AN ANALYTIC CENTER CUTTING PLANE METHOD IN CONIC PROGRAMMING

By

Vasile Lucian Basescu

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Mathematics

Approved by the
Examining Committee:

John E. Mitchell, Thesis Adviser

Kristin Bennett, Member

Joseph Ecker, Member

Joyce McLaughlin, Member

Achille Messac, Member

Rensselaer Polytechnic Institute
Troy, New York

July 2003
(For Graduation August 2003)

AN ANALYTIC CENTER CUTTING PLANE METHOD IN CONIC PROGRAMMING

By

Vasile Lucian Basescu

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Mathematics

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

John E. Mitchell, Thesis Adviser
Kristin Bennett, Member
Joseph Ecker, Member
Joyce McLaughlin, Member
Achille Messac, Member

Rensselaer Polytechnic Institute
Troy, New York

July 2003
(For Graduation August 2003)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

First, I would like to thank my thesis adviser, Professor John Mitchell for all his support and encouragements. He is the one that introduced me to the wonderful world of optimization. This thesis wouldn't exist in this form without his constant guidance and invaluable insights. I am also indebted to Professor Joyce McLaughlin for teaching me Functional Analysis.

I would like to thank to the other members of my thesis committee: Professor Kristin Bennett, Professor Joseph Ecker and Professor Achille Messac.

Finally I am deeply grateful to my wife Dana and my parents Vasile and Lucia for their love and constant support. I dedicate this thesis to them.

ABSTRACT

Conic programming has been lately one of the most dynamic area of the optimization field. Although a lot of attention was focused on designing and analyzing interior-point algorithms for solving optimization problems, the class of analytic center cutting plane methods was less investigated. These methods are designed to solve feasibility problems by finding points which are interior to different sets of interest. Although these methods can be used by themselves to solve optimization problems, most of the time they are used as an initial step in a larger interior-point scheme employed in solving optimization problems.

There are many advantages in using this class of algorithms. For these methods to work there is no need to have before hand a complete description of the set of interest. All we need is an oracle that describes the set. This feature is especially useful when such a description is either missing or it is too large to be practical.

In this thesis we present a general analytic center cutting plane method for solving feasibility problems in the context of conic programming. The set of interest is convex, bounded, fully dimensional. It is described by an oracle that either recognizes that a point is interior to the set or returns a set of constraints violated by the current point but verified by all the points of the set of interest. These violated constraints are also known as cuts.

Our approach is an extension to the analytic center methods used in linear programming, second order cone programming or semidefinite programming. We prove that our algorithm can solve any feasibility problem with a convex, bounded, fully dimensional set of interest. We derive an upper bound for the total number of iterations the algorithm requires to get the solution. Also, we analyze how expensive each iteration is.

The performance of the algorithm is analyzed by solving some feasibility problems derived from the set of problems proposed in "The Seventh DIMACS Implementation Challenge Semidefinite and Related Optimization Problems". We consider feasibility problems with the sets described only by linear and second order conic

constraints. We will also present an algorithm for solving optimization problems that incorporates our analytic center cutting plane method. In the last part of this thesis we analyze the linear programming version of this algorithm and prove that it converges. Complexity results are also presented.

CHAPTER 1
Introduction

1.1 Define the Problem

In this thesis we will analyze the problem:

"Given an m-dimensional Hilbert space $(Y, \langle \cdot, \cdot \rangle_Y)$, find a point y in the convex bounded set $\Gamma \subset Y$."

Feasibility problems can be as hard to solve as optimization problems. In fact, once we have an algorithm for solving the feasibility problem, we can use it for solving optimization problems by using binary search.

Because the set Γ is convex, the problem we analyze is of interest in the larger context of non-differentiable convex optimization.

The first assumption made in any feasibility problem is that the domain Γ is strictly included in a larger set Ω_0. This larger set can be described using a set of so called "box-constraints". These "box-constraints" have different forms, depending on the nature of the Hilbert space $(Y, \langle \cdot, \cdot \rangle_Y)$. In the most general setting, the set Ω_0 is given by

$$\Omega_0 := \{y \in Y : c_1 \leq y \leq c_2\}.$$

The inequality sign "$\leq$" used in describing Ω_0 is a partial order defined on Y. This partial order generates a cone of "positive" vectors K (hence the name of conic programming),

$$K := \{x \in Y : x \geq 0\}.$$

Note here that $u \geq v \Leftrightarrow u - v \geq 0$. This partial order is what distinguishes different classes of feasibility problems.

The most basic class of such problems is linear programming. Linear programming deals with problems that have a linear objective and linear constraints. One of the multiple equivalent forms a linear programming problem can have is:

$$\begin{aligned} \max \quad & b^T y, \\ \text{subject to} \quad & A^T y \leq c. \end{aligned}$$

In this setting, the inequality between two vectors is to be understood componentwise,

$$u \geq v \text{ iff } u_i \geq v_i \text{ for all } i.$$

This vector inequality "$\geq$" introduces a partial ordering on the vector space $\mathbb{R}^n$. The first orthant is the corresponding cone of positive vectors.

More general than linear programming is second order cone programming. The partial order involved in this case is given by

$$u \geq 0,\, u \in \mathbb{R}^n \;\Leftrightarrow u_n \geq \sqrt{\sum_{i=1}^{n-1} u_i^2}.$$

The induced cone is called the second order cone or the Lorentz cone or the ice-cream cone. Linear programming can be considered a special case for second order cone programming. To see this it is enough to observe that if $n = 1$ the second order cone is $\mathbb{R}_+$. Then the first orthant $\mathbb{R}_+^n$ can be represented as a cartesian product of n lines $\mathbb{R}_+$ or of n one dimensional second order cones.

Even more general is semidefinite programming. In this case the cone K is the cone of positive semidefinite matrices $\mathcal{S}_n$. The partial order, denoted $\succeq$ is given by

$$A \succeq B \Leftrightarrow A - B \in \mathcal{S}_n.$$

To see that second order cone programming is a subcase of semidefinite programming it is enough to notice that the second order cone can be embedded in the cone of positive semidefinite matrices because

$$u_n \geq \sqrt{\sum_{i=1}^{n-1} u_i^2} \Leftrightarrow \begin{pmatrix} u_n I & v \\ v^T & u_n \end{pmatrix} \succeq 0,$$

where v is a $n-1$ - dimensional vector with $v_i = u_i$ for $i = 1, \ldots, n-1$.

All these cases are part of the conic programming family of problems. In this general case, the cone considered is a so called self-scaled cone (it will be defined

later). The second order cone, the cone of positive semidefinite matrices and their cartesian products are examples of such cones.

This is the general context in which we intend to analyze the feasibility problem.

We assume that this problem has a solution. One way of insuring that is to require that Γ contains a small ball of radius ε. This assumption insures that the set is not too flat. This is a feasibility problem. Any point from the interior of Γ is called feasible point.

The need for finding a point interior to a set arises in various optimization problems. As an example, interior-point algorithms need an interior point (as the name suggests) in order to start. In general what they need is a point in the region described by both equality and inequality constraints. Because of the equality constraints, usually of the form

$$Ax = b \text{ with } x \geq 0,$$

the feasibility region is not fully dimensional so it cannot contain a ball of radius ε. There are different ways of handling this situation.

Every feasibility problem can be transformed into an optimization problem by minimizing 0. As an example from linear programming, let's consider the next feasibility problem:

"Find a point in the domain described by $A^T y \leq c$ and $-\tilde{c}_0 \leq y \leq \tilde{c}_0$ with A an $m \times n$ matrix, y, $\tilde{c}_0$ m-dimensional vectors and c an n-dimensional vector."

The constraints $-\tilde{c}_0 \leq y \leq \tilde{c}_0$ are the "box-constraints" for this problem.

This feasibility problem can be reformulated in a primal-dual linear optimality problem. The dual problem is actually our feasibility problem:

$$\begin{aligned} \max \quad & 0 \\ \text{subject to} \quad & A^T y + s = c, \qquad (D) \\ & -\tilde{c}_0 \leq y \leq \tilde{c}_0, \\ & s \geq 0. \end{aligned}$$

Any solution for (D) is also a solution for the feasibility problem. So these two problems are equivalent. The corresponding primal problem is:

$$\begin{aligned} \min \quad & c^T x + \tilde{c}_0^T w + \tilde{c}_0^T z, \\ \text{subject to} \quad & Ax + w - z = 0, \\ & x, w, z \geq 0. \end{aligned} \qquad (P)$$

The opposite is also true. The linear problem:

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0. \end{aligned}$$

can be written as the feasibility problem:
"Find a feasible point for the set described by:

$$\begin{aligned} & c^T x - b^T y = 0, \\ & Ax = b, \\ & A^T y + s = c, \\ & x, s \geq 0." \end{aligned}$$

Any feasible point for the previous problem is also the solution to the linear problem because (x, y, s) is feasible in the primal-dual space with no duality gap.

Next we will describe the main idea of our approach (most of the terms encountered here will be defined later on, in the second chapter).

Going back to our problem we assume that

$$\Omega_0 := \{y \in Y : -\tilde{c}_0 \preceq_{\tilde{K}_0} y \preceq_{\tilde{K}_0} \tilde{c}_0 \text{ with } \tilde{c}_0 \in int(\tilde{K}_0)\}$$

Here $\tilde{K}_0$ is a full-dimensional self-scaled cone in the Hilbert space $(\tilde{X}_0, \langle \cdot, \cdot \rangle_0)$ with $\dim(\tilde{X}_0) = m$. Also we assume that Γ contains a ball of radius ε (so the set Γ is not too flat). We assume the existence of an oracle which, given a point $\hat{y}$ either recognizes that the point is in Γ or returns a p-dimensional Hilbert space $(X, \langle \cdot, \cdot \rangle_X)$ together with an injective linear operator $A : X \to Y$ such that:

$$\Gamma \subseteq \{y \in Y : A^*(\hat{y} - y) \in K\}.$$

Here K is a full-dimensional self-scaled cone in the Hilbert space $(X, \langle \cdot, \cdot \rangle_X)$. We will say that the operator A defines p central cuts.

In solving the problem we will generate a sequence of closed, bounded sets Ω_i such that $\Gamma \subseteq \Omega_i \subset \Omega_{i-1}$ for any $i \geq 1$. Each set Ω_i is obtained from the previous set Ω_{i-1} by introducing p_i central cuts through a special point $\hat{y}_{i-1} \in \Omega_{i-1}$:

$$\Omega_i := \Omega_{i-1} \cap \{y \in Y : A_i^*(\hat{y}_{i-1} - y) \in K_i\}. \tag{1.1}$$

The operator $A_i : (X_i, \langle \cdot, \cdot \rangle_i) \rightarrow Y$ is injective and linear, X_i is a p_i-dimensional Hilbert space and K_i is a full-dimensional self-scaled cone in X_i.

The special chosen points $\hat{y}_i$ are θ - analytic centers of the corresponding domains Ω_i with respect to an intrinsically self-conjugate functional $f_i : K_i \rightarrow I\!R$.

We will prove that if the total number of cuts added is big enough then the θ - analytic center of the last generated set Ω_i is guaranteed to be in Γ. We will get an estimate on the number of cuts that are added in order to solve the problem. Also we will study the complexity of obtaining one θ - analytic center $\hat{y}_i$ from the previous one $\hat{y}_{i-1}$.

We will prove that the algorithm will stop with a solution after no more than $O^*(\frac{mP^3\Theta^3}{\varepsilon^2\Lambda^2})$ (O^* means that terms of low order are ignored) cuts are added. Here P is the maximum number of cuts added at any of the iterations, Θ is a parameter characterizing the self-concordant functionals and Λ is the minimum eigenvalue of all $A_i^* A_i$ (A_i is the injective operator describing the cuts added at step i). The complexity result we obtain is comparable with the results obtained for less general cases.

1.2 Previous Work

The notion of analytic center was introduced for the first time by Sonnevend in [22]. Atkinson and Vaidya are the ones to introduce for the first time in [2] a complete analysis of a cutting plane method using analytic centers. In their approach the cuts are introduced one by one and "short-steps" are used. Dropping cuts is also allowed.

The set Γ is included in a cube of side 2^{L+1} and contains a ball of radius 2^{-L}. The complexity obtained is $O(mL^2)$ iterations. Mitchell and Ramaswamy extended this result in [8] to "long-steps". The complexity was the same but the "long-steps" method is more promising from the computational point of view.

The first analysis of the complexity of the analytic center cutting plane method with multiple cuts was done by Ye in [25]. He proved that by adding multiple cuts, the solution to the feasibility problem can be obtained in no more than $O^*(\frac{m^2P^2}{\varepsilon^2})$ iterations. The same complexity was obtained by Goffin and Vial in [5]. They proved that the recovery of a new analytic center can be done in $O(p\ln(p+1))$ damped Newton steps. This number of steps is the same regardless of the scaling matrix that is used (primal, dual or primal-dual). In our approach we will use a primal-dual approach.

The SOCP case is treated by Oskoorouchi and Goffin in [14]. They analyze the case when one SOCP cut is added at each call of the oracle. They prove that the analytic center of the new domain can be recovered in one Newton step and the total number of analytic centers generated before getting a feasible point is fully polynomial.

The semidefinite programming case is treated by Toh et. al. in [23]. They consider the case of adding multiple central cuts. In this case the cuts are added centrally through the analytic center $\hat{Y}$. The form of these cuts is given by $\{Y \in \mathcal{S}_+^m : A_i \bullet Y \leq A_i \bullet \hat{Y}, i = 1, \ldots, p\}$. If P is the maximum of all p, the complexity they obtain is $O(\frac{m^3P}{\varepsilon^2})$. Oskoorouchi and Goffin proved in [13] that the analytic center can be recovered in $O(p\ln(p+1))$ damped Newton steps and the total number of steps required to obtain the solution is $O(\frac{m^3P^2}{\varepsilon^2})$.

O. Peton and J.-P. Vial extend the analytic center cutting plane method to the general case of convex programming. In [17] they study the introduction of multiple central cuts in a conic formulation of the analytic center cutting plane method. They prove that the new analytic center can be recovered in $O(p\ln wp)$ damped Newton iterations, where w is a parameter depending of the data.

A general survey of non-differentiable optimization problems and methods with a special focus on the analytic center cutting plane method is presented by J.-L.

Goffin and J.-Ph. Vial in [4]. This paper presents also the case of multiple cuts and the case of deep cuts.

The analytic center cutting plane class of methods is a member of the larger class of interior point cutting plane methods. Mitchell in [7] gives an overview of these methods.

We conclude this section by presenting the outline of this thesis. All the notions encountered here will be defined in the following sections.

We will start our presentation by introducing in Chapter 2 some general notions and results about self-concordant functionals. These functionals are convex, Lipschitz continuous with their Hessians Lipschitz continuous too. The exact definition will be presented in Section 1 of this chapter. The analytic center of a convex bounded set is the minimizer of such a functional defined on the set. This point is well defined (because the functional it minimizes is convex) and is strictly interior to the domain. Because these functionals become infinitely large on the boundary of their domain, the analytic center will be pushed away from the boundary of the domain. Most of the theorems presented there are taken from [18] and are introduced without proof. In Section 2.2 we introduce a special local norm and some properties that will be used later on in our analysis. Here we will see that by fixing an arbitrary element in the self-scaled cone, we can define a local inner product and a special local norm. Using the properties of this new norm, we will prove that there is a region around this arbitrary point where the Hessian of the self-conjugate functional used in defining the analytic center has all eigenvalues greater than $\frac{1}{4}$. This property will be used later in analyzing the number of steps the algorithm requires to get to the solution. After setting up the theoretical structure we will define in Section 2.3 the notion of analytic center. Because computationally it is impossible to work with exact analytic centers, the notion of an approximate analytic center will be introduced. We will analyze then some its properties. In moving from one analytic center to the other we will need to make sure that feasibility is preserved all the time. For this, Dikin's ellipsoids are introduced. In Section 2.4 we will introduce more carefully all the assumptions we make about the problem.

As the algorithm proceeds, we introduce more constraints (i.e. cuts). These con-

straints are generated by an oracle and are defined using self-scaled cones. In Section 2.5 we introduce a new operation: $\oplus$ (which is somewhat similar to a cartesian product) which describes this process of adding cuts. We will see that by introducing these new cuts, the character of the problem doesn't change. At each stage of the algorithm the outer-approximation set of Γ has as analytic center the minimizer of the sum all the previous functionals defined on the $\oplus$-sum of all self-scaled cones introduced so far.

The third chapter of this thesis is dedicated to the analysis of the algorithm. After describing the algorithm in Section 3.1, we will analyze in Section 3.2 how the feasibility can be recovered after the cuts are added centrally, right through the analytic center. In order to keep track of changes in the potentials (another name for the self-concordant functionals used in our thesis to define the analytic centers) some scaled recovery steps need to be taken.

Section 3.3 is dedicated to analysis of potentials. The main result will characterize how the potentials at two consecutive analytic centers are related.

The implications of the assumptions made about the problems are studied in Section 3.4. In Section 3.5 we will derive an upper bound for the potentials evaluated at the corresponding analytic centers. This upper bound will be the one that will be used to prove that the algorithm eventually stops with a solution. As expected, this bound depends on the radius ε of the ball we assumed that Γ contains, on the characteristics of the potentials introduced and also on the condition number of the operators describing the cuts.

In Section 3.6 we prove that the algorithm will arrive at a solution in a certain number of steps. This complexity analysis is done in Section 11. We will use the approach employed by Ye in [25] in deriving the bound for the total number of constraints that can be introduced before the algorithm stops with a solution.

In Chapter 4 we will present some numerical results. The problems we are considering are modified versions of some of the problems proposed in "The Seventh DIMACS Implementation Challenge Semidefinite and Related Optimization Problems". We will consider feasibility problems that are based on a combination of LP cones and Second Order Cones. We will study the behavior of the algorithm in

solving feasibility problems and we will interpret the results.

In the second part of this chapter we will introduce a new algorithm (based on our analytic center cutting plane method) that can be used to solve optimality problems. The analysis of this algorithm (in a simplified context) will be presented in Chapter 5. We will solve then some optimality problems (based on the ones proposed in "The Seventh DIMACS Implementation Challenge").

We conclude this chapter by solving some optimality problems that arise when solving Partial Least Squares (PLS) problems and its kernel version (KPLS).

The last chapter of this thesis is dedicated to analyzing the algorithm proposed in Chapter 4 for solving linear convex optimization problems. We will look at its LP - only version. We will prove that this algorithm converges and we will also provide an upper bound for the total number of iterations required to get the solution.

CHAPTER 2
Preliminaries

2.1 Preliminaries on Self-Concordant Functionals

Self-concordant functionals are of the utmost importance for the optimization theory. In this section we will define this notion and will give some results regarding them that are relevant for our analysis. Most of the definitions/theorems presented in this section are taken from or inspired by [18] and [15].

Let $(X, \langle \cdot, \cdot \rangle_X)$ be a finite dimensional Hilbert space and let $f : X \to \mathbb{R}$ be a strictly convex functional with the following properties: D_f, the domain of f is open and convex, $f \in C^2$ and its Hessian $H(x)$ is positive definite for all $x \in D_f$. Using the functional f we introduce for each $x \in D_f$ the local (intrinsic) inner product (at x):

$$\langle u, v \rangle_x := \langle u, v \rangle_{H(x)} = \langle u, H(x)v \rangle_X.$$

More generally, for any positive definite operator S we can define a new inner product given by

$$\langle u, v \rangle_S = \langle u, Sv \rangle. \tag{2.1}$$

Let $B_x(y, r)$ be the open ball of radius r centered at y given by:

$$B_x(y, r) = \{z : \|z - y\|_x \leq r\}. \tag{2.2}$$

Definition 1 *A functional f is said to be* (strongly nondegenerate) self-concordant *if for all $x \in D_f$ we have $B_x(x, 1) \subseteq D_f$, and if whenever $y \in B_x(x, 1)$ we have:*

$$1 - \|y - x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1 - \|y - x\|_x}, \text{ for all } v \neq 0.$$

Let SC be the family of such functionals.

Let $g(y)$ be the gradient of the functional f defined using the original inner product $\langle \cdot, \cdot \rangle$. In the local intrinsic inner product $\langle \cdot, \cdot \rangle_x$, the corresponding gradient $g_x(y)$ and Hessian $H_x(y)$ are given by:

$$g_x(y) := H(x)^{-1} g(y), \tag{2.3}$$

$$H_x(y) := H(x)^{-1} H(y). \tag{2.4}$$

Definition 2 *A functional is said to be a* (strongly nondegenerate self-concordant) barrier functional *if $f \in SC$ and*

$$\theta_f := \sup_{x \in D_f} \|g_x(x)\|_x^2 < \infty. \tag{2.5}$$

Let SCB be the family of such functionals.

Definition 3 *Let K be a closed convex cone and $f \in SCB$, $f : int(K) \to I\!R$. f is* logarithmically homogeneous *if for all $x \in int(K)$ and $t > 0$:*

$$f(tx) = f(x) - \theta_f \ln(t). \tag{2.6}$$

Equivalently, f is logarithmically homogeneous *if, for all $x \in int(K)$ and all $t > 0$:*

$$g_x(tx) = \frac{1}{t} g_x(x). \tag{2.7}$$

Theorem 1 *If f is a self-concordant logarithmically homogeneous barrier functional then:*

$$H(tx) = \frac{1}{t^2} H(x), \;\; g_x(x) = -x \;\; and \;\; \|g_x(x)\|_x = \sqrt{\theta_f}.$$

Proof: This is *Theorem* 2.3.9 from [18].

The proof for the first part follows immediately by differentiating with respect to t in the definition of logarithmic homogeneous functionals. The last part is a consequence of *Definition* 2 and the fact that $\|g_x(x)\|_x = -\langle x, g(x) \rangle$ (the gradient of the right-hand side quantity is zero) . □

In linear programming such a logarithmically homogeneous self-concordant barrier functional is: $f : I\!R^n_+ \to I\!R$ with $f(x) := -\sum_{i=1}^{n} \ln(x_i)$. In this case $\theta_f = n$. For the SOCP case, the functional is given by $f(x) := -\ln(x_1^2 - \sum_{i=2}^{n} x_i^2)$, with $\theta_f = 2$. In the case of semidefinite programming such a functional is given by $f(X) := -\ln\det(X)$, with X a positive semidefinite matrix, $X \in \mathcal{S}_n$. The corresponding value for θ_f is $\theta_f = n$.

Most of the following results (taken from [18]) are technical in nature. They are needed in our analysis of the algorithm.

Theorem 2 *Let $f \in \mathcal{C}^2$ with its domain D_f open and convex. If $x, y \in D_f$ then*

$$f(y) - f(x) = \int_0^1 \langle g(x + t(y - x)), y - x\rangle dt.$$

Theorem 3 *If $f \in SC$, $x \in D_f$ and $y \in B_x(x, 1)$, then*

$$|f(y) - f(x) - \langle g(x), y - x\rangle_X - \frac{1}{2}\|y - x\|_x^2| \leq \frac{\|y - x\|_x^3}{3(1 - \|y - x\|_x)}.$$

If we take $y = x + d$ with $\|d\|_x < 1$ then

$$f(x + d) - f(x) \leq \langle g(x), d\rangle_X + \frac{1}{2}\|d\|_x^2 + \frac{\|d\|_x^3}{3(1 - \|d\|_x)}. \tag{2.8}$$

Theorem 4 *Assume $f \in SCB$ and $x \in D_f$. If $y \in \bar{D}_f$, then for all $0 < t \leq 1$,*

$$f(y + t(x - y)) \leq f(x) - \theta_f \ln t. \tag{2.9}$$

If the functional f is also logarithmically homogeneous, then a direct consequence of *Theorem* 4 is the next lemma.

Lemma 1 *Let $f \in SCB$ be a logarithmically homogeneous functional. If $x \in D_f$, $y \in \bar{D}_f$ and for all $t \geq 0$ then*

$$f(x + ty) \leq f(x). \tag{2.10}$$

If the domain of f is a cone K then the geometrical interpretation of *Lemma* 1 is that x maximizes f over the cone $x + K$.

Definition 4 *Let K be a cone and $z \in int(K)$. The* dual cone *of K is*

$$K^* = \{s \in X : \langle x, s\rangle_X \geq 0 \text{ for all } x \in K\}. \tag{2.11}$$

The dual cone of K with respect to the local inner product $\langle \cdot, \cdot \rangle_z$ *is given by*

$$K_z^* := \{s \in X : \langle x, s\rangle_z \geq 0, \text{ for all } x \in K\}. \tag{2.12}$$

The cone K is intrinsically self-dual *if $K_z^* = K$ for all $z \in$ int (K).*

Definition 5 *The* conjugate *of $f \in SCB$ with respect to $\langle \cdot, \cdot \rangle$ is*

$$f^*(s) := -\inf_{x \in int(K)} (\langle x, s\rangle + f(x)) \text{ with } s \in int(K_z^*).$$

In particular, the conjugate *of $f \in SCB$ with respect to $\langle \cdot, \cdot \rangle_z$ is*

$$f_z^*(s) := -\inf_{x \in int(K)} (\langle x, s\rangle_z + f(x)) \text{ with } s \in int(K^*).$$

A final definition:

Definition 6 *A functional $f \in SCB$ is* intrinsically self-conjugate *if f is logarithmically homogeneous, if K is intrinsically self-dual, and for each $z \in int(K)$ there exists a constant C_Z such that $f_z^*(s) = f(s) + C_z$ for all $s \in int(K)$.*

A cone K is self-scaled *or* symmetric *if $int(K)$ is the domain of an intrinsically self-conjugate barrier functional.*

Lemma 2 *Let K be a self-scaled cone. Then*

$$K = K_z^* = H(z)^{-1}K^* = H(z)^{-1}K. \tag{2.13}$$

Hence, for any $z \in K$, $H(z)$ is a linear automorphism of K.

Lemma 3 *If $f : int(K) \to I\!R$ is an intrinsically self-conjugate barrier functional, then for all $z \in int(K)$,*

$$f_z^*(s) = f(s) - (\theta_f + 2f(z)).$$

As a direct consequence:

$$g^* \equiv g \text{ and } H^* \equiv H.$$

Theorem 5 *Assume f is self-concordant. Then $f^* \in \mathrm{C}^2$. Moreover, if x and s satisfy $s = -g(x)$, then*

$$-g^*(s) = x \text{ and } H^*(s) = H(x)^{-1}.$$

Starting now, all the functionals we will deal with will be intrinsically self-conjugate barrier functionals.

For each cone K we will consider a fixed vector $e \in \mathrm{int}(K)$ and we will take all the inner products to be scaled by e.

Starting now, unless explicitly stated otherwise, each time we deal with an intrinsic self-conjugate functional f defined on a Hilbert space $(X, \langle\cdot,\cdot\rangle_X)$, the inner product will be thought to be the one induced by e (i.e. $\langle u, v\rangle = \langle u, H(e)v\rangle_X$ where $\langle\cdot,\cdot\rangle_X$ is the original inner-product on X). Accordingly, we will denote $K^* := K_e^*$, $g(x) := g_e(x)$ to be the gradient of f, $H(x) := H_e(x)$ to be the Hessian and so on.

Also if A^* is the adjoint operator of A in the original inner product, then $H(e)^{-1}A^*$ is the adjoint operator of A in the local inner product induced by e. We will denote

$$A^* := H(e)^{-1}A^*. \tag{2.14}$$

With this notation in mind, the vector e has some immediate and useful properties:

$$\|e\| = \sqrt{\theta_f}, g(e) = -e, H(e) = I. \tag{2.15}$$

Renegar proved in [18] the following result

Theorem 6 *Let f be an intrinsically self-conjugate barrier functional. Then, for any $x \in int(K)$:*

$$H(x)^{\frac{1}{2}} e = -g(x). \tag{2.16}$$

with H and g being the Hessian and gradient of f considered in the local inner product induced by e.

2.2 On Scaled Inner Products

Let $(X, \langle \cdot, \cdot \rangle_X)$ be a finite dimensional Hilbert space, with K a self-scaled cone and $f : X \to I\!R$ the corresponding self-conjugate functional. Let $e \in \text{int}(K)$ be a fixed point chosen arbitrarily.

Define the inner product $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{H(e)}$ to be the local inner product induced by e, i.e.:

$$\langle u, v \rangle = \langle u, H(e)v \rangle_X.$$

For this point e define the set $\mathcal{B} := \{v \in X : e \pm v \in \text{int}(K)\}$. Using this set define a new norm on X:

$$|v| := \inf\{t \geq 0 : \frac{1}{t} v \in \mathcal{B}\}.$$

Lemma 4 *Assume K is self-scaled. If $x \in K$ satisfies $|x - e| < 1$, then for all $v \neq 0$:*

$$\frac{1}{1 + |x - e|} \leq \frac{\|v\|_x}{\|v\|} \tag{2.17}$$

and

$$\frac{\|v\|_{-g(x)}}{\|v\|} \leq 1 + |x - e|. \tag{2.18}$$

Note here that $\|v\|_x = \|H(x)^{\frac{1}{2}}v\|$ with $H(x)$ and $\|\cdot\|$ being the ones induced by e.

This lemma gives a lower bound on the minimum eigenvalue for the Hessian of f computed in the norm induced by e at any point x such that $|x-e|<1$:

$$\lambda_{\min}(H(x)) = \inf_{v\neq 0}\frac{\|H(x)^{\frac{1}{2}}v\|^2}{\|v\|^2} = \inf_{v\neq 0}\frac{\langle v, H(x)v\rangle}{\|v\|^2} = \inf_{v\neq 0}\frac{\|v\|_x^2}{\|v\|^2} > \frac{1}{4}. \tag{2.19}$$

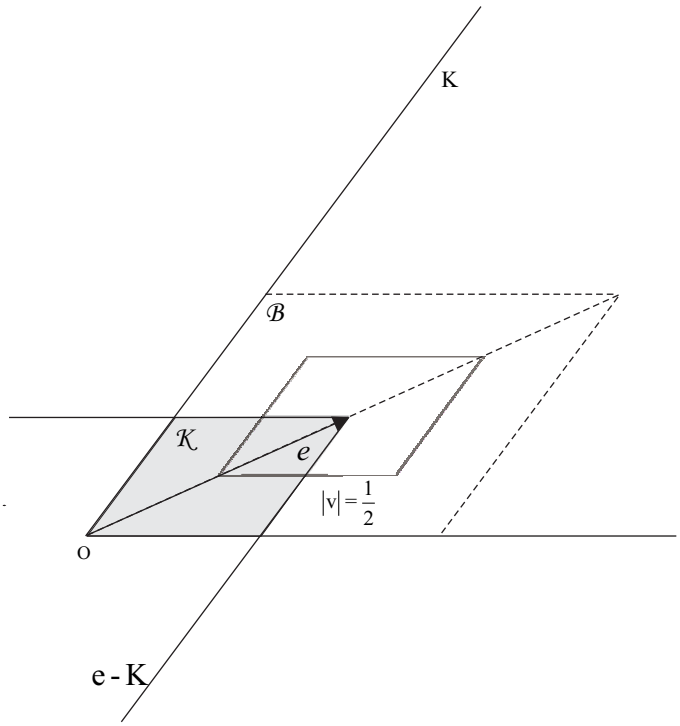

Figure 2.1: The sets $\mathcal{B}$, $\mathcal{K}$ and the level set $|v| = \frac{1}{2}$.

Now let's consider the domain described by $|x-e|<1$. We claim that:

Lemma 5 $\mathcal{K} := int(K)\cap(e-K) \subseteq \{x \in int(K) : |x-e|<1\}$.

Proof: Let $y \in \text{int}(K)\cap(e-K)$. Then $y = e-z$, with $z \in K$. The point $y-e \in \mathcal{B}$ because $e+y-e = y \in \text{int}(K)$ and $e-(y-e) = z+e \in \text{int}(K)$.

Let y' be the point of intersection between ∂K and the line that goes through e and has the direction $y-e$. Then $y'-e = t(y-e)$ for some $t>1$. The middle

point between y and y' is clearly a point interior to K. Moreover,

$$e + \frac{y+y'}{2} - e = \frac{y+y'}{2} \in \mathrm{int}(K)$$

and

$$e - \frac{y+y'}{2} + e = e + \frac{e-y}{2} + \frac{e-y'}{2} \in \mathrm{int}(K).$$

So

$$\frac{1+t}{2}(e-y) = e - \frac{y+y'}{2} \in \mathcal{B}.$$

Then:

$$|y-e| = \inf\{t \geq 0 : \frac{1}{t}(e-y) \in \mathcal{B}\} \leq \frac{2}{t+1} < 1.$$

So $|y-e| < 1$. □

As a direct consequence of the previous analysis:

Corollary 1 *Let $f \in SCB$ be intrinsically self-conjugate. Then, for any $x \in int(K) \cap (e-K)$:*

$$\lambda_{\min}(H(x)) > \frac{1}{4}. \tag{2.20}$$

2.3 Analytic Centers

Let $(X, \langle\cdot,\cdot\rangle_X)$ and $(Y, \langle\cdot,\cdot\rangle_Y)$ be two Hilbert spaces of finite dimensions: $\dim X = n$, $\dim Y = m$. In X consider a full-dimensional self-scaled cone K, pointed at zero (i.e. $K \cap -K = \{0\}$) with the corresponding intrinsically self-conjugate barrier functional $f : X \to I\!R$. Let $A : X \to Y$ be a surjective linear operator.

Using the convention from the previous section, we take an arbitrary element $e \in \mathrm{int}(K)$ and scale everything using the local inner product induced by e (hence $\langle\cdot,\cdot\rangle_X := \langle\cdot,\cdot\rangle_e$ and so on).

The analytic center (the AC) of the domain $\mathcal{F}_P := \{x \in K : Ax = 0\}$ with respect to $f(x) + \langle c, x\rangle_X$ is the exact solution to the problem:

$$\begin{array}{rll} \min & f(x) + \langle c, x\rangle_X & \\ \text{subject to} & Ax = 0 & (P_1), \\ & x \in K. & \end{array}$$

Alternatively, the analytic center can be defined using the dual formulation of the previous problem. The analytic center of $\mathcal{F}_D := \{s \in K : A^*y + s = c\}$ with respect to $f_e^*(s)$ is the solution to:

$$\begin{array}{rll} \min & f_e^*(s) & \\ \text{subject to} & A^*y + s = c, & (D_1) \\ & s \in K. & \end{array}$$

One last thing to note here. The functional f is intrinsic self-conjugate. Then, by definition, $f_e^*(s) - f(s)$ is constant. So minimizing $f_e^*(s)$ is the same with minimizing $f(s)$. In what will follow we will keep using the notation $f^*(s)$ although we are actually using $f_e^*(s)$.

Now, let's analyze the primal and dual problems. The KKT conditions for the first problem are:

$$\begin{aligned} g(x) + c + A^*\lambda &= 0, \\ Ax &= 0, \\ x &\in K. \end{aligned}$$

Because f is a self-concordant barrier functional, its gradient $g(x)$ takes the cone K into $-K$ (as shown in [18]). So $-g(x) \in K$ for any $x \in K$. With this observation, the previous system can be written as:

$$\begin{aligned} g(x) + s &= 0, \\ Ax &= 0, \\ A^*y + s &= c, \\ x, s &\in K. \end{aligned}$$

To get this formulation it is enough to take $\lambda = -y$ and $s = c - A^*y \in K$.

For the dual problem, the KKT conditions are:

$$\begin{aligned} g^*(s) + \lambda &= 0, \\ A\lambda &= 0, \\ A^*y + s &= c, \\ \lambda, s &\in K. \end{aligned}$$

Using the fact that $\lambda = -g^*(s)$ is equivalent to $s = -g(\lambda)$, it is easy to see that, in either case the KKT conditions are the same, defining the same AC.

Hence, for any analytic center the next equalities hold:

$$\begin{aligned} g(x) + s &= 0, \\ g(s) + x &= 0, \\ Ax &= 0, \\ A^*y + s &= c, \\ x, s &\in K. \end{aligned} \tag{2.21}$$

For simplicity we will say that x or y or s is an analytic center if they are the components of an analytic center.

We can introduce the notion of θ - analytic center by relaxing some of the previous equalities. First we will define this notion then, the following lemma will give an insight for this definition.

Definition 7 *(x, y, s) is a θ - analytic center for $\mathcal{F}_P$, $\mathcal{F}_D$ iff $x \in \mathcal{F}_P$, $s \in \mathcal{F}_D$ and*

$$\|I - H(x)^{-\frac{1}{2}} H(s)^{-\frac{1}{2}}\| \leq \frac{\theta}{\sqrt{\theta_f}}. \tag{2.22}$$

Lemma 6 *Let (x, y, s) be a θ - analytic center. Then:*

$$\begin{aligned} \|x + g(s)\|_{-g(s)} &\leq \theta, \\ \|s + g(x)\|_{-g(x)} &\leq \theta. \end{aligned} \tag{2.23}$$

Proof: We will prove only the first inequality. Note that the inner product $\langle \cdot, \cdot \rangle_X$ is the one induced by e. Using *Theorem* 5 :

$$\begin{aligned} \|x + g(s)\|^2_{-g(s)} = & \quad \langle x + g(s), H(-g(s))(x + g(s)) \rangle_X \\ = & \quad \langle x + g(s), H(s)^{-1}(x + g(s)) \rangle_X. \end{aligned}$$

Next we will use the fact that, as shown in *Theorem* 1 and *Theorem* 6 for any $x \in K$:

$$g(x) = H(x)^{\frac{1}{2}} e \text{ and } x = -g_x(x) = -H(x)^{-1} g(x).$$

Based on these:

$$\begin{aligned} &\|x + g(s)\|^2_{-g(s)} = \langle -H(x)^{-1} g(x) + g(s), H(s)^{-1}(-H(x)^{-1} g(x) + g(s)) \rangle_X \\ &= \langle -H(x)^{-1} H(x)^{\frac{1}{2}} e + H(s)^{\frac{1}{2}} e, H(s)^{-1}(-H(x)^{-1} H(x)^{\frac{1}{2}} e + H(s)^{\frac{1}{2}} e) \rangle_X \\ &= \langle -H(x)^{-\frac{1}{2}} e + H(s)^{\frac{1}{2}} e, H(s)^{-1}(-H(x)^{-\frac{1}{2}} e + H(s)^{\frac{1}{2}} e) \rangle_X \\ &= \langle H(s)^{\frac{1}{2}}(-H(s)^{-\frac{1}{2}} H(x)^{-\frac{1}{2}} e + e), H(s)^{-1} H(s)^{\frac{1}{2}}(-H(s)^{-\frac{1}{2}} H(x)^{-\frac{1}{2}} e + e) \rangle_X \\ &= \langle -H(s)^{-\frac{1}{2}} H(x)^{-\frac{1}{2}} e + e, -H(s)^{-\frac{1}{2}} H(x)^{-\frac{1}{2}} e + e \rangle_X. \end{aligned}$$

So:

$$\|x + g(s)\|_{-g(s)} \leq \|I - H(s)^{-\frac{1}{2}} H(x)^{-\frac{1}{2}}\| \|e\| \leq \theta.$$

□

The motivation for using this definition for a θ - analytic center should be clear if we compare it with the usual definition used in linear programming for a θ - analytic center:

$$\|e - xs\| \leq \theta,$$

with e being the vector of all ones.

Using the fact that in the linear programming case the Hessian is given by $H(x) = diag(x^{-2})$ our definition reduces to:

$$\|I - H(x)^{-\frac{1}{2}} H(s)^{-\frac{1}{2}}\| = \|diag(e - xs)\| = \max_i (1 - x_i s_i) \leq \frac{\theta}{\sqrt{\theta_f}}.$$

This is slightly different from the usual definition. Using *Lemma* 6 it is clear that our definition is close to the one used in the linear programming case:

$$\|x + g(s)\|^2_{-g(s)} = (x - s^{-1})^T diag(s^2))(x - s^{-1}) = \|e - xs\|^2$$

Next lemma is a simple exercise on the relationship between norms induced by different matrices.

Lemma 7 *Let A and B be two positive definite linear operators with B being Hermitian. Then:*

$$\|y\|_A = \|B^{-1} y\|_{B^* A B}.$$

Proof: $\|y\|_A^2 = \langle y, Ay \rangle = \langle B^{-1} y, B^* A B (B^{-1} y) \rangle = \|B^{-1} y\|^2_{B^* A B}$. □

Lemma 8 *If (x, y, s) is the analytic center for the intrinsically self-conjugate barrier functional f then $H(s)H(x) = I$.*

Proof: Note that $g^* \equiv g$ and $H^* \equiv H$ because f is intrinsically self-conjugate functional. Because (x, y, s) is an analytic center then $s = -g(x)$ so, using *Theorem* 5 we get that $H(s) = H^*(s) = H(x)^{-1}$. Hence the conclusion. □

Note here that $H(s)H(x) = I$. In a linear programming formulation this translates in $x_i s_i = 1$ for all i. This is the exact expression that defines the exact analytic center in the linear programming case.

In order to get a characterization for a θ - analytic center we need the following result.

Lemma 9 *If $\|I - A\| \leq \theta < 1$, A invertible, then*

$$1 - \theta \leq \|A\| \leq 1 + \theta.$$

and

$$\frac{1}{1+\theta} \leq \|A^{-1}\| \leq \frac{1}{1-\theta}.$$

Proof: The first inequalities are immediate. For the second set of inequalities:

$$\|A^{-1}\| = \|A^{-1} - I + I\| \leq \|A^{-1} - I\| + \|I\| \leq \|A^{-1}(I - A)\| + 1 \leq \theta\|A^{-1}\| + 1$$

and

$$1 = \|I\| = \|A^{-1}A\| \leq \|A^{-1}\|\|A\| \leq (1+\theta)\|A^{-1}\|.$$

□

Lemma 10 *Let (x, y, s) be a θ - analytic center for an intrinsically self-conjugate barrier functional f. Then:*

$$\frac{\sqrt{\theta_f}}{\sqrt{\theta_f} + \theta} \leq \|H(x)^{\frac{1}{2}} H(s)^{\frac{1}{2}}\| \leq \frac{\sqrt{\theta_f}}{\sqrt{\theta_f} - \theta}. \tag{2.24}$$

Proof: The lemma is proved immediately using the previous lemma. □

Lemma 11 *Let f be an intrinsically self-conjugate barrier functional defined on a self-scaled cone K. Let $x, s \in K$ such that $x = -g(s)$. Then:*

$$f(x) + f^*(s) = -\theta_f. \tag{2.25}$$

Proof: Because f is self-conjugate we have: $g^*(s) = g(s)$. Renegar proved in [18] that regardless of the inner product, the conjugate functional satisfies:

$$f^*(s) = \langle g^*(s), s\rangle - f(-g^*(s)).$$

So $f^*(s) = \langle g(s), s\rangle - f(x) = -\theta_f - f(x)$. □

Suppose that x is a feasible point in $\mathcal{F}_P$. If f is a self-concordant functional, then, by definition, $\|\Delta x\|_x \leq 1$ implies that $x + \Delta x$ is feasible.

This inequality describes an ellipsoid around the point x (also known as the Dikin's ellipsoid). This ellipsoid defines a region around the point x where $x+\Delta x$ is feasible too. The following lemmas will give sufficient conditions on Δx and Δs to get $x+\Delta x$, $s+\Delta s$ feasible, given that x and s are feasible in $\mathcal{F}_P$ and $\mathcal{F}_D$ respectively.

Lemma 12 *Let $\mathcal{E}_P = \{\Delta x \in X : A\Delta x = 0, \|\Delta x\|_x \leq 1\}$. Let (x, y, s) be a θ - analytic center. Then:*

$$(1 + \frac{\theta}{\sqrt{\theta_f}})^{-1}\mathcal{E}_P \subseteq \{\Delta x \in X : A\Delta x = 0, \|\Delta x\|_{H(s)^{-1}} \leq 1\} \subseteq (1 - \frac{\theta}{\sqrt{\theta_f}})^{-1}\mathcal{E}_P.$$

Proof:

$$\begin{aligned}
\|\Delta x\|_{H(s)^{-1}} &= \|H(s)^{-\frac{1}{2}}\Delta x\|_X = \|H(s)^{-\frac{1}{2}}H(x)^{-\frac{1}{2}}H(x)^{\frac{1}{2}}\Delta x\|_X \\
&\leq \|H(s)^{-\frac{1}{2}}H(x)^{-\frac{1}{2}}\|\|H(x)^{\frac{1}{2}}\Delta x\|_X \leq (1 + \frac{\theta}{\sqrt{\theta_f}})\|H(x)^{\frac{1}{2}}\Delta x\|_X \\
&= (1 + \frac{\theta}{\sqrt{\theta_f}})\|\Delta x\|_x.
\end{aligned}$$

Also:

$$\begin{aligned}
\|\Delta x\|_x &= \|H(x)^{\frac{1}{2}}\Delta x\|_X = \|H(x)^{\frac{1}{2}}H(s)^{\frac{1}{2}}H(s)^{-\frac{1}{2}}\Delta x\|_X \\
&\leq \|H(x)^{\frac{1}{2}}H(s)^{\frac{1}{2}}\|\|\Delta x\|_{H(s)^{-1}} \leq \frac{1}{(1 - \frac{\theta}{\sqrt{\theta_f}})}\|\Delta x\|_{H(s)^{-1}}.
\end{aligned}$$

□

A similar result holds for the Dikin's ellipsoid around s.

Let $\mathcal{E}_D = \{\Delta s \in X : \Delta s = -A^*\Delta y, \|\Delta s\|_s \leq 1\}$.

Lemma 13 *Let (x, y, s) be a θ - analytic center. Then:*

$$(1 + \frac{\theta}{\sqrt{\theta_f}})^{-1}\mathcal{E}_D \subseteq \{\Delta s : \Delta s = -A^*\Delta y, \|\Delta s\|_{H(x)^{-1}} \leq 1\} \subseteq (1 + \frac{\theta}{\sqrt{\theta_f}})^{-1}\mathcal{E}_D.$$

Because analytic centers are minimizers of convex functionals defined on closed, bounded, convex sets, the method of choice for computing them is the Newton method. This method is based on the second-order approximation of the functional to be minimized. Let q_x be the quadratic approximation for f. Then

$$q_x(y) = f(x) + \langle g(x), y - x\rangle + \frac{1}{2}\langle y - x, H(x)(y - x)\rangle.$$

The minimizer for $q_x(y)$ is the point $\bar{x}$ where the gradient of $q_x(y)$ becomes zero. From here we obtain immediately that:

$$\bar{x} = x - H(x)^{-1}g(x). \tag{2.26}$$

The Newton step is defined to be the vector $H(x)^{-1}g(x)$. This vector is the same with $g_x(x)$. Immediately we can see that, for logarithmically homogeneous barrier functionals the Newton step has constant length if measured in the norm induced by x: $\|g_x(x)\|_x = \sqrt{\theta_f}$. The advantage of using self concordant barrier functional is that the Newton step doesn't change when the local inner product changes. This gives us more flexibility in the way we choose the local inner product.

When computing approximate analytic centers we need a way of estimating distances to the exact analytic center. When working with general functionals it is impossible to achieve this without knowing the exact analytic center. This problem is eliminated when using self-concordant functionals. This is because we can use local inner products instead of the original one. We can compute the distance between two points x and y without knowing y. All we need to do is to use $\|x - y\|_y$ to measure the distance.

All these properties will play an important role when we will analyze the complexity of recovering the analytic center.

2.4 Assumptions and Notations

We assume that all the operators $A_i : X_i \to Y$, $i \geq 1$ defining the cuts are injective, hence the adjoint operators A_i^* are surjective. Also, wlog we assume that $\|A_i\| = 1$. The fact that A_i is injective gives also a bound on how many cuts we can add at a certain moment: $p_i \leq m$.

For each space $(X_i, \langle \cdot, \cdot \rangle_i)$ we will use the local norm induced by an arbitrary element $e_i \in \text{int}(K_i)$. So whenever we use $\langle \cdot, \cdot \rangle_i$ we will actually mean $\langle \cdot, \cdot \rangle_{e_i}$. If there is no danger for confusion, we will also use $\langle \cdot, \cdot \rangle$ instead of $\langle \cdot, \cdot \rangle_{e_i}$.

The following assumptions are not critical for our analysis. We use them just to keep the analysis simpler and easier to understand. The analysis would be the same without these assumptions but the notation would be more complicated.

We assume that $\|H_i(e_i)^{-1}\| = 1$ for $i \geq 0$, where H_i are the Hessians corresponding to the intrinsically self-conjugate functionals that are generated by the algorithm. The Hessians are computed in the original inner products (not the scaled one). To ensure this, it is enough to pick an arbitrary $e_i' \in \text{int}(K_i)$. Then take $e_i := \|H_i^{-1}(e_i')\|^{-\frac{1}{2}} e_i'$. Because f_i is logarithmically homogeneous (hence $H_i^{-1}(tx) = t^2 H_i^{-1}(x)$) for e_i we have $\|H_i^{-1}(e_i)\| = 1$. We can scale e_0 in a similar way to get $\|H_0(e_0)\| = 1$.

Let $\sigma_i := \sqrt{\frac{p_i}{\theta_i}} e_i$. The length of this vector, measured in the local inner product induced by e_i is $\|\sigma_i\| = \sqrt{p_i}$. Without loss of generality, we can assume that $f_i(\sigma_i) = 0$. We can do this easily. If f_i evaluated at this point is different from zero, then we can replace $f_i(x)$ by $f_i(x) - f_i(\sigma_i)$. Note that we can do this because the sum between a constant and an intrinsically self-conjugate barrier functional is an intrinsically self-conjugate barrier functional.

2.5 Operations with Self-Concordant Functionals

In our algorithm we start with an initial set Ω_0 containing Γ. As the algorithm proceeds we generate a sequence of sets Ω_i containing Γ such that $\Omega_i \subset \Omega_{i-1}$. Each

set Ω_i is described by all the cuts that have been added so far. In a condensed form we can represent Ω_i as:

$$\Omega_i = \{y \in Y : \bar{A}_i^* y + s = \bar{c}_i \text{ with } s \in \bar{K}_i\}.$$

In this section we will describe how $\bar{A}_i$ and $\bar{K}_i$ incorporate all the cuts added.

Note here that when the cuts are added, the dimensionality of the problem increases (i.e. the dimension of the domain of $\bar{A}_i$ increases).

Each time we add new cuts, we introduce new Hilbert spaces $(X_i, \langle\cdot,\cdot\rangle_i)$ with $\dim(X_i) = p_i$ such that $p_i < m$ for all $i \geq 1$. In each X_i we consider a full-dimensional self-scaled cone K_i. The cuts that are added are of the form

$$A_i : X_i \to Y \text{ with } A_i \text{ injective linear operators.}$$

Hence, the corresponding adjoint operators $A_i^* : Y \to X_i$ are surjective.

In the linear programming case this condition translates into asking that the matrix A_i describing the cuts be full-ranked. Together with the cone K_i we introduce an intrinsically self-conjugate barrier functional f_i:

$$f_i : K_i \to I\!R.$$

The domain of the functional is the interior of the cone K_i. When the cuts are added to the problem everything changes: the space X and its inner product, the barrier functional, and the constraints describing the outer approximation domain for Γ. Next we will describe all these changes and how are they incorporated in the structure of the algorithm. We should keep in mind that the inner products of X_i are scaled by elements $e_i \in \text{int}(K_i)$ fixed arbitrarily.

First let's define a new operation $\oplus$. This is similar to the cartesian product. This operation will be defined for Hilbert spaces, functionals and operators.

Definition 8 *Let $(X, \langle\cdot,\cdot\rangle_X)$ and $(Z, \langle\cdot,\cdot\rangle_Z)$ be two Hilbert spaces of dimensions p_X and p_Z respectively. We define $(X \oplus Z, \langle\cdot,\cdot\rangle_{X\oplus Z})$ to be an Hilbert space of dimension*

$p_X + p_Z$ *defined by:*

$$u \in X \oplus Z \text{ iff } (u_i)_{i=1,\ldots,p_X} \in X \text{ and } (u_i)_{i=p_X+1,\ldots,p_X+p_Z} \in Z. \tag{2.27}$$

We write $u = u_X \oplus u_Z$.

The corresponding inner product is given by:

$$\langle u, v\rangle_{X\oplus Z} = \langle u_X, v_X\rangle_X + \langle u_Z, v_Z\rangle_Z. \tag{2.28}$$

The $\oplus$ operation is just a concatenation, while the resulting inner product is the sum of the inner products applied to the corresponding components.

The sum of two cones can be defined in a similar manner. The same idea of concatenation applies for operators too:

Definition 9 *Let* $A_X : X \to Y$ *and* $A_Z : Z \to Y$ *be two linear operators. The* $\oplus$ *sum of these two operators is defined by:*

$$A_X \oplus A_Z : X \oplus Z \to Y, \text{ with } (A_X \oplus A_Z)(u) = A_X u_X + A_Y u_Y. \tag{2.29}$$

Observe here that the $\oplus$ sum of two linear operators is a linear operator too.

Finally, we define the $\oplus$ sum of two functionals.

Definition 10 *Let* $f_X : X \to I\!R$ *and* $f_Z : Z \to I\!R$ *be two functionals. Then the* $\oplus$ *of* f_X *and* f_Z *is given by:*

$$f_X \oplus f_Z : X \oplus Z \to I\!R, \text{ with } (f_X \oplus f_Z)(u) = f_X(u_X) + f_Z(u_Z) \tag{2.30}$$

for all $u = u_X \oplus u_Z \in X \oplus Z$.

Lemma 14 *Let* $f_i : (X_i, \langle\cdot,\cdot\rangle_i) \to I\!R$, $i = 1, 2$, *be two functionals in* $\mathcal{C}^2$, *with* g_1, g_2 *their gradients and* H_1, H_2 *their Hessians respectively. Then the gradient* g *of* $f := f_1 \oplus f_2$ *is such that:*

$$\langle u, g(v)\rangle_{X_1\oplus X_2} = \langle u_1, g_1(v_1)\rangle_1 + \langle u_2, g_2(v_2)\rangle_2 \tag{2.31}$$

and the Hessian H of f is such that:

$$\langle u, H(v)z\rangle_{X_1\oplus X_2} = \langle u_1, H_1(v_1)z_1\rangle_1 + \langle u_2, H_2(v_2)z_2\rangle_2 \tag{2.32}$$

for any $u, v, z \in X_1 \oplus X_2$ (with their components $u_i, v_i, z_i \in X_i$, $i = 1, 2$).

Proof: For the first part, the definition of the gradient of a functional gives:

$$\lim_{\|\Delta v_i\|_i \to 0} \frac{f_i(v_i + \Delta v_i) - f_i(v_i) - \langle g_i(v_i), \Delta v_i\rangle_i}{\|\Delta v_i\|_i} = 0. \tag{2.33}$$

Then:

$$\lim_{\|\Delta v\|_{X_1\oplus X_2} \to 0} \frac{\sum_{i=1}^{2} (f_i(v_i + \Delta v_i) - f_i(v_i) - \langle g_i(v_i), \Delta v_i\rangle_i)}{\|\Delta v\|_{X_1\oplus X_2}} =$$
$$= \sum_{i=1}^{2} \lim_{\|\Delta v_i\|_i \to 0} \frac{\|\Delta v_i\|_i}{\|\Delta v\|_{X_1\oplus X_2}} \frac{f_i(v_i + \Delta v_i) - f_i(v_i) - \langle g_i(v_i), \Delta v_i\rangle_i}{\|\Delta v_i\|_i}.$$

The limits inside the sum are both equal to zero because of the definition of the gradient g_i of f_i and because:

$$\frac{\|\Delta v_i\|_i}{\|\Delta v\|_{X_1\oplus X_2}} = \frac{\Delta v_i}{\sqrt{\|\Delta v_1\|^2 + \|\Delta v_2\|^2}} \leq 1 \text{ for any } \|\Delta v_i\|_i \neq 0.$$

So we proved that:

$$\lim_{\|\Delta v\|_{X_1\oplus X_2} \to 0} \frac{f(v + \Delta v) - f(v) - \sum_{i=1}^{2} \langle g_i(v_i), \Delta v_i\rangle_i}{\|\Delta v\|_{X_1\oplus X_2}} = 0. \tag{2.34}$$

This equality is exactly the definition of the gradient, hence:

$$\langle u, g(v)\rangle_{X_1\oplus X_2} = \langle u_1, g_1(v_1)\rangle_1 + \langle u_2, g_2(v_2)\rangle_2.$$

The expression for the Hessian $H(v)$ can be proved in a similar manner. It is

easy to check that the inverse $H^{-1}(v)$ is defined by:

$$\langle u, H^{-1}(v)z\rangle_{X_1\oplus X_2} = \langle u_1, H_1^{-1}(u_1)z_1\rangle_1 + \langle u_2, H_2^{-1}(v_2)z_2\rangle_2. \tag{2.35}$$

□

Theorem 7 *Let $f_1 : X_1 \to I\!R$ and $f_2 : X_2 \to I\!R$ be two intrinsically self-conjugate barrier functionals. Then $f := f_1 \oplus f_2$ is also an intrinsically self-conjugate barrier functional with $\theta_f = \theta_1 + \theta_2$.*

Proof: We will start by proving that f is self-concordant. The domain of f is the $\oplus$ - sum of the domains of f_1 and f_2, $D_f = D_{f_1} \oplus D_{f_2}$.

First, we have to check that

$$\forall x \in D_f, B_x(x, 1) \subseteq D_f.$$

Because $x \in D_f$, it can be decomposed as $x = x_1 \oplus x_2$, with $x_1 \in D_{f_1}$ and $x_2 \in D_{f_2}$. Both f_1 and f_2 are self-concordant functionals so

$$B_{x_1}(x_1, 1) \subseteq D_{f_1},$$
$$B_{x_2}(x_2, 1) \subseteq D_{f_2}.$$

For any arbitrary point $y \in B_x(x, 1)$ (with $y = y_1 \oplus y_2$),

$$\begin{aligned} \|y - x\|_x^2 &= \langle x - y, H(x)(x - y)\rangle \\ &= \langle x_1 - y_1, H_1(x_1)(x_1 - y_1)\rangle_1 + \langle x_2 - y_2, H_2(x_2)(x_2 - y_2)\rangle_2 \\ &= \|x_1 - y_1\|_{x_1}^2 + \|x_2 - y_2\|_{x_2}^2. \end{aligned}$$

This immediately implies that $y_1 \in B_{x_1}(x_1, 1) \subseteq D_{f_1}$ and $y_2 \in B_{x_2}(x_2, 1) \subseteq D_{f_2}$. Then,

$$y = y_1 \oplus y_2 \in D_{f_1} \oplus D_{f_2} = D_f.$$

Hence, we proved:

$$\forall x \in D_f,\ B_x(x,1) \subseteq D_f.$$

To complete the proof that $f \in SC$, we need to show that for any vector $y \in B_x(x,1)$ (hence its components $y_i \in B_{x_i}(x_i,1)$) and for all nonzero vectors v:

$$1 - \|y-x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1-\|y-x\|_x}.$$

Let v_1 and v_2 be the components of v, $v = v_1 \oplus v_2$. Because $v \neq 0$, then at least one of v_1, v_2 is nonzero. Suppose both are nonzero (if one of them is zero, the proof follows almost identically). For both f_1 and f_2 the previous inequality holds true:

$$1 - \|y_i - x_i\|_{x_i} \leq \frac{\|v_i\|_{y_i}}{\|v_i\|_{x_i}} \leq \frac{1}{1-\|y_i-x_i\|_{x_i}}.$$

with $i = 1, 2$.

Also,

$$1 - \|y-x\|_x = 1 - \sqrt{\|y_1 - x_1\|_{x_1}^2 + \|y_2 - x_2\|_{x_2}^2}.$$

For simplicity, let $A = \|y_1 - x_1\|_{x_1}$, $B = \|y_2 - x_2\|_{x_2}$ and $C = \|y-x\|_x$. With these notations:

$$1 - C = 1 - \sqrt{A^2 + B^2}.$$

So

$$\begin{aligned}
\frac{\|v\|_y^2}{\|v\|_x^2} = \frac{\|v_1\|_{y_1}^2 + \|v_2\|_{y_2}^2}{\|v_1\|_{x_1}^2 + \|v_2\|_{y_2}^2} &\geq \frac{(1-A)^2\|v_1\|_{x_1}^2 + (1-B)^2\|v_2\|_{x_2}^2}{\|v_1\|_{x_1}^2 + \|v_2\|_{y_2}^2} \\
&\geq \frac{(1-C)^2\|v_1\|_{x_1}^2 + (1-C)^2\|v_2\|_{x_2}^2}{\|v_1\|_{x_1}^2 + \|v_2\|_{y_2}^2} \\
&= (1-C)^2.
\end{aligned}$$

We proved the leftmost inequality. For the rightmost inequality the proof is similar. So $f \in SC$. Moreover, $f \in SCB$ with $\theta_f \leq \theta_{f_1} + \theta_{f_2}$ because:

$$\|g_x(x)\|_x^2 = \|g_{x_1}(x_1)\|_{x_1}^2 + \|g_{x_2}(x_2)\|_{x_2}^2$$

and by definition $\theta_f = \sup_{x \in D_f} \|g_x(x)\|_x^2$.

For the last part, for any $t > 0$:

$$\begin{aligned} 0 = f(tx) - f_1(tx_1) - f_2(tx_2) &= f(x) - f_1(x_1) - f_2(x_2) + (\theta_1 + \theta_2 - \theta_f) \ln t \\ &= (\theta_1 + \theta_2 - \theta_f) \ln t. \end{aligned}$$

So $\theta_f = \theta_1 + \theta_2$. □

All the previous definitions are consistent even if we use local inner products scaled by $e_i \in K_i$.

CHAPTER 3
Analysis of the Algorithm

3.1 Preliminaries

In this chapter we will analyze the performance of the algorithm. First we will analyze how new cuts are added to the problem and how the current point is moved back in the feasible region. We will see that in order to recover feasibility we need to use scaled steps. Once the point becomes feasible, we will prove that we can get to a θ - analytic center by taking two types of Newton steps.

The progress is measured here using potential functionals. First we will use Nesterov-Todd steps to move the current point closer to the exact analytic center. We will see that each such step decreases the potential by a constant value. Once, "close enough" (the exact meaning will be introduced later) we use a different type of steps to get to a θ - analytic center. Using these new type of steps the convergence towards a θ - analytic center becomes exponential.

In the second part of this chapter we will obtain an upper bound for the total number of θ - analytic centers that need to be generated in order to get feasible in Γ. The analysis employed here will follow the approach used by Yu in [25] in analyzing an analytic center cutting plane method for the LP case. Although the main steps are the same, the general character of our problem will require a different use of the initial assumptions we made about the problem.

We will prove that the algorithm stops with a solution in no more than $O^*(\frac{mP^3\Theta^3}{\varepsilon^2\Lambda^2})$ (here O^* means that terms of low order are ignored) steps, where Θ and Λ are parameters that characterize the problem and P is the largest number of cuts that are added at a given time. This result is similar to the ones obtained in less general cases.

The idea behind our study is quite simple. As the algorithm proceeds, a sequence of sets Ω_i is generated. We will use the exact analytic centers s_i^c of these sets. The main steps are:

- Get an upper bound UB_i for $f_i^*(s_i^c)$, for any i

- Compare two consecutive f_i^* at the corresponding AC s_i^c:

$$f_{i+1}^*(s_{i+1}^c) \geq f_i^*(s_i^c) + LB_i$$

- After k steps :

$$UB_k \geq f_0^*(s_0^c) + \sum_{i=0}^{k-1} LB_i$$

- We prove that $UB_k \to \infty$ slower than $\sum_{i=0}^{k-1} LB_i$ does
- The algorithm stops as soon as

$$UB_k < f_0^*(s_0^c) + \sum_{i=0}^{k-1} LB_i$$

3.2 The Algorithm

In order to get a point in Γ we generate a sequence of outer-approximations Ω_i. The algorithm stops as soon as the θ - analytic center of one set Ω_i is in Γ.

The algorithm starts with the initial set

$$\Omega_0 := \{y \in Y : -\tilde{c}_0 \preceq_{\tilde{K}_0} y \preceq_{\tilde{K}_0} \tilde{c}_0 \text{ with } \tilde{c}_0 \in int(\tilde{K}_0)\}$$

as the first outer-approximation of Γ. The cone $\tilde{K}_0$ is a self-scaled cone in $(\tilde{X}_0, \langle\cdot,\cdot\rangle_0)$ - an m - dimensional Hilbert space.

Let $X_0 := \tilde{X}_0 \oplus \tilde{X}_0$, $K_0 := \tilde{K}_0 \oplus \tilde{K}_0$ and let f_0 be the intrinsically self-conjugate barrier functional corresponding to K_0, $f_0 : \text{int}(K_0) \to I\!R$. The set Ω_0 can be described by

$$\Omega_0 := \{y \in Y : A_0^* y + s = c_0 \text{ with } s \in K_0\}.$$

Here, A_0 is a linear operator defined on X_0, $A_0 : X_0 \to Y$ such that, its adjoint $A_0^* : Y \to X_0$ describes Ω_0 (i.e. $A_0 := I_m \oplus (-I_m)$).

Let $\tilde{e}_0 \in \text{int}(\tilde{K}_0)$ be an arbitrary point chosen such that the Hessian $\tilde{H}_0$ of $\tilde{f}_0$ has unit norm at $\tilde{e}_0$: $\|\tilde{H}_0(\tilde{e}_0)\| = 1$. Let's take $e_0 := \tilde{e_0} \oplus \tilde{e}_0$. Then $e_0 \in K_0$ and $\|H_0(e_0)\| = 1$ too.

Now, we change the inner product to be the one induced by e_0. Because of this change, the adjoint of the operator A_0 changes from A_0^* to $H_0(e_0)^{-1}A_0^*$. This is because

$$\langle u, A_0 v\rangle_Y = \langle A_0^* u, v\rangle_{X_0} = \langle H_0(e_0)^{-1} A_0^* u, v\rangle_{e_0}.$$

In order not to complicate the notation, we will define A_0^* to be the adjoint of A_0 in the new inner product. Also, we will use c_0 instead of the scaled vector $H_0(e_0)^{-1}c_0$.

Using this new notation, the set Ω_0 has the same description as before:

$$\Omega_0 = \{y \in Y : A_0^* y + s = c_0, \text{ with } s \in K_0\}.$$

Let (x_0, y_0, s_0) be the θ - analytic center corresponding to f_0. In order to obtain this point, we can take a sequence of primal-dual Newton steps, starting at the strictly feasible point $(e_0, 0, c_0) \in K_0 \times \Omega_0 \times K_0$. Note that e_0 and c_0 are strictly interior to K_0. Also, the origin is a point strictly feasible in Ω_0.

Once at y_0, the oracle is called. If $y_0 \in \Gamma$ the oracle returns y_0 and the algorithm stops with the solution to our problem. If $y_0 \notin \Gamma$, the oracle returns p_1 - central cuts. That is, the oracle returns a p_1-dimensional Hilbert space $(X_1, \langle\cdot,\cdot\rangle_1)$ together with a self-scaled cone K_1, the corresponding intrinsically self-conjugate barrier functional $f_1 : K_1 \to I\!R$ and a linear injective operator $A_1 : X_1 \to Y$ such that

$$\Gamma \subseteq \{y \in Y : A_1^* y + s = A_1^* y_0 \text{ with } s \in K_1\}.$$

The equality $A_1^* y + s = A_1^* y_0$ defines a central cut. It is called central because the point $(y, s) := (y_0, 0)$ lies on the cut with s being the vertex of the cone K_1.

We change the inner product on the space X_1 with a local one induced by a vector $e_1 \in \text{int}(K_1)$ chosen arbitrarily such that the norm of the Hessian of f_1 computed

in the original norm at e_1 is unitary. Also we change the functional f_1 by adding a constant such that the modified functional:

$$f_1(\sqrt{\frac{p_1}{\theta_1}}e_1) = 0.$$

(as already discussed in *Section* 2.4).

Now we build the new instance of the algorithm. First, let $\bar{X}_1 := X_0 \oplus X_1$ be an $(2m + p_1)$ - dimensional Hilbert space with the inner product induced by the inner products of X_0 and X_1. Let $\bar{K}_1 := K_0 \oplus K_1$ be the new self-scaled cone with the corresponding intrinsically self-conjugate barrier functional $\bar{f}_1 := f_0 \oplus f_1$. After adding the new cuts Ω_0 becomes

$$\Omega_1 := \Omega_0 \cap \{y \in Y : A_1^* y + s = A_1^* y_0 \text{ with } s \in K_1\}.$$

For the new instance of the algorithm, the old θ - analytic center (x_0, y_0, s_0) becomes $(x_0 \oplus 0_{p_1}, y_0, s_0 \oplus 0_{p_1})$ (with 0_{p_1} being the zero vector in X_1).

The point y_0 lies on the boundary of the new set Ω_1. First we will take a step to recover strict feasibility for this point. After that we generate a sequence of Newton steps that will take the point to (x_1, y_1, s_1), the θ - analytic center of the new domain Ω_1.

At this point we call the oracle again. If $y_1 \in \Gamma$, we stop with the solution to our problem. If $y_1 \notin \Gamma$, the oracle returns p_2 central cuts that are added to the old instance of the algorithm, generating a new set Ω_2. Then the algorithm proceeds as before.

We will prove that the algorithm must stop with a solution after a sufficiently large number of cuts has been added.

After i iterations, the i-th instance of the algorithm is described by a Hilbert space $\bar{X}_i = \bigoplus_{j=0}^{i} X_j$ together with a self-scaled cone $\bar{K}_i = \bigoplus_{j=0}^{i} K_j$, the domain of an intrinsically self-conjugate barrier functional $\bar{f}_i = \bigoplus_{j=0}^{i} f_j$. The current set Ω_i is described by the linear operator $\bar{A}_i : \bar{X}_i \to Y$, with $\bar{A}_i = \bigoplus_{j=0}^{i} A_j$. All linear operators $A_j : X_j \to Y$, $j \geq 1$, are injective and the inner products considered

in the p_j-dimensional Hilbert spaces X_j are the ones induced by fixed elements $e_j \in \text{int}(K_j)$. These vectors e_j are strictly interior to the respective cones K_j and $\|H_j(e_j)^{-1}\| = 1, \forall j \geq 1$ (here H_j is the Hessian of f_j computed in the original norm of X_j, not in the local norm induced by e_j).

3.3 The Recovery of Feasibility

In this section we will study the impact of the central cuts added through an θ - analytic center and how feasibility can be restored.

Consider an instance of the algorithm described by an intrinsically self-conjugate functional f_1 defined on a Hilbert space $(X_1, \langle \cdot, \cdot \rangle_1)$ with the corresponding full-dimensional self-scaled cone K_1 pointed at zero (We consider here the case $i = 1$ for notational convenience. This analysis applies to any stage i of the algorithm.). The outer-approximation of the domain of interest Γ in this instance is

$$\Omega_1 = \{y \in Y : A_1^* y + s = c_1, s \in K_1\},$$

with $A_1 : X_1 \to Y$ a linear operator. Let (x_1, y_1, s_1) be the θ - analytic center for $\mathcal{F}_P$, $\mathcal{F}_D$. So its components must verify:

$$A_1 x_1 = 0, \tag{3.1}$$

$$A_1^* y_1 + s_1 = c_1, \tag{3.2}$$

$$x_1, s_1 \in K_1 \text{ and } y_1 \in Y. \tag{3.3}$$

We add p central cuts at this point: $A_2^* y + s = c_2$ with $A_2^* y_1 = c_2$. The operator A_2 is defined on a p - dimensional Hilbert space $(X_2, \langle \cdot, \cdot \rangle_2)$. We assume that A_2 is injective and linear.

The outer-approximation domain Ω_1 becomes

$$\Omega_2 := \Omega_1 \cap \{y \in Y : A_2^* y + s = c_2, s \in K_2\}.$$

K_2 is a self-scaled cone in X_2 and let $f_2 : X_2 \to I\!R$ be the corresponding intrinsically self-conjugate functional.

After adding the cuts, the primal and dual feasible sets $\mathcal{F}_P$ and $\mathcal{F}_D$ are changed:

$$\mathcal{F}_P := \{x \oplus \beta : A_1 x + A_2 \beta = 0 \text{ with } x \in K_1, \beta \in K_2\}$$

and

$$\mathcal{F}_D := \{s \oplus \gamma : A_1^* y + s = c_1, A_2^* y + \gamma = c_2 \text{ with } s \in K_1, \gamma \in K_2, y \in Y\}.$$

Let $f := f_1 \oplus f_2$, $X := X_1 \oplus X_2$. After adding the cuts the old point (x_1, y_1, s_1) becomes (x_2, y_2, s_2):

$$x_2 = x_1 \oplus \beta, y_2 = y_1, s_2 = s_1 \oplus \gamma,$$

with y_2 on the boundary of the new domain Ω_2. At this new point, $\beta = 0$ and $\gamma = 0$ hence both f and f^* are infinitely large. One step to recover feasibility is needed. Let this step be: $\Delta x \oplus \beta$, Δy and $\Delta s \oplus \gamma$. The new point must be feasible in $\mathcal{F}_P$ and $\mathcal{F}_D$ so:

$$A_1(x_1 + \Delta x) + A_2\beta = 0, \tag{3.4}$$
$$A_1^*(y_1 + \Delta y) + s_1 + \Delta s = c_1, \tag{3.5}$$
$$A_2^*(y_1 + \Delta y) + \gamma = c_2, \tag{3.6}$$

with $x_1, x_1 + \Delta x, s_1, s_1 + \Delta s \in K_1$ and $\beta, \gamma \in K_2$. So, in order to get back feasibility we need to have:

$$A_1 \Delta x + A_2 \beta = 0, \tag{3.7}$$
$$A_1^* \Delta y + \Delta s = 0, \tag{3.8}$$
$$A_2^* \Delta y + \gamma = 0. \tag{3.9}$$

Right after adding the central cuts, the current point y_2 is sitting on the boundary of Ω_2 (both $\beta = 0$ and $\gamma = 0$). Because of this the values of both f_2 and f_2^* were

equal to infinity. So in moving away from the boundary of Ω_2 we should try to minimize as much as possible the contribution of β and γ to the potential functions. One way of doing this is to set-up the next problems:

$$\begin{aligned} \min \quad & f_2(\beta) \\ \text{subject to} \quad & A_1\Delta x + A_2\beta = 0 \\ & \beta \in K_2 \end{aligned}$$

and

$$\begin{aligned} \min \quad & f_2^*(\gamma) \\ \text{subject to} \quad & A_2^*\Delta y + \gamma = 0 \\ & \gamma \in K_2 \end{aligned}$$

These two formulations do not describe completely our problem. What is needed is a constraint that insures that $x_1 + \Delta x$ and $s_1 + \Delta s$ stay feasible in K_1 too. Using the analysis of the Dikin's Ellipsoids we have already made, it is enough to add $\|\Delta x\|_{H_1(s_1)^{-1}} \leq 1 - \frac{\theta}{\sqrt{\theta_f}}$ and $\|\Delta s\|_{s_1} \leq 1$ to keep $x_1 + \Delta x$ and $s_1 + \Delta s$ feasible in K_1. Next we will analyze the problems using $\|\Delta x\|_{H_1(s_1)^{-1}} \leq 1$. We do this to keep the analysis clear. Later we will scale the steps by $\alpha < 1 - \frac{\theta}{\sqrt{\theta_f}}$ so the feasibility will be preserved.

So a good choice is to take β and γ to be the solutions to the following problems:

$$\begin{aligned} \min \quad & f_2(\beta) \\ \text{subject to} \quad & A_1\Delta x + A_2\beta = 0, \qquad (P_2) \\ & \|\Delta x\|_{H_1(s_1)^{-1}} \leq 1, \\ & \beta \in K_2 \end{aligned}$$

and

$$\begin{aligned} \min \quad & f_2^*(\gamma) \\ \text{subject to} \quad & A_2^*\Delta y + \gamma = 0, \qquad (D_2) \\ & \|\Delta s\|_{s_1} \leq 1, \\ & \gamma \in K_2. \end{aligned}$$

These two problems are well posed. The feasible regions are not empty because A_1 is surjective, A_2 is injective and the equality constraints are homogeneous. The

objectives are strongly convex functionals so, if the minimum exists, it is unique. The cone K_2 is the domain for both f_2 and f_2^*. These two functionals are self-concordant so they are infinitely large on the boundary of their domains. The second constraint in each problem ensure that the feasible sets don't contain rays. So both problems have an unique optimal value.

This approach is similar to the one proposed for the linear programming case by Goffin and Vial in [5]. It is a generalization of the approach used by Mitchell and Todd in [9] for the case $p_2 = 1$ (only one cut is added at each iteration).

Now, let's analyze (P_2). The KKT conditions are:

$$g_2(\beta) + A_2^*\lambda = 0, \tag{3.10}$$

$$A_1^*\lambda + \nu H_1(s_1)^{-1}\Delta x = 0, \tag{3.11}$$

$$\nu(1 - \langle \Delta x, H_1(s_1)^{-1}\Delta x\rangle_1) = 0, \tag{3.12}$$

$$A_1\Delta x + A_2\beta = 0. \tag{3.13}$$

If we take

$$\Delta x = -H_1(s_1)A_1^*(A_1H_1(s_1)A_1^*)^{-1}A_2\beta, \tag{3.14}$$

$$\nu = \theta_{f_2}, \tag{3.15}$$

$$\lambda = \theta_{f_2}(A_1H_1(s_1)A_1^*)^{-1}A_2\beta \tag{3.16}$$

both equations (3.11) and (3.13) are verified.

For β we use the approach used by Goffin and Vial in [5] and we will take it to be the solution to the next problem:

$$\min_{\beta\in K_2} \frac{\theta_{f_2}}{2}\langle \beta, V\beta\rangle_2 + f_2(\beta) \tag{3.17}$$

$$\text{with } V = A_2^*(A_1H_1(s_1)A_1^*)^{-1}A_2. \tag{3.18}$$

The optimality condition for this minimization problem is given by:

$$\theta_{f_2}V\beta + g_2(\beta) = 0. \tag{3.19}$$

It is easy to verify that the equation (3.10) holds true for β solution for problem (3.17). For equation (3.12) it is enough to note that

$$\|\Delta x\|^2_{H_1(s_1)^{-1}} = \langle \Delta x, H_1(s_1)^{-1}\Delta x\rangle_1 = \langle \beta, V\beta\rangle_2 = -\frac{1}{\theta_{f_2}}\langle \beta, g_2(\beta)\rangle_2 = 1. \quad (3.20)$$

Now, let's consider the second problem (D_2). The optimality conditions are:

$$\begin{aligned} g_2^*(\gamma) + \mu &= 0, && (3.21)\\ A_2\mu + \nu A_1 H_1(s_1) A_1^*\Delta y &= 0,\ \nu \geq 0, && (3.22)\\ A_2^*\Delta y + \gamma &= 0, && (3.23)\\ \nu(1 - \langle \Delta y, A_1 H_1(s_1) A_1^*\Delta y\rangle_Y) &= 0. && (3.24) \end{aligned}$$

The solution to this problem is given by:

$$\begin{aligned} \Delta y &= -(A_1 H_1(s_1) A_1^*)^{-1} A_2\beta, && (3.25)\\ \gamma &= V\beta = A_2^*(A_1 H_1(s_1) A_1^*)^{-1} A_2\beta, && (3.26)\\ \mu &= \theta_{f_2}\beta, && (3.27)\\ \nu &= \theta_{f_2}. && (3.28) \end{aligned}$$

Here β is the solution of problem (3.17). The equations (3.23) and (3.22) are obviously satisfied. For equation (3.24):

$$\langle \Delta y, A_1 H_1(s_1) A_1^*\Delta y\rangle_Y = \langle \beta, A_2^*(A_1 H_1(s_1) A_1^*)^{-1} A_2\beta\rangle_2 = \langle \beta, V\beta\rangle_2 = 1.$$

Finally, for equation (3.21) it is enough to notice that $-g_2(-g_2(\beta))$ is equal to both β (because $-g_2$ is an involution, as can be seen from *Theorem* 5) and $-g_2(\theta_{f_2}\gamma)$ (as given by equation (3.19)). Using the fact that f_2 is logarithmically homogeneous,

the conclusion follows immediately:

$$g_2(\gamma) = \theta_{f_2} g_2(\theta_{f_2}\gamma) = \theta_{f_2} g_2(-g_2(\beta)) = -\theta_{f_2}\beta = -\mu. \tag{3.29}$$

Instead of full steps Δx, Δs, some scaled steps $\alpha\Delta x$, $\alpha\Delta s$ are taken. The next lemma gives a characterization of such scaled steps.

Lemma 15 *Let Δx and Δs be the steps considered in the problems (P_2) and (D_2). For any $\alpha < (1 - \frac{\theta}{\sqrt{\theta_f}})\zeta$ with $0 < \zeta < 1$:*

$$\|\alpha\Delta x\|_{x_1} < \zeta \text{ and } \|\alpha\Delta s\|_{s_1} < \zeta.$$

Proof: Here we will use *Lemma* 12.

$$\|\alpha\Delta x\|_{x_1} = \alpha\|\Delta x\|_{x_1} \leq \alpha \frac{1}{1 - \frac{\theta}{\sqrt{\theta_f}}} \|\Delta x\|_{H(s_1)^{-1}} < \zeta.$$

The second inequality is immediate:

$$\|\alpha\Delta s\|_{s_1} = \alpha\|\Delta s\|_{s_1} \leq \alpha < \zeta.$$

□

We have that $g_2(\beta) = -\theta_{f_2}\gamma$, with $\beta, \gamma \in K_2$. So we can use *Lemma* 11:

$$f_2(\beta) + f_2^*(\theta_{f_2}\gamma) = -\theta_{f_2}. \tag{3.30}$$

The fact that f_2 and f_2^* are logarithmically homogeneous implies:

$$\begin{aligned} f_2(\alpha\beta) + f_2^*(\alpha\gamma) &= f_2(\beta) + f_2^*(\gamma) - 2\theta_{f_2}\ln\alpha & (3.31)\\ &= -\theta_{f_2} - 2\theta_{f_2}\ln\alpha + f_2^*(\gamma) - f_2^*(\theta_{f_2}\gamma). & (3.32) \end{aligned}$$

So we proved that:

$$f_2(\alpha\beta) + f_2^*(\alpha\gamma) = -\theta_{f_2} - 2\theta_{f_2}\ln\alpha + \theta_{f_2}\ln\theta_{f_2}. \tag{3.33}$$

This equality provides a measure of the influence the added cut has over the self-concordant barrier functional.

3.4 Potentials

In analyzing the complexity of the algorithm (for both local and global convergence) we will make use of primal-dual potentials. The way potentials change from one analytic center to the next one will give us a measure for the total number of cuts that can be introduced before the algorithm stops with a solution. We will also use potential functionals in finding the number of steps required to get to the θ - analytic center after new cuts are added in the problem.

Definition 11 *For an instance of the algorithm described by the functional f, the vector c and the linear operator A, we define the primal-dual potential to be:*

$$\phi_{PD}(x,s) = \langle c, x\rangle + f(x) + f^*(s) \text{ for any } x, s \in K.$$

It is customary to call $\langle c, x\rangle + f(x)$ the primal potential and $f^*(s)$ the dual potential.

Let (x_1, y_1, s_1) be the current θ - analytic center with the corresponding primal-dual potential:

$$\phi_1 := \langle c_1, x_1\rangle_1 + f_1(x_1) + f_1^*(s_1).$$

After adding the cuts described by f_2, A_2 and c_2 we take a scaled step to get back into the feasible region. At this new point, the primal-dual potential is:

$$\begin{aligned}\phi_{new} := \langle c_1, x_1 \;&+\; \alpha\Delta x\rangle_1 + \langle c_2, \alpha\beta\rangle_2 + f_1(x_1 + \alpha\Delta x) + f_2(\alpha\beta)\\ &+\; f_1^*(s_1 + \alpha\Delta s) + f_2^*(\alpha\gamma).\end{aligned}$$

Using equation (3.33) the new potential can be written as

$$\phi_{new} = \phi_1 + \theta_{f_2} \ln \frac{\theta_{f_2}}{\alpha^2} - \theta_{f_2} + \alpha(\langle c_1, \Delta x\rangle_1 + \langle c_2, \beta\rangle_2) + F, \tag{3.34}$$

with

$$F = f_1(x_1 + \alpha\Delta x) \quad - \quad f_1(x_1) + f_1^*(s_1 + \alpha\Delta s) - f_1^*(s_1). \tag{3.35}$$

Because the cuts are central: $A_2^* y_1 = c_2$, hence

$$\langle \beta, c_2\rangle_2 = \langle A_2\beta, y_1\rangle_Y = -\langle A_1\Delta x, y_1\rangle_Y = -\langle \Delta x, A_1^* y_1\rangle_1.$$

Therefore,

$$\langle c_1, \Delta x\rangle_1 + \langle c_2, \beta\rangle_2 = \langle c_1 - A_1^* y_1, \Delta x\rangle_1 = \langle s_1, \Delta x\rangle_1.$$

So, finally:

$$\phi_{new} = \phi_1 + \alpha\langle s_1, \Delta x\rangle_1 + \theta_{f_2} \ln \frac{\theta_{f_2}}{\alpha^2} - \theta_{f_2} + F.$$

Now let's evaluate $F + \alpha\langle s_1, \Delta x\rangle_1$. Let's start with $\alpha\langle s_1, \Delta x\rangle_1 + f_1(x_1 + \alpha\Delta x) - f_1(x_1)$. Note that the recovery step is scaled by $\alpha < (1 - \frac{\theta}{\sqrt{\theta_f}})\zeta$ so we can use the inequality (2.8):

$$\begin{aligned} \alpha\langle s_1, \Delta x\rangle_1 \quad &+ \quad f_1(x_1 + \alpha\Delta x) - f_1(x_1) \\ &\leq \quad \alpha\langle s_1, \Delta x\rangle_1 + \alpha\langle g_1(x_1), \Delta x\rangle_1 + \frac{1}{2}\|\alpha\Delta x\|_x^2 + \frac{\|\alpha\Delta x\|_x^3}{3(1 - \|\alpha\Delta x\|_x)}. \end{aligned}$$

Now:

$$\langle s_1 + g_1(x_1), \Delta x\rangle_1 \quad = \quad \langle s_1 + g_1(x_1), H_1(x_1)^{-\frac{1}{2}} H_1(x_1)^{\frac{1}{2}} \Delta x\rangle_1$$

$$\begin{aligned} &\leq \|H_1(x_1)^{-\frac{1}{2}}(s_1 + g_1(x_1))\|_1 \|\Delta x\|_{x_1} \\ &= \|s_1 + g_1(x_1)\|_{H_1(x_1)^{-1}} \|\Delta x\|_{x_1} \\ &= \|s_1 + g_1(x_1)\|_{-g_1(x_1)} \|\Delta x\|_{x_1} \\ &\leq \theta \|\Delta x\|_{x_1}. \end{aligned}$$

Here, we used *Lemma* 6 and the fact that $H(x)^{-1} = H(-g(x))$ (see *Theorem* 5). So:

$$\begin{aligned} \alpha\langle s_1, \Delta x\rangle_1 \quad + \quad & f_1(x_1 + \alpha\Delta x) - f_1(x_1) \leq \\ \leq \quad & \theta\|\alpha\Delta x\|_{x_1} + \frac{1}{2}\|\alpha\Delta x\|_{x_1}^2 + \frac{\|\alpha\Delta x\|_{x_1}^3}{3(1 - \|\alpha\Delta x\|_{x_1})}. \end{aligned}$$

Next we use the fact that the function $\theta x + \frac{1}{2}x^2 + \frac{x^3}{3(1-x)}$ is increasing on the open interval $(0, 1)$ and the recovery step is scaled to satisfy $\|\alpha\Delta x\|_{x_1} \leq \zeta < 1$. This implies that

$$\alpha\langle s_1, \Delta x\rangle_1 + f_1(x_1 + \alpha\Delta x) - f_1(x_1) \leq \theta\zeta + \frac{1}{2}\zeta^2 + \frac{\zeta^3}{3(1-\zeta)}. \tag{3.36}$$

Now let's consider the second part of F: $f_1^*(s_1 + \alpha\Delta s) - f_1^*(s_1)$.
Because $\|\alpha\Delta s\|_{s_1} < \zeta < 1$ we can use again inequality (2.8). So:

$$f_1^*(s_1 + \alpha\Delta s) - f_1^*(s_1) \leq \langle g_1(s_1), \alpha\Delta s\rangle_1 + \frac{1}{2}\|\alpha\Delta s\|_{s_1}^2 + \frac{\|\alpha\Delta s\|_{s_1}^3}{3(1 - \|\alpha\Delta s\|_{s_1})}.$$

Aside:
Because x_1 is in the nullspace of A_1 and Δs is in the range of A_1^*,

$$\langle x_1, \Delta s\rangle_1 = 0.$$

So

$$\begin{aligned} \langle g_1(s_1), \Delta s\rangle_1 &= \langle x_1 + g_1(s_1), \Delta s\rangle_1 \\ &= \langle H_1(s_1)^{-\frac{1}{2}}(x_1 + g_1(s_1)), H_1(s_1)^{\frac{1}{2}}\Delta s\rangle_1 \\ &\leq \|H_1(s_1)^{-\frac{1}{2}}(x_1 + g_1(s_1))\|_1 \|H_1(s_1)^{\frac{1}{2}}\Delta s\|_1 \end{aligned}$$

$$\begin{aligned} &= \|x_1 + g_1(s_1)\|_{H_1(s_1)^{-1}} \|\Delta s\|_{s_1} \\ &= \|x_1 + g_1(s_1)\|_{-g_1(s_1)} \|\Delta s\|_{s_1} \leq \theta \|\Delta s\|_{s_1}. \end{aligned}$$

So, using *Lemma* 6 and the fact that $\|\alpha \Delta s\|_{s_1} \leq \zeta < 1$,

$$\begin{aligned} f_1^*(s_1 + \alpha\Delta s) - f_1^*(s_1) &\leq \theta \|\alpha\Delta s\|_{s_1} + \frac{1}{2}\|\alpha\Delta s\|_{s_1}^2 + \frac{\|\alpha\Delta s\|_{s_1}^3}{3(1 - \|\alpha\Delta s\|_{s_1})} \\ &\leq \theta\zeta + \frac{1}{2}\zeta^2 + \frac{\zeta^3}{3(1-\zeta)}. \end{aligned} \tag{3.37}$$

Using inequalities (3.36) and (3.37) we get:

$$\phi_{new} \leq \phi_1 + \theta_{f_2} \ln \frac{\theta_{f_2}}{\alpha^2} - \theta_{f_2} + 2\theta\zeta + \zeta^2 + \frac{2\zeta^3}{3(1-\zeta)}.$$

Theorem 8 *Let (x, y, s) be a θ - analytic center corresponding to an instance of the algorithm described by the functional f, the linear operator A and the vector c. Then,*

$$\phi_{PD}(x, s) := \langle c, x\rangle + f(x) + f^*(s) \leq \frac{\theta^3}{3(1-\theta)} + \frac{\theta^2}{2}. \tag{3.38}$$

Proof: Because (x, y, s) is a θ - analytic center we can use *Lemma* 6 to get

$$\|x + g(s)\|_{-g(s)} \leq \theta.$$

This inequality implies, using (2.2), that $x \in B_{-g(s)}(-g(s), \theta)$. Because $\theta < 1$ we can use *Theorem* 3 to get:

$$\left| f(x) - f(-g(s)) + \langle -g(-g(s)), x + g(s)\rangle - \frac{1}{2}G(x,s)^2 \right| \leq \frac{G(x,s)^3}{3\,(1 - G(x,s))}.$$

where $G(x, s) = \|x + g(s)\|_{-g(s)}$.

Because f is an intrinsically self-conjugate barrier functional we have:

$$f^*(s) = \langle g(s), s\rangle - f(-g(s)) \text{ and } -g(-g(s)) = s.$$

Using these equalities together with the fact that $\langle x, s\rangle = \langle c, x\rangle$ we can write:

$$f(x) + f^*(s) + \langle c, x\rangle \leq \frac{1}{2}G(x,s)^2 + \frac{G(x,s)^3}{3(1-G(x,s))}.$$

The functional $G(x,s)$ is bounded above by θ. Using this together with the fact that the function $\frac{1}{2}x^2 + \frac{x^3}{3(1-x)}$ is increasing for $0 < x < 1$, we get the desired conclusion:

$$\langle c, x\rangle + f(x) + f^*(s) \leq \frac{\theta^3}{3(1-\theta)} + \frac{\theta^2}{2}.$$

□

Now we are ready to compare the value of the dual - potential functionals f^* at two consecutive analytic centers.

Let's consider Ω and $\tilde{\Omega}$ to be two consecutive outer-approximations of Γ. These two sets correspond to two instances of the algorithm described by (f, X, K, A, c) and $(\tilde{f}, \tilde{X}, \tilde{K}, \tilde{A}, \tilde{c})$. The second instance is obtained from the first one by adding central cuts through the θ - analytic center of Ω. Let these cuts be described by: $(\hat{f}, \hat{X}, \hat{K}, \hat{A}, \hat{c})$. So $\tilde{f} = f \oplus \hat{f}$, $\tilde{X} = X \oplus \hat{X}$ and so on. Let $(x_\theta^c, y_\theta^c, s_\theta^c)$ be the θ - analytic center for f. After adding the cuts right through $(x_\theta^c, y_\theta^c, s_\theta^c)$ a scaled step is taken to recover feasibility. Let $(x(\alpha), y(\alpha), s(\alpha))$ be the point right after this step is taken so

$$\begin{aligned} x(\alpha) &= (x_\theta^c + \alpha\Delta x) \oplus (\alpha\beta), \\ y(\alpha) &= y_\theta^c + \alpha\Delta y, \\ s(\alpha) &= (s_\theta^c + \alpha\Delta s) \oplus (\alpha\gamma). \end{aligned}$$

Using all these notations we are ready to prove the following theorem, which gives a bound for the change in the barrier functional evaluated at two consecutive exact analytic centers.

Theorem 9 *Let (x^c, y^c, s^c) and $(\tilde{x}^c, \tilde{y}^c, \tilde{s}^c)$ be two consecutive analytic centers for*

the domains Ω *and* $\tilde{\Omega}$*. Then,*

$$\tilde{f}^*(\tilde{s}^c) \geq f^*(s^c) - \hat{f}(\alpha\beta) - \theta\zeta - \frac{1}{2}\zeta^2 - \frac{\zeta^3}{3(1-\zeta)} - \frac{\theta^3}{3(1-\theta)} - \frac{1}{2}\theta^2. \tag{3.39}$$

Proof: Because $(\tilde{x}^c, \tilde{y}^c, \tilde{s}^c)$ is an exact analytic center for $\tilde{f}$, $\tilde{x}^c$ minimizes the value of $\tilde{f}(x) + \langle x, \tilde{c}\rangle$

$$\tilde{f}(\tilde{x}^c) \leq \tilde{f}(x(\alpha)) + \langle x(\alpha), \tilde{c}\rangle - \langle \tilde{x}^c, \tilde{c}\rangle. \tag{3.40}$$

Lemma 11 gives a connection between the values of $\tilde{f}(\tilde{x}^c)$ and $\tilde{f}^*(\tilde{s}^c)$:

$$\tilde{f}(\tilde{x}^c) + \tilde{f}^*(\tilde{s}^c) = -\theta_{\tilde{f}}.$$

Now let's analyze $\langle \tilde{x}^c, \tilde{c}\rangle$:

$$\langle \tilde{x}^c, \tilde{c}\rangle = \langle \tilde{x}^c, \tilde{A}^*\tilde{y}^c + \tilde{s}^c\rangle = \langle \tilde{x}^c, \tilde{s}^c\rangle = \langle \tilde{x}^c, -\tilde{g}(\tilde{x}^c)\rangle = \theta_{\tilde{f}}.$$

We can rewrite inequality (3.40) as:

$$\tilde{f}^*(\tilde{s}^c) \geq -f(x_\theta^c + \alpha\Delta x) - \hat{f}(\alpha\beta) - \langle x_\theta^c + \alpha\Delta x, c\rangle - \langle \alpha\beta, \hat{c}\rangle.$$

We can use now the bound on $f(x_\theta^c + \alpha\Delta x)$ given by the inequality (3.36). Before doing this let's notice that:

$$\begin{aligned} \langle \hat{c}, \beta\rangle &= \left\langle \hat{A}^* y_\theta^c, \beta\right\rangle = \left\langle y_\theta^c, \hat{A}\beta\right\rangle = -\langle y_\theta^c, A\Delta x\rangle = \\ &= -\langle A^* y_\theta^c, \Delta x\rangle = \langle s_\theta^c, \Delta x\rangle - \langle c, \Delta x\rangle . \end{aligned}$$

So

$$\tilde{f}^*(\tilde{s}^c) \geq -f(x_\theta^c) - \hat{f}(\alpha\beta) - \langle x_\theta^c, c\rangle - \theta\zeta - \frac{1}{2}\zeta^2 - \frac{\zeta^3}{3(1-\zeta)}.$$

In order to get the desired result we have to use *Theorem* 8 and use the fact that $f^*(s_\theta^c) \geq f^*(s^c)$ (this is because (x^c, y^c, s^c) is an exact analytic center). □

The step required to move the point back in the feasible region after the cuts are added depends upon the vector β. This vector is the solution to the minimization problem (3.17). So, using the fact that $\langle \beta, V\beta \rangle = 1$ (from equation (3.20)),

$$\hat{f}(\beta) \leq \hat{f}(\beta') + \frac{\theta_{\hat{f}}}{2}\langle \beta', V\beta' \rangle - \frac{\theta_{\hat{f}}}{2}, \text{ for any } \beta' \in \hat{K} \tag{3.41}$$

with V given by:

$$V = \hat{A}^*(AH(s_1)A^*)^{-1}\hat{A}. \tag{3.42}$$

Taking in account all these observations, the fact that $\hat{f}$ is logarithmically homogeneous and $\alpha < 1$, the previous theorem can be restated as:

$$\tilde{f}^*(\tilde{s}^c) \geq f^*(s^c) - \hat{f}(\beta') - \frac{\theta_{\hat{f}}}{2}\langle \beta', V\beta' \rangle + \frac{\theta_{\hat{f}}}{2} + \theta_{\hat{f}} \ln \alpha - \mathcal{F}(\theta, \zeta) \tag{3.43}$$

with

$$\mathcal{F}(\theta, \zeta) = \theta\zeta + \frac{1}{2}\zeta^2 + \frac{\zeta^3}{3(1-\zeta)} + \frac{\theta^3}{3(1-\theta)} + \frac{1}{2}\theta^2 \tag{3.44}$$

for any $\beta' \in \hat{K}$.

3.5 Complexity on Recovering the θ - Analytic Center

After the current point is moved back in the feasible region obtained from the old one by adding central cuts, a sequence of steps is required to get in the vicinity of the analytic center of the new domain. One way of obtaining such a point is to take some Newton steps. In this section we will prove that one way to achieve this is to use two different sequences of steps. We will use potential functionals in this analysis.

At the beginning, when the point is still far away from the analytic center, the directions used are the Nesterov-Todd directions. These directions where first used in interior-point algorithms in linear programming. Nesterov and Todd generalized them later for the general case of conic programming (see [6], [11] for more details). These directions will ensure that the primal-dual potential decreases by a fixed amount at each iteration. Once close enough to the analytic center, a different sequence of steps will bring the point to an θ - analytic center.

As before, let the primal-dual potential functional be:

$$\Phi(x, s) := \langle x, s\rangle + f(x) + f^*(s).$$

Before defining the Nesterov-Todd direction we will introduce some notations. Let L denote the null space of A (the surjective operator defining the feasible region) and $L^\perp$ the corresponding orthogonal space. Let $P_{L,v}(u)$ be the orthogonal projection of u onto L in the local inner product induced by v.

Let (x, y, s) be the current point with w the corresponding scaling point for the ordered pair (x, s) (i.e. $H(w)x = s$). Such a point is uniquely defined by x and s. Similarly we take w^* to be the scaling point for the ordered pair (s, x) (i.e. $H(w^*)s = x$).

With these notations, the primal and dual Nesterov-Todd directions are given by:

$$d_x := -P_{L,w}(x + g_w(x)), \tag{3.45}$$

$$d_s := -P_{L^\perp,w^*}(s + g_{w^*}(s)). \tag{3.46}$$

Note here that if we use the inner products induced by x and s instead of the ones induced by w and w^*, the Nesterov-Todd directions become the usual Newton directions.

One important property of these directions is that they provide an orthogonal decomposition w.r.t. $\langle\cdot,\cdot\rangle_w$ for $-(x + g_w(x))$ (see [18]) :

$$d_x + H(w)^{-1} d_s = -(x + g_w(x)). \tag{3.47}$$

Using the local inner product induced by w we define for all $\tilde{x}, \tilde{s} \in int(K)$:

$$\Phi_w(\tilde{x}, \tilde{s}) = \langle \tilde{x}, \tilde{s} \rangle_w + f(\tilde{x}) + f^*(\tilde{s}). \tag{3.48}$$

Our goal is to prove that by taking a scaled Nesterov-Todd step, the primal-dual potential functional decreases by a constant value. We will use

$$\phi(t) := \Phi(x + t d_x, s + t d_s) \tag{3.49}$$

to find the scaling parameter t that minimizes the primal-dual potential.

Let's define:

$$\breve{\phi}_w(t) := \Phi_w(x + t\breve{d}_x, x + tH(w)^{-1}\breve{d}_s), \tag{3.50}$$

$$\breve{\phi}(t) := \Phi(x + t\breve{d}_x, x + t\breve{d}_s) \tag{3.51}$$

with $\breve{d}_x$, $\breve{d}_s$, the scaled vectors:

$$(\breve{d}_x, \breve{d}_s) := \frac{1}{\|H_w(x)\|_w^{\frac{1}{2}} \|x + g_w(x)\|_w} (d_x, d_s) \tag{3.52}$$

Using the fact that $f^*(s) = f(s) - (\theta_f + 2f(e))$ (as given in Lemma 3) we can write:

$$\Phi_w(\tilde{x}, H(w)^{-1}\tilde{s}) = \langle \tilde{x}, H(w)^{-1}\tilde{s} \rangle_w + f(\tilde{x}) + f(H(w)^{-1}\tilde{s}) - \theta_f - 2f(e). \tag{3.53}$$

Now for any $x, w \in K$, $f(H(w)x) = f(x) + 2(f(w) - f(e))$ (see [18], formula (3.34)) so

$$f(\tilde{s}) = f(H(w)(H(w)^{-1}\tilde{s})) = f(H(w)^{-1}\tilde{s}) + 2f(w) - 2f(e).$$

Combining all the previous expressions we conclude that:

$$\Phi_w(\tilde{x}, H(w)^{-1}\tilde{s}) = \Phi(\tilde{x}, \tilde{s}) + 2f(e) - 2f(w). \tag{3.54}$$

Now:

$$\begin{aligned}
\breve{\phi}_w(t) &:= \Phi_w(x + t\breve{d}_x, x + tH(w)^{-1}\breve{d}_s) \\
&= \Phi_w(x + t\breve{d}_x, H(w)^{-1}(H(w)x + t\breve{d}_s)) \\
&= \Phi_w(x + t\breve{d}_x, H(w)^{-1}(s + t\breve{d}_s)) \\
&= \Phi(x + t\breve{d}_x, s + t\breve{d}_s) + 2f(e) - 2f(w) \\
&= \breve{\phi}(t) + 2f(e) - 2f(w).
\end{aligned}$$

Using the approach from [18], let's denote:

$$\psi_1(t) := \langle x + t\breve{d}_x, x + tH(w)^{-1}\breve{d}_s\rangle_w, \tag{3.55}$$

$$\psi_2(t) := f(x + t\breve{d}_x), \tag{3.56}$$

$$\psi_3(t) := f(x + tH(w)^{-1}\breve{d}_s) - \theta_f - 2f(e). \tag{3.57}$$

$$\tag{3.58}$$

With these notations:

$$\breve{\phi}_w(t) = \psi_1(t) + \psi_2(t) + \psi_3(t). \tag{3.59}$$

Because $\langle \breve{d}_x, H(w)^{-1}\breve{d}_s\rangle_w = 0$, the first functional $\psi_1(t)$ can be written as

$$\begin{aligned}
\psi_1(t) &= \psi_1(0) + t\langle x, \breve{d}_x + H(w)^{-1}\breve{d}_s\rangle_w \\
&= \psi_1(0) - t\frac{\langle x, x + g_w(x)\rangle_w}{\|H_w(x)\|_w^{\frac{1}{2}}\|x + g_w(x)\|_w}.
\end{aligned}$$

Renegar proved in [18] that:

$$\psi_2(t) \leq \psi_2(0) + t\langle g_w(x), \breve{d}_x\rangle_w + \frac{t^2}{1-t}, \tag{3.60}$$

$$\psi_3(t) \leq \psi_3(0) + t\langle g_w(x), H(w)^{-1}\breve{d}_s\rangle_w + \frac{t^2}{1-t}. \tag{3.61}$$

Using all these relations we can relate $\breve{\phi}_w(t)$ and $\breve{\phi}_w(0)$:

$$\breve{\phi}_w(t) \leq \breve{\phi}_w(0) - t\frac{\|x + g_w(x)\|_w}{\|H_w(x)\|_w^{\frac{1}{2}}} + \frac{2t^2}{1-t}. \tag{3.62}$$

Then, immediately:

$$\breve{\phi}(t) \leq \breve{\phi}(0) - t\frac{\|x + g_w(x)\|_w}{\|H_w(x)\|_w^{\frac{1}{2}}} + \frac{2t^2}{1-t}.$$

or

$$\Phi(x + t\breve{d}_x, s + t\breve{d}_s) \leq \Phi(x,s) - t\frac{\|x + g_w(x)\|_w}{\|H_w(x)\|_w^{\frac{1}{2}}} + \frac{2t^2}{1-t}.$$

Next we will introduce a theorem from [18]:

Theorem 10 *Let K be a self-scaled cone. If $x, w \in int(K)$ then:*

$$\|x + g_w(x)\|_w \geq \max\{\|H_w(x)^{\frac{1}{2}}\|_w, \|H_w(x)^{-\frac{1}{2}}\|_w\} \min\{\frac{1}{5}, \frac{4}{5}\|x - w\|_w\}. \tag{3.63}$$

We are ready now to prove the following theorem:

Theorem 11 *If $\|x - w\|_w \geq \frac{1}{4}$ then:*

$$\Phi(x + t\breve{d}_x, s + t\breve{d}_s) \leq \Phi(x,s) - \frac{1}{250}. \tag{3.64}$$

Proof: The proof is based on the previous analysis and the fact that

$$\min_{0<t<1}(\frac{2t^2}{1-t} - \frac{t}{5}) < \frac{1}{250}.$$

□

We know that, if (x, y, s) is the exact analytic center, then $x = w$. Also, the exact analytic center is the minimizer for the primal-dual potential functional $\Phi(x, s)$.

Theorem 11 says that, as long the point is sufficiently far away from the exact analytic center, the primal-dual potential is guaranteed to decrease by a constant quantity.

Because of the assumption made about the problem, the analytic center exists so, the primal-dual potential functional has a strictly feasible minimizer. This implies that, after a number of scaled Nesterov-Todd steps for the current point (x, y, s), $\|x - w\|_w < \frac{1}{4}$.

As soon as this happens, we will switch from using Nesterov-Todd steps to a new kind of step, suggested in [18]:

$$D_x := 2P_{L,w}(w - x), \tag{3.65}$$

$$D_s := 2P_{L^\perp,w^*}(w^* - s), \tag{3.66}$$

where $P_{L,w}$ is the orthogonal projection onto L (in the local product $\langle \cdot, \cdot \rangle_w$).

The key element here is the following theorem:

Theorem 12 *If at the current point (x, y, s):*

$$\|x - w\|_w < \alpha < \frac{1}{4}$$

then at the new point $(x_+, s_+) := (x + D_x, s + D_s)$:

$$\|s_+ + g(x_+)\|_{-g(x_+)} < (1 + \alpha)\frac{\alpha^2}{1 - \alpha} < \frac{1}{5}. \tag{3.67}$$

If w_+ is the scaling point for the ordered pair (x_+, s_+), then:

$$\|x_+ - w_+\|_{w_+} < \frac{5\alpha^2(1 + \alpha)}{4(1 - \alpha)} < 3\alpha^2 < \frac{1}{5}. \tag{3.68}$$

Proof:

The proof can be found in [18]. Here we will just sketch the main ideas. Similar to the Nesterov-Todd directions, D_x, D_s give an orthogonal decomposition for $w - x$

(in $\langle \cdot, \cdot \rangle_w$):

$$D_x + H(w)^{-1} D_s = 2(w - x). \tag{3.69}$$

This immediately implies that

$$\|w - x_+\|_w = \|w - x\|_w < \alpha. \tag{3.70}$$

Renegar proved in [18] that

$$\|H(w)^{-1} s_+ + g_w(x_+)\|_w \leq \frac{\|x_+ - w\|_w^2}{1 - \|x_+ - w\|_w} \leq \frac{\alpha^2}{1 - \alpha} < \frac{1}{12}. \tag{3.71}$$

On the other hand, using Lemma 4 (we have to change the local norm from $|\cdot|$ to $\|\cdot\|_w$):

$$\|v\|_{-g_w(x_+)} \leq (1 + \|x_+ - w\|_w)\|v\|_w \leq (1 + \alpha)\|v\|_w < \frac{5}{4}\|v\|_w. \tag{3.72}$$

Combining all these inequalities, we get:

$$\begin{aligned} \|s_+ + g(x_+)\|_{-g(x_+)} &= \|H(w)^{-1} s_+ + g_w(x_+)\|_{-g_w(x_+)} & (3.73) \\ &< (1 + \alpha)\frac{\alpha^2}{1 - \alpha} < \frac{1}{5}. & (3.74) \end{aligned}$$

For the second part of the theorem it is enough to use the inequality:

$$\|s + g(x)\|_{-g(x)} \geq \min\{\frac{1}{5}, \frac{4}{5}\|x - w\|_w\} \tag{3.75}$$

that holds for any $x, s \in int(K)$ (see [18]). □

It is easy to see, using Theorem 12 that, as soon

$$\|x - w\|_w < \alpha < \frac{1}{4}, \tag{3.76}$$

the sequence of points generated by using the new steps will converge exponentially

to the exact analytic center.

To be more precise, if we take (x_0, y_0, s_0) to be the first point where

$$\|x - w\|_w < \alpha < \frac{1}{4} \tag{3.77}$$

holds, then after k steps, at the point (x_k, y_k, s_k) with the corresponding scaling point w_k, we have:

$$\|x_k - w_k\|_{w_k} < 3^{2^k-1}\alpha^{2^k}. \tag{3.78}$$

This inequality together with Theorem 12 implies that:

$$\|x_k + g(s_k)\|_{-g(s_k)} < 5 \cdot 3^{2^{k+1}-3} \cdot \alpha^{2^{k+1}}. \tag{3.79}$$

In practical terms, if the parameter θ defining the θ - analytic center is of order 10^{-10}, then we need only 6 such steps to get to a θ - analytic center.

3.6 Implications of the Initial Assumptions

In this section we will derive an upper bound on the value of the dual potential f_i^* evaluated at the analytic center of the set Ω_i. This bound together with the fact that the values of the potential functionals keep increasing as the algorithm proceeds will help us prove that the algorithm will eventually stop with a solution.

Let (x^k, y^k, s^k) be the exact analytic center of Ω_k (the outer-approximation set of Γ after k iterations). This analytic center corresponds to the self-concordant barrier functional $f := f_0 \oplus f_1 \oplus f_2 \oplus \ldots \oplus f_k$ and the cone $K := K_0 \oplus K_1 \oplus K_2 \oplus \ldots \oplus K_k$ that is in the space $X = X_0 \oplus X_1 \oplus \ldots \oplus X_k$. Ω_k is described by the operator $A := A_0 \oplus A_1 \oplus \ldots \oplus A_k$, and the vector $c := c_0 \oplus c_1 \oplus \ldots \oplus c_k$. Our initial assumption that Γ contains a closed ball of radius ε implies that:

$$\mathcal{M} := \{y \in Y : y \in \Omega_k, B_Y(y, \varepsilon) \subset \Omega_k\} \neq \emptyset.$$

Because (x^k, y^k, s^k) is the analytic center of Ω_k, $s^k = c - A^* y^k$ is the minimizer

of f^* over the set of all feasible points. Then,

$$f^*(s^k) \leq f^*(s), \forall s \in \mathcal{M}_s := \{s : s = c - A^*y \text{ with } y \in \mathcal{M}\}.$$

Lemma 16 *Let s be an arbitrary point in the set $\mathcal{M}_s$, with $s_i \in K_i$, the corresponding components. Then the distance (measured using the local inner product) from s_i to the boundary of the cone K_i, for $i \geq 1$, satisfies:*

$$d(s_i, \partial K_i) \geq \varepsilon\sqrt{\lambda_{\min}(A_i^* A_i)}. \tag{3.80}$$

Here $\lambda_{\min}(A_i^* A_i)$ is the minimum eigenvalue of $A_i^* A_i$.
For the initial case $i = 0$:

$$d(s_0, \partial K_0) \geq \sqrt{2}\varepsilon. \tag{3.81}$$

Proof: Let $s \in \mathcal{M}_s$ with the corresponding $y \in \mathcal{M}$ ($s = c - A^*y$). So:

$$y + \varepsilon u \in \Omega_k, \forall u \in Y, \|u\|_Y = 1. \tag{3.82}$$

The point s is strictly interior to the cone K. This implies that each of its components s_i is strictly interior to its corresponding cone K_i.

Then

$$\exists s_\varepsilon \in K_i \text{ such that } A_i^*(y + \varepsilon u) + s_\varepsilon = c_i. \tag{3.83}$$

At the same time:

$$s_i := c_i - A_i^* y \in K_i.$$

Using the last two relations we conclude that:

$$s_\varepsilon = s_i - \varepsilon A_i^* u \text{ is feasible }, \forall u \in Y, ||u||_Y = 1.$$

Our goal is to get an estimate for the distance between s_i and the boundary of K_i. Two cases arise, one for $i = 0$ and one for $i \geq 1$. The difference between this two cases is that A_i is injective only for $i \geq 1$. However, for $i = 0$ the operator A_0 is the $\oplus$ - sum of two bijective operators I and $-I$. So, this case can be treated the same way as the general case if we are using the components of A_0.

Now let's consider the case $i \geq 1$. Let v be a vector parallel to the direction which projects s_i onto ∂K_i. The operator A_i^* is surjective so there exists a vector $u \in Y$, with $\|u\|_Y = 1$ such that $A_i^* u$ is parallel to v (for the case when $\dim(X_i) = 1$, this means $A_i^* u \neq 0$). We observe here that we can take u to be a vector in the range of A_i (because any component of u from $\mathrm{Ker}(A_i^*)$ will have no contribution to $A_i^* u$). The size of $A_i^* u$ gives a lower bound for the distance from s_i to ∂K_i.

A lower bound for the size of $\|A_i^* u\|$ is given by the solution to the next problem:

$$\begin{aligned} \min \quad & \|A_i^* u\| \\ \text{such that} \quad & u \in \mathrm{Range}(A_i), \\ & \|u\|_Y = 1. \end{aligned}$$

We can reformulate this problem as:

$$\begin{aligned} \min \quad & \|A_i^* A_i v\| \\ \text{such that} \quad & \|A_i v\|_Y = 1, \\ & v \in X_i. \end{aligned}$$

The operator $A_i^* A_i : X_i \to X_i$ is positive definite (this is because of our assumption that A_i^* is surjective which implies that A_i is injective hence $Ker(A_i) = \{0\}$).

Let $\{v_1, v_2, \ldots, v_{p_i}\}$ be an orthogonal basis formed by eigenvectors of $A_i^* A_i$ with the corresponding eigenvalues λ_i. Any vector $v \in X_i$ can be written as:

$$v = \sum_{j=1}^{p_i} \alpha_j v_j.$$

Using this decomposition:

$$\|A_i v\|_Y^2 = \sum_{j=1}^{p_i} \alpha_j^2 \lambda_j = 1 \tag{3.84}$$

and

$$\|A_i^* A_i v\|^2 = \sum_{j+1}^{p_i} \alpha_j^2 \lambda_j^2. \tag{3.85}$$

Let $\lambda_{\min}$ be the minimum eigenvalue of $A_i^* A_i$. Then, the equalities (3.84) and (3.85) imply:

$$\|A_i^* A_i v\|^2 = \sum_{j+1}^{p_i} \alpha_j^2 \lambda_j^2 \geq \lambda_{\min} \sum_{j=1}^{p_i} \alpha_j^2 \lambda_j = \lambda_{\min}.$$

Now we can conclude that:

$$\min\{\|A_i^* u\| : \|u\|_Y = 1 \text{ and } u \in \text{ Range}(A_i)\} \geq \sqrt{\lambda_{\min}}.$$

So, the distance from s_i to the boundary of the cone K_i is greater than or equal to $\varepsilon\sqrt{\lambda_{\min}}$:

$$d(s_i, \partial K_i) \geq \varepsilon \sqrt{\lambda_{\min}(A_i^* A_i)}.$$

□

Next we will analyze the implications of the assumption we made that $f_i(\sigma_i) = 0$ where σ_i is a vector of norm $\sqrt{p_i}$ described by $\sigma_i = \sqrt{\frac{p_i}{\theta_i}} e_i$, e_i being the vector in X_i that induces the scaled inner product.

Lemma 17 *Let $\sigma_i \in \partial B(0, \sqrt{\dim(X_i)}) \cap K_i$ be the point where $f_i(\sigma_i) = 0$. Then $f_i^*(\sigma_i) = \theta_i(\ln \frac{\theta_i}{p_i} - 1)$, for all $i \geq 0$ (we take here $p_0 = 2m$).*

Proof: If we use *Lemma* 3 together with $f_i(\sigma_i) = 0$:

$$f_i^*(\sigma_i) = f_i(\sigma_i) - \theta_i - 2f_i(e_i) = -\theta_i - 2f_i(e_i).$$

The functional f_i is logarithmically homogeneous and $e_i = \sqrt{\frac{\theta_i}{p_i}}\sigma_i$. So

$$f_i(e_i) = f_i(\sigma_i) - \frac{\theta_i}{2}\ln\frac{\theta_i}{p_i} = -\frac{\theta_i}{2}\ln\frac{\theta_i}{p_i}.$$

The conclusion follows immediately. □

Lemma 18 *At any instance k of the algorithm described by the Hilbert space X with the corresponding cone K and barrier functional f there exists a point $x \in \partial B(0, \sqrt{\dim(X)}) \cap K$ such that $f(x) = 0$.*

Proof: Let $x \in X$ be the vector with components $x_i = \sigma_i$, for $i \geq 0$. Clearly, $x \in K$. Also $f(x) = f_0(\sigma_0) + f_1(\sigma_1) + \ldots + f_k(\sigma_k)$. Then, immediately we can see that $f(x) = 0$.

Because $\|x\|^2 = \sum\limits_{i=0}^{k} \|\sigma_i\|_i^2 = \sum\limits_{i=0}^{k} \dim(X_i)$, it follows that $x \in \partial B(0, \sqrt{\dim(X)})$. □

Now we can prove the main result of this section:

Theorem 13 *At any instance k of the algorithm described by the space X, the cone K and the functional f (where $X := X_0 \oplus X_1 \oplus \ldots \oplus X_k$, $K := K_0 \oplus K_1 \oplus \ldots \oplus K_k$ and $f^*(s) = \sum\limits_{i=0}^{k} f_i^*(s_i), s_i \in K_i$), for all $\bar{s} \in \mathcal{M}_s$,*

$$f^*(\bar{s}) \leq \sum_{i=0}^{k} \theta_{f_i} \ln\frac{\theta_{f_i}}{\varepsilon_i}$$

where $\varepsilon_i = \varepsilon\sqrt{\lambda_{\min}(A_i^* A_i)}$ for $i \geq 1$ and $\varepsilon_0 = \varepsilon\sqrt{2}$. In particular, if s_{AC} is the analytic center,

$$f^*(s_{AC}) \leq \sum_{i=0}^{k} \theta_{f_i} \ln\frac{\theta_{f_i}}{\varepsilon_i}.$$

Proof: Let $\bar{s}$ be a point in K such that the distance from $\bar{y}$ to the boundary of Ω_k is greater than or equal to ε (i.e. $B_Y(\bar{y}, \varepsilon) \subset \Omega_k$). We have:

$$f^*(\bar{s}) = \sum_{i=0}^{k} f_i^*(\bar{s}_i),$$

where $\bar{s}_i$ are the components of $\bar{s}$ from K_i, $\bar{s}_i \in K_i$.

Using *Lemma* 16 we get $B_i(\bar{s}_i, \varepsilon_i) \subset K_i$.

For each f_i we know that there exists a point $\sigma_i \in K_i \cap \partial B_i(0, \sqrt{p_i})$ such that $f_i(\sigma_i) = 0$. It is easy to see that the point $\frac{\varepsilon_i}{\sqrt{p_i}}\sigma_i \in K_i \cap B_i(0, \varepsilon_i)$. Using *Lemma* 1, *Lemma* 17 and the fact that the functional f_i^* is logarithmically homogeneous we have:

$$f_i^*(\bar{s}_i) \leq f_i^*(\frac{\varepsilon_i}{\sqrt{p_i}}\sigma_i) = f_i^*(\sigma_i) - \theta_{f_i} \ln \frac{\varepsilon_i}{\sqrt{p_i}} = \theta_{f_i} \ln \frac{\theta_{f_i}}{\varepsilon_i \sqrt{p_i}} - \theta_{f_i} \leq \theta_{f_i} \ln \frac{\theta_{f_i}}{\varepsilon_i}.$$

So

$$f^*(\bar{s}) \leq \sum_{i=0}^{k} \theta_{f_i} \ln \frac{\theta_{f_i}}{\varepsilon_i}.$$

The last statement of the theorem is immediate because s_{AC} is the analytic center, hence it minimizes f^* over $\mathcal{M}_s$. □

Corollary 2 *Let* $\Lambda := \min\limits_{i=1,\ldots,k} \sqrt{\lambda_{\min}(A_i^* A_i)}$. *Then:*

$$f^*(s_{AC}) \leq \sum_{i=1}^{k} \theta_{f_i} \ln \frac{\theta_{f_i}}{\varepsilon \Lambda} + \theta_{f_0} \ln \frac{\theta_{f_0}}{\varepsilon \sqrt{2}}. \tag{3.86}$$

3.7 Complexity Analysis

In this section we will derive an upper bound for the number of cuts that may be added to the problem before we are guaranteed to have a solution.

First we start by getting a lower bound for the minimum eigenvalue of the Hessian of any potential functional evaluated at any feasible point.

Let $\bar{s} \in \text{int}(K)$ be any strictly feasible point for the k-th iteration of the algorithm. At this stage, the dual potential is given by

$$f^*(s) := f_0^*(s_0) + f_1^*(s_1) + f_2^*(s_2) + \ldots + f_k^*(s_k)$$

where $s = s_0 \oplus s_1 \oplus s_2 \oplus \ldots \oplus s_k$, with $s_i \in K_i$, $i = 0, \ldots, k$.

$H(\bar{s})$, the Hessian of the barrier functional f^*, has a block diagonal matrix representation, each block corresponding to a Hessian $H_i(\bar{s}_i)$. Because of this structure, the minimum eigenvalue of $H(\bar{s})$ is equal to the minimum of all eigenvalues of $H_i(\bar{s}_i)$, $i = 0, \ldots, k$.

Now let's consider the Hessian $H_i(\bar{s}_i)$, $\bar{s}_i \in \text{int}(K_i)$. The norm used is the one induced by a vector $e_i \in K_i$. In this norm $\|e_i\| = \sqrt{\theta_{f_i}}$ (see (2.15)). Moreover, the distance (measured in the norm induced by e_i) from e_i to the boundary of the cone K_i is greater than or equal to 1. Let

$$d = \max\{\|z\| : z \in \partial K_i \cap (\bar{s}_i - K_i)\},$$

$\bar{s}_i$ is strictly interior to K_i so $d \neq 0$. We define $\bar{s}_d := \frac{1}{d}\bar{s}_i$.

The next lemma will give a description for the position of $\bar{s}_d$ in the cone K_i.

Lemma 19 $\bar{s}_d \in \mathcal{K} := \{int(K_i) \cap (e_i - K_i)\}$.

Proof: Suppose $\bar{s}_d \notin \mathcal{K}$. Then, because the origin is on the boundary of the convex set $\mathcal{K}$, the line containing both $\bar{s}_d$ and the origin intersects the boundary of $e_i - K_i$ in a unique point s_e, with $\|s_e\| < \|\bar{s}_d\|$. Let $\mathcal{P}$ be the plane determined by e_i and $\bar{s}_d$ together with the origin. Then $\mathcal{P} \cap K_i = \{OA, OB\}$, with OA and OB being two rays of the cone K_i (see *Fig.* 3.1). Take OA and OB such that e_i is in the angle determined by OA and $O\bar{s}_d$. Next:

$$\begin{aligned} C_1 &= OA \cap \partial(\bar{s}_d - K_i), \\ C_2 &= OB \cap \partial(\bar{s}_d - K_i), \\ D_1 &= OA \cap \partial(e_i - K_i), \\ D_2 &= OB \cap \partial(e_i - K_i). \end{aligned}$$

With these notations we have:

$$1 \geq \|OC_2\| > \|OD_2\| = \|D_1 e_i\| \geq d(e_i, AO) \geq 1. \tag{3.87}$$

So we arrived at a contradiction. This means that $\bar{s}_d \in \mathcal{K} - \partial K_i$.

Note here that

$$\|\bar{s}_d\|^2 = \|OC_1\|^2 + \|OC_2\|^2 + 2\langle OC_1, OC_2\rangle \geq \|OC_1\|^2. \tag{3.88}$$

This inequality holds for any point $C_1 \in K_i \cap \partial(\bar{s}_d - K_i)$. Hence $\|\bar{s}_d\| \geq 1$. This implies that $\|\bar{s}_i\| \geq d$.

□

Now, as already proved in *Lemma* 5, any point $z \in \mathcal{K}$ has the property that $|z - e_i| < 1$. For such a point *Corollary* 1 shows that the minimum eigenvalue of $H_i(z)$ is greater than $\frac{1}{4}$. So:

$$\lambda_{\min}(H_i(\bar{s}_d)) > \frac{1}{4}. \tag{3.89}$$

Next:

$$\lambda_{\min}(H_i(\bar{s}_d)) = \lambda_{\min}(H_i(\frac{1}{d}\bar{s}_i)) = d^2\lambda_{\min}(H_i(\bar{s}_i)) > \frac{1}{4}.$$

So:

$$\lambda_{\min}(H_i(\bar{s}_i)) > \frac{1}{4d^2} \geq \frac{1}{4\|\bar{s}_i\|^2}. \tag{3.90}$$

In order to get a lower bound for the minimum eigenvalue of $H_i(\bar{s}_i)$ we need to find an upper bound for $\|\bar{s}_i\|$.

Because $\bar{s}$ is feasible, we have that $A^*\bar{y} + \bar{s} = c$, for some $\bar{y} \in \Omega_k$. Here $A = A_0 \oplus A_1 \oplus A_2 \oplus \ldots \oplus A_k$ and $c = c_0 \oplus c_1 \oplus c_2 \oplus \ldots \oplus c_k$. So, componentwise, for each $i = 0, \ldots, k$, $\bar{s}_i = c_i - A_i^*\bar{y}$.

Two different cases arise: one corresponding to $i = 0$ (this is right at the beginning, before adding any cuts to the initial set Ω_0) and one corresponding to $i > 0$.

Let's consider the second case. In this case there exists at least one previous θ - analytic center. Let's denote it (x_p, y_p, s_p). The cuts added through this point have the property: $A_i^* y_p = c_i$. We mention here one more time that the inner product

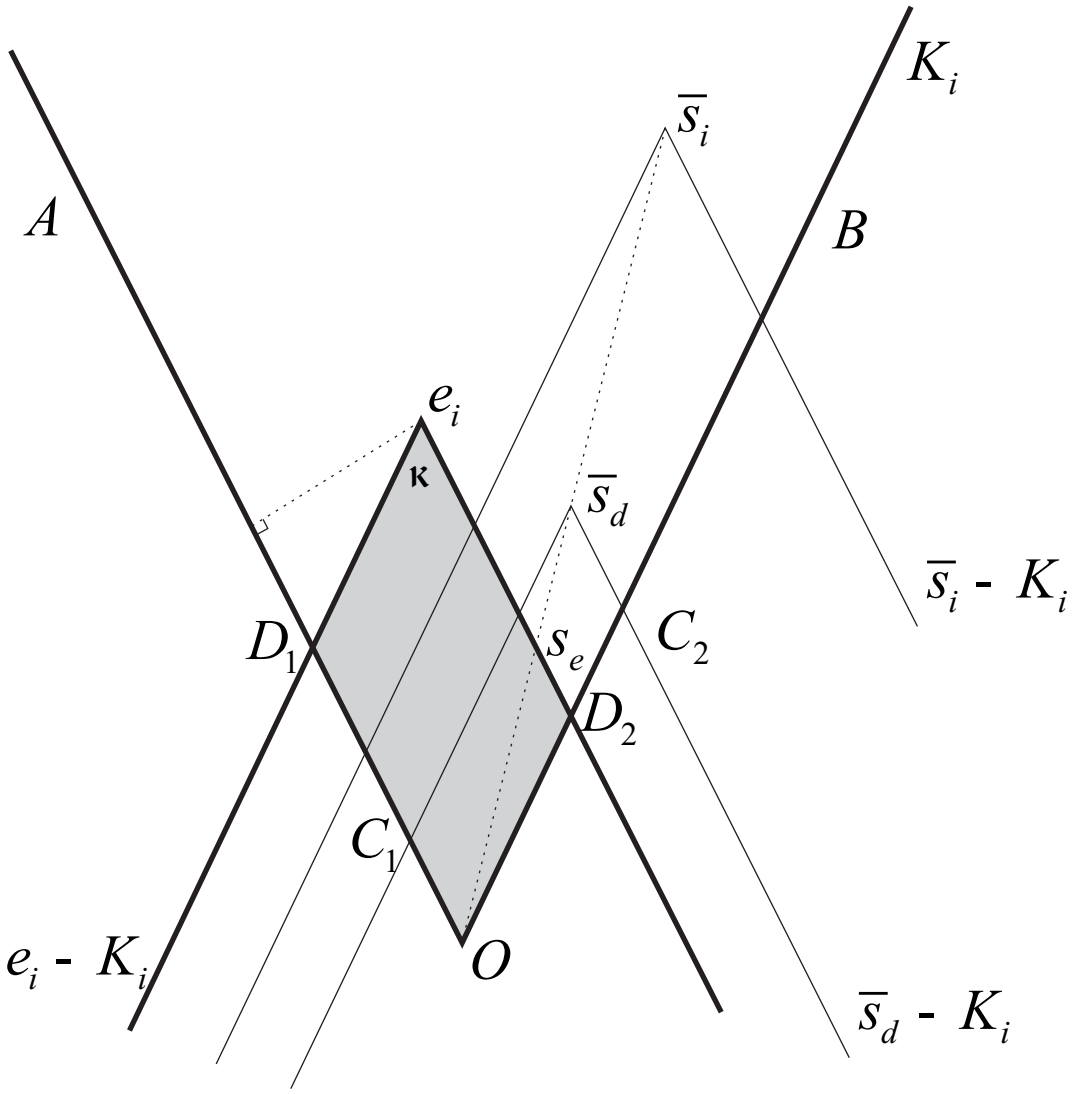

Figure 3.1: Position of $\bar{s}_d$ relative to $\mathcal{K}$.

used is the local one induced by $e_i \in K_i$ for which $\|H_i(e_i)^{-1}\| = 1$. The norm of A_i in the original inner product is one so

$$\|\bar{s}_i\|_{e_i}^2 = \|A_{e_i}^* \bar{y} - A_{e_i}^* y_p\|_{e_i}^2 \leq \|H_i(e_i)^{-1}\| \|A_i^*\|^2 \|\bar{y} - y_p\|^2 = \|\bar{y} - y_p\|^2.$$

In the above sequence of inequalities, the index e_i is for the norms induced by the local inner product. If the index e_i is missing, then the inner product used is the original one.

This implies:

$$\|\bar{s}_i\|_{e_i} \leq \|\bar{y} - y_p\|. \tag{3.91}$$

Now, both y_p and $\bar{y}$ are in Ω_i which is a subset of the initial set Ω_0. The next lemma will give a bound for the size of any point $y \in \Omega_0$.

Lemma 20 *Let* $\Omega_0 := \{y \in Y : y + s_1 = \tilde{c}_0, -y + s_2 = \tilde{c}_0 \text{ with } \tilde{c}_0, s_1, s_2 \in \tilde{K}_0\}$. *Then* $\|y\| \leq \|\tilde{c}_0\|$ *for any* $y \in \Omega_0$.

Proof: The proof is rather immediate. If we take the square of the equalities defining Ω_0,

$$\|\tilde{c}_0\|^2 = \|y\|^2 + \|s_1\|^2 + 2\langle y, s_1\rangle,$$
$$\|\tilde{c}_0\|^2 = \|y\|^2 + \|s_2\|^2 - 2\langle y, s_2\rangle.$$

This implies

$$2\|\tilde{c}_0\|^2 = 2\|y\|^2 + \|s_1\|^2 + \|s_2\|^2 + 2\langle y, s_1 - s_2\rangle. \tag{3.92}$$

Because $\tilde{K}_0$ is a self-scaled cone, and $s_1, s_2 \in \tilde{K}_0$, their inner product is positive:

$$\langle s_1, s_2\rangle \geq 0.$$

Using this observation together with the fact that $s_1 + s_2 = 2\tilde{c}_0$, we get a bound on the sum of norms of s_1 and s_2:

$$\|s_1\|^2 + \|s_2\|^2 \leq 4\|\tilde{c}_0\|^2.$$

Using this inequality and the fact that $s_1 - s_2 = -2y$ the conclusion follows immediately:

$$\|y\| \leq \|\tilde{c}_0\|.$$

□

Using the previous lemma and (3.91) we finally get:

$$\|\bar{s}_i\| \leq 2\|\tilde{c}_0\| \text{ for any } i > 0.$$

For the case $i = 0$ we have that $\bar{s}_0 \in K_0 := \tilde{K}_0 \oplus \tilde{K}_0$. So we can decompose $\bar{s}_0$ in two parts: s_1 and s_2, both elements in $\tilde{K}_0$. For $\bar{s}_0$ there exists $y \in \Omega_0$ such that:

$$y + s_1 = \tilde{c}_0,$$
$$-y + s_2 = \tilde{c}_0.$$

We know that $s_1, s_2, \tilde{c}_0 \in \tilde{K}_0$, with $\tilde{K}_0$ a self-conjugate cone. Using this and the fact that $s_1 + s_2 = 2\tilde{c}_0$ it follows that

$$\|\bar{s}_0\|^2 = \|s_1\|^2 + \|s_2\|^2 \leq 4\|\tilde{c}_0\|^2. \tag{3.93}$$

Hence,

$$\|\bar{s}_0\| \leq 2\|\tilde{c}_0\|.$$

It follows from inequality (3.90) that the smallest eigenvalue of the Hessian can be bounded away from zero.

Lemma 21 *For any strictly feasible point $\bar{s} \in K$:*

$$\lambda_{\min}(H(\bar{s})) \geq \frac{1}{16\|\tilde{c}_0\|^2}. \tag{3.94}$$

Now we are ready to get an estimate for the number of cuts required to be added in order to find an interior point in Γ.

Before this we will reintroduce some notations.

Let $(X_i, \langle\cdot,\cdot\rangle_i)$, K_i, A_i and f_i, $i = 0, \ldots, k$, be the elements that describe the initial instance of the algorithm and the cuts that are added during the first k - iterations of the algorithm. Let $\bar{X}_i = \bigoplus_{j=0}^{i} X_j$, $\bar{K}_i = \bigoplus_{j=0}^{i} K_j$, $\bar{A}_i = \bigoplus_{j=0}^{i} A_j$, $\bar{f}_i = \bigoplus_{j=0}^{i} f_j$ be the elements that describe the instance of the algorithm after adding the i-th cut.

Let $\bar{s}_i$ and s_i^θ be the exact analytic center and a θ - analytic center of the domain Ω_i respectively.

After i iterations of the algorithm, using formula (3.43) we get:

$$\bar{f}_i^*(\bar{s}_i) \geq \bar{f}_{i-1}^*(\bar{s}_{i-1}) - f_i(\beta_i') - \frac{\theta_{f_i}}{2}(\langle \beta_i', V_i\beta_i'\rangle_i - 1) + \theta_{f_i}\ln\alpha - \mathcal{F}(\theta,\zeta), \tag{3.95}$$

where $V_i = A_i^*(\bar{A}_{i-1}\bar{H}_{i-1}(\bar{s}_{i-1})\bar{A}_{i-1}^*)^{-1}A_i$, β_i' is any point in the interior of K_i and $\mathcal{F}(\theta,\zeta)$ is given in (3.44).

One of the assumptions we made about the functionals f_i was that, for each of them, there exists a point $\sigma_i \in K_i$ ($\sigma_i := \sqrt{\frac{p_i}{\theta_i}}e_i$)with norm equal to $\sqrt{p_i}$ such that $f_i(\sigma_i) = 0$.

Now, in each space X_i we can choose an orthonormal basis $\{e_j^i\}_{j=1,\dots,p_i}$ such that

$$\sigma_i = \sum_{j=1}^{p_i} e_j^i, \tag{3.96}$$

To do this it is enough to pick an orthonormal basis and then rotate it until $\sum\limits_{j=1}^{p_i} e_j^i$ overlaps with σ_i. It is clear that $\|\sigma_i\|_i = \sqrt{p_i}$, for any $i \geq 0$.

Notice here that unlike β_i (the exact solution for problem (3.17)) for which both β_i and $V_i\beta_i$ have to be in K_i, the only requirement for β_i' is to be an element from K_i. This gives us more choices for picking a suitable vector.

Now we can choose β_i' to be:

$$\beta_i' = \frac{e_i}{\sqrt{\langle e_i, V_i e_i\rangle_i}}. \tag{3.97}$$

Clearly $\langle \beta_i', V_i\beta_i'\rangle_i - 1 = 0$. Moreover, using the fact that f_i is logarithmically homogeneous:

$$f_i(\beta_i') = f_i(\sigma_i) + \frac{\theta_{f_i}}{2}\ln\langle \sigma_i, V_i\sigma_i\rangle_i = \frac{\theta_{f_i}}{2}\ln\langle \sigma_i, V_i\sigma_i\rangle_i. \tag{3.98}$$

The inequality (3.95) can be further simplified to:

$$\bar{f}_i^*(\bar{s}_i) \geq \bar{f}_{i-1}^*(\bar{s}_{i-1}) - \frac{\theta_{f_i}}{2}\ln\langle\sigma_i, V_i\sigma_i\rangle_i + \theta_{f_i}\ln\alpha - \mathcal{F}(\theta,\zeta). \tag{3.99}$$

Let's consider now $\langle\sigma_i, V_i\sigma_i\rangle_i$:

$$\langle\sigma_i, V_i\sigma_i\rangle_i = \sum_{j=1}^{p_i}\sum_{l=1}^{p_i}\langle e_j^i, V_i e_l^i\rangle_i = \sum_{j=1}^{p_i}\sum_{l=1}^{p_i}\langle e_j^i, e_l^i\rangle_{V_i} \leq \sum_{j=1}^{p_i}\sum_{l=1}^{p_i}\|e_j^i\|_{V_i}\|e_l^i\|_{V_i}. \tag{3.100}$$

Using the mean inequality:

$$\langle\sigma_i, V_i\sigma_i\rangle_i \leq \sum_{j=1}^{p_i}\sum_{l=1}^{p_i}\frac{\|e_j^i\|_{V_i}^2 + \|e_l^i\|_{V_i}^2}{2} = p_i\sum_{j=1}^{p_i}\|e_j^i\|_{V_i}^2. \tag{3.101}$$

So:

$$\bar{f}_i^*(\bar{s}_i) \geq \bar{f}_{i-1}^*(\bar{s}_{i-1}) - \frac{\theta_{f_i}}{2}\ln(p_i\sum_{j=1}^{p_i}\|e_j^i\|_{V_i}^2) + \theta_{f_i}\ln\alpha - \mathcal{F}(\theta,\zeta). \tag{3.102}$$

This inequality gives a relationship between the dual potential functionals evaluated at two consecutive exact analytic centers. A direct relationship between the potential at the initial analytic center $\bar{s}_0$ and the potential at the k-th analytic center $\bar{s}_k$ can be easily obtained by taking the sum of the previous inequalities from $i = 1$ to $i = k$:

$$\bar{f}_k^*(\bar{s}_k) \geq \bar{f}_0^*(\bar{s}_0) - \sum_{i=1}^{k}(\frac{\theta_{f_i}}{2}\ln(p_i\sum_{j=1}^{p_i}\|e_j^i\|_{V_i}^2)) + \ln\alpha\sum_{i=1}^{k}\theta_{f_i} - k\mathcal{F}(\theta,\zeta). \tag{3.103}$$

Let $P = \max_{i=1,\ldots,k} p_i$ (i.e. at each stage we do not add more than P cuts). Then:

$$\bar{f}_k^*(\bar{s}_k) \geq \bar{f}_0^*(\bar{s}_0) - \frac{1}{2}(\ln P - \ln\alpha^2)\sum_{i=1}^{k}\theta_{f_i} - \sum_{i=1}^{k}(\frac{\theta_{f_i}}{2}\ln\sum_{j=1}^{p_i}\|e_j^i\|_{V_i}^2) - k\mathcal{F}(\theta,\zeta). \tag{3.104}$$

We can simplify this inequality by using the concavity of the logarithm function together with the fact that $\theta_{f_i} \geq 1$:

$$\bar{f}_k^*(\bar{s}_k) \geq \bar{f}_0^*(\bar{s}_0) - \frac{1}{2}(\ln P - \ln \alpha^2) \sum_{i=1}^{k} \theta_{f_i} - \frac{\sum_{l=1}^{k} \theta_{f_l}}{2} \ln \frac{\sum_{i=1}^{k} (\theta_{f_i} \sum_{j=1}^{p_i} \|e_j^i\|_{V_i}^2)}{\sum_{t=1}^{k} \theta_{f_t}} - k\mathcal{F}(\theta, \zeta).$$

For any i: $\theta_{f_i} \geq 1$. So $\sum_{i=1}^{k} p_i \leq P \sum_{i=1}^{k} \theta_{f_i}$. Let $\Theta := \max_{i=1,\ldots,k} \theta_{f_i}$. Then:

$$\bar{f}_k^*(\bar{s}_k) \geq \bar{f}_0^*(\bar{s}_0) - \frac{1}{2}(\ln P - \ln \alpha^2) \sum_{i=1}^{k} \theta_{f_i} - \frac{\sum_{l=1}^{k} \theta_{f_l}}{2} \ln P\Theta \frac{\sum_{i=1}^{k} (\sum_{j=1}^{p_i} \|e_j^i\|_{V_i}^2)}{\sum_{t=1}^{k} p_t} - k\mathcal{F}(\theta, \zeta).$$

So:

$$\bar{f}_k^*(\bar{s}_k) \geq \bar{f}_0^*(\bar{s}_0) - \frac{\sum_{l=1}^{k} \theta_{f_l}}{2} (2 \ln P + \ln \frac{\Theta}{\alpha^2} + \ln \frac{\sum_{i=1}^{k} (\sum_{j=1}^{p_i} \|e_j^i\|_{V_i}^2)}{\sum_{t=1}^{k} p_t}) - k\mathcal{F}(\theta, \zeta). \quad (3.105)$$

By taking arbitrarily $\theta \leq 0.9$ and $\zeta \leq 0.9$, the value of $\mathcal{F}(\theta, \zeta)$ can be made smaller than 6.5. Then, for this choice of θ and ζ, $k\mathcal{F}(\theta, \zeta) \leq 7k \leq 7 \sum_{l=1}^{k} \theta_{f_l}$.

So:

$$\bar{f}_k^*(\bar{s}_k) \geq \bar{f}_0^*(\bar{s}_0) - \frac{\sum_{l=1}^{k} \theta_{f_l}}{2} (2 \ln P + 14 + \ln \frac{\Theta}{\alpha^2} + \ln \frac{\sum_{i=1}^{k} (\sum_{j=1}^{p_i} \|e_j^i\|_{V_i}^2)}{\sum_{t=1}^{k} p_t}). \quad (3.106)$$

Now we have to get an estimate for: $\sum_{i=1}^{k} (\sum_{j=1}^{p_i} \|e_j^i\|_{V_i}^2)$. We will take the same approach used by Ye in [25]. Because of the specifics of our problem, we will present here the entire scheme.

Let $C_0 := 16\|\tilde{c}_0\|^2$. Each term $\|e_j^i\|_{V_i}^2$ can be bounded from above if we use

Lemma 21:

$$\|e_j^i\|_{V_i}^2 \leq C_0 \langle e_j^i, A_i^*(\bar{A}_{i-1}\bar{A}_{i-1}^*)^{-1}A_i e_j^i\rangle_i.$$

Let $\mathcal{A}_i$ be the matrix representation of the operator A_i with respect to the basis $\{e_j^i\}$, $j = 1, \ldots, p_i$ for $i = 1, \ldots, k$. Let $\mathcal{A}_0$ be the matrix representation for A_0 with respect to an orthonormal basis $\{e_j^0\}, j = 1, \ldots, 2m$ of X_0. The corresponding matrix representation for $\bar{A}_i$ is given by the $m \times (2m + \sum_{i=1}^{k} p_i)$ block matrix $\bar{\mathcal{A}}_i = [\mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_k]$. Let a_j^i be the j-th column of $\mathcal{A}_i$. Using this notation we have:

$$\sum_{i=1}^{k}(\sum_{j=1}^{p_i}\|e_j^i\|_{V_i}^2) \leq C_0 \sum_{i=1}^{k}\sum_{j=1}^{p_i}({a_j^i}^T(\sum_{l=0}^{i-1}\mathcal{A}_l\mathcal{A}_l^T)^{-1}a_j^i).$$

Let $B_0 := \mathcal{A}_0\mathcal{A}_0^T$ and $B_{i+1} = B_i + \mathcal{A}_{i+1}\mathcal{A}_{i+1}^T$, for $i \geq 0$.

With this notation:

$$\sum_{i=1}^{k}(\sum_{j=1}^{p_i}\|e_j^i\|_{V_i}^2) \leq C_0 \sum_{i=1}^{k}\sum_{j=1}^{p_i}({a_j^i}^T(B_{i-1})^{-1}a_j^i). \tag{3.107}$$

Lemma 22 *Let* $C_B = \frac{1}{1+(P+3)\|B_0^{-1}\|}$. *Then*

$$\sum_{i=1}^{k}(\sum_{j=1}^{p_i}\|e_j^i\|_{V_i}^2) \leq \frac{4C_0}{3C_B}(2m \ln \frac{tr(B_0) + \sum_{i=1}^{k} p_i}{2m} - \ln(\det B_0)). \tag{3.108}$$

Proof: Notice that: $\mathcal{A}_i\mathcal{A}_i^T = \sum_{j=1}^{p_i} a_j^i {a_j^i}^T$. If we denote:

$$\omega^2 = {a_1^{i+1}}^T(B_i + \sum_{j=2}^{p_{i+1}} a_j^{i+1}{a_j^{i+1}}^T)^{-1}a_1^{i+1}$$

then, as shown by Ye in [25],

$$\det B_{i+1} = \det(B_i + \sum_{j=1}^{p_{i+1}} a_j^{i+1}{a_j^{i+1}}^T) = (1+\omega^2)\det(B_i + \sum_{j=2}^{p_{i+1}} a_j^{i+1}{a_j^{i+1}}^T).$$

We know from the initial assumptions that $\|A_i\| = 1$ for all $i > 0$ so:

$$\|a_j^{i+1}\| \leq \|\mathcal{A}_{i+1}\| = \|A_{i+1}\| = 1.$$

We can rewrite ω^2 as:

$$\omega^2 = a_1^{i+1^T} B_i^{-\frac{1}{2}} (I + \sum_{j=2}^{p_{i+1}} B_i^{-\frac{1}{2}} a_j^{i+1} a_j^{i+1^T} B_i^{-\frac{1}{2}})^{-1} B_i^{-\frac{1}{2}} a_1^{i+1}.$$

Next, for any y with $\|y\| = 1$:

$$\begin{aligned}
y^T (I + \sum_{j=2}^{p_{i+1}} B_i^{-\frac{1}{2}} a_j^{i+1} a_j^{i+1^T} B_i^{-\frac{1}{2}}) y &= 1 + \sum_{j=2}^{p_{i+1}} (y^T B_i^{-\frac{1}{2}} a_j^{i+1})^2 \\
&\leq 1 + \sum_{j=2}^{p_{i+1}} \|B_i^{-\frac{1}{2}} a_j^{i+1}\|^2 \\
&= 1 + \sum_{j=2}^{p_{i+1}} a_j^{i+1^T} B_i^{-1} a_j^{i+1} \\
&\leq 1 + \sum_{j=2}^{p_{i+1}} a_j^{i+1^T} B_0^{-1} a_j^{i+1} \\
&\leq 1 + \|B_0^{-1}\| \sum_{j=2}^{p_{i+1}} \|a_j^{i+1}\|^2 \\
&\leq 1 + (p_{i+1} - 1)\|B_0^{-1}\| \\
&\leq 1 + (P+3)\|B_0^{-1}\|.
\end{aligned}$$

So the maximum eigenvalue of $I + \sum_{j=2}^{p_{i+1}} B_i^{-\frac{1}{2}} a_j^{i+1} a_j^{i+1^T} B_i^{-\frac{1}{2}}$ is less than or equal to $1 + (P+3)\|B_0^{-1}\|$. This allows us to write:

$$\omega^2 \geq \frac{1}{1 + (P+3)\|B_0^{-1}\|} a_1^{i+1^T} B_i^{-1} a_1^{i+1}. \tag{3.109}$$

Hence:

$$\det B_{i+1} \geq (1 + \frac{a_1^{i+1^T} B_i^{-1} a_1^{i+1}}{1 + (P+3)\|B_0^{-1}\|}) \det(B_i + \sum_{j=2}^{p_{i+1}} a_j^{i+1} a_j^{i+1^T}). \tag{3.110}$$

Repeating this process inductively, we finally get

$$\ln\det B_{i+1} \geq \sum_{j=1}^{p_{i+1}} \ln(1 + \frac{{a_j^{i+1}}^T B_i^{-1} a_j^{i+1}}{1 + (P+3)\|B_0^{-1}\|}) + \ln\det(B_i).$$

or, using C_B:

$$\ln\det B_{i+1} \geq \sum_{j=1}^{p_{i+1}} \ln(1 + C_B {a_j^{i+1}}^T B_i^{-1} a_j^{i+1}) + \ln\det(B_i). \tag{3.111}$$

We know that $B_i - B_0$ is a positive semidefinite matrix for any $i \geq 1$. So:

$${a_j^{i+1}}^T B_i^{-1} a_j^{i+1} \leq {a_j^{i+1}}^T B_0^{-1} a_j^{i+1} \leq \|B_0^{-1}\| \|a_j^{i+1}\|^2 \leq \|B_0^{-1}\|.$$

Based on this, it is clear that, for any $P > 0$:

$$C_B {a_j^{i+1}}^T B_i^{-1} a_j^{i+1} \leq \frac{\|B_0^{-1}\|}{1 + (P+3)\|B_0^{-1}\|} < \frac{1}{3} < 1.$$

Now, the inequality $\ln(1+x) \geq x - \frac{x^2}{2(1-x)}$ holds true for any $x \in [0, 1)$. Using it and the fact that the function $1 - \frac{x}{2(1-x)}$ is decreasing, we get, for any $i = 0, \ldots, k-1$

$$\ln(1 + C_B {a_j^{i+1}}^T B_i^{-1} a_j^{i+1}) \geq C_B {a_j^{i+1}}^T B_i^{-1} a_j^{i+1} (1 - \frac{1/3}{2(1-1/3)})$$

or

$$\ln(1 + C_B {a_j^{i+1}}^T B_i^{-1} a_j^{i+1}) \geq \frac{3}{4} C_B {a_j^{i+1}}^T B_i^{-1} a_j^{i+1}.$$

So, for any $i = 0, \ldots, k-1$

$$\ln\det B_{i+1} \geq \ln\det(B_i) + \frac{3}{4} C_B \sum_{j=1}^{p_{i+1}} {a_j^{i+1}}^T B_i^{-1} a_j^{i+1}.$$

After we add the inequalities corresponding to $i = 0$ to $i = k - 1$, we get:

$$\ln \det B_k \geq \ln \det(B_0) + \frac{3}{4} C_B \sum_{i=1}^{k} \sum_{j=1}^{p_i} {a_j^i}^T B_{i-1}^{-1} a_j^i.$$

Now:

$$\mathrm{tr}(B_k) = \sum_{i=0}^{k} \mathrm{tr}(\mathcal{A}_i \mathcal{A}_i^T) = \mathrm{tr}(B_0) + \sum_{i=1}^{k} \sum_{j=1}^{p_i} \|a_j^i\|^2 \leq \mathrm{tr}(B_0) + \sum_{i=1}^{k} p_i.$$

Using the mean inequality (for the sum and the product of eigenvalues of B_k):

$$\ln(\det B_k) \leq 2m \ln \frac{\mathrm{tr}(B_0) + \sum_{i=1}^{k} p_i}{2m}.$$

The conclusion follows immediately.

Note that $\ln(\det B_k)$ is well defined since B_k is positive definite being the sum of the positive definite matrix $\mathcal{A}_0 \mathcal{A}_0^T$ and positive semidefinite matrices $\mathcal{A}_i \mathcal{A}_i^T$. □

Using *Lemma* 22 and inequality (3.106) we get:

$$\bar{f}_k^*(\bar{s}_k) \geq \bar{f}_0^*(\bar{s}_0) - \frac{\sum_{l=1}^{k} \theta_{f_l}}{2} \left(\ln \frac{4C_0 \Theta P^2}{3C_B \alpha} + 14 + \ln \frac{\left(2m \ln \frac{\mathrm{tr}_{(B_0)} + \sum_{i=1}^{k} p_i}{2m} - \ln(\det B_0)\right)}{\sum_{i=1}^{k} p_i} \right).$$

Corollary 2 gives an upper bound for $\bar{f}_k^*(\bar{s}_k)$:

$$\bar{f}_k^*(\bar{s}_k) \leq \sum_{i=1}^{k} \theta_{f_i} \ln \frac{\theta_{f_i}}{\varepsilon \Lambda} + \theta_{f_0} \ln \frac{\theta_{f_0}}{\varepsilon \sqrt{2}}.$$

Theorem 14 *The algorithm stops with a solution as soon as:*

$$\left(\sum_{l=1}^{k} \theta_{f_l}\right) \ln H \frac{\left(2m \ln \frac{1}{2m} (tr(B_0) + \sum_{i=1}^{k} p_i) - \ln(\det B_0)\right)}{\sum_{i=1}^{k} p_i} \leq 2\bar{f}_0^*(\bar{s}_0) - 2\theta_{f_0} \ln \frac{\theta_{f_0}}{\varepsilon \sqrt{2}}.$$

with $H = \frac{4C_0 \Theta^3 P^2 e^{14}}{3\varepsilon^2 \Lambda^2 C_B \alpha}$. *The number of cuts added is at most* $O^*(\frac{mP^3\Theta^3}{\varepsilon^2 \Lambda^2})$ *(here* O^* *means that terms of low order are ignored). Here we assumed that* $\|\tilde{c}_0\|$ *has the size of order* $\sqrt{m}$. *Also we used the fact that* $\|B_0^{-1}\|$ *has order* $O(1)$.

Proof: This result follows directly from the previous analysis. Note here that $C_B = \frac{1}{1+(P+3)\|B_0^{-1}\|}$ has a contribution in the complexity result.

□

This result is similar with the ones for linear or semidefinite programming. Θ and Λ are the only extra terms. The reason for this is straightforward. In the linear or semidefinite case the potential functions are separable. In general this is not necessarily the case. This explains the presence of Θ which characterizes the barrier functional as a whole. The only assumption we made on the cuts that are added was that the operators describing them have unit norms. This assumption is not critical. We use it only to keep the analysis simple. In the linear programming approach a similar assumption often made is that the matrices describing the cuts are assumed to have columns of norm one. This gives more structure to the cuts. In our general case we cannot work at "column" level. So we had to use an overall characterization of the cuts. The parameter Λ characterizes the quality of the cuts that are generated by the oracle.

In the next chapter we will present some numerical results and will try to interpret them. The analytic center cutting planes method can be used not only to solve feasibility problems but also to solve optimization problems. In the last chapter of our thesis we will propose an algorithm for solving linear programming problems using a mixture of an interior-point method and analytic center cutting plane method. We will determine an upper-bound for the total number of cuts that are added in order to get a solution. Also we will determine an upper-bound for the total number of iterations (i.e. the total number of analytic centers) needed to get the solution.

CHAPTER 4
Numerical Results

4.1 Preliminaries

In this chapter we will analyze the performance of the algorithm we propose. Although the method suggested by us in the previous chapters can be used for solving any type of feasibility conic programming problem, we will consider only feasibility convex problems that are based on a mixture of linear and second order cones. A similar analysis was done by Oskoorouchi in [12] for the cone of positive semidefinite matrices.

At the beginning we will take a look at some issues that arise when combining these two types of cones. Then we will present some numerical results obtained by solving a set of feasibility problems derived from the library of problems proposed in "The Seventh DIMACS Implementation Challenge Semidefinite and Related Optimization Problems". The results we obtain are in concordance with the theoretical analysis. The efficiency of the algorithm depends, as expected, on the "thickness" of the set of interest. We conclude this chapter by looking at how our algorithm might be used in solving optimality problems. Our results will show that we can use an analytic center cutting plane method approach to solve optimality problems but, in this case the performance is not very good compared to other algorithms.

4.2 LP - SOCP Feasibility Problem

In implementing the algorithm we followed all the steps described in the previous chapters. Here we will just point out the places where the specifics of the analyzed problem can be used to simplify the implementation.

The self-concordant barrier functionals corresponding to the "LP-SOCP" case have a part f_L corresponding to the linear part of the problem and a part f_Q corre-

sponding to the quadratic part. f_L has the form:

$$f_L(s) = -\sum_{i=1}^{n_L} \ln s_i. \tag{4.1}$$

The quadratic part is in fact a sum of self-concordant barrier functionals. Each term of this sum corresponds to a second order cone used in defining the current outer-approximation set. Each term has the form:

$$g_Q(s) := -\ln(s_1^2 - \sum_{j=2}^{n_Q} s_j^2), \tag{4.2}$$

where n_Q is the dimension of the cone.

For these functionals, the gradient, the Hessian and the inverse of the Hessian are easy to compute. We need all these quantities for computing the Newton steps, the recovery of feasibility direction and to measure the distance to the θ - analytic centers of the outer-approximation sets Ω_i.

For a feasible (in the current approximation set) point (x, y, s), the scaling point w is defined as the unique point such that: $H(w)x = s$. This point is important for our implementation of the algorithm because $\|x - w\|_w$ will be used to measure the distance to the θ - analytic center (as proven in *Theorem* 12).

For f_L, the gradient and the Hessians are immediate to obtain. The scaling point w is given by:

$$w_i = \sqrt{\frac{s_i}{x_i}},\ i = 1, \ldots, n_L. \tag{4.3}$$

Unlike the linear part, for the SOCP part, the Hessian of the barrier functional is no more a diagonal matrix. If the dimension of the second order cone is large, the Hessians will be fully dense causing a potential bottleneck for the algorithm. Still,

once we compute the Hessian $H_Q(z)$, its inverse $H_Q^{-1}(z)$ is easy to obtain:

$$H_Q(z) = \frac{2}{F^2}\begin{pmatrix} F + 2\sum_{i=2}^{n_Q} z_i^2 & -2z_1z_2 & -2z_1z_3 & \cdots & -2z_1z_{n_Q} \\ -2z_2z_1 & F + 2z_2^2 & 2z_2z_3 & \cdots & 2z_2z_{n_Q} \\ -2z_3z_1 & 2z_3z_2 & F + 2z_3^2 & \cdots & 2z_3z_{n_Q} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -2z_{n_Q}z_1 & 2z_{n_Q}z_2 & 2z_{n_Q}z_3 & \cdots & F + 2z_{n_Q}^2 \end{pmatrix}, \tag{4.4}$$

$$H_Q^{-1}(z) = \begin{pmatrix} F + 2\sum_{i=2}^{n_Q} z_i^2 & 2z_1z_2 & 2z_1z_3 & \cdots & 2z_1z_{n_Q} \\ 2z_2z_1 & F + 2z_2^2 & 2z_2z_3 & \cdots & 2z_2z_{n_Q} \\ 2z_3z_1 & 2z_3z_2 & F + 2z_3^2 & \cdots & 2z_3z_{n_Q} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 2z_{n_Q}z_1 & 2z_{n_Q}z_2 & 2z_{n_Q}z_3 & \cdots & F + 2z_{n_Q}^2 \end{pmatrix} \tag{4.5}$$

with $F = z_1^2 - \sum_{i=2}^{n_Q} z_i^2$.

As mentioned since the beginning, we use in our approach local-inner products instead of the original ones. This helps us in the theoretical analysis. The local-inner product is induced by an arbitrary fixed vector e strictly interior to the cone. This vector must be chosen such that, for its linear and second order conic parts e_L and e_Q, $\|H_L(e_L)^{-1}\| = 1$ and $\|H_Q(e_Q)^{-1}\| = 1$. We can choose the e_L component to be the n_L-dimensional vector of all ones. For e_Q one option is to take $e_{Q_1} = \sqrt{2}$ and $e_{Q_i} = 0$ for $i = 2, \ldots, n_Q$. For such a vector e it is easy to see that $H_L(e_L) = I_{n_L}$ and $H_Q(e_Q) = I_{n_Q}$. With this choice the Hessian matrices don't change when the inner product changes and the equalities $\|H_L(e_L)^{-1}\| = 1$ and $\|H_Q(e_Q)^{-1}\| = 1$ hold.

Once the scaling issue is solved we can start to analyze the specifics of the "LP-SOCP" problem. In this case the feasible set considered is described by both LP and SOCP inequalities. One way of defining such a set is:

$$\Gamma = \{y \in I\!R^m : A^T y + s = c, s \in I\!R^{n_0} \oplus K_1 \oplus \ldots \oplus K_n, K_i \in I\!R^{n_i} \text{ -SOCP cones } \}.$$

The algorithm starts by setting up the first outer-approximation set Ω_0:

$$\Omega_0 = \{y \in I\!R^m : -M \leq y_i \leq M, \forall i = 1, \ldots, m\}. \tag{4.6}$$

The size of Γ is unknown so we don't know before-hand what value to assign to M. One way of dealing with this issue is to choose a rather arbitrary value for M, say 1000 and to increase it if needed. We know we have to increase M if the size of the slack variables at the current θ - analytic center corresponding to Ω_0 becomes close to zero. This is a sign that the set Γ is not fully contained in Ω_0. In this case our strategy was to increase M by a factor of 10.

The most expensive part of the algorithm is to compute the Newton steps required to get close to a θ - analytic center. We used two types of Newton steps to prove that the algorithm converges: first we take a sequence of Nesterov-Todd steps to bring the current point closer to the θ - analytic center (i.e. $\|x - w\|_w \leq \alpha$). Then we proved that a different type of steps need to be taken to move the point even closer. In practice it was enough for us to use only the Nesterov-Todd directions.

The Nesterov-Todd directions are given by:

$$d_x := -P_{L,w}(x + g_w(x)), \tag{4.7}$$

$$d_s := -P_{L^\perp,w^*}(s + g_{w^*}(s)). \tag{4.8}$$

For the "LP-SOCP" case these directions become:

$$d_x = -(I - H(w)^{-1}A^T(AH(w)^{-1}A^T)^{-1}A)(w - x), \tag{4.9}$$

$$d_s = -A^T(AH(w)^{-1}A^T)^{-1}A(w - x). \tag{4.10}$$

Here A is the matrix that describes the current feasible region Ω. It is in fact the $\oplus$ sum of the matrix that describes Ω_0 and the matrices describing the cuts added during the evolution of the algorithm. The Hessian $H(w)$ is a block diagonal matrix, each block corresponding to a different cone used in describing the set Ω. So, the inverse $H(w)^{-1}$ is also a block diagonal matrix. Then, if we assume that Ω is described by n_L linear inequalities and n_Q conic ones these matrices are:

$A := A_L \oplus A_1 \oplus \ldots \oplus A_{n_Q}$ and $H(w)^{-1} = diag(H_L(w)^{-1}, H_1(w)^{-1}, \ldots, H_{n_Q}(w)^{-1})$. Computing the inverse matrix $AH(w)^{-1}A^T$ is the most numerically intensive part of the whole algorithm. Using the previous notations:

$$AH(w)^{-1}A^T = A_L H_L(w)^{-1} A_L^T + \sum_{i=1}^{n_Q} A_i H_i(w)^{-1} A_i^T. \tag{4.11}$$

One way of computing this inverse is to solve the corresponding "augmented form" system as described in [24]. This means that computing $u := (AH(w)^{-1}A^T)^{-1}v$ is equivalent to solving the system:

$$\begin{pmatrix} 0 & A \\ A^T & H(w) \end{pmatrix} \begin{pmatrix} u \\ \bar{u} \end{pmatrix} = \begin{pmatrix} -v \\ 0 \end{pmatrix} \tag{4.12}$$

The system is well-defined because $H(w)$ is positive definite (as the Hessian of a strictly convex functional) and the matrix A has full rank. In solving this system we use UMFPACK 4.1 (a set of routines for solving equations of form $Ax = b$, with A sparse). It is based on a LU decomposition that combines a column pre-ordering strategy with a right-looking unsymmetric-pattern multifrontal numeric factorization (see [3] for details). The "augmented system" method is best to be used when the problem we are considering has a sparse A and the dimensionality of the quadratic cones is low (so the block-diagonal Hessian has small blocks). This method might become unstable if the diagonal of H has both small and large elements. This usually happens in the final stages of the algorithm, when we are close to a solution and when the domain of interest Γ is very flat.

After taking a sequence of Nesterov-Todd steps, the current point (x, y, s) is at a θ - analytic center for the outer-approximation domain Ω. The primal and dual feasible sets are described by a matrix A_1 and the vectors b and c_1:

$$\mathcal{F}_P := \{x : A_1 x = b \text{ with } x \geq_K 0\}$$

and

$$\mathcal{F}_D := \{s : A_1^T y + s = c_1 \text{ with } s \geq_K 0\}.$$

Now we call the oracle. If the point is not yet strictly interior to Γ, the oracle returns a p - dimensional cut described by a matrix A_2 and a vector c_2.

After adding the cuts centrally, the point becomes $(x \oplus \beta, y, s \oplus \gamma)$, with $\gamma = 0$ and $\beta = 0$. Now a feasibility step $(\Delta x \oplus \beta, \Delta y, \Delta s \oplus \gamma)$ is required. We have already shown that:

$$A_1 \Delta x + A_2 \beta = 0, \tag{4.13}$$

$$A_1^T \Delta y + \Delta s = 0, \tag{4.14}$$

$$A_2^T \Delta y + \gamma = 0. \tag{4.15}$$

Theoretically, the best way is to choose β and γ to move the point as far as possible from the boundary. From a practical point a view, any β and γ that will bring the point inside the feasible region might be considered. Our goal is to get feasible so we could start by taking $\gamma = \beta$. If we consider the linear and SOCP components of β:

$$\beta := \beta_L \oplus \beta_{Q_1} \oplus \ldots \beta_{Q_s},$$

then one way of choosing β would be:

$$\begin{aligned} (\beta_L)_i &= 1, \text{ for all } i = 1, \ldots, n_L \\ \beta_{Q_i} &= e_1 \text{ the vector of all zeros with the first component equal to one.} \end{aligned}$$

Before adding the cuts, the point (x, y, s) was strictly feasible. So we can choose any direction for Δx and Δs as long as we scale their size to keep the point feasible. The only issue right now is to find Δy.

The cuts to be added are described by a matrix A_2 that is fully ranked and is

injective. So the system $A_2^T \Delta y + \gamma = 0$ is under-determined. If A_2 is an $m \times n$ matrix then $m \geq n$ and $rank(A_2) = n$. After choosing γ, the vector Δy is readily available from the previous system (we just need to pick n linearly independent columns of A_2^T, fix to zero the components of Δy corresponding to the other columns of A_2^T and then solve the well-defined remaining system). Once we have Δy, we take $\Delta s = -A_1^T \Delta y$.

The matrix A_1 contains all the constraints used by the algorithm in defining the outer-approximation sets. So, the first columns of A_1 correspond to the initial set Ω_0. This means that the left-most block of A_1 is an identity matrix. The requirement that $A_1 \Delta x + A_2 \beta = 0$ can be written as: $[I \ldots] \Delta x + A_2 \beta = 0$. Then one choice for Δx would be to take its first components to be equal to "$-A_2\beta$" and fix the remaining ones to zero.

Now, the last thing to do is to scale all these vectors to make sure that the new point is feasible for the new set of primal-dual problems.

The cuts returned by the oracle are central. This means that, after adding the cuts y is on the boundary of the new dual feasible set. If the cuts are linear constraints, then

$$A_2^T y = c_2.$$

After adding the cut this way all the slack corresponding to A_2 are equal to zero so y is on the boundary of the new dual feasible set.

There are two ways of adding an "SOCP" cut centrally. One is to do it as before. Because all the slack variables corresponding to A_2 are zero, the point is now at the vertex of the second order cone corresponding to A_2. Sometimes this might be too aggressive. Another way of adding the cut might be

$$A_2^T y + s = c_2 \text{ with } s_1^2 = \sum_{i \geq 2} s_i^2.$$

This way the cut is central but, after adding it the point is on a ray of the second order cone instead of being its vertex.

The second way of adding SOCP cuts is more useful when solving optimality

problems. This is because we want the vector c (used in defining the dual set) to be as close as possible to the original one, used in describing the objective function of the optimality problem. This aspect is irrelevant for the feasibility problems so, in this case we can use the first, more aggressive type of SOCP cuts.

4.3 Numerical Results

We tested our algorithm using Matlab v6.0 on a Windows XP, 1.8GHz Pentium 4 Pc with 512M RAM. For the most expensive parts of the algorithm we used mex files compiled with lcc 2.4, the Matlab's own C compiler. We used "UMFPACK 4.1" developed by Timothy A. Davis (see [3]) to compute the inverses of the matrices used by the Nesterov-Todd directions.

The test problems we solved are modified versions of the ones proposed in the "Seventh DIMACS Implementation Challenge - Semidefinite and Related Optimization Problems". This is a collection of conic programming optimization problems. We modified these problems into feasibility problems. The problems are organized in families.

Table 4.1: Details for the problems from the Seventh DIMACS Implementation Challenge

NAME	ROWS	QUADR	LIN
nql30	3680	[900; 900x3]	3602
nql60	14560	[3600;3600x3]	14402
nql180	130080	[32400; 32400x3]	129602
nql30old	3601	[900;900x3]	5560
nql60old	14401	[3600; 3600x3]	21920
nql180old	129601	[32400; 32400x3]	195360
nb-L1	915	[793; 793x3]	797
nb-L2	123	[839; 1x1667, 838x3]	4
nb	123	[793; 793x3]	4

The first column in this table contains the name of the problem. The second one indicates the number of rows used in describing the problem. The third column

gives the number of quadratic cones together with the dimension of the cones. The last column gives the number of linear constraints.

The only problems we couldn't solve were from the "qssp" and "scheduling" families of problems. This is because for our algorithm to work we need the matrices describing the cuts to have full rank. The "qssp" problems have linear dependency for the constraints corresponding to the quadratic cones. The "scheduling" problems have only one or two quadratic cones of high dimensions. Their corresponding matrices are rank deficient.

None of the problems we consider has a fully dimensional dual-feasible set. This is because all these problems contain linear constraints of the form $a^T y = 0$. This is forcing the dual feasible domain to be flat. In order to overcome this we increased all the values of c by a constant to eliminate all the zeros. This constant is in this case an upper-bound for the "thickness" of the domain. We tried different values for the constant. As expected, for bigger constants (i.e. the less flat the domain) the number of iterations required by the algorithm was smaller.

In order to improve the performance of the algorithm we are using weights for the constraints that are generated by the oracle. Usually we took the weight for a cut to be proportional to the size of the violation of that particular cut and the frequency the cut is returned by the oracle.

We are interested in the number of analytic centers generated in order to get a feasible point (the "AC's" column) and the total number of Newton steps generated (in "Newton" column). We will explore how the convergence of the algorithm depends on the number of cuts to be added, the size of the initial set Ω_0 and the thickness of the set of interest Γ (we are using the constant we used to change c as an upper bound for the thickness). Although time is important we should mention here that our code is inherently slow. This is because we are using Matlab that is much slower than the corresponding "C" or FORTRAN code. The biggest bottleneck is computing the inverse matrices for the Nesterov-Todd steps (this could take up to 90% of all running time).

We will consider three different cases for the thickness of the set Γ. The $\min |c|$ will be set to 10 (relatively "thick" set), 1 and .01 (relatively flat set). For each of

these cases we will use either a small (the size 10) starting set Ω_0 or a large one (the size 5000). Also for the number of cuts to be added at a point (in "No Cuts (%)" column) we will consider three different scenarios. We will add few cuts (the number of columns describing the cuts is no more than 5% of the number of rows of the problem), a moderate number of cuts (10% of the number of rows) and lots of cuts (50% of the total number of rows). Also we will register the total number of different cuts that are added in order to solve the problem and the highest frequency with which a cut is chosen by the oracle (in "Used (Freq)" column). We also register the percentage of the number of cuts used from the total number of constraints (in "Per." column).

As expected, the performance of the algorithm depends on the size of the first outer-approximation set Ω_0. The number of iterations is directly proportional to size of the ratio between the size of Ω_0 and the thickness of Γ. If this ratio is small, then the number of cuts added is significantly smaller than the total number of constraints.

The number of violated cuts to be added at each iteration influences the number of iterations too. It is better to add as many violated cuts as possible at a time instead of just a few ones. The quality of the oracle is really important here. If the oracle is fast in getting violated cuts then the strategy is to add many of them at each iteration, otherwise it is better to add a moderate number of them. In general, the total number of cuts used by the algorithm is the same, regardless of how many of them are added at each iteration.

The strategy of adding many cuts may backfire if the number of violated cuts is large but most of them are linearly dependent (as in the "nb" problem). In this case the oracle spends more time searching for independent cuts. So the oracle plays an important role here. That's why the structure of the problem (i.e the matrix A and vector c) should be considered when solving a particular type of problems.

As an example, all the SOCP constraints from the "nql" family of problems are linearly independent. Using this information in designing the oracle can decrease the overall running time by more than 75%.

So the best strategy to use is to start with a small initial set Ω_0 (which might be

Table 4.2: Numerical results for feasibility problems ($\min |c| = 10$)

Name	Size Ω_0	No Cuts (%)	AC's	Newton	Used (Freq)	Per.	Time
nql30	10	184 (5%)	25	87	744 (2)	0.16	29.37
		368 (10%)	14	48	744 (2)	0.16	16.84
		1840 (50%)	4	15	744 (2)	0.16	6.78
	5000	184(5%)	16	74	900 (1)	0.19	21.14
		368 (10%)	9	50	900 (1)	0.19	14.70
		1840(50%)	3	17	900 (1)	0.19	6.43
nql60	10	728 (5%)	29	114	3284 (2)	0.18	158.31
		1456 (10%)	15	59	3284 (2)	0.18	95.78
		7280 (50%)	4	16	3284 (2)	0.18	107.48
	5000	728 (5%)	16	92	3600 (1)	0.20	89.68
		1456 (10%)	9	51	3600 (1)	0.20	54.96
		7280 (50%)	3	18	3600 (1)	0.20	30.78
nql180	10	6504 (5%)	16	79	32400 (1)	0.20	1654.62
		13008 (10%)	9	44	32400 (1)	0.20	1076.68
		65040 (50%)	3	14	32400 (1)	0.20	810.10
	5000	6504 (5%)	16	154	32400 (1)	0.20	1764.11
		13008 (10%)	9	92	32400 (1)	0.20	1243.57
		65040 (50%)	3	26	32400 (1)	0.20	859.43
nb-L1	10	46 (5%)	12	44	156 (1)	0.09	298.75
		92 (10%)	7	22	161 (1)	0.10	130.94
		458 (50%)	5	18	238 (1)	0.15	53.62
	5000	46 (5%)	295	893	1448 (38)	0.91	1628.58
		92 (10%)	49	498	1448 (38)	0.91	915.47
		458 (50%)	53	158	1448 (38)	0.91	542.81
nb	10	7 (5%)	6	22	5 (1)	0.04	3.59
		13 (10%)	6	22	5 (1)	0.04	3.18
		62 (50%)	6	22	5 (1)	0.04	3.57
	5000	7 (5%)	2	8	1 (1)	0.01	1.98
		13 (10%)	2	8	1 (1)	0.01	1.85
		62 (50%)	2	8	1 (1)	0.01	1.81

expanded if needed) and to use an oracle that exploits the structure of the problem in order to generate as many linearly independent violated cuts as possible in the shortest amount of time.

Table 4.3: Numerical results for feasibility problems ($\min |c| = 1$)

Name	Size Ω_0	No Cuts (%)	AC's	Newton	Used (Freq)	Per.	Time
nql30	10	184 (5%)	13	39	744 (1)	0.16	11.09
		368 (10%)	8	23	744 (1)	0.16	7.86
		1840 (50%)	3	9	744 (1)	0.16	4.09
	5000	184(5%)	16	83	900 (1)	0.19	21.79
		368 (10%)	9	50	900 (1)	0.19	14.70
		1840(50%)	3	18	900 (1)	0.19	6.78
nql60	10	728 (5%)	15	46	3284 (1)	0.18	55.22
		1456 (10%)	8	25	3284 (1)	0.18	32.15
		7280 (50%)	3	10	3284 (1)	0.18	21.56
	5000	728 (5%)	16	106	3600 (1)	0.19	96.23
		1456 (10%)	9	57	3600 (1)	0.19	57.97
		7280 (50%)	3	19	3600 (1)	0.19	33.50
nql180	10	6504 (5%)	16	64	32400 (1)	0.19	1259.18
		13008 (10%)	9	36	32400 (1)	0.19	876.36
		65040 (50%)	3	12	32400 (1)	0.19	743.62
	5000	6504 (5%)	16	99	32400 (1)	0.19	1186.26
		13008 (10%)	9	82	32400 (1)	0.19	1096.10
		65040 (50%)	3	26	32400 (1)	0.19	983.17
nb-L1	10	46 (5%)	8	26	112 (1)	0.07	131.42
		92 (10%)	6	19	133 (1)	0.08	84.22
		458 (50%)	4	13	155 (1)	0.10	34.95
	5000	46 (5%)	428	1154	1556 (51)	0.98	2106.90
		92 (10%)	216	635	1556 (51)	0.98	1185.04
		458 (50%)	66	200	1556 (51)	0.98	625.34
nb	10	7 (5%)	2	6	1 (1)	0.01	1.84
		13 (10%)	2	6	1 (1)	0.01	1.81
		62 (50%)	2	6	1 (1)	0.01	1.71
	5000	7 (5%)	2	8	1 (1)	0.01	1.81
		13 (10%)	2	8	1 (1)	0.01	1.89
		62 (50%)	2	8	1 (1)	0.01	1.71

4.4 Solving Optimization Problems

The analytic center cutting plane method can be used not only for solving feasibility problems but also for solving optimality problems. To do this, the algorithm has to be incorporated in a larger interior-point scheme. Because we are going to use our analytic center cutting plane scheme, the problems we will consider must have a fully dimensional dual feasible set.

Table 4.4: Numerical results for feasibility problems ($\min |c| = 0.01$)

Name	Size Ω_0	No Cuts (%)	AC's	Newton	Used (Freq)	Per.	Time
nql30	10	184 (5%)	97	235	1344 (84)	0.29	116.21
		368 (10%)	92	210	1344 (84)	0.29	110.14
		1840 (50%)	87	185	1344 (84)	0.29	104.36
	5000	184(5%)	127	419	4502 (10)	1	336.12
		368 (10%)	69	272	4502 (10)	1	245.21
		1840(50%)	39	194	4502 (10)	1	222.79
nql60	10	728 (5%)	94	179	3639 (93)	0.20	394.84
		1456 (10%)	95	159	3639 (94)	0.20	370.73
		7280 (50%)	95	144	3639 (94)	0.20	392.74
	5000	728 (5%)	126	579	18002 (9)	1	2429.08
		1456 (10%)	81	595	18002 (9)	1	2592.14
		7280 (50%)	76	544	18002 (10)	1	2511.97
nql180	10	6504 (5%)	20	90	32402 (3)	0.20	2199.56
		13008 (10%)	13	55	32402 (3)	0.20	1532.60
		65040 (50%)	7	27	32402 (3)	0.20	1366.12
	5000	6504 (5%)					> 10000
		13008 (10%)					> 10000
		65040 (50%)					> 10000
nb-L1	10	46 (5%)	92	346	1576 (4)	0.99	675.67
		92 (10%)	47	176	1576 (4)	0.99	362.04
		458 (50%)	22	74	1578 (4)	0.99	296.56
	5000	46 (5%)	604	1644	1590 (66)	1	2988.47
		92 (10%)	312	936	1590 (65)	1	1723.00
		458 (50%)	87	294	1590 (66)	1	858.29
nb	10	7 (5%)	18	41	6 (10)	0.05	6.57
		13 (10%)	18	41	6 (10)	0.05	7.68
		62 (50%)	18	41	6 (10)	0.05	5.84
	5000	7 (5%)	23	68	6 (20)	0.05	5.29
		13 (10%)	23	68	6 (20)	0.05	10.09
		62 (50%)	23	68	6 (20)	0.05	6.40

The problem we are considering is given by:

$$\begin{array}{rlll} \min & c^T x & & \\ \text{subject to} & Ax & = & b, \qquad (P) \\ & x & \geq_K & 0 \end{array}$$

together with its dual:

$$\begin{array}{rlcll} \max & b^T y & & & \\ \text{subject to} & A^T y + s & = & c, & (D) \\ & s & \geq_K & 0 & \end{array}$$

Here $\geq_K$ refers to the fact that the cones used are $\oplus$ sums of second order cones and $I\!R^n_+$. In this case the feasible set of the dual problem will be the equivalent of the Γ set from a pure feasibility problem.

In our approach sometimes the current-point will be outside of Γ. We say we have a solution to the primal-dual optimization problem if the duality gap is smaller than a given tolerance τ_{dg} and the point is feasible in Γ.

As long as the current point is outside Γ we proceed with an analytic center cutting plane scheme, generating a sequence of outer-approximations sets of Γ by adding central cuts through the θ - analytic centers of those sets. Once the point becomes feasible in Γ we start an interior-point scheme. This means that we scale the logarithmic part of the barrier functional by a positive parameter μ. At the beginning the barrier functionals weren't scaled so, in this case, $\mu = 1$. As soon as the point becomes feasible in Γ we decrease μ:

$$\mu = (1 - \Theta)\mu,$$

with $0 < \Theta < 1$ a given fixed constant. After decreasing μ we compute the θ - analytic center for the same set but with a different scaled barrier functional. We keep decreasing μ until either the duality gap becomes smaller than τ_{dg} or the current point gets out of Γ. In the later case we start the new iteration with a step from the analytic center cutting plane scheme.

If Γ is not fully dimensional then the scheme will never get to decrease μ to zero, driving the point towards the solution. In this case we can start decreasing μ as soon as the distance to Γ becomes smaller than a preset value τ_{feas}.

In the final chapter we will consider the LP-only version of this algorithm (so there are no second order cones involved in describing the primal-dual optimization problem). There we will prove that this algorithm converges and we will give an

estimate for the total number of θ - analytic centers required to be generated before obtaining a solution.

For a point (x, y, s) generated by the algorithm we define the total violation to be absolute value of the sum of all negative slacks for the current point. Before presenting the results, let's introduce the algorithm.

The Algorithm

Input

setup the initial set Ω_0

initialize the point (x, y, s)

get the θ - analytic center for Ω_0

compute the duality gap

get the total violation

while |duality gap| $\geq \tau_{dg}$ or total violation $> \tau_{feas}$

call the oracle at (x, y, s)

if the point is feasible in Γ

decrease μ ($\mu := (1 - \Theta)\mu$)

get the new θ - analytic center

compute the duality gap and total violation

else the point is outside Γ

add the cuts, generate a new, smaller outer-approximation set for Γ

get the new θ - analytic center

compute the duality gap and total violation

return

STOP

The performance of the algorithm depends on the "thickness" of the dual feasible set, the size of the first outer-approximation set Ω_0 and the quality of the oracle. It is better to start with a rather small initial outer-approximation set (and expand it later if needed) instead of starting with a large one. This observations are similar to the ones for the case of pure feasibility problems.

We will attempt to solve some of the problems from the "Seventh DIMACS Implementation Challenge". This time we will consider only some of the cases we analyzed in the previous section.

As we said before, the dual-feasible sets for these problems are not fully dimensional. So we had to change the vector c to transform those sets in fully-dimensional ones.

Table 4.5: Numerical results for some modified problems suggested in the "Seventh DIMACS Implementation Challenge"

Name	Size Ω_0	No AC's	Newton sts	Cuts (Freq.)	Used/Total	Time (s)
nql30	10	326	858	2791 (20)	0.61	494.81
nql30	5000	149	275	3975 (20)	0.88	262.05
nql60	10	424	889	16830 (57)	0.93	3150.48
nql60	5000	318	654	16454 (22)	0.91	2904.20
nb	10	561	646	252 (11)	0.31	248.07
nb	500	604	1063	258 (12)	0.32	376.86
nb-L1	10	217	271	794 (20)	0.49	1331.42
nb-L1	500	322	639	1557 (51)	0.97	2562.84

There is no clear strategy on how to decrease μ. In our tests we decrease μ by 50%. The duality gap cannot become too small (less than 10^{-6} - 10^{-7}) without running in numerical problems. This happens because, when the algorithm is getting close to the optimal solution, some of the slacks become really small (at optimality they are in fact zero) and some might be quite large. This creates difficulties for the analytic center cutting plane scheme when computing the Hessian of the barrier functional and its inverse.

The number of cuts added by the algorithm is sometimes more than 80% from the total number of constraints. If the initial outer-approximation set Ω_0 is chosen carefully enough, this number can be dropped below 65%. This can save memory and space usage.

The algorithm proposed here works better when dealing with problems where the dual-feasible set is not known before hand and the oracle generates violated cuts as needed. If the dual-feasible set is completely described from the beginning, then our algorithm will systematically perform worse than some other interior-point solvers. For really big/difficult problems it is possible for the other solvers to fail while our

algorithm might be able to generate at least some strictly feasible points.

We conclude this chapter by considering different instances of a problem that appears when solving sparse Partial Least Squares (PLS) problems and its kernel version (KPLS) (see [10], [20] and [21] for more details).

The problem we are considering is given by:

$$\begin{array}{ll} \min_{w,s,\xi,\varepsilon} & \frac{1}{\nu m}\sum_i \xi_i + \varepsilon \\ \text{subject to} & \frac{1}{2}\|x_i - s - u_i w\| - \xi_i \leq \varepsilon, \qquad (P) \\ & \xi_i \geq 0, i = 1, \ldots, m. \end{array}$$

Here, x_i, s, w are points in $\mathbb{R}^n$.

Each of the constraints involved in describing the feasible region is equivalent to:

$$\begin{bmatrix} 2\varepsilon + 2\xi_i \\ x_i - s - u_i w \end{bmatrix} \in K_{n+1}, \tag{4.16}$$

with K_{n+1} an $n+1$ dimensional second order cone.

Using this observation we can write the problem in a primal-dual formulation:

$$\begin{array}{ll} \max & b^T y \\ \text{subject to} & A_i^T y + \sigma_i = c_i, \qquad (P) \\ & B^T y \leq 0, \\ & \sigma_i \in K_{n+1}, i = 1, \ldots, m. \end{array}$$

$$\begin{array}{ll} \min & \sum_{i=1}^{m} c_i^T X_i \\ \text{subject to} & \sum_{i=1}^{m} A_i X_i + B\hat{X} = b, \qquad (D) \\ & \hat{X} \in \mathbb{R}^m_+, X_i \in K_{n_i+1}, i = 1, \ldots, m. \end{array}$$

Here:

$$\begin{aligned} & y = [\epsilon, \xi_1, \ldots, \xi_m, s^T, w^T]^T \in \mathbb{R}^{2n+m+1}, \\ & X_i \in K_{n+1}, i = 1, \ldots, m, \end{aligned}$$

$$b = [-1, -\frac{1}{\nu m} e^T, 0_{1\times 2n}]^T \in I\!R^{2n+m+1},$$
$$c_i = [0,\ x_i^T]^T \in I\!R^{n+1},$$
$$B^T = [0_{m\times 1} \vdots -I_m \vdots 0_{m\times n} \vdots 0_{m\times n}],$$
$$A_i = \begin{bmatrix} -2 & \vdots & \\ & \vdots & 0_{(m+1)\times n} \\ -2e_i & \vdots & \\ \dots & & \dots \\ & \vdots & I_n \\ 0_{2n\times 1} & \vdots & \dots \\ & \vdots & u_i I_n \end{bmatrix} \in I\!R_{(2n+m+1)\times(n+1)},$$

with e the vector of all ones, e_i the vector of all zeros with the i-th component equal to 1 and I_n the identity matrix.

This problem has the dual feasible region fully dimensional so we can start solving the problem directly, without modifying the vector c (as in previous cases).

We can use our algorithm to solve the initial problem written in the above form. The performance of the algorithm when solving this type of problem depends on both, the size of m (the number of points x_i) and n (the dimensionality of the second order cones involved).

In our analysis we will look at how the size of m and n influence the performance of the algorithm.

We generate a set of problems of different sizes. The points x_i for all these problems have the coordinates between 0 and 1. The weights u_i are random numbers between 1 and 6. Also we took $\nu = 0.5$.

In solving these problems, we took the set Ω_0 to be a cube centered at 0 and with the side length equal to 10. The number of violated cuts to be added is no more than 85% of the total number of rows of the matrix A describing the problem. Also the parameter μ is decreased each time by 75%. The algorithm stops with an approximate solution when the duality gap is smaller than 10^{-6}.

For each problem we registered the number of points and their dimensionality in the "Size" column (i.e. $[100 \times 3]$ describes a problem with 100 points, each point

Table 4.6: Numerical results for PLS subproblems

Size	No AC's	Newton sts	Cuts (Freq.)	Used/Total	Time (s)
$[100 \times 3]$	132	293	171 (8)	0.85	29.43
$[100 \times 5]$	165	563	166 (12)	0.83	51.31
$[100 \times 25]$	183	504	158 (10)	0.79	185.17
$[500 \times 3]$	190	587	799 (13)	0.79	165.92
$[500 \times 5]$	185	438	871 (11)	0.87	227.71
$[500 \times 25]$	195	615	746 (9)	0.74	1396.12
$[1000 \times 3]$	159	395	1296 (9)	0.64	218.26
$[1000 \times 5]$	170	442	1381 (9)	0.69	352.26
$[1000 \times 25]$	226	746	1489 (13)	0.74	4039.45
$[5000 \times 3]$	176	462	6576 (9)	0.65	948.70
$[5000 \times 5]$	195	582	6936 (9)	0.69	1379.97
$[10000 \times 3]$	203	462	13005 (9)	0.65	2653.94
$[10000 \times 5]$	200	582	13746 (9)	0.68	3807.11

having dimension 3). The number of analytic centers generated in order to solve the problem are given in the second column. We also registered the number of Newton steps, the total number of different cuts added (together with the highest frequency a cut was used) and the time for each problem. We also registered in the "Used/Total" column the percentage of the number of cuts used by the algorithm from the total number of constraints.

As expected, the number of iterations and the number of Newton steps is proportional to the size of the problem and the dimension of the SOCP cones. The percentage of the number of cuts used from the total number of cuts used in describing a problem is slightly decreasing while the size of the problem increases. This makes the algorithm more efficient for bigger problems.

The performance of the algorithm depends on the size of the SOCP cones. This is because the Hessian associated to each cone is fully dense and its size is equal to the dimension of the cone.

CHAPTER 5
An Interior-Point Method Approach to Solving Linear Programming Problems

5.1 Introduction

In this chapter we propose an algorithm for solving linear programming problems that involves both an interior-point approach and the analytic center cutting planes method studied in the previous chapters. This algorithm is an extension of the algorithms proposed by Andersen et al. [1] and Terlaky et al. [16], [19]. The problem studied here is:

$$\begin{array}{rrcll} \min & \tilde{c}^T x & & & \\ \text{subject to} & \tilde{A}x & = & b & \qquad (\tilde{P}) \\ & x & \geq & 0 & \end{array}$$

together with its dual:

$$\begin{array}{rrcll} \max & b^T y & & & \\ \text{subject to} & \tilde{A}^T y + s & = & \tilde{c} & \qquad (\tilde{D}) \\ & s & \geq & 0 & \end{array}$$

with $\tilde{A} \in I\!R^{m \times \tilde{n}}$, full-rank matrix, $x, s, \tilde{c} \in I\!R^{\tilde{n}}$ and $\tilde{b}, y \in I\!R^m$.

Our goal is to solve the dual problem $(\tilde{D})$.

Let $\Gamma = \{y \in I\!R^m; \tilde{A}^T y < \tilde{c}\}$ be the feasible region for the dual problem. As in the general case, we will assume that Γ contains a small ball of radius 2^{-L} and it is bounded and contained in an m dimensional cube

$$C = \{y \in I\!R^m; -e \leq y \leq e\}.$$

We also assume, without loss of generality that $\tilde{c}$ and the rows of $\tilde{A}$ have the 2-norm equal to one.

The existence of an oracle is assumed. For each point $\overline{y} \in C$ the oracle either recognize that $\overline{y} \in \Gamma$ or returns a set of p central cuts:

$$B^T y \leq B^T \overline{y},\ B \in I\!R^{m \times p} \tag{5.1}$$

with $\Gamma \subseteq C \cap \{y \in I\!R^m; B^T y \leq B^T \overline{y}\}$. We assume the rows of B have norm equal to one.

In solving $(\tilde{P}) - (\tilde{D})$ we will find the μ - analytic centers (they will be defined later) for a sequence of polytopes P_i containing Γ ($\Gamma \subset P_i \subset P_{i-1}$). We start with P_0 being the m-dimensional cube C. We find its μ - analytic center (μ - AC) and call the oracle. If $AC \notin \Gamma$ add the cuts returned by the oracle to P_0 and find the AC for the new domain P_1. We keep doing this until, at the k-th iteration, the analytic center of P_k is in the interior of Γ. At this stage, we will take some primal-dual steps (with the parameter μ) along the central-path of P_k until the point gets out of Γ. We then call the oracle at the current point and use analytic centers to generate tighter outer-approximations P_i for Γ until again the analytic center of one P_i is in the interior of Γ. We will keep decreasing μ as the algorithm progresses. We stop with infeasibility if the number of cuts added exceeds a certain value or when the polytope P_i becomes too flat.

We have a solution when the current point is in Γ and $(n + 2m)^\gamma \mu$ is smaller than a small parameter ε. Here n is the total number of cuts that get added. γ is a constant between 0 and 0.5. The reason for choosing this stopping criteria instead of the classical one ($(n + 2m)\mu < \varepsilon$) will be explained towards the end of the chapter.

5.2 Notations and Conventions

Throughout this chapter the vector norm used is the usual one: $\|u\| = \sqrt{\sum_{i=1}^{n} u_i^2}$, for any vector $u \in I\!R^n$.

For a vector $u \in I\!R^n$ we define the n-dimensional vectors u^k, $\sqrt{u}$ and $\frac{1}{u}$ by:

$$(u^k)_i = u_i^k,\ (\sqrt{u})_i = \sqrt{u_i} \text{ and } (\frac{1}{u})_i = \frac{1}{u_i} \text{ for } i = 1, \ldots, n.$$

Because we use a mixture of an interior-point method and an analytic center cutting planes method we will need to have a means to estimate the distance from the current point to the analytic center of a domain or its central path. We will use two different proximity measures: δ_M that describes the distance of the point to the analytic center and δ_T describing the distance to the central path:

$$\delta_M(x, s, \mu) = \|\frac{xs}{\mu} - e\|$$

and

$$\delta_T(x, s, \mu) = \frac{1}{2}\|\sqrt{\frac{xs}{\mu}} - \sqrt{\frac{\mu}{xs}}\|$$

where $x, s \in I\!R^n_+$ and $\mu \in I\!R_+$. Here, xs is the Hadamard product of x and s (i.e. $xs \in I\!R^n$ with $(xs)_i = x_i s_i$, $i = 1 \ldots n$). Notice that $xs = XSe$ with $e \in I\!R^n$ being the vector of all ones and X and S the diagonal matrices corresponding to x and s. We will use δ_T as an overall proximity measure.

We will call a feasible point (x, s, μ) approximately centered if it satisfies the inequality $\delta_T(x, s, \mu) \leq \tau$ for some constant τ.

5.3 Dikin's Ellipsoids

In order for the analytic center cutting planes scheme to work we need to make sure that, after adding the cuts the direction we choose to recover feasibility will move the point inside the feasible region. Before the oracle is called the current point is feasible in the outer approximation set P_i. Adding the cuts does not impact the feasibility with respect to the constraints defining P_i. The point is still feasible with respect to the old constraints. The infeasibility is related only to the new added cuts.

So, when choosing the direction that brings the point inside the new outer approximation set, we need to make sure that the point remains feasible with respect to the old cuts.

It is easy to see that if we start with a feasible point $x > 0$, then $x + \Delta x > 0$ if $||x^{-1}\Delta x|| < 1$. This inequality defines the Dikin ellipsoid. So if we move along any vector from a Dikin's ellipsoid centered at a feasible point x we stay feasible.

In this section we will introduce some results regarding these ellipsoids circumscribing or contained in the feasible region $\mathcal{F}_P$. All results are similar to the ones from Goffin and Vial [5] (section 3.2), the difference being that we deal with μ - analytic centers instead of pure analytic centers. The presence of μ is the only difference.

For a point $x \in \text{int}\mathcal{F}_P$ we define an ellipsoid containing x and inscribed in $\mathcal{F}_P$ by:

$$\mathcal{E}_P = \{\Delta x : A\Delta x = 0, \|X^{-1}\Delta x\| \leq 1\}. \tag{5.2}$$

Similarly we define the Dikin ellipsoid corresponding to a point $s \in \text{int}\Gamma$, inscribed in Γ:

$$\mathcal{E}_D = \{\Delta s : \Delta s = -A^T\Delta y, \|S^{-1}\Delta s\| \leq 1\}. \tag{5.3}$$

Lemma 23 *Let (x, s) be a μ-center and $D = X^{\frac{1}{2}}S^{-\frac{1}{2}}$. Then:*

1. $\sqrt{\mu(1-\theta)}\mathcal{E}_P \subset \{\Delta x : A\Delta x = 0, \|D^{-1}\Delta x\| \leq 1\} \subset \sqrt{\mu(1+\theta)}\mathcal{E}_P,$

2. $\sqrt{\mu(1-\theta)}\mathcal{E}_D \subset \{\Delta s : \Delta s = -A^T\Delta y, \|D\Delta s\| \leq 1\} \subset \sqrt{\mu(1+\theta)}\mathcal{E}_D.$

Proof: Follows from Goffin and Vial [5]. □

5.4 Adding Cuts in a Primal-Dual Interior-Point Scheme with Long Step Updates

In solving $(\tilde{P})-(\tilde{D})$ using the algorithm we propose we will have to take sequences of interior primal-dual steps. Each time the current AC is in the interior of Γ we set up a problem $(P)-(D)$:

$$\begin{array}{rrcll} \min & c^T x & & & \\ \text{subject to} & Ax & = & b & \qquad (P) \\ & x & \geq & 0 & \end{array}$$

together with its dual:

$$\begin{array}{rrcll} \max & b^T y & & & \\ \text{subject to} & A^T y + s & = & c & \qquad (D) \\ & s & \geq & 0 & \end{array}$$

with A and c containing the constraints corresponding to the initial cube C and the cuts added as the algorithm progressed. Instead of solving $(P)-(D)$ to optimality we stop as soon as the current point gets out of Γ.

The interior-point scheme used is the one suggested by Peng et al. in [16] (a long step primal-dual scheme). Solving (P) and (D) is equivalent to solving the next system:

$$\begin{aligned} Ax &= b, \quad x \geq 0 \\ A^T y + s &= c, \quad s \geq 0 \\ xs &= 0. \end{aligned}$$

In an interior-point algorithm, the last equation is replaced by: $xs = \mu e$, with

$\mu > 0$. Next, a sequence of systems of equations is generated:

$$\begin{aligned} Ax &= b, \quad x \geq 0 \\ A^T y + s &= c, \quad s \geq 0 \\ xs &= \mu e. \end{aligned}$$

The solution $(x(\mu), y(\mu), s(\mu))$ for this system of equations describes the central path of (P) and (D) towards the solution for the original problem. As $\mu \to 0$ the solution $(x(\mu), y(\mu), s(\mu))$ moves along the central path, and at the limit, $xs = 0$, giving the optimal solution for (P) and (D).

For each μ, a sequence of damped Newton steps are taken moving the point close to the central path (i.e. $\delta_T(x, s, \mu) \leq \tau$). The Newton steps are taken along the directions Δx, Δy, Δs given by the solution to the modified Newton equation system (5.4) as given by Peng et al. in [16]:

$$\begin{aligned} A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0, \\ s\Delta x + x\Delta s &= \mu^{1+\frac{\eta}{2}} \frac{e}{(xs)^{\frac{\eta}{2}}} - xs, \eta > 0. \end{aligned} \tag{5.4}$$

The step is scaled by a damping parameter $\alpha > 0$ (the step length). In a regular interior-point scheme, the damped Newton steps are taken until $\delta_T(x, s, \mu)$ becomes smaller than a given parameter τ. Then, μ is reduced by a factor of $(1 - \Theta)$ (with $0 < \Theta < 1$) and the procedure is repeated until either the point becomes infeasible in Γ or $(n + 2m)^\gamma \mu \leq \varepsilon$ when the algorithm stops with the approximate solution for the problem.

In the approach we propose, right before μ is decreased, the oracle is called to check for the feasibility (with respect to Γ) of the current point. If the point is not feasible the oracle returns a set of p ($p < m$) violated cuts. These cuts are added right through the current iterate, increasing the dimensionality of the problem and changing the feasibility for the current point. Also we keep track of μ. When we will set up a new problem $(P) - (D)$, μ takes the last value it had when we took

primal-dual steps.

Next we will analyze how the feasibility can be recovered in an efficient way, and the impact the new cuts have on the proximity measure $\delta_T(x, s, \mu)$.

Let $(\overline{x}, \overline{y}, \overline{s})$ be the point $(x(\mu), y(\mu), s(\mu))$ right before the cuts are added. This point is strictly feasible and is close to the central path of the relaxation problem:

$$\begin{aligned} A\overline{x} &= b, \\ A^T\overline{y} + \overline{s} &= c, \\ \delta_T(\overline{x}, \overline{s}, \mu) &= \left\| \sqrt{\frac{\overline{xs}}{\mu}} - \sqrt{\frac{\mu}{\overline{xs}}} \right\| \leq \tau, \\ \overline{x}, \overline{s} &> 0. \end{aligned}$$

For this point the oracle returns p-central cuts of the form:

$$a_{m+j}^T \overline{y} \leq a_{m+j}^T y, j = 1, \ldots, p, \forall y \text{ - feasible.}$$

Let $B = (a_{m+1}, a_{m+2}, \ldots, a_{m+p})$ be the matrix corresponding to the new added cuts. After adding the cuts, the current point becomes $(\hat{x}, \hat{y}, \hat{s})$ with

$$\hat{x} = \begin{bmatrix} \overline{x} \\ 0 \end{bmatrix}, \hat{y} = \overline{y}, \hat{s} = \begin{bmatrix} \overline{s} \\ 0 \end{bmatrix}, \text{ with } \hat{x}, \hat{s} \in I\!R^{n+p} \text{ and } \hat{y} \in I\!R^m.$$

The new primal and dual feasible regions are described by:

$$\begin{aligned} A^T y + s &= c, \\ B^T y + \gamma &= B^T \overline{y}, \\ Ax + B\beta &= b, \\ x, s, \beta, \gamma &\geq 0. \end{aligned}$$

After introducing the cuts, $(\hat{x}, \hat{y}, \hat{s})$ is lying on the boundary of the new feasible region. First a step to recover strict feasibility must be taken. For more flexibility, a scaled step is taken instead of a full step.

Let $\begin{bmatrix} d_x \\ \beta \end{bmatrix}$, d_y and $\begin{bmatrix} d_s \\ \gamma \end{bmatrix}$ be the directions used for getting back into the feasible region with $d_x, d_s \in I\!R^{n+2m}$, $d_y \in I\!R^m$ and $\beta, \gamma \in I\!R^p$. Let α_P and α_D be the scaling factors. The point is moved from $(\hat{x}, \hat{y}, \hat{s})$ to:

$$\begin{bmatrix} \overline{x} + \alpha_P d_x \\ \alpha_P \beta \end{bmatrix}, \overline{y} + \alpha_D d_y, \begin{bmatrix} \overline{s} + \alpha_D d_s \\ \alpha_D \gamma \end{bmatrix}.$$

The vectors d_x, d_y and d_s are the Dikin's directions used by Goffin and Vial in [5] which are based on the directions introduced by Mitchell and Todd (for one additional constraint case). Also β and γ are similar to the ones used in the paper mentioned above. So, by using a similar approach, it can be proven that the AC of the new region can be reached in $\mu O(p(\ln(p+1))$ steps (for details see *Section* 5.6).

Next, let's analyze the recovery of feasibility step. Before taking the step, the point $(\hat{x}, \hat{y}, \hat{s})$ is on the boundary of the feasible region. By taking this step the point is moved back, inside the feasible region. So, for the new point, the following set of equalities must hold:

$$\begin{aligned} A(\overline{x} + \alpha_P d_x) + \alpha_P B\beta &= b, \\ A^T(\overline{y} + \alpha_D d_y) + \overline{s} + \alpha_D d_s &= c, \\ B^T(\overline{y} + \alpha_D d_y) + \alpha_D \gamma &= B^T \overline{y}. \end{aligned} \tag{5.5}$$

Using the fact that $(\overline{x}, \overline{y}, \overline{s})$ was feasible for the problem before adding the cuts and that the scaling factors α_P and α_D are strictly positive these equations can be simplified to:

$$\begin{aligned} Ad_x + B\beta &= 0, \\ A^T d_y + d_s &= 0, \\ B^T d_y + \gamma &= 0. \end{aligned} \tag{5.6}$$

Using the arguments of from Goffin and Vial [5] (*Theorem* 4.2) we can choose the feasibility directions:

$$\begin{aligned} d_x &= -D^2A^T(AD^2A^T)^{-1}B\beta, \\ d_y &= -(AD^2A^T)^{-1}B\beta, \\ d_s &= A^T(AD^2A^T)^{-1}B\beta, \\ \gamma &= B^T(AD^2A^T)^{-1}B\beta. \end{aligned}$$

with $D = \overline{X}^{\frac{1}{2}}\overline{S}^{-\frac{1}{2}}$ and β being defined as the unique solution to:

$$\max\left\{-\frac{p}{2}\beta^T V\beta + \sum_{i=1}^{p}\log\beta_i\right\} \text{ with } V = B^T(AD^2A^T)^{-1}B.$$

In fact it turns out that β and γ are solutions for:

$$\begin{aligned} &\max\left\{-\sum_{i=1}^{p}\log\beta_i : \beta \geq 0, Ad_x + B\beta = 0, \|D^{-1}d_x\| \leq 1\right\}, \\ &\max\left\{-\sum_{i=1}^{p}\log\gamma_i : \gamma \geq 0, B^Td_y + \gamma = 0, \|DA^Td_y\| \leq 1\right\}. \end{aligned}$$

These problems are well posed and have unique solutions. Among the KKT conditions these solutions verify are:

$$\begin{aligned} \|D^{-1}d_x\| &= 1, \\ \|Dd_s\| &= 1. \end{aligned} \tag{5.7}$$

Once back in the feasible region we take a sequence of Newton steps towards the central-path. Once close to the central-path, we call the oracle again. If the point is not in Γ then we add the p - central cuts, take the directions described before to get back strictly interior, and then get the analytic center of the new region. This procedure is repeated until the μ - analytic center gets back in Γ when we will start a new sequence of primal-dual interior steps (this time the μ - analytic center is already on the central path so we don't need to re-initialize it).

5.5 The Algorithm

The algorithm contains two main parts: one part corresponding to the Primal-Dual step, and one in which the cuts are added and feasibility is recovered.

The Algorithm

Input

a proximity parameter $\tau = \sqrt{2m}$

an accuracy parameter $\varepsilon > 0$

a scaling parameter α for the primal-dual step

scaling parameters α_P and α_D for the 'recovery of feasibility' step

a fixed barrier update parameter Θ, $0 < \Theta < 1$

a constant γ, $0 < \gamma < 0.5$

$\mu = 1$, $n = 0$, $y = 0$ (the initial analytic center)

$A = [I_m \ , -I_m]$, $c^T = [e^T \ , e^T]$

begin

while $(n + 2m)^\gamma \mu \geq \varepsilon$ and $y \notin \Gamma$

get μ -analytic center (x, y, s)

call oracle at y

if $y \in \Gamma$

$\mu = (1 - \Theta)\mu$

else

add p - central cuts

$n = n + p$

take feasibility step

get μ - analytic center

end if

end while

STOP with the solution

end

5.6 The Recovery of Feasibility Step

We encounter the problem of recovery of feasibility in two circumstances. One is after a sequence of primal-dual interior steps and the current point just got out of the feasible region. The other case is right after p - central cuts are added through a μ - analytic center. Both these circumstances are similar, so we will not differentiate between these two cases in the next analysis.

Now we are ready to analyze what the adding of cuts changes in the problem. There are two things we should worry about when choosing the directions, and the scaling factors: one is to ensure that we are getting a feasible point, and the other is to keep track of the change in the proximity measure.

It is easy to check that the directions proposed verify the system (5.6). Also, both β and γ are strictly positive (see Theorem 4.2 and formulas (11) and (12) from [5]). The only thing to ensure is:

$$\overline{x} + \alpha_P d_x > 0 \text{ and } \overline{s} + \alpha_D d_s > 0.$$

The next two lemmas will provide us with bounds for α_P and α_D that will keep the above inequalities true.

Lemma 24 *Let $\overline{\beta}$ be the maximizer for $F(\beta) = -\frac{p}{2}\beta^T V\beta + \sum_{i=1}^{p} \ln \beta_i$. Then*

$$\overline{\beta}^T V\overline{\beta} = 1 \text{ and } \overline{\beta}(V\overline{\beta}) = \frac{1}{p}e.$$

Proof: The function $F(\beta)$ is concave and it is maximized over the convex set $I\!R^m_+$. At $\overline{\beta}$ we have that: $\nabla F(\overline{\beta}) = 0$. But $\nabla F(\beta) = -pV\beta + \frac{1}{\beta}$. So, for $\overline{\beta}$ we have $pV\overline{\beta} = \frac{1}{\overline{\beta}}$ or, equivalently, $V\overline{\beta} = p^{-1}\frac{1}{\overline{\beta}}$. The result follows immediately. □

So far we did not use the fact that prior to adding the cuts

$$\overline{\delta} := \delta_T(\overline{x}, \overline{s}, \mu) \leq \tau.$$

Let's analyze how this bound affects the size of the component-wise products

$\overline{x}_i\overline{s}_i$. We have

$$\overline{\delta}^2 = \frac{1}{4}\sum_{i=1}^{n}\left(\frac{\overline{x}_i\overline{s}_i}{\mu} + \frac{\mu}{\overline{x}_i\overline{s}_i} - 2\right). \tag{5.8}$$

Each term of the sum is positive and $\overline{\delta} \leq \tau$ implies that

$$\frac{\overline{x}_i\overline{s}_i}{\mu} + \frac{\mu}{\overline{x}_i\overline{s}_i} - 2 \leq 4\tau^2, i = 1, \ldots, n. \tag{5.9}$$

We can rewrite this as:

$$(\overline{x}_i\overline{s}_i)^2 - \mu(2 + 4\tau^2)\overline{x}_i\overline{s}_i + \mu^2 \leq 0, i = 1, \ldots, n. \tag{5.10}$$

These inequalities hold if and only if:

$$\mu(1 + 2\tau^2 - 2\tau\sqrt{\tau^2 + 1}) \leq \overline{x}_i\overline{s}_i \leq \mu(1 + 2\tau^2 + 2\tau\sqrt{\tau^2 + 1}) \text{ for } i = 1, \ldots, n \tag{5.11}$$

or:

$$\mu(\sqrt{\tau^2 + 1} - \tau)^2 \leq \overline{x}_i\overline{s}_i \leq \mu(\sqrt{\tau^2 + 1} + \tau)^2 \text{ for } i = 1, \ldots, n. \tag{5.12}$$

Using these inequalities we can prove the following lemma.

Lemma 25 *1. If* $0 < \alpha_P < \sqrt{\mu}(\sqrt{\tau^2 + 1} - \tau)$ *then* $\overline{x} + \alpha_P d_x > 0$.

2. If $0 < \alpha_D < \sqrt{\mu}(\sqrt{\tau^2 + 1} - \tau)$ *then* $\overline{s} + \alpha_D d_s > 0$.

Proof: Let's notice that if $\|v\| = \alpha$ then all the components of v satisfy $-\alpha \leq v_i \leq \alpha$.

1. $\overline{x} + \alpha_P d_x = \overline{X}e - \alpha_P D^2 A^T (AD^2A^T)^{-1}B\beta$. If we substitute $D = \overline{X}^{\frac{1}{2}}\overline{S}^{-\frac{1}{2}}$ and factorize then:

$$\overline{x} + \alpha_P d_x = D(\overline{X}^{\frac{1}{2}}\overline{S}^{\frac{1}{2}}e - \alpha_P DA^T(AD^2A^T)^{-1}B\beta). \tag{5.13}$$

Now

$$
\begin{aligned}
& [DA^T(AD^2A^T)^{-1}B\beta]^T[DA^T(AD^2A^T)^{-1}B\beta] \\
= \quad & \beta^T B^T((AD^2A^T)^{-1})^T AD^2A^T(AD^2A^T)^{-1}B\beta = \beta^T V\beta = 1.
\end{aligned}
$$

So $\|DA^T(AD^2A^T)^{-1}B\beta\| = 1$. This implies that:

$$-e \leq DA^T(AD^2A^T)^{-1}B\beta \leq e.$$

Finally

$$\overline{x} + \alpha_P d_x \geq (\sqrt{\mu}(\sqrt{\tau^2+1} - \tau) - \alpha_P)De > 0 \tag{5.14}$$

if $0 < \alpha_P < \sqrt{\mu}(\sqrt{\tau^2+1} - \tau)$.

2. Similarly,

$$
\begin{aligned}
\overline{s} + \alpha_D d_s &= \overline{S}e + \alpha_D A^T(AD^2A^T)^{-1}B\beta \\
&= D^{-1}(\overline{X}^{\frac{1}{2}}\overline{S}^{\frac{1}{2}}e + \alpha_D DA^T(AD^2A^T)^{-1}B\beta) \\
&\geq (\sqrt{\mu}(\sqrt{\tau^2+1} - \tau) - \alpha_D)D^{-1}e > 0
\end{aligned}
$$

if $0 < \alpha_D < \sqrt{\mu}(\sqrt{\tau^2+1} - \tau)$.

□

Now that the feasibility is insured we have to worry about the change in $\overline{\delta}$. The next lemma will be of help for our analysis. It gives us a way to compare the proximity measures before and after adding the cuts (i.e. after changing the dimensionality of the problem).

Lemma 26 *Let* $\delta = \frac{1}{2}||v - v^{-1}||$ *and* $\delta_+ = \frac{1}{2}\left\| \begin{bmatrix} v \\ u \end{bmatrix} - \begin{bmatrix} v \\ u \end{bmatrix}^{-1} \right\|$. *Then*

$$\delta_+^2 = \delta^2 + \frac{1}{4}\left\|u - u^{-1}\right\|^2, \text{ for any } v, u. \tag{5.15}$$

Proof: The proof is almost trivial.

□

Let $\overline{\delta}$ be the proximity measure right before adding in the cuts and δ_+ the proximity value after the cuts have been added in and the step for the recovery of feasibility has been taken. We want to find an upper bound for δ_+ knowing that before adding the cuts $\overline{\delta} \leq \tau$.
If we take:

$$\delta_1 = \frac{1}{2}\left\| \sqrt{\frac{(\overline{x}+\alpha_P d_x)(\overline{s}+\alpha_D d_s)}{\mu}} - \sqrt{\frac{\mu}{(\overline{x}+\alpha_P d_x)(\overline{s}+\alpha_D d_s)}} \right\|$$

then, using the previous lemma:

$$4\delta_+^2 = 4\delta_1^2 + \left\| \sqrt{\frac{\alpha_P\alpha_D}{\mu}\beta V \beta} - \sqrt{\frac{\mu}{\alpha_P\alpha_D}\frac{1}{\beta V \beta}} \right\|^2.$$

Using *Lemma* 24, we get that:

$$4\delta_+^2 = 4\delta_1^2 + \left\| \sqrt{\frac{\alpha_P\alpha_D}{\mu}\frac{1}{p}e} - \sqrt{\frac{\mu}{\alpha_P\alpha_D}pe} \right\|^2 = 4\delta_1^2 + \frac{\alpha_P\alpha_D}{\mu} + \frac{\mu}{\alpha_P\alpha_D}p^2 - 2p.$$

Let $x^+ = \overline{x} + \alpha_P d_x$ and $s^+ = \overline{s} + \alpha_D d_s$. Then δ_1 can be written as:

$$\delta_1 = \frac{1}{2}\left\| \sqrt{\frac{x^+ s^+}{\mu}} - \sqrt{\frac{\mu}{x^+ s^+}} \right\|.$$

We already know that $d_x = -D^2A^T(AD^2A^T)^{-1}B\beta$, $d_s = A^T(AD^2A^T)^{-1}B\beta$ with $D = \overline{X}^{\frac{1}{2}}\overline{S}^{-\frac{1}{2}}$. If take $v = DA^T(AD^2A^T)^{-1}B\beta$ then $\|v\| = 1$, $d_x = -Dv$ and $d_s = D^{-1}v$. Using these notations:

$$\begin{aligned} x^+ s^+ &= \overline{xs} + \alpha_D\overline{x}(D^{-1}v) - \alpha_P\overline{s}(Dv) - \alpha_P\alpha_D(Dv)(D^{-1}v) \\ &= \overline{xs} + \alpha_D\overline{x}(\overline{X}^{-\frac{1}{2}}\overline{S}^{\frac{1}{2}}v) - \alpha_P\overline{s}(\overline{X}^{\frac{1}{2}}\overline{S}^{-\frac{1}{2}}v) - \alpha_P\alpha_D(\overline{X}^{\frac{1}{2}}\overline{S}^{-\frac{1}{2}}v)(\overline{X}^{-\frac{1}{2}}\overline{S}^{\frac{1}{2}}v). \end{aligned}$$

Componentwise:

$$x_i^+ s_i^+ = \overline{x}_i\overline{s}_i + (\alpha_D - \alpha_P)\sqrt{\overline{x}_i\overline{s}_i}v_i - \alpha_P\alpha_D v_i^2 \text{ for } i = 1, \ldots, n. \tag{5.16}$$

Substituting back in δ_1:

$$\begin{aligned}
4\delta_1^2 &= \sum_{i=1}^{n}\left(\frac{x_i^+ s_i^+}{\mu}+\frac{\mu}{x_i^+ s_i^+}-2\right)\\
&= \sum_{i=1}^{n}\left(\frac{\overline{x}_i\overline{s}_i+(\alpha_D-\alpha_P)\sqrt{\overline{x}_i\overline{s}_i}v_i-\alpha_P\alpha_D v_i^2}{\mu}+\frac{\mu}{\overline{x}_i\overline{s}_i+(\alpha_D-\alpha_P)\sqrt{\overline{x}_i\overline{s}_i}v_i-\alpha_P\alpha_D v_i^2}-2\right)\\
&= 4\overline{\delta}^2+\sum_{i=1}^{n}\left(\frac{(\alpha_D-\alpha_P)\sqrt{\overline{x}_i\overline{s}_i}v_i-\alpha_P\alpha_D v_i^2}{\mu}+\frac{\mu}{\overline{x}_i\overline{s}_i+(\alpha_D-\alpha_P)\sqrt{\overline{x}_i\overline{s}_i}v_i-\alpha_P\alpha_D v_i^2}-\frac{\mu}{\overline{x}_i\overline{s}_i}\right)
\end{aligned}$$

To simplify the analysis, let's take $\alpha_P = \alpha_D := \alpha$. The expression for δ_1 is significantly simplified:

$$4\delta_1^2 = 4\overline{\delta}^2+\sum_{i=1}^{n}\left(\frac{-\alpha^2 v_i^2}{\mu}+\frac{\mu}{\overline{x}_i\overline{s}_i-\alpha^2 v_i^2}-\frac{\mu}{\overline{x}_i\overline{s}_i}\right). \tag{5.17}$$

Using (5.12), the fact that $\|v\|=1$ and $\alpha>0$ we get:

$$\begin{aligned}
4\delta_1^2 &= 4\overline{\delta}^2-\frac{\alpha^2}{\mu}+\mu\sum_{i=1}^{n}\left(\frac{1}{\overline{x}_i\overline{s}_i-\alpha^2 v_i^2}-\frac{1}{\overline{x}_i\overline{s}_i}\right)\\
&= 4\overline{\delta}^2-\frac{\alpha^2}{\mu}+\mu\sum_{i=1}^{n}\frac{\alpha^2 v_i^2}{\overline{x}_i\overline{s}_i(\overline{x}_i\overline{s}_i-\alpha^2 v_i^2)}\\
&\leq 4\overline{\delta}^2-\frac{\alpha^2}{\mu}+\frac{\mu}{\mu(\sqrt{\tau^2+1}-\tau)^2-\alpha^2}.
\end{aligned}$$

Now we can relate the measures of proximity before adding the cuts and after they are added and one step is taken inside the feasible region:

$$4\delta_+^2 \leq 4\overline{\delta}^2-\frac{\alpha^2}{\mu}+\frac{\mu}{\mu(\sqrt{\tau^2+1}-\tau)^2-\alpha^2}+\frac{\alpha^2}{\mu}+\frac{\mu}{\alpha^2}p^2-2p.$$

with $0<\alpha<\sqrt{\mu}(\sqrt{\tau^2+1}-\tau)$.

We should choose the scalar α such that the step is as deep as possible and the bound for δ_+ is as tight as possible. Possible choices for α might be $\sqrt{\frac{\mu}{2}}(\sqrt{\tau^2+1}-\tau)$ or $\sqrt{\frac{\mu}{4}}(\sqrt{\tau^2+1}-\tau)$.

As μ decreases from one iteration to another, it seems that smaller values for α

give a good upper-bound for δ_+. We can get a bound for δ_+:

$$\delta_+^2 \leq \overline{\delta}^2 + 3(2\tau+1)^2 p^2 \tag{5.18}$$

if we take $\alpha = \sqrt{\frac{\mu}{2}}(\sqrt{\tau^2+1} - \tau)$ and use the fact that $\frac{1}{2\tau+1} < \sqrt{\tau^2+1} - \tau < \frac{1}{2\tau}$ for any positive τ.

This inequality holds for any $\tau > 0$. Actually, because we use the long step primal-dual algorithm proposed by Terlaky et al. in [16], in order to get a better complexity, τ should be of order of $\sqrt{n+2m}$.

In fact choosing $\tau = \sqrt{n+2m}$ is useful in one more way (as we will see shortly). In the next section we will prove that if a μ - analytic center is feasible in Γ then, it is also approximately centered. Hence, we can start the sequence of long primal-dual steps by directly decreasing μ (because we are already close to the central path).

5.7 Potential Functions and Analytic Centers

Let m be the dimension of the dual-space and n the total number of cuts that have been added so far by the algorithm. In analyzing the algorithm, potential functions are used. The functions used here are: the primal potential

$$\varphi_P(x) = -c^T x + \mu \sum_{i=1}^{2m+n} \ln x_i, \tag{5.19}$$

the dual potential

$$\varphi_D(s) = b^T y + \mu \sum_{i=1}^{2m+n} \ln s_i \tag{5.20}$$

and the primal-dual potential

$$\varphi_{PD}(x,s) = \varphi_P(x) + \varphi_D(s). \tag{5.21}$$

Here $\mu \in (0,1)$.

We define the *exact μ - analytic center* of $\mathcal{F}_D = \{s > 0 : A^T y + s = c\}$ to be the unique point maximizing φ_D over $\mathcal{F}_D$.

If we consider the problem

$$\max\{\varphi_D(s) : A^T y + s = c, s > 0\}$$

then the first-order optimality conditions are:

$$\begin{aligned} xs &= \mu e, \\ A^T y + s &= c, s > 0, \\ Ax &= b, x > 0. \end{aligned} \tag{5.22}$$

Equivalently, the exact μ - analytic center may be defined as the optimal solution to

$$\max\{\varphi_P(x) : Ax = b, x > 0\}.$$

A μ - *analytic center* is a point (x, y, s) for which the next relations hold:

$$\begin{aligned} \|\tfrac{xs}{\mu} - e\| &\leq \theta < 1, \\ A^T y + s &= c, \quad s > 0, \\ Ax &= b, \quad x > 0. \end{aligned}$$

Now we can prove:

Lemma 27 *If $\theta < \frac{3}{4}$, each μ - analytic center feasible in $\mathcal{F}_D$, is also approximately centered (i.e. $\delta_T(x, s, y) \leq \tau = \sqrt{n + 2m}$).*

Proof: If a μ - analytic center is approximately centered then $\delta_M(x, s, \mu) \leq \theta$. So, $1 - \theta \leq \frac{x_i s_i}{\mu} \leq 1 + \theta$. Then

$$\delta_T(x, s, \mu) \leq \sqrt{(n + 2m)\frac{2\theta - \theta^2}{1 - \theta}} < \sqrt{n + 2m}$$

for any $\theta < \frac{3}{4}$. □

This lemma allows us to call the oracle as soon as the μ - analytic center becomes infeasible. So no extra centering steps are required.

We now introduce a result that relates the potential functions evaluated at an exact μ - analytic center with the values of the same potentials at an μ - analytic center (similar to the *Corollary* 3.2 from [5]). Let's start by noticing that the primal-dual potential value at any feasible point (x, y, s) is bounded and the bound depends only on the dimensionality of the space and the parameter μ.

Lemma 28 *Let $x \in int\mathcal{F}_P$ and $s \in int\mathcal{F}_D$. Then*

$$\varphi_{PD}(x, s) \leq -\mu(2m + n)$$

with equality if and only if $\mu = 1$ and $xs = e$.

Proof: We will use the following inequality:

$$\ln(t) \leq t - 1, \forall t > 0,$$

with equality if and only if $t = 1$. Using this inequality and the fact that $x \in \text{int}\mathcal{F}_P$ and $s \in \text{int}\mathcal{F}_D$ (which implies that $x^T s = c^T x - b^T y$) we get:

$$\varphi_{PD}(x, s) = \varphi_P(x) + \varphi_D(s) = \mu \sum_{i=1}^{2m+n} \ln x_i s_i + b^T y - c^T x \leq (\mu - 1)x^T s - \mu(2m + n).$$

Because $0 < \mu < 1$ and $x, s > 0$ the conclusion follows. □

Lemma 29 *Let (x^C, s^C) be the exact μ - analytic center and (x, s) be a μ - analytic center. Let $\varphi_P^C = \varphi_P(x^C)$ and $\varphi_D^C = \varphi_D(s^C)$. Then:*

$$-\mu(2m + n) \geq \varphi_{PD}(x, s) \geq -\mu(2m + n) - \frac{\theta^2}{1 - \theta^2}\mu + \mu(2m + n) \ln \mu$$

and

$$\varphi_P(x) \geq \varphi_P^C - \mu\frac{\theta^2}{1 - \theta^2}.$$

Proof: The first inequality is already proved.

For the second one we will minimize $\varphi_{PD}(x,s)$ over the set of all μ - analytic centers. We can get a lower bound for $\varphi_{PD}(x,s)$ by minimizing it over the set $\{(x,s); \|\frac{xs}{\mu} - e\| \leq \theta\}$. By taking $u_i := x_i s_i$ and using the fact that $x^T s = c^T x - b^T y$ we can rewrite the problem as:

$$\begin{aligned} \min \quad & -\sum_{i=1}^{2m+n} u_i + \mu \sum_{i=1}^{2m+n} \ln u_i, \\ s.t. \quad & \sum_{i=1}^{2m+n} (\tfrac{u_i}{\mu} - 1)^2 \leq \theta^2. \end{aligned}$$

The objective function is concave and the feasible region is convex and compact (it is a sphere). So the solution of this problem is among the KKT points. It turns out that the KKT point that minimizes the objective function is the vector u with one component equal to $\mu\,(1-\theta)$ and all other $2m+n-1$ components equal to μ. So

$$\varphi_{PD} \geq \mu(\theta + \ln(1-\theta)) - (2m+n)\mu + (2m+n)\mu \ln(\mu).$$

Using the inequality

$$t + \ln(1-t) \geq -\frac{t^2}{1-t^2}, \forall t \in (0,1),$$

we get the first result.

At the μ - analytic center (x^C, s^C) we have: $x_i^C s_i^C = \mu$, $\forall\ i = 1, \ldots, 2m+n$. So

$$\varphi_P^C + \varphi_D^C = -(x^C)^T s^C + \mu \sum_{i=1}^{2m+n} \ln(x_i^C s_i^C) = -\mu(2m+n) + \mu(2m+n)\ln(\mu).$$

If we substitute in the first result we get:

$$\varphi_P(x) + \varphi_D(s) = \varphi_{PD}(x,s) \geq \varphi_P^C + \varphi_D^C - \mu\frac{\theta^2}{1-\theta^2}.$$

Because $\varphi_D^C \geq \varphi_D(s)$ we conclude that:

$$\varphi_P(x) \geq \varphi_P^C - \mu\frac{\theta^2}{1-\theta^2}.$$

□

Another useful result from [5] is

Lemma 30 *Let h be any point in $\mathbb{R}^m$ such that $\|h\| < 1$. Then,*

$$\sum_{i=1}^{m} \ln(1+h_i) \geq e^T h + \|h\| + \ln(1-\|h\|).$$

5.8 Algorithm Analysis

Let (x, y, s) be a μ - analytic center corresponding to an outer-approximation set P_i. Let's assume that this point is infeasible so we call the oracle and add the cuts centrally. Let d_x, d_y, d_s, β and γ be the vectors describing the feasibility step. As already proven in *Section* 5.4, d_x and d_s inherently verify (5.7):

$$\|D^{-1}d_x\| = 1,$$
$$\|Dd_s\| = 1,$$

with $D = X^{-1/2}S^{1/2}$.

For this point and these directions the following lemma holds.

Lemma 31 *For any α with $\alpha < \mu(1-\theta)$, the next inequalities hold:*

$$\|\alpha X^{-1}d_x\| < \frac{\alpha}{\mu(1-\theta)} < 1 \text{ and } \|\alpha S^{-1}d_s\| < \frac{\alpha}{\mu(1-\theta)} < 1.$$

Proof: Using *Lemma* 23 and $\mu < 1$ we get:

$$\|\alpha X^{-1}d_x\| = \alpha\|X^{-1}d_x\| \leq \alpha\sqrt{\mu(1+\theta)} \leq \frac{\alpha}{\sqrt{\mu(1-\theta)}} < \frac{\alpha}{\mu(1-\theta)} < 1.$$

Similarly:

$$\|\alpha S^{-1}d_s\| = \alpha\|S^{-1}d_s\| \leq \frac{\alpha}{\sqrt{\mu(1+\theta)}} < \frac{\alpha}{\sqrt{\mu(1-\theta)}} < \frac{\alpha}{\mu(1-\theta)} < 1.$$

□

Lemma 32 *The following inequalities hold:*

$$|c^T d_x + y^T B\beta - \mu e^T X^{-1} d_x| \leq \frac{\theta}{1-\theta}$$

and

$$|(d_y)^T b + \mu e^T S^{-1} d_s| \leq \frac{\theta}{1-\theta}.$$

Proof: Similar to *Lemma* 5.2 from [5]. □

Using these results, we can relate the values of the potential functions right before adding the cuts through the μ - analytic center, and after the feasibility is recovered (by taking the feasibility step scaled by a factor α).

Lemma 33 *The following inequalities hold:*

$$\begin{aligned}
\tilde{\varphi}_P(x^+(\alpha)) &\geq \varphi_P(x) + \mu p \ln\alpha + \alpha + \mu \ln\left(1 - \frac{\alpha}{\mu(1-\theta)}\right) + \mu \sum_{i=1}^{p} \ln\beta_i, \\
\tilde{\varphi}_D(s^+(\alpha)) &\geq \varphi_D(s) + \mu p \ln\alpha + \alpha + \mu \ln\left(1 - \frac{\alpha}{\mu(1-\theta)}\right) + \mu \sum_{i=1}^{p} \ln(\gamma_i), \\
\tilde{\varphi}_{PD}(x^+(\alpha), s^+(\alpha)) &\geq \varphi_{PD}(x,s) + 2\mu p \ln\alpha + 2\alpha + 2\mu \ln\left(1 - \frac{\alpha}{\mu(1-\theta)}\right) - \mu p \ln p.
\end{aligned}$$

Here $x^+(\alpha) = x + \alpha d_x$ and $s^+(\alpha) = s + \alpha d_s$ correspond to the new point after the scaled feasibility step was taken.

Proof: Let's define u_i to be equal to either $\alpha x_i^{-1} d_{xi}$ or $\alpha s_i^{-1} d_{si}$. In either case, using *Lemma* 31, it turns out that $\|u\| \leq \frac{\alpha}{\mu(1-\theta)} < 1$. Note that $t + \ln(1-t)$ is a

decreasing function of t. Then using *Lemma* 30:

$$\sum_{i=1}^{m} \ln(1+u_i) \geq e^T u + \|u\| + \ln(1-\|u\|) \geq e^T u + \frac{\alpha}{\mu(1-\theta)} + \ln(1 - \frac{\alpha}{\mu(1-\theta)}).$$

The first relation is proved immediately.

$$\begin{aligned}
&\tilde{\varphi}_P(x^+(\alpha)) = -\tilde{c}^T x^+(\alpha) + \mu \sum_{i=1}^{m} \ln x_i^+(\alpha) + \mu \sum_{i=1}^{p} \ln \alpha\beta_i \\
&= -c^T x - \alpha c^T d_x - \alpha y^T B\beta + \mu \sum_{i=1}^{m} \ln x_i^+(\alpha) + \mu \sum_{i=1}^{p} \ln \alpha\beta_i \\
&= -c^T x - \alpha c^T d_x - \alpha y^T B\beta + \mu \sum_{i=1}^{m} \ln x_i + \mu \sum_{i=1} \ln\left(1 + \alpha x_i^{-1} d_{xi}\right) + \mu \sum_{i=1}^{p} \ln \alpha\beta_i \\
&= \varphi_P(x) - \alpha c^T d_x - \alpha y^T B\beta + \mu \sum_{i=1}^{p} \ln \alpha\beta_i + \mu \sum_{i=1} \ln\left(1 + \alpha x_i^{-1} d_{xi}\right) \\
&\geq \varphi_P(x) - \alpha c^T d_x - \alpha y^T B\beta + \mu \sum_{i=1}^{p} \ln \alpha\beta_i + \mu\alpha e^T X^{-1} d_x + \tfrac{\alpha}{1-\theta} + \mu \ln(1 - \tfrac{\alpha}{\mu(1-\theta)}).
\end{aligned}$$

Now we just have to use *Lemma* 32.

The second statement follows similarly. The last inequality follows by adding the other two inequalities and using $\beta\gamma = \frac{1}{p}e$ (from *Lemma* 24). □

Once feasible a sequence of Newton steps is taken toward the μ - analytic center of the new outer approximation set. A standard result in linear optimization (see [26] or [19]) gives a description of the change in the primal-dual potential after a scaled Newton step.

Theorem 15 *Let (x, y, s) be a feasible point with $\|\frac{xs}{\mu} - e\| \geq \theta > 0$. Let $\tilde{x}(\alpha) = x + \alpha\Delta x$ and $\tilde{s}(\alpha) = s + \alpha\Delta s$ with Δx and Δs being the primal-dual Newton directions. Then there exists a step size $\alpha > 0$ and a constant σ such that:*

$$\varphi_{PD}(\tilde{x}(\alpha), \tilde{s}(\alpha)) \geq \varphi_{PD}(x, s) + \sigma. \tag{5.23}$$

So for each scaled Newton step the primal-dual potential changes by a constant σ. This helps us in estimating the number of steps required to get to an μ - analytic center.

Lemma 34 *The number of Newton steps required to compute the updated*

μ - analytic center is bounded by:

$$\nu = \frac{-\mu p - \rho}{\sigma} = \mu O(p \ln(p+1)) \tag{5.24}$$

where

$$\rho = -\frac{\theta^2}{1-\theta^2}\mu - \mu p \ln \mu + 2p\mu \ln \alpha + 2\alpha - \mu p \ln p + 2\mu \ln (1 - \frac{\alpha}{\mu(1-\theta)}). \tag{5.25}$$

Proof: Let's consider the potential function gap $\Delta\tilde{\varphi}_{PD}$:

$$\Delta\tilde{\varphi}_{PD} = \tilde{\varphi}_P^C + \tilde{\varphi}_D^C - \tilde{\varphi}_{PD}(x^+(\alpha), s^+(\alpha)).$$

Here $\tilde{\varphi}_P^C$ and $\tilde{\varphi}_D^C$ are the potential functions evaluated at the exact μ - analytic center of the region obtained after adding p - cuts. We know (see (5.22)) that, at an exact μ - analytic center:

$$x_i s_i = \mu.$$

Substituting this in the definition for potential functions it turns out that:

$$\tilde{\varphi}_P^C + \tilde{\varphi}_D^C = (2m + n + p)(-\mu + \mu \ln(\mu)). \tag{5.26}$$

Using the bounds for $\tilde{\varphi}_{PD}(x^+(\alpha), s^+(\alpha))$ (from *Lemma* 33) and for $\varphi_{PD}(x, s)$ (from *Lemma* 29) we get:

$$\Delta\tilde{\varphi}_{PD} \leq -p(\mu - \mu \ln(\mu)) + \frac{\theta^2}{1-\theta^2}\mu - 2p\mu \ln \alpha - 2\alpha + p\mu \ln p - 2\mu \ln(1 - \frac{\alpha}{\mu(1-\theta)}).$$

By taking:

$$\rho = -\frac{\theta^2}{1-\theta^2}\mu - \mu p \ln \mu + 2p \ln \alpha + 2\alpha - \mu p \ln p + 2\mu \ln (1 - \frac{\alpha}{\mu(1-\theta)})$$

we obtain the result. □

Notice that the number of Newton steps required to move from one analytic

center to the next one decreases as the algorithm proceeds (because μ decreases geometrically).

Now that we know the number of Newton steps required to move from one analytic center to another, the only thing we need is an estimate for the total number of μ - analytic centers generated by the algorithm before we are guaranteed to get the solution of the problem.

The key observation here is that for each μ -analytic center generated by the algorithm, there exists a unique exact μ - analytic center corresponding to the same outer-approximation set. So the total number of μ - analytic centers is equal to the total number of exact μ - analytic centers.

Following the approach in Goffin and Vial [5], the analysis of the dual potential will be used in estimating the total number of exact μ - analytic centers.

First we will establish how the dual potential changes from one exact μ - analytic center to the next one.

Two different situations arise. One is when the exact μ - analytic centers correspond to different outer-approximation sets (so for in this case these AC correspond to the same μ but different sets). This case is similar to the analysis from [5] and we will just cite it for the results we need.

The second case, carefully analyzed here, is when these two exact μ - analytic centers correspond to the same set but the μ parameter is different (i.e. they are arising in a sequence of Primal-Dual steps).

For the first case we have:

Lemma 35 *Let s^C and $\tilde{s}^C$ be two consecutive exact μ - analytic centers corresponding to two different outer-approximation sets. Let φ_D^C and $\tilde{\varphi}_D^C$ be the values of the dual potentials at these points. For all $0 < \alpha < 1 - \theta$*

$$\tilde{\varphi}_D^C \leq \varphi_D^C + \mu \sum_{i=1}^{p} \ln \tau_i + \mu\kappa(\theta, \alpha, p)$$

where

$$\kappa(\theta,\alpha,p)=p\ln p-p(1-\ln\mu)-p\ln\alpha-\frac{\alpha}{\mu}+\frac{\theta^2}{1-\theta^2}-\ln\left(1-\frac{\alpha}{\mu(1-\theta)}\right)$$

and τ_i *are the components of the vector* τ *given by* $\tau^2=diagV$*. That is:*

$$\tau_i=\sqrt{a_{m+i}^T(AX^2A^T)^{-1}a_{m+i}}=\mu^{-1}\sqrt{a_{m+i}^T(AS^{-2}A^T)^{-1}a_{m+i}},i=1,\ldots,p.$$

Proof: Similar to Theorem 5.5 from [5]. The only difference is the presence of the parameter μ. Although this difference might seem to be a small, it has a big impact in analyzing the complexity of the algorithm. □

For $\mu=1$ the exact formulation of Theorem 5.5 from [5] is obtained. If we choose α and θ carefully enough then, it turns out that

$$\kappa(\theta,\alpha,p)\leq\mu p\ln(p+1)$$

(see Theorem 5.5 from [5]) . So:

$$\tilde{\varphi}_D^C\leq\varphi_D^C+\mu\sum_{i=1}^{p}\ln\tau_i+\mu p\ln(p+1). \tag{5.27}$$

We now need some upper bound for τ_i. Ye in [25] (see Theorem 10) gives a global bound when all n added cuts are taken into account (not only the last p cuts added at the last call of the oracle). Using a similar approach, the following inequality can be proved:

$$\sum_{i=1}^{n}\tau_i^2\leq\frac{24m^2}{5\mu^2}\ln\left(1+\frac{n}{8m^2}\right). \tag{5.28}$$

Here the μ parameter corresponds to the smallest value it had at the end of the last sequence of Primal-Dual interior point steps. Using the concavity of the logarithm

function:

$$\sum_{i=1}^{n} \ln \tau_i = \frac{1}{2} \sum_{i=1}^{n} \frac{\ln \tau_i^2}{n} \leq \frac{1}{2} \ln(\frac{24m^2}{5\mu^2 n} \ln (1 + \frac{n}{8m^2})). \tag{5.29}$$

Let's consider now the case when the current μ - analytic center is obtained after a sequence of interior-point steps followed by a feasibility step.

Using the same notation as before, let s^C and $\tilde{s}^C$ be the two consecutive exact μ - AC and φ_D^C and $\tilde{\varphi}_D^C$ be the values of the dual potentials. In this case the outer-approximation set doesn't change but $\tilde{\mu} = (1 - \Theta)\mu$. So:

$$\begin{aligned} \varphi_D^C &= b^T y + \mu \sum_{i=1}^{2m+n} \ln s_i, \\ \tilde{\varphi}_D^C &= b^T \tilde{y} + \tilde{\mu} \sum_{i=1}^{2m+n} \ln \tilde{s}_i = b^T \tilde{y} + (1 - \Theta)\mu \sum_{i=1}^{2m+n} \ln \tilde{s}_i. \end{aligned}$$

Because these μ - analytic centers are exact, they maximize the dual potential functions (with parameters μ and $(1 - \Theta)\mu$). This observation leads to:

$$\sum_{i=1}^{2m+n} \ln s_i \geq \sum_{i=1}^{2m+n} \ln \tilde{s}_i.$$

So:

$$\frac{\tilde{\varphi}_D^C}{\tilde{\mu}} \leq \frac{\varphi_D^C}{\mu} + \frac{b^T \tilde{y}}{\tilde{\mu}} - \frac{b^T y}{\mu}. \tag{5.30}$$

Let (x_j, y_j, s_j) be an μ_j - analytic center. Let us introduce the following notation:

$$\varphi_j(s_j) = \mu_j \Psi_j(s_j), \tag{5.31}$$

$$l_j = b^T y_j. \tag{5.32}$$

Here, φ_j is the dual potential function at the j - th iteration.

Using this notation, the relationship between the dual-potential functions at two

consecutive exact μ - analytic centers (x_j, y_j, s_j) and $(x_{j+1}, y_{j+1}, s_{j+1})$ become:

$$\Psi_{j+1}(s_{j+1}) \leq \Psi_j(s_j) + \sum_{i=1}^{p} \ln \tau_i + \kappa(\theta, \alpha, p) \tag{5.33}$$

for the first case and

$$\Psi_{j+1}(s_{j+1}) \leq \Psi_j(s_j) + \frac{l_{j+1}}{\mu_{j+1}} - \frac{l_j}{\mu_j}. \tag{5.34}$$

Now, before analyzing the complexity let's review the algorithm, introducing some useful notations at the same time.

5.9 Algorithm Overview

We are interested in finding an upper bound for the total number of cuts that are added before we get the solution. This means that we should keep track only of the number of calls of the oracle. We start with the big cube C as the first outer approximation for Γ. Let's call it P_1. Let $\mu_1 = 1$ and AC_1 be its $\mu = \mu_1$ - analytic center. Call the oracle. Assuming AC_1 is not in Γ, the oracle returns p_1 central cuts. Introduce the cuts; P_1 becomes P_2 with the μ_1 - analytic center AC_2. Call the oracle, get p_2 central cuts, add them and obtain the new region P_3. Keep doing this until, at the $(k_1 - 1)$-th iteration, the μ_1 - analytic center AC_{k_1} of P_{k_1} is strictly interior to Γ. Now take a sequence of primal-dual steps. Initialize μ with μ_1. Take Newton steps until the point gets close to the central path (i.e. $\delta_T(x, s, \mu) < \tau$, for some τ). Change μ into $\mu(1 - \Theta)$ (with $0 < \Theta < 1$). Keep taking primal-dual interior steps until at the end of the $\overline{k}_1 + 1$-th iteration (when $\mu = \mu_1(1 - \Theta)^{\overline{k}_1}$) the point gets out of Γ. Take $\mu_2 = \mu_1(1 - \Theta)^{\overline{k}_1}$. Call the oracle that returns p_{k_1} central cuts. Add them to P_{k_1} that becomes P_{k_1+1} with the μ_2 - analytic center AC_{k_1+1}. Call the oracle at AC_{k_1+1} and find the next μ_2 - analytic center. Do this until after $k_2 - 1$ outside steps the μ_2 - analytic center $AC_{k_1+k_2}$ is again inside Γ. Initialize the sequence of interior primal-dual steps with $\mu = \mu_2$. Take the steps until, the point gets out of Γ (with a corresponding $\mu = \mu_2(1 - \Theta)^{\overline{k}_2}$). Call the oracle, add p_{k_2} cuts and get the new domain $P_{k_1+k_2+1}$. Take now $\mu_3 = \mu_1(1 - \Theta)^{\overline{k}_2}$. Now generate

a sequence of μ_3 - analytic centers until, after $k_3 - 1$ iterations, the AC gets back into Γ. We keep doing this until $(n + 2m)^\gamma \mu < \varepsilon$ (n being the total number of constraints) while the current point is in Γ that is the solution.

We assumed that Γ contains a ball of radius 2^{-L}. In the following section we will prove that the number of μ - analytic centers generated by the algorithm is finite. So, sooner or later the μ - analytic center will end-up in Γ. Hence at least one sequence of interior primal-dual steps will be generated. In a limit case we might get feasibility at the same time with the solution.

5.10 Complexity

Using the notations we just introduced and the relations between dual potentials at different μ - analytic centers we get, after the algorithm just finished the s-th sequence of Primal-Dual sequence of steps:

$$\begin{aligned}
\Psi_{k_1+\ldots+k_s+1} \quad \leq \quad & \Psi_1 + \sum_{i=1}^{n_s} \ln(\tau_i) + \sum_{i=1}^{k_1+\ldots+k_s-1} (p_i \ln(p_i+1)) - \\
- \quad & \frac{l_{k_1}}{\mu_1} + \frac{l_{k_1+1} - l_{k_1+k_2}}{\mu_2} + \frac{l_{k_1+k_2+1} - l_{k_1+k_2+k_3}}{\mu_3} + \\
+ \quad & \ldots + \frac{l_{k_1+\ldots+k_{s-1}+1} - l_{k_1+\ldots+k_s}}{\mu_s} + \frac{l_{k_1+\ldots+k_s+1}}{\mu_{s+1}},
\end{aligned}$$

with $\Psi_i = \Psi(s_i)$, p_i - the number of cuts returned by the oracle at AC_i and n_s the total number of cuts generated by the algorithm until $AC_{k_1+\ldots+k_s}$.

Let p be the maximum number of cuts that the oracle generated at any call. All the μ - analytic centers generated by the algorithm are in the initial cube

$$C = \{y \in I\!R^m; -e \leq y \leq e\}.$$

This helps provides us with bounds for l_i and $l_i - l_j$:

$$\begin{aligned}
l_i \quad & \leq \quad \sqrt{m}\|b\|, \\
l_i - l_j \quad & \leq \quad 2\sqrt{m}\|b\|.
\end{aligned}$$

So:

$$\begin{aligned}\Psi_{k_1+\ldots+k_s} &\leq \Psi_1 + \sum_{i=1}^{n_s} \ln(\tau_i) + \sum_{i=1}^{k_1+\ldots+k_s-1} (p_i \ln(p_i+1)) + \\ &+ 2\|b\|\sqrt{m}(\tfrac{1}{\mu_1} + \tfrac{1}{\mu_2} + \ldots + \tfrac{1}{\mu_{s+1}}).\end{aligned}$$

Because the algorithm did not stop at the last sequence of primal-dual interior steps, $(n_s + 2m)^\gamma \mu_{s+1} \geq \varepsilon$. On the other hand $\mu_s = (1-\Theta)^t$ for some integer t, so we can write:

$$\frac{1}{\mu_1} + \frac{1}{\mu_2} + \ldots + \frac{1}{\mu_{s+1}} \leq \sum_{i=0}^{t} \frac{1}{(1-\Theta)^i} \leq \frac{1}{(1-\Theta)^t \Theta} \leq \frac{1}{\mu_s \Theta} \leq \frac{(n_s+2m)^\gamma}{\varepsilon\Theta}.$$

Thus:

$$\Psi_{k_1+\ldots+k_s+1} \leq \Psi_1 + \sum_{i=1}^{n_s} \ln(\tau_i) + (n_s+2m)\ln(1+p) + 2\|b\|\sqrt{m}\frac{(n_s+2m)^\gamma}{\varepsilon\Theta}.$$

The first μ - analytic center generated by the algorithm is the one corresponding to the fixed, given cube C. So Ψ_1 can be considered a constant (its value does not depend on the problem).

At the beginning we made the assumption that the domain Γ contains a ball of radius 2^{-L}. Let $\bar{y}$ be the center of this ball with the corresponding slack vector $\bar{s}$. So $\bar{s}_i > 2^{-L}$. The last exact μ - analytic center considered in our analysis is $(x_{k_1+\ldots+k_s+1}, y_{k_1+\ldots+k_s+1}, s_{k_1+\ldots+k_s+1})$. So:

$$\Psi_{k_1+\ldots+k_s+1} = \frac{\varphi_D(s_{s+1})}{\mu_{s+1}} \geq \frac{b^T\bar{y}}{\mu_{s+1}} + \sum_{i=1}^{2m+n_s} \ln \bar{s}_i.$$

Or:

$$\Psi_{k_1+\ldots+k_s+1} \geq -2\frac{\|b\|\sqrt{m}}{\varepsilon}(n_s+2m)^\gamma + (n_s+2m)\ln 2^{-L}.$$

Finally, using the concavity of the logarithm function, (5.29) and the above inequalities:

$$-2\|b\|\sqrt{m}(\frac{1}{\varepsilon\Theta} + \frac{1}{\varepsilon})(n_s+2m)^{\gamma-1} + \ln 2^{-L} - \ln(1+p) - \frac{\Psi_1}{n_s+2m} \leq$$

$$\leq \frac{1}{2}\ln(\frac{24m^2}{5\varepsilon^2(n_s+2m)^{1-2\gamma}}\ln(1+\frac{n_s}{8m^2})). \qquad (5.35)$$

Now it is clear why we need to have $(n+2m)^\gamma\mu < \varepsilon$, $\gamma < 0.5$, as a stopping criteria for the primal-dual part of the algorithm (instead of the usual one: $(n+2m)\mu < \varepsilon$). Using this last inequality it is clear that the algorithm converges. This is because, in time n_s - the total number of cuts that are added increases. At the limit, as n_s goes to infinity, the left hand side of the inequality converges to a constant number while the right-hand side converges to negative infinity making the inequality untrue. So the algorithm converges.

By ignoring the lower order terms the total number of cuts that are added during the evolution of the algorithm is of order:

$$O((\frac{pm}{2^{-L}\varepsilon})^{\frac{2}{1-2\gamma}}). \qquad (5.36)$$

By taking $\gamma = \frac{1}{16}$ then $\frac{2}{1-2\gamma}$ is about 2.3. In implementing this algorithm we need to choose γ such that to get a small number of analytic centers without being too aggressive in the primal-dual interior point scheme.

Our results compare favorably with results obtained in different instances.

In a pure analytic center cutting planes method scheme applied to a linear feasibility problem (see [5]), the number of cuts added before the algorithm generates an strictly feasible point is of order

$$O((\frac{pm}{2^{-L}})^2). \qquad (5.37)$$

For a pure analytic center cutting plane method applied to a semidefinite feasibility problem (see [12]) the result is

$$O(\frac{m^3p^2}{2^{-2L}}). \qquad (5.38)$$

CHAPTER 6
Conclusions and Future Work

In this thesis we proposed and analyzed an algorithm for solving feasibility problems that arise in conic programming. The approach is based on an analytic center cutting plane method. We generalized here the particular cases of linear programming, second order cone programming and semidefinite programming. Our algorithm can be easily adjusted to these particular cases.

The assumptions we made about the problem are usual ones. Although we are dealing with a general case we didn't need to impose any extra conditions on the problems. The feasibility problems have convex, closed, bounded, fully dimensional sets of interest. These sets are described by an oracle that either recognizes that a point is strictly interior to the set or returns a set of violated constraints. Multiple cuts are added centrally when the current point is infeasible. These cuts can be linear, quadratic, semidefinite or any combination of these types.

The complexity results are similar to the ones obtained for less general cases. We proved that our algorithm generates no more than $O^*(\frac{mP^3\Theta^3}{\varepsilon^2\Lambda^2})$ analytic centers before a solution is obtained. This result compares favorably with $O^*(\frac{m^2P^2}{\varepsilon^2})$ (obtained for the linear case) and $O(\frac{m^3P^2}{\varepsilon^2})$ (for the semidefinite case). The extra terms we have are Θ and Λ, which characterize the self-concordant functionals and the cuts that are introduced, respectively.

The numerical results we obtained are encouraging and are in line with the theoretical ones.

We also proposed a new algorithm for solving optimality problems. This algorithm incorporates the analytic center cutting plane method we proposed for feasibility problems. We completely analyzed the complexity of this algorithm in the linear case.

Open questions remain to be addressed in future work. It would be interesting to analyze how the algorithm changes if deep cuts are used (instead of central ones) or if some of them are dropped. In our analysis the operators describing the cuts had

to be injective. This requirement limits the size of second order cones that can be added by the oracle. Also we didn't analyze here the complexity of the algorithm we proposed for solving optimality problems.

Finally, from a practical perspective, a better implementation of the algorithm is required to make possible a fair comparison (from the time point of view) with other existing solvers.

All these issues remain to be addressed in our future work.

BIBLIOGRAPHY

[1] E.D. Andersen, J. E. Mitchell, C. Roos and T. Terlaky. *A Homogenized Cutting Plane Method to Solve the Convex Feasibility Problem.* Optimization Methods and Applications, Chapter 10, Kluwer Academic Publishers, April 2001.

[2] D.S. Atkinson and P.M. Vaidya. *A Cutting Plane Algorithm for Convex Programming That Uses Analytic Centers.* Mathematical Programming, 69:1-43, 1995.

[3] T. A. Davis. *A Column Pre-Ordering Strategy for the Unsymmetric-Pattern Multifrontal Method.* May 6, 2003.

[4] J.-L. Goffin and J.-Ph. Vial. *Convex nondifferentiable optimization: a survey focused on the analytic center cutting plane method.* Technical report 99.02, McGill University, Canada, February 1999.

[5] J.-L. Goffin and J.-Ph. Vial. *Multiple Cuts in the Analytic Center Cutting Plane Method.* SIAM Journal on Optimization, 11:266-288, 2000.

[6] M. Kojima, S. Mizuno and A. Yoshise *A primal-dual interior point algorithm for linear programming.* Progress in Mathematical Programming - Interior Point and Related Methods, N. Megido, ed., Springer-Verlag, Berlin, 1989, pp. 29-47.

[7] J.E. Mitchell. *Polynomial interior point cutting plane methods.* Technical report, Mathematical Sciences, Rensselaer Polytcehnic Institute, Troy, NY 12180, April 1, 2003.

[8] J.E. Mitchell, S. Ramaswamy. *A Long-Step, Cutting Plane Algorithm for Linear and Convex Programming.* Annals of OR, Volume 99, 2000, pages 95-122.

[9] J.E. Mitchell, M. Todd. *Solving combinatorial optimization problems using Karmarkar's algorithm.* Mathematical Programming, 1992, Vol. 56, pp. 245-284.

[10] M. Momma, K. P. Bennett. *Sparse Kernel Partial Least Squares Regression.* Technical report, Mathematical Sciences, Rensselaer Polytcehnic Institute, Troy, NY 12180, June 2003.

[11] R.D.C. Monteiro and I. Adler. *Interior path following primal-dual algorithms. Part I: Linear programming.* Mathematical Programming, 44(1989), pp 27-41.

[12] M.-R. Oskoorrouchi. *The Analytic Center Cutting Plane Method with Semidefinite Cuts.* Doctoral Thesis, May 2002.

[13] M.-R. Oskoorouchi and J.-L. Goffin. *The Analytic Center Cutting Plane Method with Semidefinite Cuts.* Les Cahiers du GERAD, G-2000-32, 2000.

[14] M.-R. Oskoorouchi and J.-L. Goffin. *An interior point cutting plane method for convex feasibility problem with second-order cone inequalities.* Technical report, April 2003.

[15] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms iin Convex Programming.* SIAM studies in applied mathematics; vol.13, 1993.

[16] J. Peng, C. Roos, T. Terlaky. *A New Class of Polynomial Primal-Dual Methods for Linear and Semidefinite Optimization.* Preprint. August 18, 2000.

[17] O. Peton and J.-Ph. Vial. *Multiple Cuts with a Homogeneous Analytic Center Cutting Plane Method.* HEC/Logilab Technical Report 2001.1 40, bd du Pont d'Arve CH-1211 Geneva 4, Switzerland. May 2001.

[18] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* MPS-SIAM Series on Optimization, 2001.

[19] C. Roos, T.Terlaky and J.-Ph. Vial. *Theory and Algorithms for Linear Optimization: An Interior Point Approach*, Wiley-Intersci. Ser. Discrete Math. Optim., John Wiley & Sons, Chichester, UK, 1997.

[20] R. Rosipal and L. Trejo. *Kernel partial least squares regression in reproducing kernel Hilbert space.* Journal of Machine Learning Research, 2:97-123, 2001.

[21] B. Schölkopf, A. Smola and K.-R. Müller. *Kernel principal component analysis.* Adv. in Kernel Methods - Support Vector Learning Research, MIT Press, 1999.

[22] G. Sonnevend. *New Algorithms in Convex Programming Basesd on a Notion of "Centre" (for Systems of Analytic Inequalities) and on Rational Expectations.* Trends in Mathematical Optimization: Proceedings of the 4th French-German Conference on Optimization in Irsee, West Germany, April 1986, volume 84, pp 311-327, of International Series of Numerical Mathematics, Birkhauser, Basel, Switzerland.

[23] K.-C. Toh, G. Zhao and J. Sun. *A Multiple-Cut Analytic Center Cutting Plane Method for Semidefinite Feasibility Problems.* SIAM Journal on Optimization, 12: 1126-1146, 2002.

[24] S. J. Wright. *Primal-Dual Interior-Point Methods*, SIAM, 1997.

[25] Y. Ye. *Complexity Analysis of the Analytic Center Cutting Plane Method That Uses Multiple Cuts.* Mathematical Programming 78:85-104, 1996.

[26] Y. Ye. *Interior Point Algorithms: Theory and Analysis*, John Wiley & Sons, New York, 1997.