following norms on the space $\mathcal{S}_n$:

$$\|U\|_1 := \sum_{j=1}^{n} |\lambda_j(U)|,$$

(8.2)

$$\|U\|_{\infty} := \max_{1 \leq j \leq n} |\lambda_j(U)|,$$

(8.3)

for all $U$ in $\mathcal{S}_n$. Notice that from the inequality

$$|U \cdot V| \leq \|U\|_{\infty}\|V\|_1,$$

it follows that the dual norm $\|\cdot\|_*$ of the norm $\|\cdot\|_1$ defined on the space $\mathcal{S}_n$ is the norm $\|\cdot\|_{\infty}$ defined on the dual space $\mathcal{S}_n^* = \mathcal{S}_n$. Furthermore, observe that the polar $(\mathcal{S}_n^+)^*$ of the cone $\mathcal{S}_n^+$ is the cone $\mathcal{S}_n^+$ itself.

In this chapter we consider data instances of the form $d = (A_1, \ldots, A_m, b, C)$, where $A_1, \ldots, A_m$ and $C$ are matrices in $\mathcal{S}_n$, and $b$ is a vector in $\mathbb{R}^n$. We denote by $\mathcal{D}$ the space of data instances, that is, $\mathcal{D} = \{d = (A_1, \ldots, A_m, b, C) : A_i \in \mathcal{S}_n, i = 1, \ldots, m, b \in \mathbb{R}^m, C \in \mathcal{S}_n\}$. Using the array $(A_1, \ldots, A_m)$, $A_i \in \mathcal{S}_n$ for $i = 1, \ldots, m,$
we can define a linear operator $[A_1, \ldots, A_m]$ from $\mathcal{S}_n$ to $\mathbb{R}^m$ given by:

$$[A_1, \ldots, A_m]X := (A_1 \bullet X, \ldots, A_m \bullet X)^T.$$ 

We define the rank of this operator as the dimension of the subspace generated by the matrices $A_1, \ldots, A_m$, that is,

$$\text{rank}([A_1, \ldots, A_m]) := \dim(\langle A_1, \ldots, A_m \rangle).$$

Given an array $(A_1, \ldots, A_m)$, we define the norm of the corresponding linear operator $[A_1, \ldots, A_m]$ to be the operator norm:

$$\|\|\|A_1, \ldots, A_m\|\| = \max \{\|\|A_1, \ldots, A_m\|X\|_1 : X \in \mathcal{S}_n, \|X\|_1 \leq 1\}.$$  

Using the norms (8.2), (8.3), and (8.4), we provide the space of data instances $\mathcal{D}$ with the following norm:

$$\|d\| = \|(A_1, \ldots, A_m, b, C)\| := \max \{\|\|A_1, \ldots, A_m\|, \|b\|_1, \|C\|_\infty\},$$

for all $d \in \mathcal{D}$. For $d \in \mathcal{D}$, we define the ball centered at $d$ with radius $\delta$ as:

$$B(d, \delta) := \{d + \Delta d \in \mathcal{D} : \|\Delta d\| \leq \delta\},$$

for all scalars $\delta \geq 0$.

For a given data instance $d = (A_1, \ldots, A_m, b, C) \in \mathcal{D}$, we study the following semi-definite optimization problems:

$$PSD(d) : \min \{C \bullet X : A_i \bullet X = b_i, i = 1, \ldots, m, X \succeq 0\},$$
8.1 Overview

\[
DSD(d) : \max \left\{ b^T y : \sum_{i=1}^{m} y_i A_i + S = C, S \succeq 0 \right\}, \\
PSD_\mu(d) : \min \left\{ C \cdot X + \mu p(X) : A_i \cdot X = b_i, i = 1, \ldots, m, X \succcurlyeq 0 \right\}, \\
DSD_\mu(d) : \max \left\{ b^T y - \mu p(S) : \sum_{i=1}^{m} y_i A_i + S = C, S \succcurlyeq 0 \right\}, \\
ASDE(d) : \min \left\{ p(X) : A_i \cdot X = b_i, i = 1, \ldots, m, X \succcurlyeq 0 \right\}, \\
ASDI(d) : \min \left\{ p(S) : \sum_{i=1}^{m} y_i A_i + S = C, S \succcurlyeq 0 \right\},
\]

where, for \( U \succcurlyeq 0 \), \( p(U) = -\ln \det U \). As proven in [NN94], \( p(\cdot) \) is a strictly convex \( n \)-normal barrier for the cone \( S_n^+ \) (see Definition 7.6). Furthermore, the Legendre transformation \( p^* \) of the barrier \( p \) is the functional \( p \) itself. Finally, the variable \( \mu \) is a positive parameter that will remain independent with respect to the data instance \( d \in \mathcal{D} \).

It is not difficult to show that \( PSD(d) \) and \( DSD(d) \) are a pair of dual (Lagrangian) problems. Similarly, \( PSD_\mu(d) \) and \( DSD_\mu(d) \) are a pair of dual (Lagrangian) problems. Moreover, the sequence \( \{ X(\mu) : \mu > 0 \} \) of optimal solutions to \( PSD_\mu(d) \) (when they exist) defines a smooth mapping from \( (0, \infty) \) to \( S_n^+ \) [VB96] and it is called the primal central trajectory. Analogously, for \( DSD_\mu(d) \) the sequence \( \{ (y(\mu), S(\mu)) : \mu > 0 \} \) of optimal solutions to \( DSD_\mu(d) \) (when they exist) also defines a smooth mapping from \( (0, \infty) \) to \( \mathbb{R}^m \times S_n^+ \) and it is called the dual central trajectory. Finally, the problems \( ASDE(d) \) and \( ASDI(d) \) correspond to analytic center problems associated with the feasible regions of the \( PSD(d) \) and \( DSD(d) \), respectively. Observe that \( ASDE(d) \) and \( ASDI(d) \) are not dual programs of each other.

Consider the following subset of the data set \( \mathcal{D} \):

\[
\mathcal{F} = \{(A_1, \ldots, A_m, b, C) \in \mathcal{D} : \text{there exists } (X, y) \text{ such that } \}
\]
A_i \cdot X = b_i, i = 1, \ldots, m, X \succeq 0, \sum_{i=1}^{m} y_i A_i \preceq C \},

that is, the elements in $\mathcal{F}$ correspond to those data instances in $\bar{\mathcal{D}}$ for which $PSD(d)$ and $DSD(d)$ are feasible. The complement of $\mathcal{F}$, denoted by $\mathcal{F}^C$, is the set of data instances $d = (A_1, \ldots, A_m, b, C)$ for which $PSD(d)$ or $DSD(d)$ is infeasible. The boundary of $\mathcal{F}$, denoted by $\mathcal{B}$, is called the set of ill-posed data instances. This is because arbitrarily small changes in a data instance $d = (A_1, \ldots, A_m, b, C) \in \mathcal{B}$ can yield data instances in $\mathcal{F}$ as well as data instances in $\mathcal{F}^C$.

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

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

see [Ren94, Ren95a, Ren95b], and so $\rho(d)$ is the distance of the data instance $d = (A_1, \ldots, A_m, b, C)$ to the set of ill-posed instances $\mathcal{B}$. It is straightforward to show that

$$\rho(d) = \begin{cases} 
\sup\{\delta : B(d, \delta) \subset \mathcal{F}\} & \text{if } d \in \mathcal{F}, \\
\sup\{\delta : B(d, \delta) \subset \mathcal{F}^C\} & \text{if } d \in \mathcal{F}^C,
\end{cases} \tag{8.5}$$

so that we could also define $\rho(d)$ by employing (8.5). The “condition number” $C(d)$ of the data instance $d$ is defined as

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

when $\rho(d) > 0$, and $C(d) = \infty$ when $\rho(d) = 0$. The condition number $C(d)$ can be viewed as a scale-invariant reciprocal of $\rho(d)$, as it is elementary to demonstrate that $C(d) = C(\alpha d)$ for any positive scalar $\alpha$. Moreover, for $d = (A_1, \ldots, A_m, b, C) \notin \mathcal{B}$ let $\Delta d = (-A_1, \ldots, -A_m, -b, -C)$. Observe that $d + \Delta d = (0, \ldots, 0, 0, 0) \in \mathcal{B}$ and since $\mathcal{B}$ is a closed set, we have $\|d\| = \|\Delta d\| \geq \rho(d) > 0$, so that $C(d) \geq 1$. The value of
8.2 Semi-Definite Analytic Center Problems

\( C(d) \) is a measure of the relative conditioning of the data instance \( d \).

We also consider the following subset of data instances in \( \bar{D} \):

\[
\mathcal{F}_S = \{ (A_1, \ldots, A_m, b, C) \in \bar{D} : \text{there exists } (X, y) \text{ such that} \\
A_i \cdot X = b_i, i = 1, \ldots, m, X \succ 0, \sum_{i=1}^{m} y_i A_i \prec C \},
\]

corresponding to data instances that have strictly primal and dual feasible solutions. As will be shown in Lemma 8.6, we have \( \partial \mathcal{F}_S = \mathcal{B} \), and so the distance to ill-posedness associated with the set \( \mathcal{F}_S \) is the same as the distance to ill-posedness associated with the set \( \mathcal{F} \), that is,

\[
\rho(d) = \inf \{ \| \Delta d \| : d + \Delta d \in \partial \mathcal{F}_S \},
\]

for all \( d \in \bar{D} \). We will use this fact throughout the remainder of this chapter.

8.2 Semi-Definite Analytic Center Problems

In this section we study the semi-definite analytic center problems \( ASDE(d) \) and \( ASDI(d) \), for a given data instance \( d \in \bar{D} \). Observe that \( ASDE(d) \) has a unique solution when its feasible region is bounded and non-empty. We call this unique solution the analytic center of the semi-definite program in equality form. Similarly, the program \( ASDI(d) \) has a unique solution when its feasible region is bounded and non-empty, and we call this unique solution the analytic center of the semi-definite program in inequality form. As mentioned before, note in particular that the two programs \( ASDE(d) \) and \( ASDI(d) \) are not Lagrangian duals of each other. In particular, both problems cannot be feasible at the same time.
Since we are interested in studying these analytic center problems when they have solutions, we introduce the following sets of feasible data instances to both $ASDE(d)$ and $ASDI(d)$ respectively:

$$ F_E = \{(A_1, \ldots, A_m, b, C) \in \bar{D} : \text{there exists } (X, u) \text{ such that } A_i \cdot X = b_i, i = 1, \ldots, m, X \succ 0, \sum_{i=1}^{m} u_i A_i \prec 0\}, $$

$$ F_I = \{(A_1, \ldots, A_m, b, C) \in \bar{D} : \text{there exists } (V, y) \text{ such that } A_i \cdot V = 0, i = 1, \ldots, m, V \succ 0, \sum_{i=1}^{m} y_i A_i \prec C\}. $$

Observe that $F_E$ consists of data instances $d$ for which $ASDE(d)$ is feasible and attains its optimal solution. Similarly, $F_I$ consists of data instances $d$ for which $ASDI(d)$ is feasible and attains its optimal solution. As usual, we introduce sets of "ill-posed" data instances corresponding to these analytic center problems: $B_E = \partial F_E$ and $B_I = \partial F_I$. For a data instance $d \in \bar{D}$, we define the following distances to ill-posedness as:

$$ \rho_E(d) = \inf\{\|\Delta d\| : d + \Delta d \in B_E\}, $$

$$ \rho_I(d) = \inf\{\|\Delta d\| : d + \Delta d \in B_I\}. $$

For instance, if $d \in F_E$, then $\rho_E(d)$ can be interpreted as a measure of how much the data instance $d$ can be perturbed before the feasible region of the problem $ASDE(d)$ becomes either empty or unbounded. Corresponding to each distance to ill-posedness we have a condition number:

$$ C_E(d) = \begin{cases} \frac{\|d\|}{\rho_E(d)} & \text{if } \rho_E(d) > 0, \\
\infty & \text{if } \rho_E(d) = 0. \end{cases} $$
8.2 Semi-Definite Analytic Center Problems

\[ C_I(d) = \begin{cases} \frac{\|d\|}{\rho_I(d)} & \text{if } \rho_I(d) > 0, \\ \infty & \text{if } \rho_I(d) = 0. \end{cases} \]

The value of \( C_E(d) \) is a measure of the relative conditioning of the data instance \( d \). A similar argument applies to \( C_I(d) \).

The following lemmas present some particular upper bounds on the norms of feasible solutions of the semi-definite analytic center problems \( ASDE(d) \) and \( ASDI(d) \). These lemmas are immediate consequences of Lemma 7.1 and Lemma 7.2, respectively.

**Lemma 8.1** Let \( d = (A_1, \ldots, A_m, b, C) \in F_E \) and \( \rho_E(d) > 0 \). Then

\[ \|X\|_1 \leq C_E(d) \]

for any feasible solution \( X \) of \( ASDE(d) \).

**Lemma 8.2** Let \( d = (A_1, \ldots, A_m, b, C) \in F_I \) and \( \rho_I(d) > 0 \). Then

\[ \|y\|_\infty \leq C_I(d), \]
\[ \|S\|_\infty \leq 2\|d\|C_I(d), \]

for any feasible solution \((y, S)\) of \( ASDI(d) \).

From the arithmetic-geometric mean inequality we have for \( X > 0 \)

\[ \frac{\det(X)}{(\text{trace}(X))^n} \leq \left(\frac{1}{n}\right)^n, \]
from which we obtain

\[ \ln(\det(X)) \leq n \ln(\text{trace}(X)) - n \ln(n). \]

Hence, for \( X \succ 0 \),

\[ -p(X) \leq n \ln(\|X\|_1) - n \ln(n). \]

Similarly, using that \( \|V\|_1 \leq n\|V\|_\infty \) for all \( V \in \mathcal{S}_n \), it follows that for any \( S \succ 0 \),

\[ -p(S) \leq n \ln(\|S\|_\infty). \]

Combining these results with the bounds from Lemmas 8.1 and 8.2, we obtain the following corollary:

**Corollary 8.1** Let \( d = (A_1, \ldots, A_m, b, C) \in \tilde{\mathcal{D}} \). If \( d \in \mathcal{F}_E \) and \( \rho_E(d) > 0 \), then

\[ p(X) \geq -n \ln(C_E(d)) + n \ln(n), \]

for all feasible solution \( X \) of \( \text{ASDE}(d) \). If \( d \in \mathcal{F}_I \) and \( \rho_I(d) > 0 \), then

\[ p(S) \geq -n \ln(2\|d\|C_I(d)), \]

for all feasible solution \((y, S)\) of \( \text{ASDI}(d) \).

As in previous chapters, we introduce the dual programs \( \text{ASDED}(\cdot) \) and \( \text{ASDID}(\cdot) \) of the semi-definite analytic center problems \( \text{ASDE}(\cdot) \) and \( \text{ASDI}(\cdot) \), respectively:

\[
\text{ASDED}(d) : \max \left\{ b^T u - p(T) : \sum_{i=1}^m u_i A_i + T = 0, T \succ 0 \right\},
\]

\[
\text{ASDID}(d) : \max \left\{ -C \cdot V - p(V) : A_i \cdot V = 0, i = 1, \ldots, m, V \succ 0 \right\},
\]

for \( d = (A_1, \ldots, A_m, b, C) \in \tilde{\mathcal{D}} \).
The following two propositions are immediate consequences of Proposition 7.4 and Proposition 7.5, respectively. The propositions show upper bounds on the optimal solutions to the programs $AED(\cdot)$ and $AID(\cdot)$, respectively.

**Proposition 8.1** Let $d = (A_1, \ldots, A_m, b, C) \in \mathcal{F}_E$ and $\rho_E(d) > 0$. Then

\[ \|u\|_\infty \leq \frac{n}{\rho_E(d)}, \]

\[ \|T\|_\infty \leq nC_E(d), \]

where $(u, T)$ is the optimal solution of the program $ASDED(d)$.

**Proposition 8.2** Let $d = (A_1, \ldots, A_m, b, C) \in \mathcal{F}_I$ and $\rho_I(d) > 0$. Then

\[ \|V\|_1 \leq \frac{n}{\rho_I(d)}, \]

where $V$ is the optimal solution of the program $ASDID(d)$.

Next, from Propositions 8.1 and 8.2, we prove several results concerning lower bounds on solutions to $ASDE(d)$ and $ASDI(d)$.

**Lemma 8.3** If $d = (A_1, \ldots, A_m, b, C) \in \mathcal{F}_E$ and $\rho_E(d) > 0$, then

\[ \|X\|_1 \geq \frac{1}{C_E(d)}, \]

for all feasible solution $X$ of $ASDE(d)$. Moreover, if $\hat{X}$ is the optimal solution to $ASDE(d)$ and $(\hat{u}, \hat{T})$ is the optimal solution to $ASDED(d)$ then

\[ \lambda_j(\hat{X}) \geq \frac{1}{nC_E(d)}, \]

\[ \lambda_j(\hat{T}) \geq \frac{1}{C_E(d)}, \]
for all \( j = 1, \ldots, n \).

**Proof:** Let \((\hat{u}, \hat{T})\) be the optimal solution to \(ASDED(d)\). Hence, for any feasible solution \(X\) to \(ASDE(d)\) we have, from the optimality conditions and (7.28), \(X \bullet \hat{T} = n\). Therefore, \(\|X\|_1 \cdot \|\hat{T}\|_\infty \geq n\), that is, \(\|X\|_1 \geq n / \|\hat{T}\|_\infty\), and by Proposition 8.1 the first inequality follows. For the second inequality, observe that because of the optimality conditions we have \(\hat{X} \hat{T} = I\). Since \(\hat{X} \in S_n\), there exists an orthogonal matrix \(U\) such that \(\hat{X} = UDUT^T\), where \(D = \text{diag}(\lambda(\hat{X}))\). Then, if \(\hat{T} = U^T \hat{T} U\), we have \(I = DU^T \hat{T} U = DT\), and so, \(\lambda_j(\hat{X}) \hat{T}_{jj} = 1\), for \(j = 1, \ldots, n\). Thus, \(\lambda_j(\hat{X}) = 1 / \hat{T}_{jj} \geq 1 / \|\hat{T}\|_\infty = 1 / \|\hat{T}\|_\infty\). Therefore, the result follows again from Proposition 8.1. By using a similar argument and Lemma 8.1, we can show the lower bound on \(\lambda_j(\hat{T})\), thus completing the proof.

q.e.d.

**Lemma 8.4** If \(d = (A_1, \ldots, A_m, b, C) \in F_I\) and \(\rho_I(d) > 0\), then

\[
\|S\|_\infty \geq \rho_I(d),
\]

for all feasible solution \((y, S)\) of \(ASDI(d)\). Moreover, if \((\hat{y}, \hat{S})\) is the optimal solution to \(ASDI(d)\) and \(\hat{V}\) the optimal solution to \(ASDID(d)\), then

\[
\lambda_j(\hat{S}) \geq \frac{\rho_I(d)}{n},
\]

\[
\lambda_j(\hat{V}) \geq \frac{1}{2 \|d\|C_I(d)},
\]

for all \(j = 1, \ldots, n\).

**Proof:** Let \(\hat{V}\) be an optimal solution to \(ASDID(d)\). Hence, for any feasible solution \((y, S)\) to \(ASDI(d)\) we have, from the optimality conditions and (7.33), \(S \bullet \hat{V} = n\). Therefore, \(\|S\|_\infty \|\hat{V}\|_1 \geq n\), that is, \(\|S\|_\infty \geq n / \|\hat{V}\|_1\), and by Proposition 8.2 the
first inequality follows. For the second inequality, observe that because of the optimality conditions we have \( \hat{\mathbf{V}} \hat{\mathbf{S}} = I \). Since \( \hat{\mathbf{S}} \in \mathcal{S}_n \), there exists an orthogonal matrix \( U \) such that \( \hat{\mathbf{S}} = UDU^T \), where \( D = \text{diag}(\lambda(\hat{\mathbf{S}})) \). Then, if \( \tilde{\mathbf{V}} = U^T \hat{\mathbf{V}} U \), we have \( I = DU^T \tilde{\mathbf{V}} U = D \tilde{\mathbf{V}} \), and so, \( \lambda_j(\hat{\mathbf{S}}) \tilde{V}_{jj} = 1 \), for \( j = 1, \ldots, n \). Thus, \( \lambda_j(\hat{\mathbf{S}}) = 1/\tilde{V}_{jj} \geq 1/\|\hat{\mathbf{V}}\|_\infty = 1/\|\tilde{\mathbf{V}}\|_\infty \geq 1/\|\tilde{\mathbf{V}}\|_1 \). Therefore, the result follows again from Proposition 8.2. By using a similar argument and Lemma 8.2, we can show the lower bound on \( \lambda_j(\hat{\mathbf{V}}) \), thus completing the proof.

q.e.d.

The following corollary follows immediately from Lemmas 8.3 and 8.4.

**Corollary 8.2** If \( d \in \mathcal{F}_E \) and \( \rho_E(d) > 0 \), then

\[
p(\hat{\mathbf{X}}) \leq n \ln(n \mathcal{C}_E(d)),
\]

where \( \hat{\mathbf{X}} \) is the optimal solution to \( \text{ASDE}(d) \). If \( d \in \mathcal{F}_I \) and \( \rho_I(d) > 0 \), then

\[
p(\hat{\mathbf{S}}) \leq n \ln \left( \frac{n}{\rho_I(d)} \right),
\]

where \((\hat{\mathbf{y}}, \hat{\mathbf{S}})\) is the optimal solution to \( \text{ASDI}(d) \).

### 8.3 Bounds on Solutions Along the Central Trajectory

This section presents results on lower and upper bounds on sizes of optimal solutions along the central trajectory, for the pair of dual logarithmic barrier problems \( \text{PSD}_\mu(d) \) and \( \text{DSD}_\mu(d) \). As in Chapter 7, before presenting our results, it is convenient to introduce the following constant, denoted \( \mathcal{K}(d, \mu) \), which arises in many of the results
to come:
\[ \mathcal{K}(d, \mu) = C(d)^2 + \frac{\mu n}{\rho(d)}. \]

The following theorem is an immediate consequence of Theorem 7.2.

**Theorem 8.1** If \( d \in \mathcal{F} \) and \( \rho(d) > 0 \), then

\[
\|X(\mu)\|_1 \leq \mathcal{K}(d, \mu), \\
\|y(\mu)\|_\infty \leq \mathcal{K}(d, \mu), \\
\|S(\mu)\|_\infty \leq 2\|d\|\mathcal{K}(d, \mu),
\]

for the optimal solution \( X(\mu) \) to \( PSD_\mu(d) \) and the optimal solution \( (y(\mu), S(\mu)) \) to the dual problem \( DSD_\mu(d) \).

This theorem states that the norms of optimal solutions along the central trajectory of a semi-definite program are bounded above by quantities only involving the condition number \( C(d) \) and the distance to ill-posedness \( \rho(d) \) of the data \( d \), as well as the dimension \( n \) and the barrier parameter \( \mu \). In particular, the theorem shows that the norm of the optimal primal solution along the central trajectory grows at most linearly in the barrier parameter \( \mu \), and at a rate no larger than \( n/\rho(d) \).

The next result presents lower bounds on the norms of the optimal solutions to the central trajectory problems \( PSD_\mu(d) \) and \( DSD_\mu(d) \), respectively.

**Theorem 8.2** If the program \( d \in \mathcal{F} \) and \( \rho(d) > 0 \), then

\[
\|X(\mu)\|_1 \geq \frac{\mu n}{2\|d\|\mathcal{K}(d, \mu)}, \\
\|S(\mu)\|_\infty \geq \frac{\mu n}{\mathcal{K}(d, \mu)}, \\
\lambda_j(X(\mu)) \geq \frac{\mu}{2\|d\|\mathcal{K}(d, \mu)},
\]
8.3 Bounds on Solutions Along the Central Trajectory

\[ \lambda_j(S(\mu)) \geq \frac{\mu}{\mathcal{K}(d, \mu)}, \]

where \( X(\mu) \) is the optimal solution to \( PSD_\mu(d) \) and \( (y(\mu), S(\mu)) \) is the optimal solution to \( DSD_\mu(d) \).

**Proof:** The first two inequalities are immediate consequences of Theorem 7.3. For the other two inequalities, observe because of the optimality conditions that \( X(\mu)S(\mu) = \mu I \). Since \( X(\mu) \in S_n \), there exists an orthogonal matrix \( U \) such that \( X(\mu) = UDUT^T \), where \( D = \text{diag}(\lambda(X(\mu))) \). Then, if \( \tilde{S} = U^T S(\mu) U \), we have \( \mu I = DU^T S(\mu) U = D\tilde{S} \), and so, \( \lambda_j(X(\mu))\tilde{S}_{jj} = \mu \), for \( j = 1, \ldots, n \). Thus, \( \lambda_j(X(\mu)) = \mu/\tilde{S}_{jj} \geq \mu/\|\tilde{S}\|_\infty = \mu/\|S(\mu)\|_\infty \), and the result for \( \lambda_j(X(\mu)) \) follows from Theorem 8.1. On the other hand, there exists an orthogonal matrix \( V \) such that \( S(\mu) = VDV^T \), where \( D = \text{diag}(\lambda(S(\mu))) \). Then, if \( \tilde{X} = V^T X(\mu)V \), we have \( \mu I = DV^T X(\mu)V = DX \), and so, \( \lambda_j(S(\mu))\tilde{X}_{jj} = \mu \), for \( j = 1, \ldots, n \). Thus, \( \lambda_j(S(\mu)) = \mu/\tilde{X}_{jj} \geq \mu/\|\tilde{X}\|_\infty = \mu/\|X(\mu)\|_\infty \geq \mu/\|X(\mu)\|_1 \), and the result for \( \lambda_j(S(\mu)) \) follows again from Theorem 8.1.

q.e.d.

This theorem shows that \( \|X(\mu)\|_1 \) and \( \lambda_j(X(\mu)) \) are bounded from below by functions only involving the quantities \( \|d\|, C(d), \rho(d), n, \) and \( \mu \). Furthermore, the theorem shows that for \( \mu \) close to zero, that \( \lambda_j(X(\mu)) \) grows at least linearly in \( \mu \), and at a rate that is at least \( 1/(2\|d\|C(d)^2) \). The theorem offers less insight when \( \mu \to \infty \), since the lower bound on \( \|X(\mu)\|_1 \) presented in the theorem converges to \( (2C(d))^{-1} \) as \( \mu \to \infty \). When the feasible region is unbounded, it is well known that \( \|X(\mu)\|_1 \to \infty \) as \( \mu \to \infty \), so that as \( \mu \to \infty \) the lower bound of Theorem 8.2 does not adequately capture the behavior of the sizes of optimal solutions to \( PSD_\mu(d) \) when the feasible region is unbounded.
8.4 Further Properties

The following corollaries are immediate consequences of Corollary 7.1 and Corollary 7.2. They show upper and lower bounds on optimal solutions to $PSD_\mu(d + \Delta d)$ and $DSD_\mu(d + \Delta d)$, where $d + \Delta d$ is a data instance representing a small perturbation of the data instance $d$.

**Corollary 8.3** Let $d \in \mathcal{F}$ be such that $\rho(d) > 0$. If $\|\Delta d\| \leq \rho(d)/3$, then

$$
\|X(\mu)\|_1 \leq 4\mathcal{K}(d, \mu), \\
\|y(\mu)\|_\infty \leq 4\mathcal{K}(d, \mu), \\
\|S(\mu)\|_\infty \leq 16\|d\|\mathcal{K}(d, \mu),
$$

where $X(\mu)$ is the optimal solution to $PSD_\mu(d + \Delta d)$ and $(y(\mu), S(\mu))$ is the optimal solution to $DSD_\mu(d + \Delta d)$.

**Corollary 8.4** Let $d \in \mathcal{F}$ be such that $\rho(d) > 0$. If $\|\Delta d\| \leq \rho(d)/3$, then

$$
\|X(\mu)\|_1 \geq \frac{\mu n}{8(\|d\| + \rho(d))\mathcal{K}(d, \mu)}, \\
\|S(\mu)\|_\infty \geq \frac{\mu n}{4\mathcal{K}(d, \mu)}, \\
\lambda_j(X(\mu)) \geq \frac{\mu}{8(\|d\| + \rho(d))\mathcal{K}(d, \mu)}, \\
\lambda_j(S(\mu)) \geq \frac{\mu}{4\mathcal{K}(d, \mu)},
$$

for $j = 1, \ldots, n$, where $X(\mu)$ is the optimal solution to $PSD_\mu(d + \Delta d)$ and $(y(\mu), S(\mu))$ is the optimal solution to $DSD_\mu(d + \Delta d)$.

8.4 Further Properties

We first show that the boundaries of the sets $\mathcal{F}$ and $\mathcal{F}_S$ are the same. To do so, we show that these sets have the same interiors and closures. Before doing so, we prove
the following proposition that characterizes the interior of the set $\mathcal{F}$. Observe that this proposition generalizes Lemma 3.1 which characterizes the interior of $\mathcal{F}$ in the context of linear programming in standard form (Lemma 3.1 was originally proven by Robinson [Rob77] and Ashmanov [Ash81]).

**Proposition 8.3** Let $\mathcal{H}$ be the following subset of data instances in $\bar{D}$:

$$
\mathcal{H} = \{(A_1, \ldots, A_m, b, C) \in \bar{D} : \text{there exists } (X, y) \text{ such that } A_i \cdot X = b_i, i = 1, \ldots, m, X \succ 0, \sum_{i=1}^{m} y_i A_i < C, \\
\text{and rank}([A_1, \ldots, A_m]) = m\}.
$$

Then $\text{int}(\mathcal{F}) = \mathcal{H}$.

**Proof:** It is not difficult to show that $\mathcal{H}$ is an open set contained in $\mathcal{F}$. Therefore, $\mathcal{H} \subset \text{int}(\mathcal{F})$.

On the other hand, let $d = (A_1, \ldots, A_m, b, C) \in \text{int}(\mathcal{F})$. We first assume that $\text{rank}([A_1, \ldots, A_m]) < m$. Then there exists a vector $u \in \mathbb{R}^m$ such that $\sum_{i=1}^{m} u_i A_i = 0$, $b^T u \geq 0$, and $u \neq 0$. Since $u \neq 0$, there exists $\bar{u}$ such that $\bar{u}^T u > 0$. For a given $\epsilon > 0$, let $\Delta d(\epsilon) = (0, \ldots, 0, \epsilon \bar{u}, 0)$. Then clearly $d + \Delta d(\epsilon) \notin \mathcal{F}$ for all $\epsilon > 0$. In other words, $d \in B$, a contradiction. Therefore, $\text{rank}([A_1, \ldots, A_m]) = m$.

Second, suppose that for all $X$ such that $A_i \cdot X = b_i$, for $i = 1, \ldots, m$, and $X \succeq 0$ we have $X \neq 0$. Thus, $\lambda_1(X) = 0$ for all feasible solutions $X$ of $PSD(d)$. Let $r \in \mathbb{R}^m$ be the vector defined as $r = (A_1 \cdot I, \ldots, A_m \cdot I)^T$, where $I$ is the $n \times n$ identity matrix. For a given $\epsilon > 0$, let $\Delta d(\epsilon) = (0, \ldots, 0, -\epsilon r, 0)$. Then $d + \Delta d(\epsilon) \notin \mathcal{F}$ for all $\epsilon > 0$ because the problem $PSD(d + \Delta d(\epsilon))$ is not feasible. In other words, again $d \in B$, a contradiction. Therefore, $A_i X = b_i$, for $i = 1, \ldots, m$, and $X \succ 0$ is a feasible system.

Finally, suppose that for all $(y, S)$ such that $\sum_{i=1}^{m} y_i A_i + S = C$ and $S \succeq 0$ we have $S \neq 0$. Thus, as before $\lambda_1(S) = 0$ for all feasible solution $(y, S)$ of $DSD(d)$. For a given $\epsilon > 0$, let $\Delta d(\epsilon) = (0, \ldots, 0, 0, -\epsilon I)$. Then $d + \Delta d(\epsilon) \notin \mathcal{F}$ for all $\epsilon > 0$ because
the problem $DSD(d + \Delta d(\epsilon))$ is not feasible. In other words, once again $d \in B$, a contradiction. Therefore, $\sum_{i=1}^{m} y_i A_i < C$, is a feasible system, thus completing the proof.

q.e.d.

Lemma 8.5

$$\text{int}(\mathcal{F}) = \text{int}(\mathcal{F}_S), \quad (8.6)$$
$$\text{cl}(\mathcal{F}) = \text{cl}(\mathcal{F}_S). \quad (8.7)$$

Proof: Since $\mathcal{F}_S \subset \mathcal{F}$, clearly $\text{int}(\mathcal{F}_S) \subset \text{int}(\mathcal{F})$ and $\text{cl}(\mathcal{F}_S) \subset \text{cl}(\mathcal{F})$. From Proposition 8.3, it is also clear that $\text{int}(\mathcal{F}) \subset \text{int}(\mathcal{F}_S)$, and so (8.6) follows.

Given $d = (A_1, \ldots, A_m, b, C) \in \mathcal{F}$, there exist $X$ and $y$ such that $A_i X = b_i$, for $i = 1, \ldots, m$, $X \succeq 0$, and $\sum_{i=1}^{m} y_i A_i \preceq C$. Let $r = (A_1 \bullet I, \ldots, A_m \bullet I)^T$. For a given $\epsilon > 0$, we have $A_i \bullet (X + \epsilon I) = b_i + \epsilon r_i$, for $i = 1, \ldots, m$, $X + \epsilon I \succ 0$, and $\sum_{i=1}^{m} y_i A_i \prec C + \epsilon I$. Hence, $d + \Delta d(\epsilon) \in \mathcal{F}_S$ for all $\epsilon > 0$, where $\Delta d(\epsilon) = (0, \ldots, 0, \epsilon r, \epsilon I)$. Therefore, $d \in \text{cl}(\mathcal{F}_S)$, and so $\mathcal{F} \subset \text{cl}(\mathcal{F}_S)$, which implies that $\text{cl}(\mathcal{F}) \subset \text{cl}(\text{cl}(\mathcal{F}_S)) = \text{cl}(\mathcal{F}_S)$, thus proving (8.7).

q.e.d.

The following lemma is an immediate consequence of Lemma 8.5, by noticing that $\partial \mathcal{G} = \text{cl}(\mathcal{G}) \cap \text{int}(\mathcal{G})^C$ for any set $\mathcal{G}$.

Lemma 8.6

$$\partial \mathcal{F} = \partial \mathcal{F}_S.$$
feasible region is bounded.

**Corollary 8.5** Let $X(\mu)$ be the optimal solution to $PSD_\mu(d)$ and suppose that the feasible region of $DSD_\mu(d)$ is bounded. Then

$$\lambda_j(X(d)) \geq \frac{\mu}{2\|d\|C_I(d)},$$

for $j = 1, \ldots, n$.

**Proof:** By Lemma 8.2, $\|S(\mu)\|_{\infty} \leq 2\|d\|C_I(d)$, where $(y(\mu), S(\mu))$ is the optimal solution to $DSD_\mu(d)$.

Because of the optimality conditions we have $X(\mu)S(\mu) = \mu I$. Since $X(\mu) \in S_n$, there exists an orthogonal matrix $V$ such that $X(\mu) = VDV^T$, where $D = \text{diag}(\lambda(X(\mu)))$. Then, if $\tilde{S} = V^T S(\mu)V$, we have that $\mu I = DV^T S(\mu)V = D\tilde{S}$, and so, $\lambda_j(X(\mu))\tilde{S}_{jj} = \mu$, for $j = 1, \ldots, n$. Thus, $\lambda_j(X(\mu)) = \mu/\tilde{S}_{jj} \geq \mu/\|\tilde{S}\|_{\infty} = \mu/\|S(\mu)\|_{\infty}$, and the result follows.

q.e.d.

From this result it follows that, under the conditions of Corollary 8.5, the eigenvalues of $X(\mu)$ diverge at least linearly to $\infty$ as the parameter $\mu$ grows to $\infty$.

Motivated by the preceding corollary, our next task is to present two conjectures concerning the behavior of the primal and dual semi-definite central trajectories, as the parameter $\mu$ tends to $\infty$. However, before stating these conjectures, we introduce some notation and prove a partial result towards this goal.

Given a data instance $d = (A_1, \ldots, A_m, b, C)$, we define the following two sets of primal and dual rays:

$$R_P(d) = \{ X \in S_n : A_i \cdot X = 0, i = 1, \ldots, m, X \succeq 0 \},$$
8.4 Further Properties

\[ R_D(d) = \left\{ S \in S_n : S = \sum_{i=1}^{m} y_i A_i \succeq 0, \text{ for some } y \in \mathbb{R}^m \right\}. \]

Observe that \( X \cdot S = 0 \) for all \( X \in R_P(d) \) and \( S \in R_D(d) \). Associated with these two sets, we define the maximal ranks \( N_P(d) \) and \( N_D(d) \) as follows,

\[ N_P(d) = \max \{ \text{rank}(X) : X \in R_P(d) \}, \]
\[ N_D(d) = \max \{ \text{rank}(S) : S \in R_D(d) \}. \]

In contrast to the linear programming context, in semi-definite programming we do not necessarily have the strict complementarity property for the eigenvalues of primal and dual rays, as the following example shows.

Consider the data instance \( d = (A_1, A_2, b, C) \), where

\[
A_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\]

Then, it is easy to verify that

\[ R_P(d) = \left\{ X \in S_3 : X = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \alpha \geq 0 \right\}, \]
\[ R_D(d) = \left\{ S \in S_3 : S = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \geq 0 \end{bmatrix} \right\}, \]

and so there is no strict complementarity for eigenvalues of rays in \( R_P(d) \) and \( R_D(d) \). Moreover, \( N_P(d) = 1 \) and \( N_D(d) = 1 \), so that \( N_P(d) + N_D(d) = 2 < 3 = n. \)
Lemma 8.7 Given a data instance \( d = (A_1, \ldots, A_m, b, C) \), we have

1. If \( R_P(d) = \emptyset \), then there exists \( y \in \mathbb{R}^m \) such that \( \sum_{i=1}^m y_i A_i \succ 0 \).

2. If \( R_D(d) = \emptyset \), then there exists \( X \in S_n \) such that \( A_i \cdot X = 0, \; i = 1, \ldots, m \), and \( X \succ 0 \).

Proof: (This proof is based on the research notes of Freund [Fre94]) To prove (1), consider the set \( G = \{ (r, z) \in \mathbb{R}^{m+1} : A_i \cdot X = r_i, i = 1, \ldots, m, \text{trace}(X) = z \}, \) for some \( X \succeq 0 \}. \) Clearly, \( G \) is a convex set. Moreover, \( G \) is a closed set. To see this, let \( \{ (r_k, z_k) : k \geq 1 \} \) be a sequence of points in \( G \) such that \( (r_k, z_k) \to (\bar{r}, \bar{z}) \) as \( k \to \infty \). Hence, there exists a sequence \( \{ X_k : k \geq 1 \} \) such that \( A_i \cdot X_k = (r_k)_i, i = 1, \ldots, m, \text{trace}(X_k) = z_k \), and \( X_k \succeq 0 \), for all \( k \geq 1 \). It is easy to see that if the sequence \( \{ X_k : k \geq 1 \} \) has a limit point, then \( (\bar{r}, \bar{z}) \in G \), and so \( G \) is closed. If \( \{ X_k : k \geq 1 \} \) does not have a limit point, then \( \|X_k\| \to \infty \) as \( k \to \infty \). Thus, given an arbitrary \( \epsilon > 0 \) and for \( k \) large enough, we have \( X_k \in H = \{ X : \| (A_1 \cdot X, \ldots, A_m \cdot X) - \bar{r} \| \leq \epsilon, X \succeq 0 \} \). Since \( H \) is a closed convex set and the sequence \( \{ X_k : k \geq 1 \} \) is unbounded, then \( H \) has a ray, that is, there exists a symmetric matrix \( V \) such that \( A_i \cdot V = 0, \; i = 1, \ldots, m, \; V \succeq 0 \), and \( V \neq 0 \). But then, \( V \in R_P(d) = \emptyset \), which is a contradiction.

Since \( (0, 1) \notin G \), there exists \( (u, \beta) \in \mathbb{R}^{m+1} \) and \( \alpha \in \mathbb{R} \) such that \( u^T 0 + \beta \times 1 \geq \alpha, u^T r + \beta z < \alpha \) for all \( (r, z) \in G \), and \( (u, \beta) \neq 0 \). By noticing that \( (0, 0) \in G \), we have \( \alpha > 0 \), and so, \( \beta \geq \alpha > 0 \). Without lost of generality we assume that \( \beta = 1 \). Hence, for all \( X \succeq 0 \) we have \( \alpha > u^T r + z = \sum_{i=1}^m u_i A_i \cdot X + \text{trace}(X) = (\sum_{i=1}^m u_i A_i + I) \cdot X \). Therefore \( \sum_{i=1}^m u_i A_i + I \preceq 0 \), that is, \( \sum_{i=1}^m u_i A_i < 0 \), and so, by taking \( y = -u \), we obtain the proof of (1).

To prove (2), consider the set \( G = \{ (V, z) \in S_n \times \mathbb{R} : V = \sum_{i=1}^m y_i A_i \text{ for some } y \in \mathbb{R}^m, V \succeq 0, \text{trace}(V) = z \} \). Clearly, \( G \) is a convex set. By using a similar argument to the one used to prove (1), we obtain that \( G \) is also a closed set. Hence, by observing
that $(0,1) \notin G$, we can again use a separation argument to obtain a symmetric matrix $X$ satisfying the required properties.

\textbf{q.e.d.}

Since a pair of rays $X \in R_P(d)$ and $S \in R_D(d)$ are complementary, that is, $X \bullet S = 0$, it follows that $N_P(d) + N_D(d) \leq n$. Moreover, from Lemma 8.7, it follows that when $d \in \mathcal{F}$, the primal feasible region or the dual feasible region is unbounded. Hence, we must have $N_P(d) + N_D(d) > 0$. Furthermore, again from Lemma 8.7, if the primal feasible region is bounded, then $N_P(d) = 0$ and $N_D(d) = n$; and if the dual feasible region is bounded, then $N_P(d) = n$ and $N_D(d) = 0$. Therefore, if $N_P(d) + N_D(d) < n$, then both the primal and dual feasible region are unbounded.

Now, we are in position to propose the following conjectures:

\textbf{Conjecture 8.1 (Weak version)} Let $d = (A_1, \ldots, A_m, b, C)$ be in $\mathcal{F}$ and let $X(\mu)$ denote an optimal solution of $PSD_\mu(d)$.

1. At least $N_P(d)$ eigenvalues of $X(\mu)$ diverge to $\infty$ at a rate $\Theta(\mu)$ as $\mu \to \infty$.

2. At least $N_D(d)$ eigenvalues of $X(\mu)$ remain bounded as $\mu \to \infty$.

3. At most $n-(N_P(d)+N_D(d))$ eigenvalues of $X(\mu)$ diverge to $\infty$ at a rate $\Theta(\mu^{2^{-r}})$, for some integer $r \geq 1$.

\textbf{Conjecture 8.2 (Strong version)} Let $d = (A_1, \ldots, A_m, b, C)$ be in $\mathcal{F}$ and let $X(\mu)$ denote an optimal solution of $PSD_\mu(d)$.

1. Exactly $N_P(d)$ eigenvalues of $X(\mu)$ diverge to $\infty$ at a rate $\Theta(\mu)$ as $\mu \to \infty$.

2. Exactly $N_D(d)$ eigenvalues of $X(\mu)$ remain bounded as $\mu \to \infty$.

3. Exactly $n-(N_P(d)+N_D(d))$ eigenvalues of $X(\mu)$ diverge to $\infty$ at a rate $\Theta(\mu^{2^{-r}})$, for some integer $r \geq 1$. 
Chapter 9

Conclusions and Future Work

In the previous chapters, we have shown that condition numbers are inherently related to many properties of important geometric objects in different convex programming contexts. These properties are not only useful in understanding the geometry and stability of numerical algorithms in optimization, but also are useful in understanding the complexity of such algorithms. In particular, our results on analytic centers and central trajectories have implications for the implementation of interior-point methods for linear programming, semi-definite programming, and conic linear systems in general.

Although we have addressed many issues in this thesis, there still remain many other research topics of great interest in this theory. The following is a list of open research topics derived from the study presented in this thesis:

- From Chapter 3 on the characterization of ill-posed data instances for linear programs in standard form and analytic center problems, there still remains open the issue of characterizing ill-posed data instances for other optimization problems such as semi-definite programming or, more generally, conic linear programs. This is a very difficult task to accomplish because it closely depends
on the degree in which we can extend "theorems of the alternative" to the more
general optimization contexts. We were able to generalize the characteriza-
tion of the interior of the set of feasible data instances for a primal and dual
semi-definite program. Nevertheless, we have not been able to characterize the
closure of this set.

Another related issue is the following: Since the conditioning depends on the
form of the program (for instance, a condition number for a linear program
in standard form does not necessarily have the same meaning as the condition
number for a linear program in symmetric inequality form), even if we stay
within the realm of linear programming, the characterization of ill-posed data
instances for linear programs (not in standard form) is still an open question.

- Since the central trajectory of a linear program converges to the analytic cen-
ter of the optimal face of such program, it is possible to use the results from
Chapter 4 concerning properties of analytic center problems to study the con-
vergence of the central trajectory to the optimal face and, hopefully, obtain new
results relating the complexity of path-following interior-point methods to the
conditioning of the optimal face of a linear program.

In order to do this, one possible approach is to find an adequate stability mea-
sure of the optimal face of a linear program to determine the complexity of
interior-point path-following methods, under the assumption that the program
does not necessarily have unique optimal solutions. We already have such a
measure in the case when a linear program has both unique primal and dual
optimal solutions. By using this particular stability measure, we were able to
derive analogous results concerning the complexity of interior-point methods
to those results derived from the traditional Kramer's rule approach and the
bit-length complexity approach. It remains to be seen if these results based on perturbation theory can be generalized to the case of multiple-optima.

- In Chapter 5, we studied the sensitivity of the optimal solutions to central trajectory problems as either the data changes or the barrier parameter changes. In the conic linear context, it is interesting to determine whether we can state similar bounds on the sizes of the changes in optimal solutions as well as in the optimal objective function values as the data or the barrier parameter is changed. In the linear programming case, we obtained our results by using specific properties of the log barrier function for the non-negative orthant. Although many of these properties generalize to logarithmically homogeneous barriers in the conic linear context, the proofs for the linear programming case are not easily carried over to the more general case.

- At the end of Chapter 8, we proposed two conjectures dealing with the behavior of the central trajectory of a semi-definite program as the barrier parameter tends to infinity. As shown in that chapter, this behavior is quite different from the behavior of the central trajectory of a linear program. In order to prove these conjectures, it is possible that the condition measures based on the perturbation model used in this thesis are not enough to assert the degree of truth of these conjectures. A related issue is studying the rate of convergence of the eigenvalues of the solutions on the central trajectory as the barrier parameter goes to zero, that is, as the central trajectory converges to the optimal face. The lack of strict complementarity in semi-definite programs makes this issue particularly difficult.

- Finally, throughout this thesis we have assumed that data instances are subject to deterministic perturbations. In some applications, it is more realistic to assume that data perturbations are stochastic, that is, that the data instance
associated with a convex program is a random variable having a known probability distribution on the space of data instances. Under these circumstances, we identify two open issues:

1. In the deterministic case, given a data instance $d$ of a linear program, we know that it is necessary to have $O(\log(C(d)))$ bits of precision to obtain an approximate data instance $d + \Delta d$ with the same feasibility properties as $d$. In the stochastic context, an analogous issue would be to determine how many bits of precision are necessary to obtain an approximate data instance $d + \Delta d$ that has the same feasibility properties as $d$ with a prespecified probability or degree of confidence.

2. In the deterministic case, we have several results concerning bounds in terms of condition numbers on changes of optimal solutions to several optimization problems as the data changes. In the stochastic case, an analogous issue would be to determine bounds on the expected error of optimal solutions as the data is randomly perturbed.
Bibliography


[Fre95] Robert M. Freund. Complexity of an infeasible interior-point algorithm for finding an approximate solution of a semi-definite program with no


Bibliography


Bibliography


