The Analytic Center Cutting Plane Method with Semidefinite Cuts

Mohammad R. OSKOOROUCHI

Faculty of Management
McGill University, Montreal

July 2002

A thesis submitted to the Faculty of Graduate Studies and Research In partial fulfillment of the requirements of the degree of Doctor of Philosophy

Your file Votre référence
ISBN: 0-612-85729-8
Our file Notre référence
ISBN: 0-612-85729-8

Abstract

We propose an analytic center cutting plane algorithm for semidefinite programming (SDP). Reformulation of the dual problem of SDP into an eigenvalue optimization, when the trace of any feasible primal matrix is a positive constant, is well known. We transform the eigenvalue optimization problem into a convex feasibility problem. The problem of interest seeks a feasible point in a bounded convex set, which contains a full dimensional ball with $\varepsilon(< 1)$ radius and is contained in a compact convex set described by matrix inequalities, known as the set of localization. At each iteration, an approximate analytic center of the set of localization is computed. If this point is not in the solution set, an oracle is called to return a p-dimensional semidefinite cut. The set of localization then, is updated by adding the semidefinite cut through the center. We prove that the analytic center is recovered after adding a p-dimensional semidefinite cut in $O(p \log(p + 1))$ damped Newton's iteration and that the ACCPM with semidefinite cuts is a fully polynomial approximation scheme. We report the numerical result of our algorithm when applied to the semidefinite relaxation of the Max-Cut problem.

Dedicated to the memory of my father

Résumé

Nous proposons une méthode de centre analytique par plan coupé pour la programmation semi-définie (SDP). La formulation du problème dual du SDP en un problème d'optimisation de valeur propre, lorsque la trace de toute matrice primale réalisable est une constante positive, est bien connue. Nous transformons le problème d'optimisation de valeur propre en un problème de faisabilité convexe. Le problème en question cherche un point réalisable dans un ensemble convexe borné, lequel contient une sphère de pleine dimension avec un rayon de $\varepsilon(< 1)$, et est contenu dans un ensemble convexe compact décrit par des inégalités de matrices, appelé ensemble de localisation. À chaque itération, un centre analytique approximatif de l'ensemble de localisation est calculé. Si ce point n'est pas dans l'ensemble de solution, un oracle est appelé, lequel retourne une coupe semi-définie de dimension p. L'ensemble de localisation est par la suite mis à jour en ajoutant au centre la coupe semi-définie. Nous démontrons que le centre analytique est reconstitué en $O(plog(p+1))$ itérations de Newton suite à l'addition d'une coupe semi-définie de dimension p. Ns démontrons aussi que la méthode de centre analytique par plan coupé avec des coupes semi-définies est une stratégie d'approximation pleinement polynomiale. Nous rapportons les résultats numériques de notre algorithme appliqué à la relaxation semi-définie du problème Max-cut.

Contents

Acknowledgments

First and foremost, I wish to thank my advisor, Professor Jean-Louis Goffin, for his long-lasting guidance and fruitful cooperation during my Ph.D. studies. Without his generous support, this thesis would not have been possible.

I would like to express my gratitude to my thesis committee, Prof. David Avis, Prof. Pierre Hansen, Prof. Richard Loulou, and Prof. Henry Wolkowicz for their thoughtful suggestions throughout my research in this field.

In the past five years, my daily life in Montreal would not have been so wonderful without the support from several people in the Faculty of Management. Among them are my friends Abdullah Dasci, Tripat Gill, Troy Anderson, Arif Butt and Kamal Munir.

I enjoyed fruitful discussions with my colleagues Samir Elhedhli and Fatma Gzara about the ACCPM and with Javad Mashreghi, a former Ph.D. student in the Department of Mathematics about some of the mathematical proofs of this dissertation. Many thanks goes to my colleague Catherine Paquet for writing the French abstract of this thesis.

I would like to thank Mr. Pierre Cambron, a member of the technical support in the Faculty of Management, who was always ready to provide immediate help when there was a computer problem.

I am grateful to the following programs and research centers for providing me with the financial support of my Ph.D. studies and travel grants for conferences: Ph.D. program of the Faculty of Management, Faculty of Graduate Studies, Management Science Research Center (MSRC), Natural Science and Engineering Research Council of Canada (NSERC) and FCAR of Quebec.

Above all, I would like to thank my mother and my brothers and sisters

for their endless support throughout my education.

Chapter 1

Introduction

Semidefinite Programming (SDP) is an extension of linear programming where vector variable is replaced by matrix variable and nonnegativity constraint is replaced by positive semidefiniteness. Semidefinite Programming can also be studied as a convex optimization problem as the objective function is linear and the constraints form a convex feasible set. Theoretical properties of SDP, as a convex optimization problem have attracted the researchers' attention for the past few decades [7], [10], [51], [12], [5], and [27].

However, efficient algorithms for semidefinite programs were developed at the end of 1980s, following the development of interior point algorithms for linear programming. The first polynomial time algorithm was proposed by Nesterov and Nemirovsky [42, 43, 44]. Their algorithm is based on a self-concordant barrier function. They proved that interior point methods for linear programming can be extended to all convex optimization problems. Independently, Alizadeh [1] extended the potential reduction algorithm from linear programming to SDP.

The algorithms proposed by Nesterov and Nemirovski and by Alizadeh were applicable to only primal (or only dual) problem. The extension of

efficient primal-dual algorithms such as path-following methods, from linear programming to SDP, took the researchers a few years. The first extension of the Newton direction in primal-dual algorithms for SDP was proposed by Alizadeh, Haeberly and Overton [2], and independently by Helmberg, Rendl, Vanderbei, and Wolkowicz [24], Kojima, Shindoh and Hara [28] and Monteiro [40]. A different direction was later introduced by Nesterov and Todd [45, 46]. For a survey on the primal-dual search directions, see Vandenberghe and Boyd [58] and for the references on the complexity of these algorithms see [3], [4]. For a comprehensive reference of the theory, algorithms and applications of SDP see Wolkowicz, Saigal and Vandenberghe [60].

Efficient algorithms and applications in areas such as control theory, statistics, probability, combinatorial optimization, structural design, and nonconvex quadratic optimization made SDP the topic of concentrated research in 1990s. The most efficient interior point algorithms for semidefinite programs as for linear programs are the primal-dual methods. The first strong numerical result based on these methods, was reported in 1996 in [24]. The result of this paper was very encouraging for many practical problems. However, semidefinite relaxations arising from combinatorial applications were still out of reach. Part of the reason for this incapability was that due to the nature of the combinatorial optimization, the semidefinite relaxation on these problems is of large size and often has a sparse structure and the primal-dual algorithm does not exploit this sparsity. On the other hand at each iteration of the interior point algorithms, a Cholesky factorization of a dense matrix, known as the Gram matrix, has to be computed. Benson, Ye and Zhang [8] propose a dual scaling algorithm for the problems with rank one matrix coefficients. In their algorithm, they keep the sparsity of the problem throughout of the algorithm, but they still need to compute the Gram matrix at each iteration.

Helmberg and Rendl [23] reformulate the dual problem of semidefinite program as an eigenvalue optimization problem and apply a spectral bundle method as a nonsmooth optimization technique to solve it. Their algorithm is, by far, the most efficient one for large-scale SDP [22]. However, they don't report any theoretical complexity bound on the number of iterations of their algorithm. Another interior point cutting plane algorithm for the maxcut problem based on a linear programming approach is discussed in Krishnan and Mitchell [30] and Krishnan [29]. Other references in interior point cutting plane algorithm are [20, 37, 38, 39].

An alternative technique for nonsmooth optimization is the analytic center cutting plane method (ACCPM). This method was introduced by Sonnevend [53], Ye [61] and Goffin, Haurie and Vial [15]. ACCPM has been successfully implemented in a wide variety of applications, as for instance in [14] and [19]. The complexity of the method has been analyzed in case of single cuts by Atkinson and Vaidya [6], Nesterov [41] and Goffin, Luo and Ye [16], in the case of multiple cuts by Ye [63] and Goffin and Vial [18], and in the case of quadratic cuts by Luo and Sun [33], Lüthi and Büeler [34] and Sharifi Mokhtarian and Goffin [52].

For the purpose of proving complexity results, ACCPM is more clearly described in the context of a convex feasibility problem: find a point in a bounded convex set Ω^*, with a nonempty interior. The solution set Ω^* is assumed to contain a ball $\mathcal{N}_\varepsilon$, with radius $\varepsilon < 1$ and is contained in a compact convex set described by matrix inequalities. At each iteration the analytic center of the set of localization is computed and a separation oracle is called: the oracle determines if either the center is in Ω^*, thus solving the problem, or returns a cut which cuts off the current point and contains the solution set. A special updating step is then needed to get as close as possible to the next analytic center, as first suggested by Mitchell and Todd [36].

ACCPM is well developed for linear and quadratic programming. In this thesis we employ, for the first time, a nonpolyhedral model into the ACCPM. We transform the eigenvalue min-max problem arising from semidefinite program into a *convex feasibility problem* and propose the analytic center cutting surface algorithm with *semidefinite cuts* to solve this problem. A semidefinite cut contains as special cases, single and multiple linear cuts, as well as quadratic cuts. At each step of the algorithm, an oracle returns a p-dimensional semidefinite cut. We add the cut at the center, and derive the *optimal updating direction* by maximizing the "log det" of the new slack matrix. The optimal updating direction is used to obtain an interior point of the updated set of localization as an initial point to compute the next analytic center. This is an extension of the direction obtained by Goffin and Vial [18] for the multiple linear cuts to the semidefinite cuts. For alternative approaches to solving determinant maximization problems see [56, 59].

The restoration procedure is discussed in detail. We prove that the number of Newton steps needed to recover the analytic center from the interior point obtained by the optimal updating direction is of order of $p\log(p+1)$. We call this interior point, a *warm start*. Moreover, we show that the analytic center cutting plane algorithm stops with a point in the solution set when the dimension of the accumulated block diagonal cut matrix reaches to the bound of $O^*(p_{\max}^2 m^3/\mu^2\varepsilon^2)$, where $p_{\max}$ is the maximum dimension of the cut matrices and $\mu > 0$ is a condition number of the field of cuts. Furthermore, we prove that the Newton method finds the optimal updating direction in at most $O\left(\frac{p\log\frac{1}{\varepsilon^*}+\log\frac{1}{\mu}}{\beta-\log(1+\beta)}\right)$ iterations, where β is the Newton decrement and $\varepsilon^* = \frac{(1-\theta)\varepsilon}{(1+\theta)(1+n)}$ and θ is a positive constant less than 1.

From the application point of view, our algorithm can be applied to a large class of semidefinite programs. More precisely, the semidefinite problems in which the trace of any feasible primal matrix is constant. We see that many

semidefinite relaxations arising from the combinatorial optimization possess this property. In particular, we apply our algorithm to the Max-Cut problem. We develop the ACCPM algorithm in primal and dual settings and discuss the implementation issues for shallow and deep cuts in both cases. Our numerical results are created using the primal algorithm.

It is important to notice that, the main intention of this thesis is not to develop an algorithm for large-scale semidefinite programs. It is rather, the extension of the ACCPM by means of the semidefinite cuts for the convex feasibility problem. Nevertheless, we illustrate that ACCPM with semidefinite cuts efficiently solves moderate-size semidefinite programs. To the best of our knowledge, this algorithm is the second nonsmooth approach to semidefinite programs after the Bundle method [23]. The advantage of the algorithm developed here over the Bundle method is however, its strong theoretical background and complexity results. Independent of our work in this thesis, very recently, there has been some interests in semidefinite feasibility problem [54, 57, 9]. However, their setting is different and they target different applications.

The thesis is organized as follows: In Chapter 2 we introduce the primal and dual problems of semidefinite programming and state the duality theorems. We also define the max-cut problem and derive the semidefinite relaxation on this problem. Chapter 3 is devoted to the most important properties of the analytic center of a convex set of linear matrix inequalities. This includes the primal, dual and primal-dual potential functions, the optimality conditions for the exact and approximate analytic center and primal and dual algorithms for the computation of the analytic center. We introduce the convex feasibility problem and semidefinite cuts in Chapter 4, and derive the optimal updating direction to restore the analytic center after adding a semidefinite cut. In Section 4.3 we present the ACCPM algorithm

and in Section 4.4 we derive the complexity of our algorithm. We also derive an upper bound on the number of damped Newton steps to compute the optimal updating direction. Chapter 5 deals with the applications. We reformulate the dual problem of SDP into an eigenvalue optimization problem and then reduce the optimization problem to a feasibility problem. We define the weighted analytic center in Section 5.3. Primal and dual directions for computing the weighted analytic center are derived in Sections 5.4 and 5.5. We also discuss the implementations of primal and dual algorithms when the set of localization is composed of both linear and semidefinite cuts and in case of the deep cuts. Our numerical results for a number of random graphs generated by graph generator "Rudy" are presented in Section 5.7. We also discuss the computational difficulties for both primal and dual algorithms.

Chapter 2

Preliminaries

In this chapter we provide some basic facts that are fundamental for this thesis. We introduce the standard formulations of primal and dual semidefinite program and state basic definitions of the analytic center of a compact set of matrix inequalities. We begin by introducing our notations:

Lower case letters are used to show vectors and upper case letters are used for matrices. I and I_n are identity matrices of appropriate size or of size n. The i^{th} column of I is shown by e_i, and $Diag(e_i)$ is a diagonal matrix with e_i on its main diagonal.

We refer to the space of $n \times n$ symmetric matrices by $\mathcal{S}^n$, positive semidefinite matrices by $\mathcal{S}^n_+$, and positive definite matrices by $\mathcal{S}^n_{++}$. We denote the j^{th} eigenvalue of a symmetric matrix A, by $\lambda_j(A)$ in decreasing order.

For square matrix A, $\mathbf{tr}(A)$ is the trace of A, $diag(A)$ is a column vector made up of the diagonal elements of A, the Frobenius-norm of A is defined via

$$\|A\|^2 = \mathbf{tr}A^TA = \sum_{j=1}^{n}(\lambda_j(A^TA)),$$

and if A is symmetric, the ∞-norm of A is defined by

$$\|A\|_\infty = \max|\lambda_j(A)|,\ j = 1, ..., n.$$

The operator "$\bullet$" indicates the inner product of two matrices:

$$A \bullet B = trA^T B = \sum_{i,j} a_{ij} b_{ij}.$$

With the abuse of notation we indicate the component-wise product of two vectors x and s by

$$xs = \begin{pmatrix} x_1 s_1 \\ \vdots \\ x_n s_n \end{pmatrix}, \ x, s \in R^n,$$

and the component-wise inverse of vector $x \in R^n$ by

$$x^{-1} = \begin{pmatrix} x_1^{-1} \\ \vdots \\ x_n^{-1} \end{pmatrix}.$$

For symmetric matrices A_i, $i = 1, ..., m$, we define the m-vector a_q^l by

$$a_q^l = ((A_1)_{lq}, (A_2)_{lq}, ..., (A_m)_{lq}). \tag{2.1}$$

This vector is denoted by b_q^l when we deal with symmetric matrices B_i.

The Löwner partial order on the symmetric matrices is defined by $A \succeq B$ ($A \succ B$) if $A - B \in \mathcal{S}_+^n$ ($A - B \in \mathcal{S}_{++}^n$).

2.1 Semidefinite Programming

Semidefinite programming (SDP) is an extension of linear programming where vector variable is replaced by matrix variable and nonnegativity constraint is replaced by positive semidefiniteness. The primal problem of SDP can be expressed in the form

$$\begin{aligned} \max \quad & C \bullet X \\ s.t. \quad & \\ & A_i \bullet X = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0, \end{aligned} \tag{2.2}$$

where $X \in \mathcal{S}^n$ is the matrix variable and $n \times n$ real matrices A_i and C and vector $b \in R^m$ are the parameters of the problem. Without loss of generality we assume that $A_i \in \mathcal{S}^n$ and $C \in \mathcal{S}^n$. Note that if C is not symmetric, since matrix X is symmetric then $C \bullet X = C^T \bullet X$, and thus C can be replaced by $\frac{1}{2}(C + C^T)$. The same argument holds for matrices A_i.

Let $\mathcal{A}$ be a linear operator from $\mathcal{S}^n$ to m-vector R^m defined by

$$\mathcal{A}X = \begin{pmatrix} A_1 \bullet X \\ \vdots \\ A_m \bullet X \end{pmatrix} \tag{2.3}$$

for any symmetric matrix X. Then problem (2.2) can be rewritten via

$$\begin{aligned} \max \quad & C \bullet X \\ s.t. \quad & \\ & \mathcal{A}X = b \\ & X \succeq 0, \end{aligned} \tag{2.4}$$

Corresponding to the primal problem, the dual problem is defined as the minimization of a linear function over a convex set of matrix inequalities:

$$\begin{aligned} \min \quad & b^T y \\ s.t. \quad & \\ & \mathcal{A}^T y \succeq C, \end{aligned} \tag{2.5}$$

where $\mathcal{A}^T : R^m \to \mathcal{S}^n$ is the adjoint operator of $\mathcal{A}$, defined by

$$\mathcal{A}^T y = \sum_{j=1}^{m} y_j A_j. \tag{2.6}$$

Using the slack matrix S the dual problem can be written as

$$\begin{array}{ll} \min & b^T y \\ s.t. & \\ & \mathcal{A}^T y - S = C \\ & S \succeq 0. \end{array} \tag{2.7}$$

The duality gap of the objective values at a dual feasible solution (y, S) and a primal feasible solution X is

$$\begin{aligned} b^T y - C \bullet X &= (\mathcal{A}X)^T y - C \bullet X \\ &= (\mathcal{A}^T y - C) \bullet X \\ &= S \bullet X \\ &\geq 0. \end{aligned}$$

The strong duality theorem in SDP is weaker than that of in linear programs. In other words, a Slater's condition is needed in order to guarantee a zero duality gap at the optimal solution.

Theorem 1 (Strong Duality Theorem) *$S^* \bullet X^* = 0$ if the primal problem (2.4) is strictly feasible, i.e., there exists $X \succ 0$ such that $\mathcal{A}X = b$, or the dual problem (2.7) is strictly feasible, i.e., there exists y such that $\mathcal{A}^T y \succ C$.*

If both conditions satisfy then the optimal solution is bounded.

Proof. See Alizadeh [1] ■

Without strict feasibility assumption the duality gap at the optimal solution may not be zero. We illustrate this by modifying an example of Vandenberghe and Boyd [58].

Example 2 *Consider problem (2.4), with $m = 2$, $n = 3$, $b^T = (1, 0)$, $c_{33} = -1$ and the coefficient matrix C is zero elsewhere,*

$$A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

From the constraint $A_1 \bullet X = b_1$, *we have* $2x_{12} + x_{33} = 1$ *and from constraint* $A_2 \bullet X = b_2$, *we get* $x_{22} = 0$ *and since* X *is a positive semidefinite matrix the entire row and column* 2 *must be zero. Therefore the problem is reduced to:*

$$\begin{aligned} \max \quad & -x_{33} \\ s.t. \quad & \\ & \begin{pmatrix} x_{11} & 0 & x_{13} \\ 0 & 0 & 0 \\ x_{13} & 0 & 1 \end{pmatrix} \succeq 0 \end{aligned}$$

This problem is feasible but not strictly feasible and the optimal objective value is $z_p^* = -1$. *The dual problem for this example as defined in (2.5) is as follows:*

$$\begin{aligned} \min \quad & y_1 \\ s.t. \quad & \\ & \begin{pmatrix} 0 & y_1 & 0 \\ y_1 & y_2 & 0 \\ 0 & 0 & y_1 + 1 \end{pmatrix} \succeq 0, \end{aligned}$$

this problem is feasible for any $y_1 = 0$ *and* $y_2 \geq 0$ *but not strictly feasible. The dual optimal objective value is* $z_d^* = 0$. *Thus* $z_p^* \neq z_d^*$ *or, the duality gap* η *is not zero.*

The following theorem provides the optimality conditions for semidefinite programming:

Theorem 3 (Complementary Slackness) *Assume that one of the primal or dual problems is strictly feasible. Then primal and dual feasible points* X^* *and* (y^*, S^*) *are optimal if and only if*

$$X^* S^* = 0.$$

Proof. See Alizadeh [1]. ■

2.2 Maximum Cut Problem

Applications of semidefinite programming in combinatorial optimization and nonconvex quadratic optimization problems have been developed in the past few years. In what follows we study semidefinite relaxations for Max-Cut Problem. The material of this section is mostly selected from Helmberg [21].

Semidefinite relaxation for the Maximum cut (Max-Cut) problem gives the tightest bound on the optimal objective value amongst that of other relaxations. Here we briefly review the derivation of this relaxation and its approximation result derived by Goemans and Williamson [13].

Max-Cut problem is one of the standard NP-hard problems and is defined as follows. Let $G(V, E)$ be an undirected weighted graph without loop or multiple edges. Let $|V| = n$ and $|E| = m$. The goal is to divide the node set V into two sets S and $V \backslash S$ such that the total weight of edges that connect these two sets is maximized:

$$\max_{S \subseteq V} \sum_{i \in S, j \notin S} w_{ij}.$$

The quadratic formulation of the problem is

$$\max \ \tfrac{1}{2} \textstyle\sum_{i<j} w_{ij}(1 - x_i x_j), \tag{2.8}$$

Where w_{ij} is the weight associated with edge ij and $x_i = 1$, where $i \in S$ and $x_i = -1$ otherwise.

Formulation of the max-cut problem can be improved using the Laplace Matrix of the graph G. Let $C = \frac{1}{4}L$, where $L = Diag(Ae) - A$ and A is the weighted adjacency matrix, then problem (2.8) can be reformulated as

$$\begin{aligned} \max \ & x^T C x \\ s.t. \ & \\ & x \in \{-1, 1\}^n . \end{aligned} \tag{2.9}$$

Now let $\mathcal{C}$ be the convex hull of $\left\{xx^T : x \in \{-1,1\}^n\right\}$. Since the extreme points of $\mathcal{C}$ are in $\{-1,1\}^n$, problem (2.9) can be written as follows:

$$\begin{aligned} \max \quad & C \bullet X \\ s.t. \quad & \\ & X \in \mathcal{C}. \end{aligned} \tag{2.10}$$

For any matrix C, an optimal extreme point solution of problem (2.10) is an optimal solution of problem (2.9). This problem would be a linear programming problem if we had an explicit description of the cut polytope $\mathcal{C}$. However, unfortunately for NP-hard problems, there is almost no hope of finding a good description. Although we do not have $\mathcal{C}$ explicitly but a tight upper bound can be found by an approximation of the cut polytope. Laurent and Poljak [31] proved that the following formulation is equivalent to the max-cut problem (2.10):

$$\begin{aligned} \max \quad & C \bullet X \\ s.t. \quad & \\ & diag(X) = e \\ & X \succeq 0 \\ & \text{rank}(X) = 1. \end{aligned} \tag{2.11}$$

Removing the rank constraint we come up with a semidefinite relaxation for the max-cut problem:

$$\begin{aligned} \max \quad & C \bullet X \\ s.t. \quad & \\ & diag(X) = e \\ & X \succeq 0. \end{aligned} \tag{2.12}$$

Using random hyperplane rounding scheme, Goemans and Williamson [13] obtain the following result:

Let $G(V, E)$ be an undirected graph with nonnegative edge weights and let m^* be the optimal objective value of problem (2.9). Then for any feasible solution of (2.12),

$$m^* \geq \alpha C \bullet X,$$

with $\alpha > 0.87856$. In particular, for the optimal matrix X^*,

$$\alpha C \bullet X^* \leq m^* \leq C \bullet X^*.$$

Moreover, Delorme and Poljak [11] prove that the semidefinite relaxation of the max-cut problem is asymptotically optimal for a large class of random graphs.

Chapter 3

The Analytic Center of Linear Matrix Inequalities

In this chapter we review the definition of the *analytic center for linear matrix inequalities* and introduce primal, dual and primal-dual barrier functions to derive the optimality conditions for the analytic center. We present two algorithms for computing an approximate analytic center of a convex compact set described by matrix inequalities based on primal and dual settings.

The chapter is organized as follows: Section 3.1 defines the exact analytic center and derives its optimality conditions based on primal, dual and prima-dual parameterization. In Section 3.2 we introduce an approximate analytic center and establish lower and upper bounds on the potential functions at an approximate center. We also define the Dikin's ellipsoids in primal and dual spaces. Section 3.3 is devoted to the primal algorithm for computation of an approximate center, and finally in Section 3.4 we present the dual algorithm for computing an approximate analytic center.

3.1 The Analytic Center: Optimality Conditions

Before we begin we state a very important lemma which plays a key role in the analysis of the interior point methods.

Lemma 4 *For a symmetric matrix X we have,*

$$\log\det X \leq I \bullet (X - I).$$

and if $||X||_\infty < 1$, then,

$$\log\det X \geq I \bullet (X - I) - \frac{||X - I||^2}{2(1 - ||X - I||_\infty)},$$

Proof. See Alizadeh [1]. ■

3.1.1 Dual parameterization

Consider the following set:

$$\Omega_D = \{y \in R^m : C - \mathcal{A}^T y \succeq 0\},$$

where $\mathcal{A}^T$ is the operator defined by (2.6) and $C \in \mathcal{S}_n$. We assume that Ω_D is a convex compact set and contains a strictly feasible point. That is

$$\Omega_D^\circ = \{y \in R^m : C - \mathcal{A}^T y \succ 0\},$$

is nonempty. Given a point y in Ω_D°, the *dual potential function* is defined via

$$\begin{aligned} \phi_D(y) &= \log\det(C - \mathcal{A}^T y)^{-1} \\ &= \log\det S^{-1} \end{aligned}$$

The minimizer of the dual potential function is called the analytic center:

$$y^a = \arg\min \phi_D(y). \tag{3.1}$$

Since $\phi_D(y)$ is strictly convex on Ω_D°, the analytic center is well defined and unique. Abusing notation somewhat, we also denote the dual potential function by $\phi_D(S)$.

For small symmetric matrix dS, the second order expansion of $\log\det S^{-1}$ is:

$$\begin{aligned}&\log\det(S+dS)^{-1} = \\ &\quad \log\det S^{-1} - \mathbf{t}rS^{-1}dS + \frac{1}{2}\mathbf{t}rS^{-1}(dS)S^{-1}(dS) + o(||dS||^2).\end{aligned} \tag{3.2}$$

Now the first optimality conditions of problem (3.1) can be derived by setting $\nabla\phi_D(S) = 0$. That is

$$\mathbf{t}rA_iS^{-1} = 0, \quad i = 1, ..., m,$$

or

$$A_i \bullet S^{-1} = 0, \quad i = 1, ..., m.$$

Let $X = S^{-1}$, then y^a is the analytic center of Ω_D iff there exist matrices $S^a \succ 0$ and $X^a \succ 0$ such that

$$\begin{aligned}\mathcal{A}X^a &= 0 \\ \mathcal{A}^Ty^a + S^a &= C \\ X^aS^a &= I\end{aligned} \tag{3.3}$$

3.1.2 Primal parameterization

The analytic center can also be derived using the primal potential function. Let

$$\Omega_P = \{X \in S^n : \mathcal{A}X = 0, X \succeq 0\},$$

and define the algebraic interior of Ω_P by

$$\Omega_P^\circ = \{X \in S^n : \mathcal{A}X = 0, X \succ 0\}.$$

Suppose that Ω_P° is nonempty. We show that the minimizer of the *primal potential function*

$$\phi_P(X) = C \bullet X - \log\det X,$$

over Ω_P satisfies the optimality conditions (3.3).

We are interested in solving the following convex optimization problem:

$$\begin{array}{ll} \min & C \bullet X - \log\det X \\ s.t. & \\ & \mathcal{A}X = 0 \\ & X \succ 0. \end{array} \tag{3.4}$$

The first order optimality conditions for this problem read

$$C - X^{-1} - \mathcal{A}^T y = 0,$$

letting $X^{-1} = S$, (3.3) is immediate.

3.1.3 Primal-Dual parameterization

The analytic center can alternatively be characterized as the minimizer of the *primal-dual potential function*:

$$\begin{aligned} \phi_{PD}(X,S) &= \phi_P(X) + \phi_D(S) \\ &= C \bullet X - \log\det X - \log\det S, \end{aligned}$$

over $\Omega_{PD} = \Omega_P \times \Omega_D$. Let us show that (3.3) is the optimality condition for this minimization problem. First observe that

$$\begin{aligned} \phi_{PD}(X^a, S^a) &= C \bullet X^a - \log\det X^a S^a \\ &= X^a \bullet S^a - \log\det X^a S^a \\ &= trI - \log\det I \\ &= n. \end{aligned}$$

On the other hand from Lemma 4

$$\begin{aligned}\phi_{PD}(X,S) &\geq X \bullet S - \mathrm{tr}(XS - I) \\ &= \mathrm{tr}XS - \mathrm{tr}XS + \mathrm{tr}I \\ &= n,\end{aligned}$$

for all $(X,S) \in \Omega_{PD}$, with equality iff $XS = I$. Thus (X^a, y^a, S^a) is the (unique) minimizer of $\phi_{PD}(X,S)$.

3.2 Approximate Analytic Centers

In practice, however, computation of the exact analytic center is impossible due to the round-off error. Approximate analytic centers are defined for computational reasons. A θ-approximate analytic center is denoted by $(\bar{X}, \bar{y}, \bar{S})$ and defined via

$$\begin{aligned}\mathcal{A}\bar{X} &= 0 \\ \mathcal{A}^T\bar{y} + \bar{S} &= C \\ \|\bar{X}\bar{S} - I\| &\leq \theta < 1.\end{aligned} \tag{3.5}$$

The next lemma provides lower and upper bounds on the potential functions at a θ-approximate center.

Lemma 5 *Let $(\bar{X}, \bar{y}, \bar{S})$ be a θ-approximate center. Then*

1. $\phi_P(X^a) \leq \phi_P(\bar{X}) \leq \phi_P(X^a) + \frac{\theta^2}{2(1-\theta)}$
2. $\phi_D(S^a) \leq \phi_D(\bar{S}) \leq \phi_D(S^a) + \frac{\theta^2}{2(1-\theta)}$
3. $n \leq \phi_{PD}(\bar{X}, \bar{S}) \leq n + \frac{\theta^2}{1-\theta}$

Proof. The left-hand side inequalities are trivial. We prove the upper bounds. From Lemma 4

$$\begin{aligned}\phi_{PD}(\bar{X}, \bar{S}) &\leq \bar{X} \bullet \bar{S} - I \bullet (\bar{X}\bar{S} - I) + \frac{\|\bar{X}\bar{S} - I\|^2}{2(1 - \|\bar{X}\bar{S} - I\|)} \\ &\leq n + \frac{\theta^2}{2(1-\theta)},\end{aligned}$$

and therefore

$$\phi_{PD}(\bar{X}, \bar{S}) - \phi_{PD}(X^a, S^a) \leq \frac{\theta^2}{2(1-\theta)},$$

or

$$(\phi_P(\bar{X}) - \phi_P(X^a)) + (\phi_D(\bar{S}) - \phi_D(S^a)) \leq \frac{\theta^2}{2(1-\theta)}.$$

Since (X^a, y^a, S^a) is the analytic center, then

$$\phi_D(\bar{S}) - \phi_D(S^a) \geq 0,$$

and

$$\phi_P(\bar{X}) - \phi_P(X^a) \geq 0.$$

Thus,

$$\phi_P(\bar{X}) - \phi_P(X^a) \leq \frac{\theta^2}{2(1-\theta)},$$

and

$$\phi_D(\bar{S}) - \phi_D(S^a) \leq \frac{\theta^2}{2(1-\theta)}.$$

Summing up the two inequalities, proves the upper bound in (3). ■

Computational algorithms for the analytic center of a polytope have been devised in primal, dual, and primal-dual settings based on Newton method. These algorithms can be extended to compute the analytic center of a convex body described by matrix inequalities [58]. In the next two sections we discuss the extension of the primal and the dual algorithms to compute an approximate analytic center. We refer the reader to [62, chapter 3] for a comprehensive analysis of the computational algorithms for the analytic center in the linear case. Before we start we define the Dikin ellipsoid:

3.2.1 Dikin ellipsoid

Let $\tilde{y} \in \Omega_D^\circ$ and $\tilde{S}$ be its slack matrix. A coordinate-aligned ellipsoid centered at $\tilde{y}$, and contained in the interior of Ω_D, i.e., the set

$$\{y : \|\tilde{S}^{-.5}\mathcal{A}^T(y - \tilde{y})\tilde{S}^{-.5}\| \leq 1\}$$

is called the dual Dikin ellipsoid. Similarly, we can define the Dikin ellipsoid for the primal feasible set. Let $\tilde{X} \succ 0$ be in Ω_P. Then, the primal Dikin ellipsoid centered at $\tilde{X}$ is

$$\{X : \|\tilde{X}^{-.5}(X - \tilde{X})\tilde{X}^{-.5}\| \leq 1\}.$$

3.3 Primal Algorithm

In this section we derive the Newton direction to solve problem (3.4) from a strictly primal feasible matrix. Let $X \in \Omega_P^\circ$, for small symmetric matrix dX one has,

$$\begin{aligned} \phi_P(X + dX) &= C \bullet (X + dX) + \log\det(X + dX)^{-1} \\ &\approx \phi_P(X) + (C - X^{-1}) \bullet (dX) + \frac{1}{2}trX^{-1}(dX)X^{-1}(dX). \end{aligned}$$

We minimize $\phi_P(X + dX)$ over the feasible directions $dX = (dX)^T$:

$$\begin{aligned} \min \quad & (C - X^{-1}) \bullet (dX) + \tfrac{1}{2}(dX)X^{-1} \bullet X^{-1}(dX) \\ s.t. \quad & \\ & \mathcal{A}(dX) = 0. \end{aligned} \tag{3.6}$$

Notice that the constraint $\mathcal{A}(dX) = 0$ guarantees the feasibility of the updated matrix $X^+ = X + dX$. The first order optimality conditions for problem (3.6) can be stated as

$$C - X^{-1} + X^{-1}(dX)X^{-1} - \mathcal{A}^T y = 0. \tag{3.7}$$

By multiplying X from the right side and from the left side to (3.7) and then applying the operator $\mathcal{A}$, with the feasibility of X and dX one has,

$$\mathcal{A}(X(\mathcal{A}^T y)X) = \mathcal{A}(XCX),$$

or

$$(\mathcal{A}_P\mathcal{A}_P^T)y = \mathcal{A}_P C_P, \tag{3.8}$$

where $\mathcal{A}_P : \mathcal{S}^n \to \Re^m$ is a linear operator and $\mathcal{A}_P^T : \Re^m \to \mathcal{S}^n$ is its adjoint operator, defined via

$$\mathcal{A}_P Y = \begin{pmatrix} X^{.5}A_1X^{.5} \bullet Y \\ \vdots \\ X^{.5}A_mX^{.5} \bullet Y \end{pmatrix}, \text{ and } \mathcal{A}_P^T y = \sum_{i=1}^{m} y_i X^{.5}A_iX^{.5},$$

and $C_P = X^{.5}CX^{.5}$. Note that $(\mathcal{A}_P\mathcal{A}_P^T) \in \mathcal{S}^m$ with $(\mathcal{A}_P\mathcal{A}_P^T)_{ij} = tr A_iXA_jX$. We assume that matrices A_i are linearly independent and therefore $\mathcal{A}_P\mathcal{A}_P^T \succ 0$.

From (3.8) (we denote y by $y(X)$ as it is a function of X)

$$y(X) = (\mathcal{A}_P\mathcal{A}_P^T)^{-1}\mathcal{A}_P C_P, \tag{3.9}$$

and in view of (3.7) we have

$$dX = X - XS(X)X, \tag{3.10}$$

where

$$S(X) = C - \mathcal{A}^T y(X). \tag{3.11}$$

3.3.1 Projection of dX on to the null space of $\mathcal{A}$

A feasible direction dX is a symmetric matrix such that $\mathcal{A}dX = 0$. This condition may not be satisfied after computing dX due to the computational round-off error. We therefore project the direction dX on to the null space of $\mathcal{A}$ at each iteration:

From (3.10) we have

$$d\bar{X} = X^{-.5}(dX)X^{-.5} = I - X^{.5}S(X)X^{.5}.$$

Let $\mathcal{P}_{\mathcal{A}_P}(d\bar{X})$ be the projection of $d\bar{X}$ on to the null space of $\mathcal{A}_P$, and $\mathcal{P}_{\mathcal{A}}(dX)$ be the projection of dX on to the null space of $\mathcal{A}$. Then

$$\mathcal{P}_{\mathcal{A}_P}(d\bar{X}) = \left(I - \mathcal{A}_P^T(\mathcal{A}_P\mathcal{A}_P^T)^{-1}\mathcal{A}_P\right) d\bar{X}$$

and thus

$$\begin{aligned}\mathcal{P}_{\mathcal{A}}(dX) &= X^{.5}\mathcal{P}_{\mathcal{A}_P}(d\bar{X})X^{.5} \\ &= dX - X^{.5}\left(\mathcal{A}_P^T(\mathcal{A}_P\mathcal{A}_P^T)^{-1}\mathcal{A}dX\right)X^{.5}\end{aligned}$$

Now let

$$q = (\mathcal{A}_P\mathcal{A}_P^T)^{-1}\mathcal{A}dX.$$

Thus

$$\begin{aligned}\mathcal{P}_{\mathcal{A}}(dX) &= dX - X^{.5}\mathcal{A}_P^T q X^{.5} \\ &= dX - X\mathcal{A}^T q X.\end{aligned}$$

or

$$\mathcal{P}_{\mathcal{A}}(dX) = dX - \sum_{i=1}^{m} q_i(XA_iX). \tag{3.12}$$

Note that $\mathcal{A}\mathcal{P}_{\mathcal{A}}(dX) = 0$.

Observe that

$$\begin{aligned}X^{-.5}(dX)X^{-.5} &= \mathcal{A}_P^T y - C_P + I &(3.13)\\ &= \mathcal{A}_P^T(\mathcal{A}_P\mathcal{A}_P^T)^{-1}\mathcal{A}_P C_P - C_P + I \\ &= -\left(I - \mathcal{A}_P^T(\mathcal{A}_P\mathcal{A}_P^T)^{-1}\mathcal{A}_P\right)(C_P - I) &(3.14)\\ &= -\mathcal{P}_{\mathcal{A}_P}(C_P - I),\end{aligned}$$

where $\mathcal{P}_{\mathcal{A}_P}(C_P - I)$ is the projection of $(C_P - I)$ on the null space of $\mathcal{A}_P$. Therefore, the primal direction can alternatively be stated via

$$dX = -X^{.5}\mathcal{P}_{\mathcal{A}_P}(C_P - I)X^{.5}.$$

This observation can be also seen from (3.13):

$$I - C_P = X^{-.5}(dX)X^{-.5} - \mathcal{A}_P^T y,$$

since $X^{-.5}(dX)X^{-.5}$ is in the null space of $\mathcal{A}_P$ and $\mathcal{A}_P^T y$ is in the range of $\mathcal{A}_P^T$ and this decomposition is unique, then $X^{-.5}(dX)X^{-.5}$ is the projection of $I - C_P$ on the null space of $\mathcal{A}_P$.

The following lemma shows that the slack matrix $S(X)$ can be characterized as a least square problem:

Lemma 6 *Let $X \in \Omega_P^\circ$. Then the dual solution $S(X)$ is the minimizer of the following least square problem:*

$$\begin{array}{ll} \min & ||X^{.5}SX^{.5} - I|| \\ s.t. & \\ & \mathcal{A}^T y + S = C \end{array} \tag{3.15}$$

Proof.

$$\begin{aligned} & ||X^{.5}SX^{.5} - I||^2 \\ &= \mathbf{tr}XSXS - 2\mathbf{tr}XS + \mathbf{tr}I \\ &= \mathbf{tr}X(C - \mathcal{A}^T y)X(C - \mathcal{A}^T y) - 2\mathbf{tr}X(C - \mathcal{A}^T y) + n \\ &= \mathbf{tr}\left(X^{.5}(C - \mathcal{A}^T y)X^{.5}\right)^2 - 2\mathbf{tr}X^{.5}(C - \mathcal{A}^T y)X^{.5} + n \\ &= \mathbf{tr}(C_P - \mathcal{A}_P^T y)^2 - 2\mathbf{tr}(C_P - \mathcal{A}_P^T y) + n \\ &= \mathbf{tr}C_P^2 - 2C_P \bullet \mathcal{A}_P^T y + \mathcal{A}_P^T y \bullet \mathcal{A}_P^T y - 2\mathbf{tr}C_P + 2\mathbf{tr}\mathcal{A}_P^T y + n \\ &= \mathbf{tr}C_P^2 - 2\mathbf{tr}C_P - 2y^T \mathcal{A}_P C_P + y^T(\mathcal{A}_P\mathcal{A}_P^T)y + 2\mathbf{tr}\mathcal{A}_P^T y + n. \end{aligned}$$

The first order optimality conditions read

$$2(\mathcal{A}_P\mathcal{A}_P^T)y - 2\mathcal{A}_P C_P + 2\mathcal{A}_P I = 0.$$

The proof follows from the fact that $\mathcal{A}_P I = \mathcal{A}X = 0$. ■

Let $P(X)$ be the optimal objective value of problem (3.15):

$$P(X) := X^{.5}S(X)X^{.5} - I, \tag{3.16}$$

and define the primal measure of proximity

$$\eta_p(X) = \|P(X)\|.$$

Observe that

$$\begin{aligned} P(X) &= X^{.5}(C - \mathcal{A}^T y(X))X^{.5} - I \\ &= X^{.5}(C - X^{-1})X^{.5} - X^{.5}\mathcal{A}^T y(X)X^{.5} \\ &= X^{.5}(C - X^{-1})X^{.5} - \mathcal{A}_P^T y(X) \\ &= \left(I - \mathcal{A}_P^T(\mathcal{A}_P\mathcal{A}_P^T)^{-1}\mathcal{A}_P\right)X^{.5}(C - X^{-1})X^{.5} \\ &= \mathcal{P}_{\mathcal{A}_P}X^{.5}(C - X^{-1})X^{.5} \\ &= \mathcal{P}_{\mathcal{A}_P}(C_P - I) \end{aligned}$$

The following lemma shows the strict feasibility of the primal direction and the rate of convergence when $\|P(X)\| < 1$.

Lemma 7 *If $\eta_p(X) \leq \alpha < 1$, then,*

$$X^+ \succ 0 \quad \textit{and} \quad \eta_p(X^+) \leq \eta_p(X)^2 < 1$$

Proof.

$$\begin{aligned} X^+ &= X + dX \\ &= X - X^{.5}\mathcal{P}_{\mathcal{A}_P}(C_P - I)X^{.5} \\ &= X - X^{.5}P(X)X^{.5} \\ &= X^{.5}(I - P(X))X^{.5}, \end{aligned}$$

and since $\|P(X)\| < 1$, then $\max|\lambda_j(P(X))| < 1$ and therefore $I - P(X) \succ 0$, i.e., $X^+ \succ 0$. To prove the rate of convergence, first observe that from Lemma 3.15

$$\begin{aligned} \|P(X^+)\| &= \|(X^+)^{.5}S(X^+)(X^+)^{.5} - I\| \\ &\leq \|(X^+)^{.5}S(X)(X^+)^{.5} - I\|. \end{aligned} \tag{3.17}$$

On the other hand,

$$dX = -X^{.5}P(X)X^{.5} = X - XS(X)X, \tag{3.18}$$

and thus $X^+ = 2X - XS(X)X$. Now from (3.17) and (3.18): (in what follows we denote $S(X)$ by S)

$$\begin{aligned} \|P(X^+)\|^2 &\leq \|S^{.5}X^+S^{.5} - I\|^2 \\ &= \|S^{.5}(2X - XSX)S^{.5} - I\|^2 \\ &= \|(S^{.5}XS^{.5} - I)^2\|^2 \\ &= \mathbf{t}r(S^{.5}XS^{.5} - I)^4 \\ &= \sum(\lambda_j(S^{.5}XS^{.5}) - 1)^4 \\ &\leq (\sum(\lambda_j(S^{.5}XS^{.5}) - 1)^2)^2 \\ &= \left(\|S^{.5}XS^{.5} - I\|^2\right)^2 \\ &= \|P(X)\|^4. \end{aligned}$$

The proof now follows. ■

Lemma (7) guarantees quadratic convergence within the primal Dikin ellipsoid. We now derive the complexity of the primal algorithm:

3.3.2 Complexity of Primal Algorithm

Let a strict feasible matrix X with $\|P(X)\| \geq 1$ is available. We apply a step size $\alpha < 1$ to reduce the primal potential function by a constant amount at

each iteration. Let $d'X = X^+ - X$ be the primal direction with step size α, and consider the reduction in the potential function when $\eta_p(X) \geq 1$:

$$\begin{aligned}
\phi_P(X^+) - \phi_P(X) &= C \bullet X^+ - C \bullet X - \log\det X^+ + \log\det X \\
&= C \bullet (X^+ - X) - \log\det X^{-1}X^+ \\
&\leq C \bullet d'X - X^{-1} \bullet d'X + \frac{||X^{-1}(d'X)||^2}{2(1 - ||X^{-1}(d'X)||)} \\
&\leq (C - X^{-1}) \bullet d'X + \frac{\alpha^2}{2(1-\alpha)}. \qquad (3.19)
\end{aligned}$$

From lemma (4) the last two inequalities are true if $\|X^{-1}(d'X)\| \leq \alpha < 1$. The best bound can be found by minimizing (3.19) over the feasible directions $d'X$:

$$\begin{aligned}
&\min \quad (C - X^{-1}) \bullet d'X \\
&s.t. \\
&\qquad \mathcal{A}(d'X) = 0 \\
&\qquad \|X^{-.5}(d'X)X^{-.5}\| \leq \alpha < 1.
\end{aligned} \qquad (3.20)$$

We employ a transformation technique to solve this problem. The idea is to transform problem (3.20) to an equivalent problem where the identity matrix I is feasible. To this end we define a new variable $X' := X^{-.5}X^+X^{-.5}$ and observe that the constraints can be written in terms of the new variable as follows,

$$X^{-.5}(d'X)X^{-.5} = X^{-.5}(X^+ - X)X^{-.5} = X' - I,$$

and,

$$\mathcal{A}(d'X) = X^{.5}\mathcal{A}X^{.5}X^{-.5}(d'X)X^{-.5} = \mathcal{A}_P(X' - I),$$

and the objective function,

$$\begin{aligned}
(C - X^{-1}) \bullet d'X &= X^{.5}(C - X^{-1})X^{.5} \bullet X^{-.5}(d'X)X^{-.5} \\
&= (C_P - I) \bullet (X' - I).
\end{aligned}$$

Therefore the transformed problem can be expressed as,

$$\begin{aligned} \min \quad & (C_P - I) \bullet (X' - I) \\ s.t. \quad & \\ & \mathcal{A}_P(X' - I) = 0, \\ & \|X' - I\| \leq \alpha < 1. \end{aligned} \tag{3.21}$$

The optimal solution to problem (3.21) can be computed from the identity matrix I. The direction towards the optimality must be on the null space of $\mathcal{A}_P$ and of the opposite direction of $C_P - I$ in order to minimize the problem. Thus,

$$X' = I - \frac{\alpha}{\|P(X)\|} P(X),$$

where,

$$P(X) = \mathcal{P}_{\mathcal{A}_P}(C_P - I).$$

Consequently,

$$X' - I = -\frac{\alpha}{\|P(X)\|} \mathcal{P}_{\mathcal{A}_P}(C_P - I),$$

and

$$\begin{aligned} (C - X^{-1}) \bullet d'X &= (C_P - I) \bullet (X' - I) \\ &= -\frac{\alpha}{\|P(X)\|} \|P(X)\|^2 \\ &= -\alpha \|P(X)\|. \end{aligned}$$

Therefore in view of (3.19) we have,

$$\begin{aligned} \phi_P(X^+) - \phi_P(X) &\leq -\alpha \|P(X)\| + \frac{\alpha^2}{2(1-\alpha)} \\ &\leq -\alpha + \frac{\alpha^2}{2(1-\alpha)} \\ &= \delta. \end{aligned}$$

This shows that by taking the step size α when the iteration is outside the Dikin ellipsoid the potential function is reduced by $\delta > 0$ at each step.

Therefore after a finite number of iterations, $\|P(X)\|$ satisfies the desired condition (< 1) and then by taking the pure Newton's steps the algorithm quadratically converges to the analytic center. The complexity of the primal algorithm can be obtained from the fact that the primal potential function at the analytic center is a lower bound for $\phi_P(X)$. This implies that after at most

$$O(\phi_P(X) - \phi_P(X^a))$$

Newton steps the algorithm stops with an approximate analytic center.

In practice, however, we find the step size by a line search method to get as close as possible to the Dikin ellipsoid. We now describe this procedure:

Suppose that the primal direction dX is obtained by (3.10). We want to find the step size α such that $\phi_P(X + \alpha dX)$ is minimized. Observe that

$$\begin{aligned}\phi_P(X + \alpha dX) &= C \bullet (X + \alpha dX) + \log\det(X + \alpha dX)^{-1} \\ &= C \bullet X + \alpha C \bullet dX - \log\det X - \sum_{j=1}^{n} \log(1 + \alpha\lambda_j),\end{aligned}$$

where λ_j are the eigenvalues of symmetric matrix $X^{-.5}dXX^{-.5}$. Now by taking the differential with respect to α and setting it to zero, one has

$$C \bullet dX - \sum_{j=1}^{n} \frac{\lambda_j}{1 + \alpha\lambda_j} \equiv 0.$$

We solve the above equation by a line search. The initial interval is set such that

$$X^+ = X + \alpha dX = X^{.5}\left(I + \alpha X^{-.5}dXX^{-.5}\right)X^{.5} \succ 0.$$

Note that, if $\alpha\lambda_{\min}\left(X^{-.5}dXX^{-.5}\right) > -1$, then $X + \alpha dX \succ 0$. Thus

If $\lambda_{\min}\left(X^{-.5}dXX^{-.5}\right) < 0$, we choose

$$\alpha < \frac{-1}{\lambda_{\min}\left(X^{-.5}dXX^{-.5}\right)}$$

Otherwise, any $\alpha > 0$ would serve our goal.

3.3.3 Primal Algorithm

The algorithm can be summarized as follows:

Algorithm 1 (Primal Algorithm) *Given* $X^0 \succ 0$, $\epsilon > 0$, *and* $0 < \alpha < .075$, *set k=0.*

Step 1. *Compute* $y(X^k)$ *from (3.9) and* $S(X^k)$ *from (3.11)*

Step 2. *Compute primal direction* dX *from (3.10)*

Step 3. *Compute the projection of* dX *on to the null space of* $\mathcal{A}$ *by (3.12)*

Step 4. *Compute* $P(X^k)$ *from (3.16) and the measure of proximity* $\eta_p(X^k) = \|P(X^k)\|$,

Step 5. *If* $\eta_p(X^k) < 0.75$ *take the pure Newton's directions and set* $X^{k+1} = X^k + dX$

Step 6. *If* $\eta_p(X^k) \geq 0.75$, *compute the step size* α *and set* $X^{k+1} = X^k + \alpha dX$,

Step 7. *If* $\phi_P(X^{k+1}) - \phi_P(X^k) < \epsilon$ *go to step 8,*

Step 8. *Set* $k = k + 1$ *and return to step 1.*

Step 9. *Update* $y(X)$ *and* $S(X)$ *and stop.*

3.4 Dual Algorithm

In this section we discuss the dual algorithm for solving problem (3.1). Recall that

$$\Omega_D = \{y \in R^m : C - \mathcal{A}^T y \succeq 0\}.$$

and Ω_D° is the strict feasible points of Ω_D. Let $y \in \Omega_D^\circ$ and dy (and $dS = dS^T$) be the dual direction. Consider the quadratic approximation of $\phi_D(S)$:

$$\begin{aligned}\phi_D(S+dS) &= \log\det(S+dS)^{-1} \\ &\approx \log\det S^{-1} - \mathbf{t}rS^{-1}dS + \frac{1}{2}\mathbf{t}rS^{-1}(dS)S^{-1}(dS).\end{aligned}$$

By feasibility of dy one has:

$$\mathcal{A}^T(y+dy) + (S+dS) = C$$

or,

$$\mathcal{A}^T dy + dS = 0$$

Therefore,

$$\phi_D(y+dy) = \phi_D(y) + (\mathcal{A}S^{-1})^T dy + \frac{1}{2}dy^T(\mathcal{A}_D\mathcal{A}_D{}^T)dy,$$

where $\mathcal{A}_D : \mathcal{S}^n \to \Re^m$ is a linear operator and $\mathcal{A}_D^T : \Re^m \to \mathcal{S}^n$ is its adjoint operator, defined via

$$\mathcal{A}_D X = \begin{pmatrix} S^{-.5}A_1S^{-.5} \bullet X \\ \vdots \\ S^{-.5}A_mS^{-.5} \bullet X \end{pmatrix}, \text{ and } \mathcal{A}_D^T y = \sum_{i=1}^m y_i S^{-.5}A_iS^{-.5}$$

Note that $(\mathcal{A}_D\mathcal{A}_D^T) \in \mathcal{S}^m$ with $(\mathcal{A}_D\mathcal{A}_D^T)_{ij} = \mathbf{t}rA_iS^{-1}A_jS^{-1}$. Since A_i are linearly independent, then $\mathcal{A}_D\mathcal{A}_D^T \succ 0$.

Note that,

$$\begin{aligned}\frac{1}{2}\mathbf{t}rS^{-1}(dS)S^{-1}(dS) &= \frac{1}{2}\mathbf{t}rS^{-1}(-\mathcal{A}^T dy)S^{-1}(-\mathcal{A}^T dy) \\ &= \frac{1}{2}\mathbf{t}r\left(\sum_i dy_i S^{-1}A_i\right)\left(\sum_j dy_j S^{-1}A_j\right) \\ &= \frac{1}{2}\sum_{i,j} dy_i dy_j \mathbf{t}r\left(S^{-1}A_iS^{-1}A_j\right) \\ &= \frac{1}{2}dy^T(\mathcal{A}_D\mathcal{A}_D{}^T)dy.\end{aligned}$$

By minimizing the quadratic approximation of $\phi_D(y)$

$$\min_{dy \in R^m} \quad (\mathcal{A}S^{-1})^T dy + \tfrac{1}{2} dy^T (\mathcal{A}_D \mathcal{A}_D{}^T) dy. \tag{3.22}$$

one has

$$dy = -(\mathcal{A}_D \mathcal{A}_D{}^T)^{-1} \mathcal{A} S^{-1} \tag{3.23}$$

and

$$dS = \mathcal{A}^T (\mathcal{A}_D \mathcal{A}_D{}^T)^{-1} \mathcal{A} S^{-1} \tag{3.24}$$

Now let

$$X(S) = S^{-1} \left(\mathcal{A}^T dy + S \right) S^{-1} \tag{3.25}$$

The following lemma shows that $X(S)$ is a solution of a least square problem:

Lemma 8 *Let $y \in \Omega_D^\circ$ and let S be the slack matrix. Then the primal solution $X(S)$ of the analytic center is the minimizer of the following least square problem:*

$$\begin{aligned} \min \quad & ||S^{.5} X S^{.5} - I|| \\ s.t. \quad & \\ & \mathcal{A}X = 0 \end{aligned} \tag{3.26}$$

Proof. The objective function can be written as,

$$\begin{aligned} \left|\left| S^{.5} X S^{.5} - I \right|\right|^2 &= \mathbf{t}r(SXSX - 2SX + I) \\ &= \mathbf{t}r(SXSX) - 2\mathbf{t}rSX + n \end{aligned}$$

The KKT condition for this problem is

$$2SXS - 2S - \mathcal{A}^T v = 0, \tag{3.27}$$

where $v \in R^m$. By multiplying the matrix S^{-1} from the right side and from the left side to equation (3.27) and then applying the operator $\mathcal{A}$ and noting that $\mathcal{A}X = 0$, we have:

$$-2\mathcal{A}S^{-1} - (\mathcal{A}_D \mathcal{A}_D{}^T) v = 0$$

or

$$\begin{aligned} v &= -2(\mathcal{A}_D \mathcal{A}_D{}^T)\mathcal{A}S^{-1} \\ &= 2dy \end{aligned}$$

and thus from (3.27),

$$X(S) = S^{-1}\left(\mathcal{A}^T dy + S\right)S^{-1}$$

The proof follows. ■

Let $P(S)$ be the optimal objective value of problem (3.26):

$$P(S) = S^{.5}X(S)S^{.5} - I, \tag{3.28}$$

note that

$$\begin{aligned} P(S) &= S^{.5}\left(S^{-1}\left(\mathcal{A}^T dy + S\right)S^{-1}\right)S^{.5} - I \\ &= S^{-.5}\mathcal{A}^T dy S^{-.5} \\ &= -S^{-.5}(dS)S^{-.5}. \end{aligned}$$

The following lemma shows that if $\|P(S)\| < 1$ then the updated slack matrix $S^+ = S + dS$ is strictly feasible and the dual algorithm converges to an approximate analytic center quadratically.

Lemma 9 *If $\eta_d(S) = \|P(S)\| < 1$ for some interior point $y \in \Omega_D^\circ$ and its slack matrix S, then*

$$S^+ \succ 0 \quad \textit{and} \quad \eta_d(S^+) \leq \eta_d(S)^2 < 1$$

Proof. Note that

$$\begin{aligned} S^+ &= S + dS \\ &= S - S^{.5}P(S)S^{.5} \\ &= S^{.5}(I - P(S))S^{.5}, \end{aligned}$$

and since $||P(S)|| < 1$, then $\lambda_j P(S) < 1$, for $j = 1, ..., n$, which implies that $S^+ \succ 0$.

To prove the second part of the lemma first observe that $X(S)$ is the minimizer of problem (3.26) and thus

$$\begin{aligned} \eta_d(S^+) &= \left\| P(S^+) \right\| \\ &= \left\| (S^+)^{.5} X(S^+)(S^+)^{.5} - I \right\| \\ &\leq \left\| (S^+)^{.5} X(S)(S^+)^{.5} - I \right\|, \end{aligned} \tag{3.29}$$

for all X with $\mathcal{A}X = 0$. On the other hand,

$$S^+ = S - \mathcal{A}^T dy = 2S - (\mathcal{A}^T dy + S) = 2S - SX(S)S. \tag{3.30}$$

Now from (3.29) and (3.30) (denoting $X(S)$ by X)

$$\begin{aligned} \left\| (S^+)^{.5} X(S)(S^+)^{.5} - I \right\|^2 &= \left\| X^{.5} S^+ X^{.5} - I \right\|^2 \\ &= \left\| X^{.5}(2S - SX(S)S)X^{.5} - I \right\|^2 \\ &= \left\| X^{.5} SX(S)SX^{.5} - 2X^{.5} SX^{.5} + I \right\|^2 \\ &= \left\| (X^{.5} SX^{.5} - I)^2 \right\|^2 \\ &= \mathbf{tr}(X^{.5} SX^{.5} - I)^4 \\ &= \sum (\lambda_j (X^{.5} SX^{.5}) - 1)^4 \\ &\leq \left(\sum (\lambda_j (X^{.5} SX^{.5}) - 1)^2 \right)^2 \\ &= \left(\left\| X^{.5} SX^{.5} - I \right\|^2 \right)^2 \\ &= \left\| S^{.5} XS^{.5} - I \right\|^4 \\ &= \eta_d(S)^4, \end{aligned}$$

The proof now follows. ■

3.4.1 Complexity of Dual Algorithm

Assume that (y, S), with $S \succ 0$ is given such that $\eta_d(S) \geq 1$, we consider the new point y^+ with step size $\alpha/\eta_d(S)$, for $0 < \alpha < 1$:

$$y^+ = y - \frac{\alpha}{\eta_d(S)}(\mathcal{A}_D\mathcal{A}_D{}^T)^{-1}\mathcal{A}S^{-1}, \tag{3.31}$$

Let

$$d'y = -\tfrac{\alpha}{\eta_d(S)}(\mathcal{A}_D\mathcal{A}_D{}^T)^{-1}\mathcal{A}S^{-1}, \quad \text{and} \quad d'S = -\mathcal{A}^T d'y,$$

We prove that by taking the direction $d'y$ the dual potential function $\phi_D(y)$ is reduced by a constant amount at each iteration and after a finite number of Newton steps the iteration lies within the quadratic convergence region.

Consider the reduction in the potential function when $\|S^{-1}S^+ - I\| \leq \alpha < 1$,

$$\begin{aligned}
\phi_D(y^+) - \phi_D(y) &= -\log\det S^{-1}S^+ \\
&\leq -I \bullet (S^{-1}S^+ - I) + \frac{\|S^{-1}S^+ - I\|^2}{2(1 - \|S^{-1}S^+ - I\|)} \qquad (3.32)\\
&= -\mathbf{tr}S^{-1}d'S + \frac{\|S^{-1}d'S\|^2}{2(1 - \|S^{-1}d'S\|)} \\
&= \mathbf{tr}S^{-1}\mathcal{A}^T d'y + \frac{\|S^{-1}d'S\|^2}{2(1 - \|S^{-1}d'S\|)} \\
&= (\mathcal{A}S^{-1})^T d'y + \frac{\|S^{-1}d'S\|^2}{2(1 - \|S^{-1}d'S\|)}.
\end{aligned}$$

Thus

$$\begin{aligned}
\phi_D(y^+) - \phi_D(y) &\leq \nabla^T\phi_D(y)d'y + \frac{\left\|S^{-1}\mathcal{A}^T d'y\right\|^2}{2(1 - \|S^{-1}\mathcal{A}^T d'y\|)}. \qquad (3.33)\\
&= (\mathcal{A}S^{-1})^T d'y + \frac{\alpha^2}{2(1-\alpha)} \qquad (3.34)
\end{aligned}$$

The best bound on the reduction of potential function can be obtained by minimizing the right hand side of (3.33).

$$\begin{aligned} \min \quad & \nabla^T \phi_D(y) d'y \\ s.t \quad & \\ & \left\| S^{-1} \mathcal{A}^T d'y \right\| \leq \alpha, \quad \alpha < 1 \end{aligned}$$

or, in the symmetric form

$$\begin{aligned} \min \quad & (\mathcal{A}S^{-1})^T d'y \\ s.t \quad & \\ & \left\| S^{-.5} \mathcal{A}^T d'y S^{-.5} \right\| \leq \alpha, \quad \alpha < 1 \end{aligned} \tag{3.35}$$

The optimality condition for problem (3.35) is

$$\mathcal{A}S^{-1} + \mu \mathcal{A}S^{-1}(\mathcal{A}^T d'y)S^{-1} = 0, \quad \mu \geq 0$$

or

$$d'y = -\frac{1}{\mu}(\mathcal{A}_D \mathcal{A}_D{}^T)^{-1} \mathcal{A}S^{-1}.$$

The correct value of μ can be found by the constraint

$$\begin{aligned} \alpha^2 &= \left\| S^{-.5} \mathcal{A}^T d'y S^{-.5} \right\|^2 \\ &= \mathrm{tr} S^{-1}(\mathcal{A}^T d'y) S^{-1}(\mathcal{A}^T d'y) \\ &= \frac{1}{\mu^2} \left\| S^{-1} \mathcal{A}^T (\mathcal{A}_D \mathcal{A}_D{}^T)^{-1} \mathcal{A}S^{-1} \right\|^2 \\ &= \frac{1}{\mu^2} \left\| S^{-.5} dSS^{-.5} \right\|^2 \\ &= \frac{1}{\mu^2} \left\| P(S) \right\|^2, \end{aligned}$$

and thus $1/\mu = \alpha/\eta_d(S)$. Thus the optimal objective value of this problem can be bounded above by $-\alpha$:

$$\nabla^T \phi_D(y) d'y = \frac{-\alpha}{\eta_d(S)} (\mathcal{A}S^{-1})^T (\mathcal{A}_D \mathcal{A}_D{}^T)^{-1} \mathcal{A}S^{-1}$$

$$
\begin{aligned}
&= \frac{-\alpha}{\eta_d(S)} tr S^{-1} \mathcal{A}^T (\mathcal{A}_D \mathcal{A}_D{}^T)^{-1} \mathcal{A} S^{-1} \\
&= \frac{-\alpha}{\eta_d(S)} \eta_d^2(S) \\
&= -\alpha \eta_d(S) \\
&\leq -\alpha.
\end{aligned}
$$

Consequently

$$\phi_D(y^+) - \phi_D(y) \leq -\alpha + \frac{\alpha^2}{2(1-\alpha)}.$$

Let $\delta = \alpha - \alpha^2/2(1-\alpha)$. Thus

$$\phi_D(y^+) \leq \phi_D(y) - \delta$$

where $\delta > 0$ for $\alpha < 1/3$.

Like in primal case, in practice, we find the step size α by a line search method. Suppose that the dual direction dS is obtained by (3.24)

$$
\begin{aligned}
\phi_D(S + \alpha dS) &= \log\det(S + \alpha dS)^{-1} \\
&= -\log\det S - \sum_{j=1}^{n} \log(1 + \alpha\lambda_j),
\end{aligned}
$$

where λ_js are the eigenvalues of symmetric matrix $S^{-.5} dS S^{-.5}$. The minimizer of the dual potential function can be found by solving

$$\sum_{j=1}^{n} \frac{\lambda_j}{1 + \alpha\lambda_j} \equiv 0$$

for α. The initial interval of the line search is set such that

$$S^+ = S + \alpha dS = S^{.5} \left(I + \alpha S^{-.5} dS S^{-.5} \right) S^{.5} \succ 0.$$

If $\alpha\lambda_{\min}\left(S^{-.5} dS S^{-.5}\right) > -1$, then $S + \alpha dS \succ 0$. Thus, If $\lambda_{\min}\left(S^{-.5} dS S^{-.5}\right) < 0$, we choose

$$\alpha < \frac{-1}{\lambda_{\min}\left(S^{-.5} dS S^{-.5}\right)}$$

Otherwise, any $\alpha > 0$ would serve our goal.

The general complexity of the algorithm can be obtained from the fact that the dual potential function at the analytic center is a lower bound for $\phi_D(y)$ and hence after at most

$$O(\phi_D(y) - \phi_D(y^a))$$

iterations the algorithm stops with an approximation of the analytic center.

The analysis of the primal-dual algorithm is more or less the same as the analysis presented for primal and dual cases. The complexity result, however, for the primal-dual case is more specific since the potential function at the analytic center is known in advance. That is the complexity of the primal-dual algorithm is

$$O(\phi_{PD}(X^0, y^0, S^0) - n).$$

3.4.2 Dual Algorithm

The dual algorithm for computing an approximate analytic center of Ω_D can now be presented:

Algorithm 2 (Dual Algorithm) *Given* (y^0, S^0) *strictly feasible,* $\epsilon > 0$, *and* $0 < \alpha < .075$, *set k=0.*

Step 1. *Compute the dual directions dy and dS from (3.23) and (3.24),*

Step 2. *Compute* $X(S)$ *from (3.25) and the measure of proximity* $\eta_d(S^k)$ *from (3.28),*

Step 3. *If* $\eta_d(S^k) < 0.75$ *take the pure Newton's directions and set* $y^{k+1} = y^k + dy$ *and* $S^{k+1} = S^k + dS$

Step 4. *If* $\eta_d(S^k) \geq 0.75$ *set* $y^{k+1} = y^k + \frac{\alpha}{\eta_d(S^k)} dy$ *and* $S^{k+1} = S^k + \frac{\alpha}{\eta_d(S^k)} dS$,

Step 5. *If* $\phi_D(y^{k+1}) - \phi_D(y^k) < \epsilon$ *go to step 7,*

Step 6. *Set* $k = k + 1$ *and return to step 1.*

Step 7. *Update* $X(S)$ *from (3.25) and stop.*

Chapter 4

Analytic Center Cutting Plane Method and its Complexity

In this chapter we present an algorithm based on the *analytic center cutting plane method (ACCPM) with semidefinite cuts* and discuss its complexity in detail. We start with introducing the convex feasibility problem in Section 4.1. We define the semidefinite cuts and derive the primal and dual directions after adding a central semidefinite cut. In Section 4.2 we analyze the restoration procedure and establish an upper bound on the number of damped Newton steps after adding cuts. The ACCPM algorithm is presented in Section 4.3. Section 4.4 is devoted to the convergence analysis of the algorithm. We derive the complexity of our algorithm in this section. We complete our analysis by setting up the complexity of the Newton algorithm for computing a strictly feasible point after adding cuts in Section 4.5.

4.1 Convex Feasibility Problem

For the purpose of proving complexity results, ACCPM is more clearly described in the context of a convex feasibility problem: find a point in a

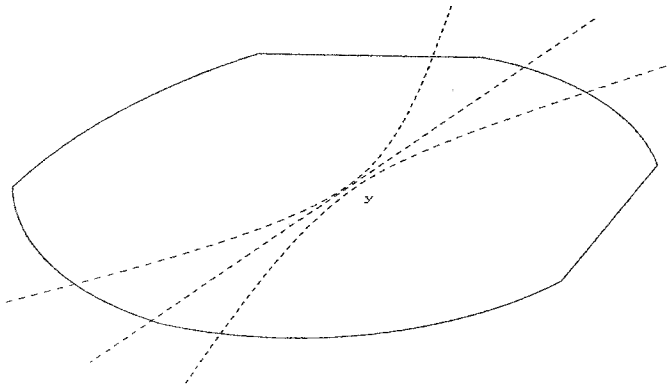

Figure 4.1: A 2-dimensional semidefinite cut

bounded convex set Ω^*, with a nonempty interior. The solution set Ω^* is assumed to contain a ball $\mathcal{N}_\varepsilon$, with radius $\varepsilon < 1$ and is contained in a compact convex set described by matrix inequalities, so called the *set of localization*. At each iteration the analytic center of the set of localization is computed and a separation oracle is called: the oracle determines if either the center is in Ω^*, thus solving the problem, or returns a cut which cuts off the current point and contains the solution set. A special updating step is then needed to get as close as possible to the next analytic center, as first suggested by Mitchell and Todd [36]. First, we define the semidefinite cuts.

Definition 10 *A p-dimensional semidefinite cut is a cut of the form*

$$\mathcal{B}^T y \preceq D,$$

where $D \in \mathcal{S}^p$ and $\mathcal{B} : \mathcal{S}^p \to \Re^m$ is a linear operator defined by $(\mathcal{B}X)_i = B_i \bullet X$, with $B_i \in \mathcal{S}^p$; and $\mathcal{B}^T y = \sum_{i=1}^m y_i B_i$ is its adjoint operator. The matrices B_i are called the cut matrices, and if $D = \mathcal{B}^T \bar{y}$, where $\bar{y}$ is an approximate analytic center, then the cut is called a central semidefinite cut.

Figure 4.1 shows a semidefinite cut in $\mathcal{S}^2$. Notice that the semidefinite cut $\mathcal{B}^T y \preceq D$ is a generalization of linear, multiple, and quadratic cuts. If

B_i and D are scalar, then $\mathcal{B}^T y \preceq D$ is reduced to a single cut $b^T y \leq b_0$ and if they are diagonal matrices, then the cut is reduced to a set of multiple linear cuts $B^T y \leq diag(D)$, where the columns of matrix B are $diag(B_i)$. Furthermore, if the cut matrices B_i and the constant matrix D are of the form of

$$B_i = \begin{pmatrix} 0 & -b_i \\ -b_i^T & q_i \end{pmatrix}, \quad D = \begin{pmatrix} \mathrm{I} & 0 \\ 0 & d \end{pmatrix},$$

then $\mathcal{B}^T y \preceq D$ is reduced to a quadratic cut $y^T(B^T B)y + q^T y \leq d$, where the vectors b_i form the columns of matrix B.

Let

$$\Omega_D = \{y \in R^m : \mathcal{A}^T y \preceq C\}$$

be the current set of localization. Then the updated set $\Omega_D^+ \subset \Omega_D$ after adding a central semidefinite cut is

$$\Omega_D^+ = \{y \in R^m : \mathcal{A}^T y \preceq C,\ \mathcal{B}^T y \preceq \mathcal{B}^T \bar{y}\}.$$

To compute an approximate center of the updated set of localization, we need a strict interior point of Ω_D^+. We start from $\bar{y}$ and choose the direction $dy = y - \bar{y}$ towards the interior of the set of localization as the maximizer of the determinant of the new slack matrix to the boundary of the dual Dikin ellipsoid centered at $\bar{y}$.

$$\begin{aligned} \min \quad & -\log\det \Lambda \\ s.t. \quad & \\ & \|\bar{S}^{-.5}\mathcal{A}^T dy \bar{S}^{-.5}\| \leq 1 \\ & \mathcal{B}^T dy + \Lambda = 0 \\ & \Lambda \succeq 0. \end{aligned} \tag{4.1}$$

We call the optimal solution of Problem (4.1), the *optimal updating direction.* An interior point of Ω_D^+ is obtained from the analytic center of Ω_D and the optimal updating direction. We call this point the *warm start* and it will

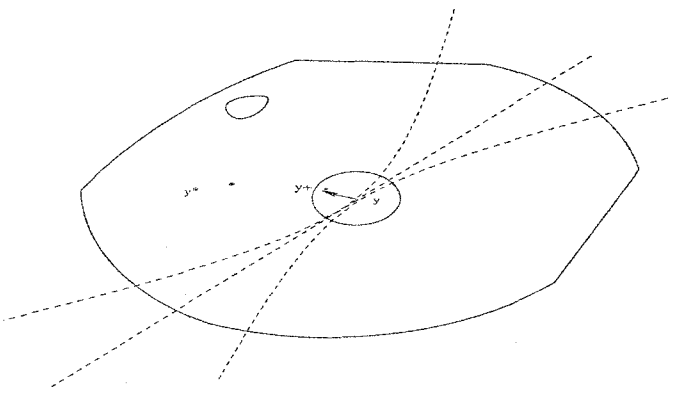

Figure 4.2: Optimal updating direction

be used as an initial point to compute the analytic center of Ω_D^+. Figure 4.2 gives a pictorial view of Problem 4.1. It shows the updated set of localization and the optimal updating direction over Ω_D^+, as well as the warm start.

Problem (4.1) can be reformulated as follows:

$$\begin{aligned} \min \quad & -\log\det(-\mathcal{B}^T dy) \\ s.t. \quad & \\ & dy^T(\mathcal{A}_D\mathcal{A}_D^T)dy \leq 1. \end{aligned}$$

By the KKT optimality conditions, $\tilde{dy}$ and $\tilde{\Lambda}$ are optimal iff there exists unique multiplier $\sigma \geq 0$ such that,

$$\mathcal{B}\tilde{\Lambda}^{-1} + \sigma(\mathcal{A}_D\mathcal{A}_D^T)\tilde{dy} = 0 \tag{4.2}$$

$$\mathcal{B}^T\tilde{dy} + \tilde{\Lambda} = 0, \tag{4.3}$$

From (4.2)

$$\tilde{dy} = -\frac{1}{\sigma}(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{\Lambda}^{-1} \tag{4.4}$$

and from (4.3)

$$\tilde{\Lambda} = \frac{1}{\sigma}\mathcal{B}^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{\Lambda}^{-1}.$$

Define the operator $\mathcal{V} : \mathcal{S}^p \to \mathcal{S}^p$ by $\mathcal{V} = \mathcal{B}^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}$, then

$$\tilde{\Lambda} = \frac{1}{\sigma}\mathcal{V}\tilde{\Lambda}^{-1}. \tag{4.5}$$

If $\mathcal{V}$ is nonsingular, then the dual direction $\tilde{\Lambda}$ can be uniquely computed by solving the following optimization problem

$$\min_{\Lambda \succeq 0} \quad \tfrac{\sigma}{2}\mathbf{tr}\Lambda\mathcal{V}^{-1}\Lambda - \log\det\Lambda,$$

The correct value of the Lagrange multiplier $\sigma \geq 0$ is known in advance:

$$\begin{aligned}
\|S^{-.5}\mathcal{A}^T\tilde{dy}S^{-.5}\|^2 &= \tilde{dy}^T(\mathcal{A}_D\mathcal{A}_D^T)\tilde{dy} \\
&= \frac{1}{\sigma^2}(\mathcal{B}\tilde{\Lambda}^{-1})^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{\Lambda}^{-1} \qquad \text{(from (4.4))} \\
&= \frac{1}{\sigma^2}\mathbf{tr}\tilde{\Lambda}^{-1}\mathcal{B}^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{\Lambda}^{-1} \\
&= \frac{1}{\sigma^2}\mathbf{tr}\tilde{\Lambda}^{-1}\mathcal{V}\tilde{\Lambda}^{-1} \\
&= \frac{1}{\sigma}\mathbf{tr}\tilde{\Lambda}^{-1}\tilde{\Lambda} \qquad \text{(from (4.5))} \\
&= \frac{p}{\sigma}.
\end{aligned}$$

On the other hand, $\|\bar{S}^{-.5}\mathcal{A}^T\tilde{dy}\bar{S}^{-.5}\| = 1$ and thus $\sigma = p$. Consequently,

$$\tilde{dy} = -\tfrac{1}{p}(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{\Lambda}^{-1} \quad \text{and} \quad \tilde{dS} = -\mathcal{A}^T\tilde{dy}$$

where,

$$\tilde{\Lambda} = \arg\min_{\Lambda \succeq 0}\{\frac{p}{2}\mathbf{tr}\Lambda\mathcal{V}^{-1}\Lambda - \log\det\Lambda\}.$$

To update the primal direction, observe that the updated primal feasible region Ω_P^+ is

$$\Omega_P^+ = \left\{ \begin{pmatrix} X & \\ & T \end{pmatrix} \succeq 0 : \mathcal{A}X + \mathcal{B}T = 0 \right\},$$

and the primal direction $d\tilde{X}$ is obtained by maximizing $\log\det T$ while respecting primal feasibility and remaining in the primal Dikin ellipsoid centered at $\bar{X}$:

$$\begin{aligned} \min \quad & -\log\det T \\ s.t. \quad & \\ & \mathcal{A}dX + \mathcal{B}T = 0 \\ & \|\bar{S}dX\| \leq 1 \\ & T \succeq 0. \end{aligned} \tag{4.6}$$

The optimality conditions of problem (4.6) are

$$-\tilde{T}^{-1} + \mathcal{B}^T v = 0 \tag{4.7}$$

$$\mathcal{A}^T v + \sigma' \bar{S}(d\tilde{X})\bar{S} = 0 \tag{4.8}$$

$$\sigma'(1 - \|\bar{S}d\tilde{X}\|) = 0 \tag{4.9}$$

$$\mathcal{A}(d\tilde{X}) + \mathcal{B}\tilde{T} = 0, \tag{4.10}$$

where $\sigma' \geq 0$ is the Lagrange multiplier associated with the norm constraint. By multiplying equation (4.8) from the left and from the right by $\bar{S}^{-1}$ and then applying the operator $\mathcal{A}$ we have,

$$(\mathcal{A}_D\mathcal{A}_D^T)v + \sigma'\mathcal{A}(d\tilde{X}) = 0,$$

using (4.10)

$$v = \sigma'(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{T},$$

and again from (4.8)

$$\begin{aligned} d\tilde{X} &= -\frac{1}{\sigma'}\bar{S}^{-1}(\mathcal{A}^T v)\bar{S}^{-1} \\ &= -\bar{S}^{-1}\mathcal{A}^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{T}\bar{S}^{-1}. \end{aligned}$$

Since $\mathcal{A}^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{T}$ is symmetric, then $d\tilde{X}$ is symmetric.

Finally from (4.7)

$$\begin{aligned}\tilde{T}^{-1} &= \mathcal{B}^T v \\ &= \sigma' \mathcal{B}^T (\mathcal{A}_D \mathcal{A}_D^T)^{-1} \mathcal{B}\tilde{T} \\ &= \sigma' \mathcal{V}\tilde{T}, \end{aligned} \tag{4.11}$$

and $\tilde{T}$ is the unique solution of the following optimization problem:

$$\tilde{T} = \arg\min_{T \succeq 0} \{ \frac{\sigma'}{2} \mathbf{tr} T\mathcal{V}T - \log\det T \}. \tag{4.12}$$

Let us find the Lagrange multiplier σ'.

$$\begin{aligned}\|\bar{S}(d\tilde{X})\|^2 &= \mathbf{tr}\bar{S}^{.5}(d\tilde{X})\bar{S}(d\tilde{X})\bar{S}^{.5} \\ &= \mathbf{tr}\bar{S}^{-1}\mathcal{A}^T \underbrace{(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{T}}_{u} \bar{S}^{-1}\mathcal{A}^T \underbrace{(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{T}}_{u} \\ &= u^T(\mathcal{A}_D\mathcal{A}_D^T)u \\ &= (\mathcal{B}\tilde{T})^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}\tilde{T} \\ &= \mathbf{tr}\tilde{T}\mathcal{V}\tilde{T} \\ &= \frac{p}{\sigma'}. \end{aligned}$$

Hence $\sigma' = p$.

Now for $\alpha < 1 - \theta$, let $y^+ = \bar{y} + \alpha\tilde{dy}$, and

$$X^+ = \begin{pmatrix} \bar{X} + \alpha d\tilde{X} & 0 \\ 0 & \alpha\tilde{T} \end{pmatrix}, \quad S^+ = \begin{pmatrix} \bar{S} + \alpha\tilde{dS} & 0 \\ 0 & \alpha\tilde{\Lambda} \end{pmatrix},$$

be the warm start. Since $\tilde{T}$ is uniquely defined, in view of (4.5) and (4.11) one can easily prove that $\tilde{T}\tilde{\Lambda} = \frac{1}{p}I$ and therefore computing $\tilde{T}$ suffices to compute S^+. We postpone the complexity analysis of the optimal updating direction (problem (4.12)) to section 4.5. The following lemma guarantees the strict feasibility of the warm start.

Lemma 11 *The updated points X^+ and (y^+, S^+) are strictly feasible for Ω_P^+ and Ω_D^+ respectively, and therefore they can be used as the starting point to recover the analytic center.*

Proof. First observe that

$$\begin{aligned}
\|\bar{X}^{-1}d\tilde{X}\|^2 &= \|(\bar{S}^{-.5}\bar{X}^{-1}\bar{S}^{-.5})(\bar{S}^{.5}d\tilde{X}\bar{S}^{.5})\|^2 \\
&= \mathbf{t}r(\bar{S}^{-.5}\bar{X}^{-1}\bar{S}^{-.5})^2(\bar{S}^{.5}d\tilde{X}\bar{S}^{.5})^2 \\
&\leq \sum_{j=1}^{n}\lambda_j^2(\bar{S}^{-.5}\bar{X}^{-1}\bar{S}^{-.5})\lambda_j^2(\bar{S}^{.5}d\tilde{X}\bar{S}^{.5}) \qquad (4.13) \\
&\leq \lambda_1^2(\bar{S}^{-.5}\bar{X}^{-1}\bar{S}^{-.5})\sum_{j=1}^{n}\lambda_j^2(\bar{S}^{.5}d\tilde{X}\bar{S}^{.5}) \\
&= \|\bar{S}^{-1}\bar{X}^{-1}\|_\infty^2\|\bar{S}d\tilde{X}\|^2 \qquad (4.14)
\end{aligned}$$

where inequality (4.13) is due to Theobald [55] (see also Marshall and Olkin [35]). From (4.14) and noting that $d\tilde{X}$ is optimal for problem (4.6), and $\bar{X}$ and $\bar{S}$ are approximate centers, one has

$$\|\bar{X}^{-1}d\tilde{X}\| \leq \frac{1}{1-\theta}.$$

On the other hand $\alpha < 1-\theta$, and

$$\bar{X} + \alpha d\tilde{X} = \bar{X}^{.5}(I + \alpha\bar{X}^{-.5}d\tilde{X}\bar{X}^{-.5})\bar{X}^{.5}.$$

Thus $\bar{X} + \alpha d\tilde{X} \succ 0$. Moreover $\tilde{T}$ is positive definite by construction, and hence $X^+ \succ 0$. Since $\bar{X}$ is primal feasible then $\mathcal{A}(\bar{X} + \alpha d\tilde{X}) + \mathcal{B}(\alpha\tilde{T}) = 0$. That is X^+ is strictly feasible for the updated primal set Ω_P^+.

To prove the strict feasibility of the dual iteration, we have

$$\bar{S} + \alpha d\tilde{S} = \bar{S}^{.5}(I - \alpha\bar{S}^{-.5}\mathcal{A}^T d\tilde{y}\bar{S}^{-.5})\bar{S}^{.5},$$

and since $d\tilde{y}$ is optimal for problem (4.1), then $\|\bar{S}^{-.5}\mathcal{A}^T d\tilde{y}\bar{S}^{-.5}\| = 1$. Thus $\bar{S} + \alpha d\tilde{S} \succ 0$. ■

4.2 Analysis of Restoration

Before getting started, we state a lemma similar to Lemma 4:

Lemma 12 *Let $S \in \mathcal{S}^n$ be such that $\|S\| < 1$. Then*

$$\log\det(I+S) \geq I \bullet S + \|S\| + \log(1-\|S\|).$$

Proof. The following inequality is well known (for a proof see Roos, Terlaky, and Vial [50], page 439)

$$\sum_{j=1}^{n} \log(1+s_j) \geq e^T s + \|s\| + \log(1-\|s\|).$$

The lemma follows by letting $\lambda_j(S) = s_j$. ■

The following lemma bounds the potential functions at the warm start:

Lemma 13 *Let $(\bar{X}, \bar{y}, \bar{S})$ be a θ-approximate analytic center. Then*

$$\phi_D(S^+) \leq \phi_D(\bar{S}) - \alpha(1-\theta) - \log(1-\alpha) - \log\det\alpha\tilde{\Lambda}, \tag{4.15}$$

$$\phi_P(X^+) \leq \phi_P(\bar{X}) - \alpha(1-\theta) - \log(1-\alpha) - \log\det\alpha\tilde{T}, \tag{4.16}$$

and

$$\phi_{PD}(X^+, S^+) \leq \phi_{PD}(\bar{X}, \bar{S}) - 2\alpha(1-\theta) - 2\log(1-\alpha) - 2p\log\alpha + p\log p \tag{4.17}$$

Proof.

$$\begin{aligned}
\phi_D(S^+) &= -\log\det\bar{S}(I+\alpha\bar{S}^{-1}\tilde{dS}) - \log\det\alpha\tilde{\Lambda} \\
&= \phi_D(\bar{S}) - \log\det(I+\alpha\bar{S}^{-1}\tilde{dS}) - \log\det\alpha\tilde{\Lambda}
\end{aligned}$$

By Lemma 12

$$\begin{aligned}
&\phi_D(S^+) \leq \phi_D(\bar{S}) + \\
&\quad -I \bullet \alpha\bar{S}^{-1}\tilde{dS} - \|\alpha\bar{S}^{-1}\tilde{dS}\| - \log(1-\|\alpha\bar{S}^{-1}\tilde{dS}\|) - \log\det\alpha\tilde{\Lambda}
\end{aligned} \tag{4.18}$$

Since $\bar{X} \bullet \tilde{dS} = -(\mathcal{A}\bar{X})^T \tilde{dy} = 0$,

$$\begin{aligned}
\left| I \bullet \alpha \bar{S}^{-1} \tilde{dS} \right| &= \alpha \left| (\bar{S}^{-1} - \bar{X}) \bullet \tilde{dS} \right| \\
&= \alpha \left| (I - \bar{S}^{.5} \bar{X} \bar{S}^{.5}) \bullet \bar{S}^{-.5} (\tilde{dS}) \bar{S}^{-.5} \right| \\
&\leq \alpha \| \bar{S}^{.5} \bar{X} \bar{S}^{.5} - I \| \| \bar{S}^{-.5} (\tilde{dS}) \bar{S}^{-.5} \| \\
&\leq \alpha\theta.
\end{aligned}$$

The first inequality, thus, follows from the inequality (4.18), the above fact and noting that $f(t) = -t - \log(1-t)$ is an increasing function over its domain.

To prove the second inequality note that

$$\phi_P(X^+) = \phi_P(\bar{X}) + \alpha C \bullet \tilde{dX} + \alpha \bar{y}^T \mathcal{B}\tilde{T} - \log\det(I + \alpha \bar{X}^{-1} \tilde{dX}) - \log\det \alpha \tilde{T},$$

again by Lemma 12

$$\begin{aligned}
\phi_P(X^+) \leq\ & \phi_P(\bar{X}) + \alpha C \bullet \tilde{dX} + \alpha \bar{y}^T \mathcal{B}\tilde{T} - I \bullet \alpha \bar{X}^{-1} \tilde{dX} \\
& - \|\alpha \bar{X}^{-1} \tilde{dX}\| - \log(1 - \|\alpha \bar{X}^{-1} \tilde{dX}\|) - \log\det \alpha \tilde{T}.
\end{aligned} \tag{4.19}$$

On the other hand from $\mathcal{A}\tilde{dX} + \mathcal{B}\tilde{T} = 0$

$$\begin{aligned}
\left| \alpha C \bullet \tilde{dX} + \alpha \bar{y}^T \mathcal{B}\tilde{T} - \alpha \bar{X}^{-1} \bullet \tilde{dX} \right| &= \left| \alpha \bar{S} \bullet \tilde{dX} - \alpha \bar{X}^{-1} \bullet \tilde{dX} \right| \\
&= \alpha \left| (\bar{X}^{.5} \bar{S}\ \bar{X}^{.5} - I) \bullet \bar{X}^{-.5} \tilde{dX} \bar{X}^{-.5} \right| \\
&\leq \alpha \| \bar{X}^{.5} \bar{S}\ \bar{X}^{.5} - I \| \| \bar{X}^{-.5} \tilde{dX} \bar{X}^{-.5} \| \\
&\leq \alpha\theta.
\end{aligned}$$

The primal inequality, therefore, follows from the inequality (4.19), the above fact, and the property of increasing function $f(t)$. Finally the last inequality is obtained by adding up (4.15) and (4.16). ■

The following theorem provides the complexity of updating the analytic center after adding a p-dimensional semidefinite cut:

Theorem 14 *Starting from the strict interior point (X^+, S^+), the number of iterations to update an approximate analytic center is bounded by $O(p\log(p+1))$, where p is the dimension of the central semidefinite cut $\mathcal{B}^T y \preceq \mathcal{B}^T \bar{y}$.*

Proof. Since $(\bar{X}, \bar{S})$ is a θ-approximate center of the current set of localization, from Lemma 4 we have

$$\begin{aligned}\phi_{PD}(\bar{X}, \bar{S}) &\leq \bar{X} \bullet \bar{S} - I \bullet (\bar{X}\bar{S} - I) + \frac{\|\bar{X}\bar{S} - I\|^2}{2(1 - \|\bar{X}\bar{S} - I\|)} \\ &\leq n + \frac{\theta^2}{2(1-\theta)}.\end{aligned}$$

Now, let the analytic center of the new convex body be $((X^a)^+, (S^a)^+)$. Since $\phi_{PD}\left((X^a)^+, (S^a)^+\right) = n + p$, from (4.17)

$$\phi_{PD}(X^+, S^+) - \phi_{PD}\left((X^a)^+, (S^a)^+\right) \leq \kappa(\alpha, \theta, p) + p\log p,$$

where

$$\kappa(\alpha, \theta, p) = \frac{\theta^2}{2(1-\theta)} - 2\alpha(1-\theta) - 2\log(1-\alpha) - 2p\log\alpha - p.$$

At each iteration of the Newton's method the potential function is reduced by a constant amount δ. Therefore after at most

$$\lceil \frac{\kappa(\alpha, \theta, p) + p\log p}{\delta} \rceil \sim O(p\log(p+1))$$

iterations the algorithm stops with an updated analytic center. ■

Figure 4.3 illustrates the recentering iterations from the warm start after adding a 2-dimensional semidefinite cut. As the figure shows, the number of iterations to recover the centrality is only one. Alternatively, one can compute the analytic center of the updated set of localization without using the optimal updating direction. In this case, one should make the cut shallow and use the previous analytic center as the starting point. As Figure 4.4

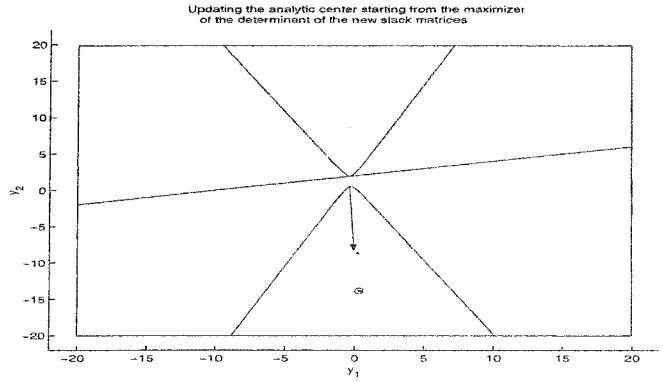

Figure 4.3: Recentering steps from the optimal updating direction

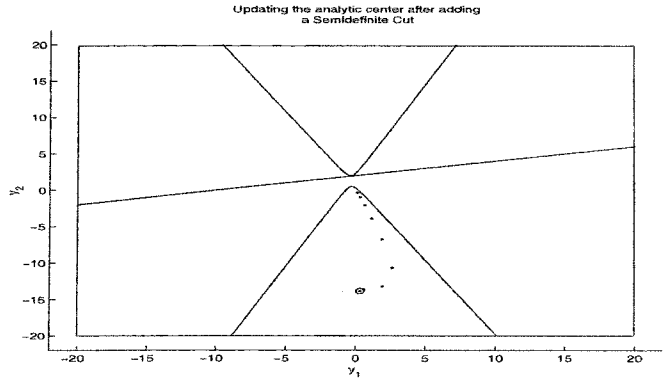

Figure 4.4: Recentering steps without using the optimal updating direction.

shows this procedure drastically increases the number of steps to update the analytic center.

In the next section we present an ACCPM algorithm for the convex feasibility problem with semidefinite cuts.

4.3 The ACCPM Algorithm

The ACCPM algorithm attempts to find a feasible point in $\Omega^* \subset \Omega_D$, where Ω^* is the solution set and contains a full dimensional ball $\mathcal{N}_\varepsilon$ with radius ε. We make the following assumptions:

Assumption 1 $\Omega_D \subset [0,1]^m$

Assumption 2 *Ω_D is described by an oracle. That is, the oracle determines if either the center is in Ω^*, thus solving the problem, or returns a p-dimensional semidefinite cut which contains Ω^*.*

Assumption 3 *For the semidefinite cut $\mathcal{B}^T$, we assume that*

$$\max_{i,l,q} \left(\mathbf{t}r B_i, \|b_q^l\| \right) = 1, \tag{4.20}$$

where b_q^l is the m-vector defined in (2.1).

For the next assumption we need to define a condition number on the semidefinite cut.

Definition 15 *At any point $z \notin \Omega^*$, let $\mathcal{B}_z^T y \preceq \mathcal{B}_z^T z$ be the cut generated by the oracle. The condition number of the cut $\mathcal{B}_z^T$ is defined via:*

$$\mu_z = \max\{\det \mathcal{B}_z^T u : \ \mathcal{B}_z^T u \succeq 0, \ \|u\| = 1\} \tag{4.21}$$

and the condition number of the field of cuts $\{\mathcal{B}_z^T, \ \forall \ z \notin \Omega^*\}$ is defined by:

$$\mu = \inf_{z \notin \Omega^*} \mu_z. \tag{4.22}$$

Assumption 4

$$\mu > 0.$$

Now let $S_z(y) = \mathcal{B}_z^T(z - y) \succeq 0$ be the slack matrix corresponding to one of the cuts, y^c be the center of $\mathcal{N}_\varepsilon$, and u be a vector such that $\|u\| = 1$. Then

$$S_z(y^c + \varepsilon u) = S_z(y^c) - \varepsilon \mathcal{B}_z^T u,$$

and as $y^c + \varepsilon u \in \mathcal{N}_\varepsilon \subset \Omega^*$,

$$S_z(y^c + \varepsilon u) \succeq 0,$$

and thus

$$S_z(y^c) \succeq \varepsilon \mathcal{B}_z^T u$$

In view of Assumption 4 now the following lemma is clear:

Lemma 16 *For any $z \notin \Omega^*$,*

$$\det S_z(y^c) \geq \varepsilon^p \mu,$$

where p is the dimension of the cut.

Now we present the ACCPM algorithm:

Algorithm 3 *Given $\Omega_D^0 = \{y \in R^m : (\mathcal{A}^0)^T y \preceq C^0\}$, where $(\mathcal{A}^0)^T y = \sum_i y_i A_i^0$, with*

$$A_i^0 = \begin{pmatrix} Diag(e_i) & 0 \\ 0 & -Diag(e_i) \end{pmatrix}, \text{ and } C^0 = \begin{pmatrix} I_m & 0 \\ 0 & 0_m \end{pmatrix}.$$

Let $k = 0$

1. *Compute an approximate analytic center $\bar{y}^k$ for Ω_D^k.*

2. *If* $\bar{y}^k \in \Omega^*$, *stop.*

3. *Otherwise, call the oracle for the* p_k*-dimensional cut* $(\mathcal{B}^k)^T y \preceq (\mathcal{B}^k)^T \bar{y}^k$.

4. *Update the set of localization:* $\Omega_D^{k+1} = \{y \in R^m : (\mathcal{A}^{k+1})^T y \preceq C^{k+1}\}$, *where* $(\mathcal{A}^{k+1})^T y = \sum_i y_i A_i^{k+1}$, *with*

$$A_i^{k+1} = \begin{pmatrix} A_i^k & 0 \\ 0 & B_i^k \end{pmatrix}, \text{ and } C^{k+1} = \begin{pmatrix} C^k & 0 \\ 0 & (\mathcal{B}^k)^T \bar{y}^k \end{pmatrix}. \tag{4.23}$$

Set $k = k + 1$ *and go to step 1.*

It is worth mentioning that at each iteration k, we enlarge the dimension of the cut matrices A_i, by p_k when adding the semidefinite cut as a block diagonal. That is, for all k

$$dim(A_i^k) = 2m + n_k = 2m + \sum_{i=0}^{k-1} p_i,$$

and $n_0 = 0$.

4.4 Convergence of the Algorithm

Let us bound the potential function at the new center. We first define the min-potential functions.

Let Ω_P and Ω_D be the current primal and dual feasible sets respectively. The primal (dual) min-potential function denoted by $\mathcal{P}(\Omega_P)$ $(\mathcal{D}(\Omega_D))$ is the value of the primal (dual) potential function at the analytic center of Ω_P (Ω_D). We have the following theorem

Theorem 17 *Let* $\mathcal{D}(\Omega_D)$ *be the dual min-potential function at the current set of localization* Ω_D *and let* Ω_D^+ *be the updated set after adding the p-dimensional semidefinite cut* $\mathcal{B}^T y \preceq \mathcal{B}^T \bar{y}$ *at a* θ*-approximate center* $\bar{y}$. *Then*

$$\mathcal{D}(\Omega_D^+) \geq \mathcal{D}(\Omega_D) - \sum_{i=1}^{p} \log t_i - \mathcal{C}(p, \theta, \alpha), \tag{4.24}$$

where

$$\mathcal{C}(p,\theta,\alpha) = \frac{\theta^2}{2(1-\theta)} - \alpha(1-\theta) - \log(1-\alpha) - p(1+\log\alpha) + p\log p,$$

and

$$t_i = \sqrt{(b_i^i)^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}b_i^i}, \tag{4.25}$$

where b_i^i is the m-vector defined in (2.1).

Proof. Let $\mathcal{P}(\Omega_P)$ be the primal min-potential function at Ω_P and let Ω_P^+ be the updated primal feasible set after adding the cut. From the properties of the primal-dual potential function and (4.16)

$$\begin{aligned}\mathcal{D}(\Omega_D^+) &= n+p-\mathcal{P}(\Omega_P^+)\\ &\geq n+p-\phi_P(\bar{X})+\alpha(1-\theta)+\log(1-\alpha)+\log\det\alpha\tilde{T}.\end{aligned}$$

In view of Lemma 5 and the above inequality

$$\mathcal{D}(\Omega_D^+) \geq \mathcal{D}(\Omega_D)+p-\frac{\theta^2}{2(1-\theta)}+\alpha(1-\theta)+\log(1-\alpha)+\log\det\alpha\tilde{T}. \tag{4.26}$$

Recall that

$$\tilde{T} = \arg\min\{\frac{p}{2}\mathbf{tr}T\mathcal{V}T - \log\det T\}, \text{ and } \mathbf{tr}\tilde{T}\mathcal{V}\tilde{T} = 1.$$

Thus $\log\det\tilde{T} \geq \log\det T'$ for any positive semidefinite matrix T' with $\mathbf{tr}T'\mathcal{V}T' = 1$. Let

$$T' = \frac{T^{-1}}{\sqrt{\mathbf{tr}T^{-1}\mathcal{V}T^{-1}}},$$

where T is a diagonal matrix made up of $t_k > 0$, defined in (4.25). First we prove that $\mathbf{tr}T^{-1}\mathcal{V}T^{-1} \leq p^2$:

$$\begin{aligned}\mathbf{tr}T^{-1}\mathcal{V}T^{-1} &= \mathbf{tr}T^{-1}\mathcal{B}^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}T^{-1}\\ &= (\mathcal{B}T^{-1})^T(\mathcal{A}_D\mathcal{A}_D^T)^{-1}\mathcal{B}T^{-1}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i,j=1}^{m} (\mathbf{t}r B_i T^{-1})(\mathbf{t}r B_j T^{-1})(\mathcal{A}_D \mathcal{A}_D^T)^{-1}_{ij} \\
&= \sum_{i,j=1}^{m} \left(\sum_{k=1}^{p} \frac{(B_i)_{kk}}{t_k} \right) \left(\sum_{k=1}^{p} \frac{(B_j)_{kk}}{t_k} \right) (\mathcal{A}_D \mathcal{A}_D^T)^{-1}_{ij} \\
&= \sum_{l,q=1}^{p} \frac{1}{t_l t_q} \sum_{i,j=1}^{m} (B_i)_{ll} (B_j)_{qq} (\mathcal{A}_D \mathcal{A}_D^T)^{-1}_{ij} \\
&= \sum_{l,q} \frac{1}{t_l t_q} (b_l^l)^T (\mathcal{A}_D \mathcal{A}_D^T)^{-1} b_q^q,
\end{aligned}
$$

clearly $(b_l^l)^T (\mathcal{A}_D \mathcal{A}_D^T)^{-1} b_q^q \leq t_l t_q$. Thus

$$\mathbf{t}r T^{-1} \mathcal{V} T^{-1} \leq p^2. \tag{4.27}$$

Now

$$
\begin{aligned}
\log\det \tilde{T} &\geq -p \log \sqrt{\mathbf{t}r T^{-1} \mathcal{V} T^{-1}} - \log\det T \\
&\geq -p \log p - \sum_{i=1}^{p} \log t_i.
\end{aligned}
$$

From (4.26), the inequality (4.24) is immediate now. ■

Theorem 17 establishes a bound on the potential function at the new center in terms of p as well as θ and α. Since the values of θ and α are arbitrary within their limit, we can simplify the bound by choosing fixed values for them. Let $\theta = 0.01$ and $\alpha = 0.9$. One can check that

$$\mathcal{C}(p, \theta, \alpha) \leq p \log(p+1),$$

and therefore the inequality (4.24) is reduced to

$$\mathcal{D}(\Omega_D^+) \geq \mathcal{D}(\Omega_D) - p \log(p+1) - \sum_{i=1}^{p} \log t_i. \tag{4.28}$$

We note that (4.24) is valid for moderate values of θ and α, i.e, for θ close to zero, one should not choose α very close to 1 (e.g. $\alpha < 0.9$ does the job).

At iteration k, let $p = \max\{p_i, i = 1, ..., k\}$ and $\mathcal{D}(\Omega_D^k)$ be the dual min-potential function at Ω_D^k; using (4.28), we have

$$\begin{aligned}
\mathcal{D}(\Omega_D^{k+1}) &\geq \mathcal{D}(\Omega_D^k) - p_k \log(p_k + 1) - \sum_{i=1}^{p_k} \log t_i \\
&\geq \mathcal{D}(\Omega_D^k) - p_k \log(p + 1) - \sum_{i=1}^{p_k} \log t_i \\
&\vdots \\
&\geq \mathcal{D}(\Omega_D^0) - n_{k+1} \log(p + 1) - \sum_{i=1}^{n_{k+1}} \log t_i \qquad (4.29)
\end{aligned}$$

Now we state a series of technical lemmas to construct a bound on the summation term in (4.29).

Lemma 18 *Let $\mathcal{A}\mathcal{A}^T \in \mathcal{S}^m$, with $(\mathcal{A}\mathcal{A}^T)_{ij} = A_i \bullet A_j$, where $A_i \in \mathcal{S}^n$. Then*

$$\mathcal{A}\mathcal{A}^T = \sum_{l,q=1}^{n} a_q^l (a_q^l)^T,$$

where a_q^l is the m-vector defined in (2.1).

Proof. First observe that

$$A_i \bullet A_j = \sum_{q=1}^{n} (\bar{a}_q^i)^T \bar{a}_q^j,$$

where $A_i = (\bar{a}_1^i, \bar{a}_2^i, ..., \bar{a}_n^i)$, $\bar{a}_q^i \in \mathcal{R}^n$, $i = 1, ..., m$. Now consider n Gram matrices G^q, $q = 1, ..., n$ defined by $G_{ij}^q = (\bar{a}_q^i)^T \bar{a}_q^j$. Thus

$$\begin{aligned}
\mathcal{A}\mathcal{A}^T &= \sum_{q=1}^{n} G^q \\
&= \sum_{q=1}^{n} \bar{A}_q^T \bar{A}_q,
\end{aligned}$$

where $\bar{A}_q = \left(\bar{a}_q^1, \bar{a}_q^2, ..., \bar{a}_q^m\right)$. $\bar{A}_q^T \bar{A}_q$ can alternatively be expressed by the summation of a number of rank one matrices:

$$G^q = \sum_{l=1}^{n} a_q^l (a_q^l)^T,$$

where a_q^l is the row l of $\bar{A}_q$, i.e., $a_q^l = ((A_1)_{lq}, (A_2)_{lq}, ..., (A_m)_{lq})$. ■

Lemma 19

$$\mathcal{A}_D\mathcal{A}_D^T \succeq \frac{1}{(\mathbf{tr}S)^2}\mathcal{A}\mathcal{A}^T$$

Proof. Consider the quadratic form associated with $\mathcal{A}_D\mathcal{A}_D^T - \frac{1}{(\mathbf{tr}S)^2}\mathcal{A}\mathcal{A}^T$.

For $y \in R^m$

$$f(y) = y^T\left(\mathcal{A}_D\mathcal{A}_D^T - \frac{1}{(\mathbf{tr}S)^2}\mathcal{A}\mathcal{A}^T\right)y.$$

Let $\mathcal{A}_D^T y = W$, then $\mathcal{A}^T y = S^{.5}WS^{.5}$ and we have

$$\begin{aligned} f(y) &= W \bullet W - \frac{1}{(\mathbf{tr}S)^2}(S^{.5}WS^{.5}) \bullet (S^{.5}WS^{.5}) \\ &= \|W\|^2 - \frac{1}{(\mathbf{tr}S)^2}\|S^{.5}WS^{.5}\|^2 \\ &\geq \|W\|^2 - \frac{1}{(\mathbf{tr}S)^2}\|S^{.5}\|^4\|W\|^2 \\ &= 0. \end{aligned}$$

Therefore the quadratic form is nonnegative for any $y \in R^m$. ■

Lemma 20 *At the k^{th} iteration of the ACCPM algorithm*

$$\mathcal{A}_D^k(\mathcal{A}_D^k)^T \succeq 8I + \frac{1}{m^2}\sum_{l,q=1}^{n_k} b_q^l(b_q^l)^T, \tag{4.30}$$

where $b_q^l = ((B_1)_{lq}, (B_2)_{lq}, ..., (B_m)_{lq})$, and matrices B_i are block diagonal matrices composed of cut matrices B_i^r, for $r = 0, 1, ..., k-1$.

Proof. From algorithm 3 since $\Omega_D^0 = [0,1]^m$

$$\mathcal{A}_D^0(\mathcal{A}_D^0)^T \succeq 8I,$$

and after adding $k+1$ semidefinite cuts $(\mathcal{B}^r)^T y \preceq (\mathcal{B}^r)^T y^r$, $r = 0, 1, ..., k$ we have

$$\begin{aligned} \mathcal{A}_D^{k+1}(\mathcal{A}_D^{k+1})^T &= \mathcal{A}_D^k(\mathcal{A}_D^k)^T + \mathcal{B}_D^k(\mathcal{B}_D^k)^T \\ &\succeq 8I + \sum_{r=0}^{k}\mathcal{B}_D^r(\mathcal{B}_D^r)^T, \end{aligned} \tag{4.31}$$

where $(\mathcal{B}_D^r(\mathcal{B}_D^r)^T)_{ij} = \mathbf{t}r(S^r)^{-1}B_i^r(S^r)^{-1}B_j^r$ and $S^r = (\mathcal{B}^r)^T(y^r - y)$.

On one hand from Lemma 19

$$\mathcal{B}_D^r(\mathcal{B}_D^r)^T \succeq \frac{1}{(\mathbf{t}rS^r)^2}\mathcal{B}^r(\mathcal{B}^r)^T,$$

where $(\mathcal{B}^r(\mathcal{B}^r)^T)_{ij} = \mathbf{t}rB_i^rB_j^r$.

From (4.20), we have $\mathbf{t}rS^r = \sum_{i=1}^m (y^r - y)_i \mathbf{t}rB_i^r \leq m$ and therefore

$$\mathcal{B}_D^r(\mathcal{B}_D^r)^T \succeq \frac{1}{m^2}\mathcal{B}^r(\mathcal{B}^r)^T. \tag{4.32}$$

On the other hand by Lemma 18

$$\mathcal{B}^r(\mathcal{B}^r)^T \succeq \sum_{l,q=1}^{p_r} b_q^l(b_q^l)^T. \tag{4.33}$$

The lemma follows now from 4.31, 4.32, and 4.33. ■

The next lemma is essential to bound (4.29). This lemma is due to Ye [63] with some changes to suit our case.

Lemma 21 *If $p \leq m$, then*

$$\sum_{i=1}^{n_{k+1}} t_i^2 \leq 2m^3 \log\left(8 + \frac{n_{k+1}^2}{m^3}\right), \tag{4.34}$$

for t_i defined in (4.25).

Proof. Define

$$\mathcal{H}^{k+1} = \mathcal{H}^k + \frac{1}{m^2}\sum_{i,j=1}^{p_k} b_j^i(b_j^i)^T,$$

where b_j^i is the m-vector defined in (2.1); and let $\mathcal{H}^0 = 8I$.

$$\det \mathcal{H}^{k+1} = \det\left(\mathcal{H}^k + \frac{1}{m^2}\sum_{i,j\in\mathcal{I}_1} b_j^i(b_j^i)^T\right)\left(1 + \frac{r^2}{m^2}\right), \tag{4.35}$$

where $\mathcal{I}_1 = \{i, j = 1, ..., p_k \backslash (i,j) = (1,1)\}$ and

$$r^2 = (b_1^1)^T\left(\mathcal{H}^k + \frac{1}{m^2}\sum_{i,j\in\mathcal{I}_1} b_j^i(b_j^i)^T\right)^{-1} b_1^1.$$

Now we establish a lower bound on r. To this end we study the eigenvalues of

$$\mathcal{G} = I + \frac{1}{m^2} \sum_{i,j \in \mathcal{I}_1} (\mathcal{H}^k)^{-.5} b_j^i (b_j^i)^T (\mathcal{H}^k)^{-.5}.$$

Let $x \in R^m$ with $\|x\| = 1$, then

$$\begin{aligned} x^T \mathcal{G} x &= \|x\|^2 + \frac{1}{m^2} \sum_{i,j \in \mathcal{I}_1} (x^T (\mathcal{H}^k)^{-.5} b_j^i)^2 \\ &\leq \|x\|^2 + \frac{1}{m^2} \sum_{i,j \in \mathcal{I}_1} \|x\|^2 \|(\mathcal{H}^k)^{-.5} b_j^i\|^2 \\ &= 1 + \frac{1}{m^2} \sum_{i,j \in \mathcal{I}_1} (b_j^i)^T (\mathcal{H}^k)^{-1} b_j^i. \end{aligned}$$

Since $\mathcal{H}^k \succeq 8I$

$$x^T \mathcal{G} x \leq 1 + \frac{1}{m^2} \sum_{i,j \in \mathcal{I}_1} \frac{1}{8} \|b_j^i\|^2.$$

From Assumption (4.20)

$$x^T \mathcal{G} x \leq 1 + \frac{p_k^2 - 1}{8m^2},$$

and since $p_k \leq p \leq m$, then

$$x^T \mathcal{G} x \leq \frac{9}{8}.$$

That is $\mathcal{G}^{-1} \succeq (8/9)I$ and therefore

$$\begin{aligned} r^2 &= (b_1^1)^T (\mathcal{H}^k)^{-.5} \mathcal{G}^{-1} (\mathcal{H}^k)^{-.5} b_1^1 \\ &\geq (8/9) r_{11}^2, \end{aligned}$$

where $r_{11}^2 = (b_1^1)^T (\mathcal{H}^k)^{-1} b_1^1$. Now from (4.35)

$$\det \mathcal{H}^{k+1} \geq \left(1 + \frac{8 r_{11}^2}{9m^2}\right) \det \left(\mathcal{H}^k + \frac{1}{m^2} \sum_{i,j \in \mathcal{I}_1} b_j^i (b_j^i)^T\right).$$

By repeating this procedure for each i and j one has

$$\det \mathcal{H}^{k+1} \geq \prod_{i,j=1}^{p_k} \left(1 + \frac{8 r_{ij}^2}{9m^2}\right) \det \mathcal{H}^k,$$

where $r_{ij}^2 = (b_j^i)^T(\mathcal{H}^k)^{-1}b_j^i$.

By taking logarithm from both sides of the above inequality we have

$$\begin{aligned}\log\det\mathcal{H}^{k+1} &\geq \sum_{i,j=1}^{p_k}\log\left(1+\frac{8r_{ij}^2}{9m^2}\right)+\log\det\mathcal{H}^k\\ &\geq \sum_{i=1}^{p_k}\log\left(1+\frac{8r_{ii}^2}{9m^2}\right)+\log\det\mathcal{H}^k.\end{aligned}$$

Since $r_{ii}^2 \leq 1/8$, then $\frac{8r_{ii}^2}{9m^2} \leq \frac{1}{9}$ and

$$\log\left(1+\frac{8r_{ii}^2}{9m^2}\right) \geq \frac{8r_{ii}^2}{9m^2} - \frac{\left(\frac{8r_{ii}^2}{9m^2}\right)^2}{2\left(1-\frac{8r_{ii}^2}{9m^2}\right)} \geq \frac{r_{ii}^2}{2m^2}$$

and therefore

$$\begin{aligned}\log\det\mathcal{H}^{k+1} &\geq \sum_{i=1}^{p_k}\frac{r_{ii}^2}{2m^2}+\log\det\mathcal{H}^k\\ &\geq \sum_{i=1}^{n_{k+1}}\frac{r_{ii}^2}{2m^2}+\log\det\mathcal{H}^0. \qquad (4.36)\end{aligned}$$

On the other hand, using the arithmetic–geometric inequality

$$\log\det\mathcal{H}^{k+1} = \log\prod_{j=1}^{m}\lambda_j(\mathcal{H}^{k+1}) \leq m\log\frac{\mathbf{tr}\mathcal{H}^{k+1}}{m}$$

and from the definition of $\mathcal{H}^k$ and Assumption (4.20)

$$\mathbf{tr}\mathcal{H}^{k+1} = \mathbf{tr}\left(8I+\frac{1}{m^2}\sum_{i,j=1}^{n_{k+1}}b_j^i(b_j^i)^T\right) = 8m+\frac{n_{k+1}^2}{m^2},$$

thus

$$\log\det\mathcal{H}^{k+1} \leq m\log\left(8+\frac{n_{k+1}^2}{m^3}\right).$$

This inequality together with (4.36) give

$$\sum_{l=1}^{n_{k+1}}r_{ii}^2 \leq 2m^3\log\left(8+\frac{n_{k+1}^2}{m^3}\right). \qquad (4.37)$$

In view of Lemma 20, $\mathcal{A}_D^k(\mathcal{A}_D^k)^T \succeq \mathcal{H}^k$ and thus

$$(b_i^i)^T \left(\mathcal{A}_D^k(\mathcal{A}_D^k)^T\right)^{-1} b_i^i \leq (b_i^i)^T(\mathcal{H}^k)^{-1}b_i^i$$

or $t_i^2 \leq r_{ii}^2$. The proof follows form (4.37) now. ■

In the next theorem we derive a bound on n_k, the dimension of the accumulated block diagonal cut matrix.

Theorem 22 *The ACCPM algorithm stops with a solution in Ω^* when*

$$n_k \sim O^*(\frac{p^2m^3}{\mu^2\varepsilon^2}),$$

where in O^* the lower-order terms are ignored.

Proof. Consider the k^{th} iteration of the algorithm. Since the analytic center is the minimizer of the dual potential function and since $\bar{y}^j \notin \Omega^*$, for $j = 0, 1, ..., k-1$, in view of Lemma (16) we have

$$\begin{aligned}
\mathcal{D}(\Omega_D^k) &\leq -\log\det(C^k - (\mathcal{A}^k)^T y^c) \\
&= -\log\det(C^0 - (\mathcal{A}^0)^T y^c) - \sum_{j=0}^{k-1} \log\det((\mathcal{B}^j)^T(\bar{y}^j - y^c)) \\
&\leq -(2m + n_k)\log\varepsilon - k\log\mu.
\end{aligned}$$

Notice that if $\mu \geq 1$, this parameter can simply be eliminated from the above inequality. We therefore consider the worst case complexity where $\mu < 1$. Now from inequality (4.29)

$$(2m + n_{k+1})\log\mu\varepsilon \leq 2m\log\frac{1}{2} + n_{k+1}\log(p+1) + \frac{1}{2}\sum_{i=1}^{n_{k+1}} \log t_i^2,$$

or

$$\begin{aligned}
\log\mu\varepsilon - \log(p+1) &\leq \frac{1}{2(2m + n_{k+1})}\left(2m\log\frac{1}{4} + \sum_{i=1}^{n_{k+1}} \log t_i^2\right) \\
&\leq \frac{1}{2}\log\frac{\frac{m}{2} + \sum_{i=1}^{n_{k+1}} t_i^2}{2m + n_{k+1}}.
\end{aligned}$$

Note that the second inequality is due to the arithmetic–geometric mean. Finally by Lemma 21

$$\left(\frac{\mu\varepsilon}{p+1}\right)^2 \leq \frac{\frac{m}{2} + 2m^3 \log\left(8 + \frac{n_{k+1}^2}{m^3}\right)}{2m + n_{k+1}}.$$

The algorithm stops with a solution in Ω^* when this inequality is violated. In other words, when $n_k \sim O^*(p^2m^3/\mu^2\varepsilon^2)$. ■

The next section completes our analysis by bounding the number of damped Newton steps needed to solve problem (4.12).

4.5 Complexity of the recentering direction

Let $\bar{y}$ be an approximate center of Ω_D and consider a p-dimensional semidefinite cut at $\bar{y}$. Let

$$F(T) = \frac{p}{2}\mathbf{t}rT\mathcal{V}T - \log\det T.$$

Recall that the optimal restoration direction is obtained by minimizing this function over the positive semidefinite cone. In this section we analyze the behavior of the Newton method as applied to F.

We first prove that the (dual) feasible region is contained in an enlarged Dikin ellipsoid. This result is used to construct an upper bound on the functional gap of F at its optimal and initial points.

Lemma 23 *Let $(\bar{X}, \bar{y}, \bar{S})$ be a θ-approximate analytic center of Ω_D. Then*

$$\|\bar{S}^{-.5}\mathcal{A}^T(y - \bar{y})\bar{S}^{-.5}\| \leq \frac{1+\theta}{1-\theta}(n+1),$$

for any $y \in \Omega_D$. In other words, the current set of localization is contained in a Dikin ellipsoid centered at $\bar{y}$, and enlarged by a factor of $\frac{(1+\theta)(n+1)}{1-\theta}$.

Proof. Let $y \in \Omega_D$ be dual feasible and $S = C - \mathcal{A}^T y$. From the properties of matrix norm one can prove that (see the proof of (4.14))

$$\|\bar{S}^{-1}(S - \bar{S})\| \leq \|\bar{S}^{-1}\bar{X}^{-1}\|_\infty \|\bar{X}(S - \bar{S})\|. \tag{4.38}$$

Since $\bar{X} \bullet (S - \bar{S}) = 0$,

$$\|\bar{X}(S - \bar{S}) + I\|^2 = \|\bar{X}(S - \bar{S})\|^2 + n,$$

and therefore

$$\|\bar{X}(S - \bar{S})\| \leq \|\bar{X}S\| + \|\bar{X}\bar{S} - I\|. \tag{4.39}$$

But,

$$\begin{aligned} \|\bar{X}S\|^2 &= tr(\bar{X}^{.5}S\bar{X}^{.5})^2 \\ &= \sum_i \lambda_i^2(\bar{X}^{.5}S\bar{X}^{.5}) \\ &\leq \left(\sum \lambda_i(\bar{X}^{.5}S\bar{X}^{.5})\right)^2 \\ &= (\bar{X} \bullet S)^2, \end{aligned}$$

and since $\mathcal{A}\bar{X} = 0$, then $\|\bar{X}S\| = \bar{X} \bullet \bar{S}$.

Now from $\|\bar{X}\bar{S} - I\| \leq \theta$ and

$$\begin{aligned} \|\bar{X}\bar{S} - I\|^2 &= \|\bar{S}^{.5}\bar{X}\bar{S}^{.5} - I\|^2 \\ &= \sum_i \lambda_i^2(\bar{S}^{.5}\bar{X}\bar{S}^{.5} - I), \end{aligned}$$

one has

$$1 - \theta \leq \lambda_i(\bar{X}\bar{S}) \leq 1 + \theta,$$

and thus

$$\bar{X} \bullet \bar{S} \leq (1 + \theta)n \quad \text{and} \quad \|\bar{S}^{-1}\bar{X}^{-1}\|_\infty \leq \tfrac{1}{1-\theta}.$$

The above inequalities along with (4.38) and (4.39) prove the lemma. ■

In the next theorem we derive a bound on the number of iterations of the Newton method as applied to $F(T)$.

Theorem 24 *Let* $T^0 = \frac{T^{-1}}{\sqrt{\mathbf{tr}T^{-1}\mathcal{V}T^{-1}}}$, *where* T *is the diagonal matrix defined in Theorem 17. Then starting from* T^0 *the Newton method finds the optimal updating direction in at most*

$$O\left(\frac{p\log\frac{1}{\varepsilon^*}+\log\frac{1}{\mu}}{\beta-\log(1+\beta)}\right)$$

iterations, where

$$\varepsilon^* = \frac{(1-\theta)\varepsilon}{(1+\theta)(1+n)},$$

β *is the Newton decrement, and* $\mu > 0$ *is the condition number for the field of cuts defined by (4.22).*

Proof. Let $\tilde{T}$ and $\tilde{\Lambda}$ be the optimal solutions of problems (4.12) and (4.1) respectively. We first derive an upper bound on the functional gap of F at T^0 and $\tilde{T}$. Observe that

$$\begin{aligned} F(T^0) &= \frac{p}{2}\mathbf{tr}T^0\mathcal{V}T^0 - \log\det T^0 \\ &= \frac{p}{2} + p\log\sqrt{\mathbf{tr}T^{-1}\mathcal{V}T^{-1}} - \log\det T^{-1}, \end{aligned}$$

and from (4.27) and definition of T

$$F(T^0) \le \frac{p}{2} + p\log p + \sum_{i=1}^{p}\log t_i.$$

Since $\mathcal{A}_D\mathcal{A}_D^T \succeq 8I$, then $\sum\log t_i \le (p/2)\log(1/8) \le 0$. Thus

$$F(T^0) \le \frac{p}{2} + p\log p. \tag{4.40}$$

On the other hand, recall that $p\tilde{T}\tilde{\Lambda} = I$; thus

$$F(\tilde{T}) - \log\det\tilde{\Lambda} = \frac{p}{2} + p\log p. \tag{4.41}$$

Let us construct an upper bound on $-\log\det\tilde{\Lambda}$. From Lemma 23, the updated set of localization Ω_D^+ is contained in a Dikin ellipsoid enlarged by a

factor of $\frac{(1+\theta)(n+1)}{1-\theta}$. By shrinking the Dikin ellipsoid with a factor of $\frac{1-\theta}{(1+\theta)(n+1)}$ at $\bar{y}$ and noting that Ω_D^+ contains a ball with radius ε, one can prove that

$$\Omega_D^+ \cap \{y \in \mathcal{R}^m : \|\bar{S}^{-.5}\mathcal{A}^T(y-\bar{y})\bar{S}^{-.5}\| \leq 1\}$$

contains a ball $\mathcal{N}_{\varepsilon^*}$ with radius $\varepsilon^* = \frac{(1-\theta)\varepsilon}{(1+\theta)(n+1)}$. This set is the feasible region of problem (4.1). Let y^* be the center of $\mathcal{N}_{\varepsilon^*}$. Then $y^* + \varepsilon^* u \in \mathcal{N}_{\varepsilon^*}$ for any u such that $\|u\| = 1$. In view of Assumption 4, following the same line of argument as in Lemma 16, we have

$$-\log\det\tilde{\Lambda} \leq p\log\frac{1}{\varepsilon^*} + \log\frac{1}{\mu}.$$

Now from (4.40), (4.41), and the above inequality, one has

$$F(T^0) - F(\tilde{T}) \leq p\log\frac{1}{\varepsilon^*} + \log\frac{1}{\mu}. \tag{4.42}$$

Now observe that $F(T)$ is composed of a self-concordant barrier and a convex quadratic function and due to the stability of the self-concordant functions under summation [44, proposition 2.1.1], $F(T)$ is a self-concordant function on $\mathcal{S}_+^p$. Using Theorem 2.2.3 in [44] one can prove that the Newton algorithm with step size $\frac{1}{1+\beta}$ reduces the value of $F(T)$ by a constant amount $(\beta - \log(1+\beta))$ at each iteration, where $\beta \geq 1$ is the Newton decrement; and the convergence rate becomes quadratic when the iteration is close to the optimal solution.

Thus, we have

$$F(T^+) \leq F(T) - (\beta - \log(1+\beta)), \tag{4.43}$$

where $T^+ = T + \frac{1}{1+\beta}dT$, and $\beta \geq 1$.

The theorem now follows from (4.42) and (4.43). ■

Chapter 5

Applications and implementations

In this chapter we discuss the applications and implementation issues of the algorithm developed in Section 4.3. We apply the ACCPM as a nondifferentiable technique to solve a special case of semidefinite programming. ACCPM is an alternative technique for solving nonsmooth optimization problems. The advantage of this method, as we saw in the previous chapters, is that the computation of an approximate analytic center of a compact convex set is relatively simple and the recentering procedure can be done very efficiently. This is the first integration of a nonpolyhedral cone with the ACCPM.

Our work was motivated by the Spectral Bundle Method for semidefinite programming which was introduced by Helmberg and Rendl [23]. This method is an extension of the algorithm proposed by Kiwiel [26] into the cone of semidefinite matrices, where the new iterate is updated based on the bundle of subgradient information collected from the previous iterates. In this chapter we use their idea in reformulating dual problem of semidefinite program into an eigenvalue optimization. Then we transform the optimization problem into a convex feasibility problem and apply the ACCPM to find

the solution.

The chapter is organized as follows: In Section 5.1 we reformulate the dual problem of a semidefinite program into an eigenvalue optimization and study the maximum eigenvalue function. We then reduce the eigenvalue optimization problem to a convex feasibility problem with semidefinite cuts in Section 5.2. Section 5.3 introduces the weighted analytic center when the set of localization is composed of linear and semidefinite cuts. Sections 5.4 and 5.5 extend the primal and dual algorithms presented in Sections 3.3 and 3.4 for the weighted analytic center and derive the primal and dual directions corresponding to the linear and the semidefinite cuts. We describe recovering the primal feasibility after adding linear and semidefinite cuts in Section 5.4.1 and discuss the difficulties when dealing with the deep cuts in dual space in Section 5.5.1. We also present a path-following algorithm for recovering dual feasibility in such cases. In Section 5.6 we explain how a feasible solution, near the optimal solution, for the primal semidefinite program can be created. Finally, Section 5.7 is devoted to the numerical results of our algorithm when applied to the semidefinite relaxation of the max-cut problem.

5.1 Eigenvalue Optimization

Recall the primal problem of semidefinite programming defined in Section 2.1.

$$\begin{aligned} \max \quad & C \bullet X \\ s.t. \quad & \\ & \mathcal{A}X = b \\ & X \succeq 0, \end{aligned} \tag{5.1}$$

and the dual problem

$$\begin{aligned} \min \quad & b^T y \\ s.t. \quad & \\ & \mathcal{A}^T y - S = C \\ & S \succeq 0. \end{aligned} \tag{5.2}$$

In the rest of this chapter we assume that $\mathbf{tr}(X)$ for any primal feasible matrix X is a positive constant. This assumption is satisfied by many semidefinite relaxations arising from combinatorial optimization. In Section 2.2 we studied the max-cut problem and derived a semidefinite relaxations for this problem. The above assumption on the trace of primal feasible matrix X is clearly satisfied by this relaxation in (2.12).

Reformulation of the dual problem into an eigenvalue optimization problem when $\mathbf{tr}(X)$ for any primal feasible matrix X is a positive constant is well known (see Helmberg and Rendl [23]). We demonstrate the most important steps of this procedure. For a comprehensive analysis of this reformulation see Helmberg [21].

First observe that the slack matrix S is positive semidefinite if the largest eigenvalue of $-S$ is nonpositive. Moreover, since the optimal primal matrix X is nonzero, then due to Theorem 1 the optimal matrix S must be singular. That is $\lambda_{\max}(-S) = 0$. Thus problem (5.2) is equivalent to

$$\begin{aligned} \min \quad & b^T y \\ s.t. \quad & \lambda_{\max}(C - \mathcal{A}^T y) = 0. \end{aligned}$$

Using Lagrange multiplier τ we lift the constraint into the objective function and obtain

$$\min \quad \tau \lambda_{\max}(C - \mathcal{A}^T y) + b^T y. \tag{5.3}$$

In general, the optimal value of Lagrange multiplier is not known in advance. However, with our assumption it is not difficult to verify that the Lagrange

multiplier τ at the optimal solution is equal to $\mathbf{tr}(X)$. (see [21, §5.1] for proof).

Problem (5.3) is an eigenvalue optimization problem. More precisely, it is the minimization of the maximum eigenvalue of an affine combination of symmetric matrices. Although the affine combination is differentiable with respect to y, the maximum eigenvalue of this matrix is not differentiable. Therefore we need more sophisticated methods for solving this problem.

Let us first study the eigenvalue function and its properties in detail. Consider the following function:

$$g(y) = \lambda_{\max}(C - \mathcal{A}^T y).$$

This function is well known to be continuous, convex, nondifferentiable and nonpolyhedral cone. Moreover, it cannot be written as the point-wise maximum of finite number of convex smooth functions. Thus, the standard nondifferentiable techniques cannot be applied to minimize this function. The following example from Overton [47] gives a clear pictorial view of the types of function we are dealing with.

Example 25 *Let* $A_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, $A_2 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$, *and* $C = I$, *then*

$$C - \mathcal{A}^T y = \begin{pmatrix} 1 + y_1 & y_2 \\ y_2 & 1 - y_1 \end{pmatrix},$$

and $\lambda_{\max} = 1 + \sqrt{y_1^2 + y_2^2}$.

Figure 5.1 shows that $\lambda_{\max}$ possess all properties mentioned above.

Note that the maximum eigenvalue function in the above example is a convex cone that is nondifferentiable at $y = 0$ which happens to be the minimum. In general, $g(y)$ is differentiable at y if the maximum eigenvalue

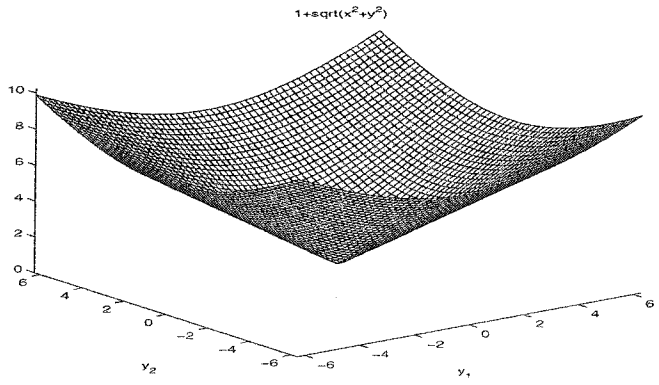

Figure 5.1: Graph of the maximum eigenvalue of the affine combination of symmetric matrices A_1, A_2 and C

has multiplicity one. However, in practice we are dealing with functions that do not have this smooth property. In such cases we work with the set of subgradients of g at each point.

The subdifferential of function g at point $\bar{y}$ can be obtained using chain rule and the Clark generalized gradient:

Theorem 26 (Overton [47]) *Let $\bar{y}$ be in domain of g and the maximum eigenvalue of $g(\bar{y})$ has multiplicity p, with a corresponding orthonormal basis of eigenvectors $\bar{Q} = [\bar{q}_1, ..., \bar{q}_p]$. The generalized gradient of g at $\bar{y}$ is*

$$\partial g(\bar{y}) = \left\{ \mathcal{A}\left(\bar{Q}U\bar{Q}^T\right) : U \succeq 0, \mathbf{t}r(U) = 1 \right\}. \tag{5.4}$$

The eigenvector matrix $\bar{Q}$ plays a key role in our algorithm. In the next section we discuss transformation of the optimization problem into the convex feasibility problem.

5.2 From Optimization to Feasibility

The goal is to solve problem (5.3). Let

$$f(y) = \tau\lambda_{\max}(C - \mathcal{A}^T y) + b^T y.$$

First, note that the maximum eigenvalue of a symmetric matrix can be computed by solving a semidefinite program. If A is a symmetric matrix, then the maximum eigenvalue of A can be obtained via

$$\lambda_{\max}(A) = \max\{A \bullet U : \mathbf{tr}(U) = 1, U \succeq 0\},$$

and consequently, $\lambda_{\max}$ is a convex function of A (see [47] for the proof). Thus

$$\begin{aligned} f(y) &= \tau \max\left\{\left(C - \mathcal{A}^T y\right) \bullet U : \mathbf{tr}(U) = 1, U \succeq 0\right\} + b^T y \\ &= \tau \max\left\{\left(C - \mathcal{A}^T y\right) \bullet U + b^T y : \mathbf{tr}(U) = 1, U \succeq 0\right\} \end{aligned}$$

Now consider a restriction of the space of positive semidefinite matrices to a subcone generated by the faces of the cone at $\bar{y}$. That is, let $\bar{y}$ be a point in domain of f, and the maximum eigenvalue of $C - \mathcal{A}^T\bar{y}$ has multiplicity p and $\bar{Q} \in R^{n\times p}$ be a matrix whose columns form a basis for the eigenspace of the maximum eigenvalue:

$$(C - \mathcal{A}^T\bar{y})\bar{Q} = \lambda_{\max}(C - \mathcal{A}^T\bar{y})\bar{Q} \tag{5.5}$$

and define

$$\begin{aligned} \bar{f}(y) &= \tau \max\left\{\left(C - \mathcal{A}^T y\right) \bullet \bar{Q}U\bar{Q}^T + b^T y : \mathbf{tr}(U) = 1, U \succeq 0\right\} \\ &= \tau\lambda_{\max}\left(\bar{Q}^T(C - \mathcal{A}^T y)\bar{Q}\right) + b^T y \end{aligned}$$

$\bar{f}(y)$ is a convex function which establishes a lower bound on $f(y)$. That is

$$\bar{f}(y) \leq f(y), \text{ for all } y \in R^m.$$

Moreover

$$\begin{aligned} \bar{f}(\bar{y}) &= \tau\lambda_{\max}\left(\bar{Q}^T(C - \mathcal{A}^T\bar{y})\bar{Q}\right) + b^T\bar{y} \\ &= \tau\lambda_{\max}\left(C - \mathcal{A}^T\bar{y}\right) + b^T\bar{y} \\ &= f(\bar{y}). \end{aligned}$$

Let θ be an upper bound on the optimal objective value and

$$\bar{\Omega} = \left\{(y,z) \in R^{m+1} : \bar{f}(y) \leq z \leq \theta\right\},$$

be bounded. $\bar{\Omega}$ contains the optimal solution since

$$\bar{f} \leq f^* = \min f(y) \leq \theta.$$

Note that $\bar{\Omega}$ is the area bounded by a set of subgradients of f at $\bar{y}$ and the hyperplane $z = \theta$.

$$\bar{\Omega} \;=\; \left\{(y,z) \in R^{m+1} : \tau\lambda_{\max}\left(\bar{Q}^T\left(C - \mathcal{A}^T y\right)\bar{Q}\right) + b^T y \leq z \leq \theta\right\}.$$

Observe that

$$\tau\lambda_{\max}\left(\bar{Q}^T\left(C - \mathcal{A}^T y\right)\bar{Q}\right) + b^T y \leq z,$$

implies that

$$\tau\lambda_{\max}\left(\bar{Q}^T\left(C - \mathcal{A}^T y\right)\bar{Q} + \frac{1}{\tau}(b^T y)I\right) \leq z,$$

and therefore for all $i = 1, ..., m$

$$\tau\lambda_i\left(\bar{Q}^T\left(C - \mathcal{A}^T y\right)\bar{Q} + \frac{1}{\tau}(b^T y)I\right) \leq z,$$

or

$$\tau\left(\bar{Q}^T\left(C - \mathcal{A}^T y\right)\bar{Q} + \frac{1}{\tau}(b^T y)I\right) \preceq zI,$$

and thus

$$\tau\bar{Q}^T C\bar{Q} - \tau\sum_{i=1}^{m} y_i\bar{Q}^T A_i\bar{Q} + \sum_{i=1}^{m} y_i b_i I \preceq zI.$$

By rearranging

$$\sum_{i=1}^{m} y_i\left(-\tau\bar{Q}^T A_i\bar{Q} + b_i I\right) - zI \preceq -\tau\bar{Q}^T C\bar{Q}.$$

Thus

$$\bar{\Omega} = \left\{y_z \in R^{m+1} : \bar{A}^T y_z \preceq \bar{C}, z \leq \theta\right\}$$

where $y_z = (y, z) \in R^{m+1}$ and

$$\bar{A}^T y_z = \sum_{i=1}^{m} y_i \bar{A}_i + z\bar{A}_{m+1},$$

and

$$\bar{A}_i = -\tau \bar{Q}^T A_i \bar{Q} + b_i I, \quad i = 1, ..., m,$$

and

$$\bar{A}_{m+1} = -I,$$

and

$$\bar{C} = -\tau \bar{Q}^T C \bar{Q}.$$

Note that $\bar{\Omega}$ is of the form of Ω_D defined in Chapter 4. That is, $\bar{\Omega}$ is a compact convex set which contains the optimal solution set. Therefore Algorithm 3 can be used to locate a point in the solution set.

So far we demonstrated that the optimization problem (5.3) can be transformed to a convex feasibility problem. In order to make sure that the set of localization is bounded, in practice we initiate it by a *box constraint*. That is

$$\begin{aligned} \Omega_D^0 &= \{y_z \in R^{m+1} : l \leq y_i \leq u \text{ and } \underline{z} \leq z \leq \bar{z}\} \\ &= \{y_z \in R^{m+1} : (A^0)^T y_z \leq c^0\}, \end{aligned}$$

where

$$A^0 = [I_{m+1} \quad -I_{m+1}]$$

and

$$c^0 = \begin{pmatrix} u \\ \bar{z} \\ -l \\ -\underline{z} \end{pmatrix}$$

The criterion for choosing lower bounds l and $\underline{z}$, and upper bounds u and $\bar{z}$ will be discussed later.

If f is differentiable at y^k then the oracle returns the gradient of f at y^k, which is a linear cut. Otherwise, it returns a semidefinite cut. Therefore the set of localization is composed of both linear and semidefinite cuts. In order to be able to treat these cuts separately, we make the distinction here. At the k-th iteration of the ACCPM algorithm the set of localization is

$$\Omega_D^k = \left\{ y_z \in R^{m+1} : (\mathcal{A}^k)^T y_z \preceq C^k,\ (A_{lp}^k)^T y_z \leq c_{lp}^k,\ z \leq \theta^k \right\},$$

where A_{lp}^k is a matrix whose columns are linear cuts and $(\mathcal{A}^k)^T y_z \preceq C^k$ represents the semidefinite cuts. We compute y^k, an approximate analytic center Ω_D^k. If y^k is in the solution set, stop. Otherwise, a separation oracle is called to evaluate the function f at y^k and return an orthonormal matrix $Q^k \in R^{n \times p_k}$, where p_k is the multiplicity of the maximum eigenvalue of $C - \mathcal{A}^T y^k$. The matrix Q^k is used to create a new cut and to update the set of localization. Ω_D^{k+1} should contain the solution set and cuts off the current point. If $p_k > 1$, then the new cut is a semidefinite cut and the set of localization is updated via

$$\Omega_D^{k+1} = \Omega_D^k \cap \left\{ y_z \in R^{m+1} : (\mathcal{B}^k)^T y_z \preceq D^k,\ z \leq \theta^{k+1} \right\},$$

where

$$(\mathcal{B}^k)^T y_z = \sum_{i=1}^{m} y_i B_i^k + z B_{m+1}^k,$$

$$B_i^k = -\tau (Q^k)^T A_i Q^k + b_i I, \quad i = 1, ..., m,$$

$$B_{m+1}^k = -I,$$

$$D^k = -\tau (Q^k)^T C Q^k,$$

and

$$\theta^{k+1} = \min\{\theta^k, f(y^k)\}. \tag{5.6}$$

The semidefinite cut matrix A_i^k and the matrix C^k, in this case are updated as in (4.23)

$$A_i^{k+1} = \begin{pmatrix} A_i^k & 0 \\ 0 & B_i^k \end{pmatrix}, \text{ and } C^{k+1} = \begin{pmatrix} C^k & 0 \\ 0 & D^k \end{pmatrix},$$

That is, the dimension of the semidefinite cut matrices A_i^k is enlarged, by p_k when adding a p_k-dimensional semidefinite cut as a block diagonal.

If $p_k = 1$ (f is differentiable at y^k), then B_i^k, for $i = 1, ..., m+1$ are scalars and therefore the new cut is a single linear cut. In this case we store B_i^k, for $i = 1, ..., m+1$ in a column vector b_{lp} and update the linear cut matrix A_{lp}^k via

$$A_{lp}^{k+1} = \begin{bmatrix} A_{lp}^k & b_{lp} \end{bmatrix}.$$

Likewise, D^k is a scalar and is used to update c_{lp}^k.

$$c_{lp}^{k+1} = \begin{pmatrix} c_{lp}^k \\ D^k \end{pmatrix}$$

Thus the updated set of localization in this case is

$$\Omega_D^{k+1} = \Omega_D^k \cap \left\{ y_z \in R^{m+1} : (A_{lp}^{k+1})^T y_z \leq c_{lp}^{k+1},\ z \leq \theta^{k+1} \right\},$$

where θ^{k+1} is as defined in (5.6).

In practice, as k increases the dimension of the cut matrix also increases. This, pushes the analytic center of Ω_D^k to get closer and closer to the upper bound θ. To avoid this tragedy, we put a weight on the upper bound cut $z \leq \theta$, i.e., we repeat this constraint ρ times, ($\rho \geq 1$), and compute the analytic center. By experience we learned that the best value for this weight is equal to the current dimension of the cut matrix. In the next section we define the weighted analytic center and derive its optimality conditions.

5.3 Weighted Analytic Center

As mentioned before, the set of localization is composed of linear and semidefinite cuts. Let n_{sd} and n_l be the number of semidefinite and linear cuts respectively and let $N_{sd} = \sum_{j=1}^{n_{sd}} p_j$ be the dimension of the current semidefinite cut matrix. Dropping the index we indicate $\mathcal{A}^k$ by $\mathcal{A}$, A_i^k by A_i, C^k by C, A_{lp}^k by A, c_{lp}^k by c and Ω_D^k by Ω_D. Thus the set of localization in its general form in R^{m+1} is

$$\Omega_D = \left\{ y_z \in R^{m+1} : \mathcal{A}^T y_z \preceq C,\ A^T y_z \leq c,\ z \leq \theta \right\}.$$

The linear cut $A^T y_z \leq c$ in this formulation contains the box constraint $(A^0)^T y_z \leq c^0$.

Consider Ω_D when we repeat $z \leq \theta$, ρ times, where $\rho \geq 1$:

$$\Omega_D = \left\{ y_z \in R^{m+1} : \mathcal{A}^T y_z \preceq C,\ A^T y_z \leq c,\ \underbrace{z \leq \theta, \ldots, z \leq \theta}_{\rho \text{ times}} \right\}.$$

The dual potential function for this set is

$$\phi_D(y_z) = \log\det S^{-1} + \sum_{j=1}^{n_l} \log s_j^{-1} + \rho \log \sigma^{-1},$$

where $S = C - \mathcal{A}^T y_z$, $s = c - A^T y_z$ and $\sigma = \theta - z$. The analytic center of Ω_D is obtained by minimizing the dual potential function over Ω_D:

$$\begin{aligned} \min\ & \log\det S^{-1} + \textstyle\sum_{j=1}^{n_l} \log s_j^{-1} + \rho \log \sigma^{-1} \\ s.t.\ & \\ & \mathcal{A}^T y_z + S = C \\ & A^T y_z + s = c \\ & z + \sigma = \theta \\ & S \succeq 0, s \geq 0,\ \sigma \geq 0 \end{aligned} \tag{5.7}$$

The first order optimality condition for Problem (5.7) is

$$\begin{pmatrix} A_1 \bullet S^{-1} \\ \vdots \\ A_m \bullet S^{-1} \\ A_{m+1} \bullet S^{-1} \end{pmatrix} + As^{-1} + \rho\sigma^{-1}e_{m+1} = 0.$$

Let $X = S^{-1}$, $x = s^{-1}$ and $\xi = \rho\sigma^{-1}$, then the optimality conditions read

$$\begin{aligned} \mathcal{A}X + Ax + \xi e_{m+1} &= 0 \\ \mathcal{A}^T y_z + S &= C \\ A^T y_z + s &= c \\ z + \sigma &= \theta \\ XS &= I \\ xs &= e \\ \sigma\xi &= \rho. \end{aligned} \tag{5.8}$$

Note that xs is the coordinate-wise product of vectors x and s, and s^{-1} is the component-wise inverse of vector s.

The optimality conditions for the weighted analytic center can also be derived by the primal potential function. Let

$$\Omega_P = \left\{ X \in \mathcal{S}_+^{N_{sd}}, x \in R_+^{n_l}, \xi \in R_+ : \mathcal{A}X + Ax + \xi e_{m+1} = 0 \right\},$$

and let

$$\phi_P(X, x, \xi) = C \bullet X + c^T x + \theta\xi - \log\det X - \sum_{j=1}^{n_l} \log x_j - \rho \log \xi.$$

Then the optimal solution of the following problem

$$\begin{aligned} \min \quad & C \bullet X + c^T x + \theta\xi - \log\det X - \textstyle\sum_{j=1}^{n_l} \log x_j - \rho \log \xi \\ s.t. \quad & \\ & \mathcal{A}X + Ax + \xi e_{m+1} = 0 \\ & X \succeq 0, x \geq 0, \xi \geq 0, \end{aligned}$$

satisfies system (5.8).

5.4 Primal Algorithm for Weighted Analytic Center

Like in Section 3.3, consider the quadratic approximation of the the primal potential function

$$\begin{aligned}
&\phi_P(X+dX, x+dx, \xi+d\xi) = \\
&\quad C \bullet (X+dX) + c^T(x+dx) + \theta(\xi+d\xi) - \log\det(X+dX) \\
&\quad - \sum \log(x_j + dx_j) - \rho \log(\xi + d\xi), \\
\\
&= \phi_P(X, x, \xi) + (C - X^{-1}) \bullet dX + (c - x^{-1})^T dx + (\theta - \rho\xi^{-1})d\xi \\
&\quad + \frac{1}{2}\mathbf{t}rX^{-1}(dX)X^{-1}(dX) + \frac{1}{2}dx^T X_{lp}^{-2} dx + \frac{\rho}{2}\xi^{-2}d\xi^2,
\end{aligned}$$

where X_{lp} is a diagonal matrix made up of x. Feasible directions dX, dx and $d\xi$ should satisfy

$$\mathcal{A}(X+dX) + A(x+dx) + (\xi+d\xi)e_{m+1} = 0,$$

or

$$\mathcal{A}dX + Adx + d\xi e_{m+1} = 0.$$

Thus, we solve

$$\begin{aligned}
\min \quad & \phi_P(X+dX, x+dx, \xi+d\xi) \\
s.t. \quad & \\
& \mathcal{A}dX + Adx + d\xi e_{m+1} = 0.
\end{aligned} \tag{5.9}$$

The KKT conditions are

$$C - X^{-1} + X^{-1}(dX)X^{-1} - \mathcal{A}^T y_z = 0 \tag{5.10}$$

$$c - x^{-1} + X_{lp}^{-2}dx - A^T y_z = 0 \tag{5.11}$$

$$\theta - \rho\xi^{-1} + \rho\xi^{-2}d\xi - z = 0, \tag{5.12}$$

by multiplying X from the right side and from the left side to (5.10) and then applying operator $\mathcal{A}$ one has

$$\mathcal{A}(XCX) - \mathcal{A}X + \mathcal{A}dX - (\mathcal{A}_P\mathcal{A}_P^T)y_z = 0. \tag{5.13}$$

By multiplying AX_{lp}^2 to (5.11) from the left side, we have

$$AX_{lp}^2c - Ax + Adx - AX_{lp}^2A^Ty_z = 0, \tag{5.14}$$

and by multiplying $\rho^{-1}\xi^2e_{m+1}$ to (5.12), we have

$$\rho^{-1}\theta\xi^2e_{m+1} - \xi e_{m+1} + d\xi e_{m+1} - \rho^{-1}z\xi^2e_{m+1} = 0. \tag{5.15}$$

Let

$$G^P = \mathcal{A}_P\mathcal{A}_P^T + A(X_{lp})^2A^T + \rho^{-1}\xi^2e_{m+1}e_{m+1}^T,$$

and

$$g^P = \mathcal{A}(XCX) + AX_{lp}^2c + \rho^{-1}\theta\xi^2e_{m+1}.$$

Then summing up (5.13), (5.14) and (5.15), implies

$$G^Py_z = g^P,$$

and thus

$$y_z = (G^P)^{-1}g^P.$$

Substituting y_z into (5.13), (5.14) and (5.15) we derive the primal directions for computing the weighted analytic center

$$\begin{aligned} dX &= (X\mathcal{A}^TX)y_z + X - XCX \\ &= X - X(C - \mathcal{A}^Ty_z)X \end{aligned} \tag{5.16}$$

$$dx = x - X_{lp}^2s \tag{5.17}$$

$$d\xi = \rho^{-1}z\xi^2 + \xi - \rho^{-1}\theta\xi^2. \tag{5.18}$$

We refer to G^P as the primal Gram matrix.

As in Section 3.3, we project the dX, dx and $d\xi$ such that

$$\mathcal{A}dX + Adx + d\xi e_{m+1} = 0.$$

Let $\bar{G}^P$ and $\bar{g}^P$ be the same as G^P and g^P where $\rho = 1$. That is

$$\bar{G}^P = \mathcal{A}_P\mathcal{A}_P^T + A(X_{lp})^2A^T + \xi^2 e_{m+1}e_{m+1}^T,$$

and

$$\bar{g}^P = \mathcal{A}(XCX) + AX_{lp}^2c + \theta\xi^2 e_{m+1},$$

and let $q \in R^{m+1}$ be defined via

$$q = (\bar{G}^P)^{-1}\left(\mathcal{A}dX + Adx + d\xi e_{m+1}\right).$$

Let the projection of the primal directions be $d\bar{X}$, $\bar{dx}$, and $\bar{d\xi}$. Then with the same line of proof as in derivation of (3.12) one can show that

$$d\bar{X} = dX - (X\mathcal{A}^TX)q,$$

$$\bar{dx} = dx - X_{lp}^2A^Tq,$$

and

$$\bar{d\xi} = d\xi - \xi^2 q.$$

Observe that

$$\begin{aligned}
&\mathcal{A}d\bar{X} + A\bar{dx} + \bar{d\xi}e_{m+1} \\
&= \mathcal{A}dX + Adx + d\xi e_{m+1} - (\mathcal{A}_P\mathcal{A}_P^T)q - (AX_{lp}^2A^T)q - (\xi^2 e_{m+1})q \\
&= \mathcal{A}dX + Adx + d\xi e_{m+1} - \left(\mathcal{A}_P\mathcal{A}_P^T + AX_{lp}^2A^T + \xi^2 e_{m+1}e_{m+1}^T\right)q \\
&= \mathcal{A}dX + Adx + d\xi e_{m+1} - \bar{G}^Pq \\
&= 0.
\end{aligned}$$

In our algorithm we apply a step size α to move as far as possible along with the primal direction while respecting primal feasibility. Consider the quadratic approximation of the primal potential function with the step size:

$$\begin{aligned}
&\phi_P(X+\alpha dX, x+\alpha dx, \xi+\alpha d\xi) \\
&= C\bullet(X+\alpha dX)+c^T(x+\alpha dx)+\theta(\xi+\alpha d\xi)-\log\det(X+\alpha dX) \\
&\quad -\sum\log(x_j+\alpha dx_j)-\rho\log(\xi+\alpha d\xi), \\
&= C\bullet X+\alpha C\bullet dX+c^Tx+\alpha c^Tdx+\theta\xi+\alpha\theta d\xi-\log\det X \\
&\quad -\sum\log(1+\alpha\lambda_j)-\sum\log x_j-\sum\log(1+\alpha\gamma_j) \\
&\quad -\rho\log\xi-\rho(1+\alpha\xi^{-1}d\xi),
\end{aligned}$$

where λ_j's are the eigenvalues of symmetric matrix $X^{-.5}(dX)X^{-.5}$ and $\gamma_j = x_j^{-1}dx_j$. Taking the derivative of ϕ_P with respect to α and setting it to zero one has

$$C\bullet dX+c^Tdx+\theta d\xi-\sum\frac{\lambda_j}{1+\alpha\lambda_j}-\sum\frac{\gamma_j}{1+\alpha\gamma_j}-\frac{\rho\xi^{-1}d\xi}{1+\alpha\xi^{-1}d\xi}=0.$$

We solve the above equation by a line search, where we initiate the step size α to be in the feasible interval $[0,\alpha_{\max}]$. The upper bound of the step size interval will be determined such that the updated matrix $X(\alpha)$ is positive definite and the updated vector $x(\alpha)$ and $\xi(\alpha)$ are positive.

Lemma 27 *Let $X(\alpha)=X+\alpha dX$, $x(\alpha)=x+\alpha dx$ and $\xi(\alpha)=\xi+\alpha d\xi$, where dX, dx and $d\xi$ are primal directions computed by the primal algorithm, and let λ_i and γ_j be as defined above. Then $X(\alpha)\succ 0$, $x(\alpha)>0$ and $\xi(\alpha)>0$ for any $\alpha\in[0,\alpha_{\max}]$, where*

$$\alpha_{\max}=\frac{-1}{\min_{i,j}(\lambda_i,\gamma_j,\xi^{-1}d\xi)}$$

Proof. Note that

$$X(\alpha) = X + \alpha dX = X^{.5}(1 + \alpha X^{-.5} dX X^{-.5}) X^{.5},$$

and $X(\alpha) \succ 0$ only if $\alpha\lambda_i > -1$, for all $i = 1, ..., n_{sd}$, or

$$\alpha(\min_i \lambda_i) > -1.$$

Similarly,

$$x(\alpha) = x + \alpha dx = x(e + \alpha x^{-1} dx),$$

and $x(\alpha) > 0$ only if $\alpha\gamma_j > -1$, for all $j = 1, ..., n_{lp}$, or

$$\alpha(\min_j \gamma_j) > -1,$$

and finally, $\xi(\alpha) = \xi + \alpha d\xi = \xi(1 + \xi^{-1} d\xi) > 0$, if

$$\alpha\xi^{-1} d\xi > -1.$$

Thus a feasible step size should satisfy

$$\alpha \left(\min_{i,j} \left(\lambda_i, \gamma_j, \xi^{-1} d\xi \right) \right) > -1$$

The proof is immediate now. ■

Thus far, we showed that we can transform the optimization problem 5.3 into a feasibility problem. We discussed the issues related to the computational algorithm for the weighted analytic center of a convex set composed of linear and semidefinite cuts in primal setting. We derived the primal directions from a strict interior point and showed that how we project the direction to recover the feasibility. The primal algorithm for computing the weighted analytic center, as we mentioned before, starts with a strict feasible point as its initial point. We now address this issue.

5.4.1 Adding Cuts in Primal Algorithm

In this section we deal with the issues in recovering the primal feasibility after adding a linear or a semidefinite cut for computing the weighted analytic center. Theoretically, a linear cut can be considered as a semidefinite cut, as we did in the previous chapters. However, in practice the distinction between the two cuts is important. Treating a linear cut as a semidefinite cut would drastically increase the computation time of both primal and dual algorithms. It is mostly because of the Cholesky factorization of the Gram matrices $\mathcal{A}_P\mathcal{A}_P^T$ and $\mathcal{A}_D\mathcal{A}_D^T$ are computationally very expensive. We first discuss adding a semidefinite cut:

Adding a p- dimensional Semidefinite Cut

Let

$$\Omega_P = \left\{ X \in \mathcal{S}_+^{n_{sd}+n_{lp}}, \xi \in R_+ : \mathcal{A}X + \xi e_{m+1} = 0 \right\},$$

be the current localization set, where $(\mathcal{A}X)_i = A_i \bullet X$, for $i = 1, ..., m+1$, and $A_i \in \mathcal{S}^{n_{sd}+n_{lp}}$ are the block diagonal matrices composed of the semidefinite and the linear cuts. We make this combination for the sake of simplicity in our analysis below. Separate directions for linear and semidefinite cuts will be extracted from the general form. The optimality conditions for the weighted analytic center of Ω_P are

$$\begin{aligned} \mathcal{A}X + \xi e_{m+1} &= 0 \\ z + \sigma &= \theta \\ XS &= I \\ \mathcal{A}^T y_z + S &= C \\ \sigma\xi &= \rho. \end{aligned} \tag{5.19}$$

Now let

$$\bar{A}_i = \begin{pmatrix} A_i & \\ & 0 \end{pmatrix}, \text{ for } i = 1, ..., m \quad \text{and } \bar{A}_{m+1} = \begin{pmatrix} A_{m+1} & \\ & 1 \end{pmatrix},$$

and

$$\bar{X} = \begin{pmatrix} X & \\ & \xi \end{pmatrix}.$$

Then

$$\Omega_P = \left\{ X \in \mathcal{S}_+^{n_{sd}+n_{lp}+1} : \bar{\mathcal{A}}\bar{X} = 0 \right\},$$

and optimality conditions 5.19 can be rewritten via

$$\begin{aligned} \bar{\mathcal{A}}\bar{X} &= 0 \\ \bar{\mathcal{A}}^T y_z + \bar{S} &= \bar{C} \\ \bar{S}\bar{X} &= \bar{I}, \end{aligned}$$

where

$$\bar{S} = \begin{pmatrix} S & \\ & \sigma \end{pmatrix} \quad \text{and } \bar{C} = \begin{pmatrix} C & \\ & \theta \end{pmatrix},$$

and

$$\bar{I} = \begin{pmatrix} I & \\ & \rho \end{pmatrix}.$$

Assume that the oracle returns a p-dimensional semidefinite cut $\mathcal{B}^T$ at the current iteration. We update Ω_P by adding this cut:

$$\Omega_P^+ = \left\{ X \in \mathcal{S}_+^{n_{sd}+n_{lp}+1}, T \in \mathcal{S}_+^p : \bar{\mathcal{A}}\bar{X} + \mathcal{B}T = 0 \right\}.$$

As in Section 4.1 we compute the optimal updating direction $d\tilde{\bar{X}}$ by maximizing $\log\det$ of the new slack matrix over the primal feasible region and the primal Dikin ellipsoid:

$$\begin{aligned} \max \quad & \log\det T \\ s.t. \quad & \\ & \bar{\mathcal{A}} d\bar{X} + \mathcal{B}T = 0 \\ & \|\bar{X}^{-1} d\bar{X}\| \leq 1 \\ & T \succeq 0. \end{aligned} \tag{5.20}$$

The optimality conditions of problem (5.20) are

$$\begin{aligned} -\tilde{T}^{-1} + \mathcal{B}^T v &= 0 && (5.21)\\ \bar{\mathcal{A}}^T v + \sigma \bar{X}^{-1}(\tilde{dX})\bar{X}^{-1} &= 0 && (5.22)\\ \sigma(1 - \|\bar{X}^{-1}\tilde{dX}\|) &= 0 && (5.23)\\ \bar{\mathcal{A}}(\tilde{dX}) + \mathcal{B}\tilde{T} &= 0, && (5.24) \end{aligned}$$

where $\sigma \geq 0$ is the Lagrange multiplier associated with the norm constraint. By multiplying equation (5.22) from the left and from the right by $\bar{X}$ and then applying the operator $\mathcal{A}$ we have,

$$(\bar{\mathcal{A}}_P \bar{\mathcal{A}}_P^T)v + \sigma \bar{\mathcal{A}}(\tilde{dX}) = 0,$$

using (5.24)

$$v = \sigma(\bar{\mathcal{A}}_P \bar{\mathcal{A}}_P^T)^{-1}\mathcal{B}\tilde{T},$$

and again from (5.22)

$$\begin{aligned} \tilde{dX} &= -\frac{1}{\sigma}\bar{X}(\bar{\mathcal{A}}^T v)\bar{X} \\ &= -\bar{X}\bar{\mathcal{A}}^T(\bar{\mathcal{A}}_P \bar{\mathcal{A}}_P^T)^{-1}\mathcal{B}\tilde{T}\bar{X}. && (5.25) \end{aligned}$$

$\tilde{dX}$ is symmetric since $\bar{\mathcal{A}}^T(\bar{\mathcal{A}}_P \bar{\mathcal{A}}_P^T)^{-1}\mathcal{B}\tilde{T}$ is symmetric. Finally from (5.21) $\tilde{T}$ is the unique solution of the following optimization problem:

$$\tilde{T} = \arg\min_{T \succeq 0}\{\frac{p}{2}\mathbf{t}rT\mathcal{V}T - \log\det T\}, \qquad (5.26)$$

where $\mathcal{V} = \mathcal{B}^T(\bar{\mathcal{A}}_P \bar{\mathcal{A}}_P^T)^{-1}\mathcal{B}$.

Now the primal direction $d\bar{X}$ from 5.25 can be decomposed as follows:

$$\tilde{dX} = -X\mathcal{A}^T(\bar{\mathcal{A}}_P \bar{\mathcal{A}}_P^T)^{-1}\mathcal{B}\tilde{T}X,$$

and

$$\tilde{dx} = -X_{lp}^2 A^T(\bar{\mathcal{A}}_P \bar{\mathcal{A}}_P^T)^{-1}\mathcal{B}\tilde{T},$$

and

$$d\tilde{\xi} = -\xi^2(\bar{A}_P\bar{A}_P^T)^{-1}\mathcal{B}\tilde{T}e_{m+1},$$

where $\bar{A}_P\bar{A}_P^T = A_PA_P^T + AX_{lp}^2A^T + \xi^2$. In view of Lemma 11, the updated point

$$X^+ = \begin{pmatrix} X + \alpha d\tilde{X} & \\ & \alpha\tilde{T} \end{pmatrix},$$

$x^+ = x + \alpha d\tilde{x}$ and $\xi^+ = \xi + \alpha d\tilde{\xi}$, for $\alpha < 1$ is a strict feasible point of Ω_P^+ or the warm start.

We mentioned in Section 4.1 that Problem (5.26) can be solved using a Newton method and we derived a complexity bound on the number of Newton steps in Section 4.5. In practice, however, an exact Newton direction seems difficult to achieve. Let

$$F(T) = \frac{p}{2}\mathbf{tr}T\mathcal{V}(T) - \log\det T.$$

Let $T \succ 0$ be given. For small symmetric dT

$$F(T+dT) = \frac{p}{2}\mathbf{tr}(T+dT)\mathcal{V}(T+dT) - \log\det(T+dT).$$

Using the quadratic approximation of $\log\det(T+dT)^{-1}$, one has

$$F(T+dT) - F(T) =$$
$$p\mathbf{tr}(dT)\mathcal{V}(T) + \frac{p}{2}\mathbf{tr}(dT)\mathcal{V}(dT) - \mathbf{tr}T^{-1}dT + \frac{1}{2}\mathbf{tr}T^{-1}(dT)T^{-1}(dT).$$

Note that

$$\begin{aligned}
\mathbf{tr}T\mathcal{V}(dT) &= \mathbf{tr}T\mathcal{B}^T(\bar{A}_P\bar{A}_P^T)^{-1}\mathcal{B}dT \\
&= (\mathcal{B}T)^T(\bar{A}_P\bar{A}_P^T)^{-1}\mathcal{B}dT \\
&= (\mathcal{B}dT)^T(\bar{A}_P\bar{A}_P^T)^{-1}\mathcal{B}T \\
&= \mathbf{tr}(dT)\mathcal{B}^T(\bar{A}_P\bar{A}_P^T)^{-1}\mathcal{B}T \\
&= \mathbf{tr}(dT)\mathcal{V}(T).
\end{aligned}$$

The Newton step is obtained by setting the gradient of $F(T+dT)$ with respect to dT, to zero. That is

$$p\mathcal{V}(T) + p\mathcal{V}(dT) - T^{-1} + T^{-1}(dT)T^{-1} = 0.$$

By multiplying the above equation from the left side and from the right side by T, we have

$$pT\mathcal{V}(T)T + pT\mathcal{V}(dT)T - T + dT = 0. \tag{5.27}$$

An explicit form of dT cannot be obtained from (5.27) and therefore computing an exact Newton direction seems to be impossible. To over pass this problem, in our algorithm, we approximate the quadratic term $\mathbf{t}rT\mathcal{V}T$ in $F(T)$ by a linear term. That is we ignore $\mathbf{t}r(dT)\mathcal{V}(dT)$ in $F(T+dT)$. Consequently (5.27) becomes

$$pT\mathcal{V}(T)T - T + dT = 0,$$

and hence

$$dT = T - pT\mathcal{V}(T)T.$$

This approximation does not significantly change the direction. Our numerical results show that the rate of convergence is still quadratic in most cases and super linear in some. As an initial point we use T^0 defined in Theorem 24 and we apply a line search to compute the step size.

Adding a Single Linear Cut

Let Ω_P, the current set of localization, be as defined in the previous section and assume that the oracle returns a single linear cut $b^T y_z \leq c$. We update Ω_P by adding this cut

$$\Omega_P^+ = \left\{ X \in \mathcal{S}_+^{n_{sd}+n_{lp}+1}, x \in R_+ : \bar{A}\bar{X} + bx = 0 \right\}.$$

To find a warm start for computing an approximate analytic center for the updated set, we maximize $\log x$ subject to the primal feasibility within the Dikin ellipsoid:

$$\begin{aligned} \max \quad & \log x \\ s.t. \quad & \\ & \bar{A}d\bar{X} + bx = 0 \\ & \|\bar{X}^{-1}d\bar{X}\| \leq 1 \\ & x \geq 0. \end{aligned} \tag{5.28}$$

Using the first order optimality conditions, $\tilde{x}$ and $\tilde{d\bar{X}}$ are solutions to Problem (5.28) if and only if

$$-\tilde{x}^{-1} + bv = 0 \tag{5.29}$$

$$\bar{A}^T v + \sigma \bar{X}^{-1}(\tilde{d\bar{X}})\bar{X}^{-1} = 0 \tag{5.30}$$

$$\sigma\left(1 - \|\bar{X}^{-1}\tilde{d\bar{X}}\|\right) = 0 \tag{5.31}$$

$$\bar{A}(\tilde{d\bar{X}}) + b\tilde{x} = 0. \tag{5.32}$$

From (5.30), (5.31) and (5.32)

$$v = p(\bar{A}_P\bar{A}_P^T)^{-1}b\tilde{x},$$

and

$$\tilde{d\bar{X}} = -\bar{X}\left(\bar{A}^T(\bar{A}_P\bar{A}_P^T)^{-1}b\tilde{x}\right)\bar{X},$$

and from (5.29)

$$\tilde{x}^{-1} = p\left(b(\bar{A}_P\bar{A}_P^T)^{-1}b\right)\tilde{x}.$$

or

$$\tilde{x} = \left(\sqrt{b(\bar{A}_P\bar{A}_P^T)^{-1}b}\right)^{-1}.$$

Similar to the previous case we can decompose $\tilde{d\bar{X}}$ to linear and semidefinite directions:

$$\tilde{dX} = -X\mathcal{A}^T(\bar{A}_P\bar{A}_P^T)^{-1}b\tilde{x}X,$$

and

$$\tilde{dx} = -X_{lp}^2 A^T (\bar{A}_P \bar{A}_P^T)^{-1} b\tilde{x},$$

and

$$\tilde{d\xi} = -\xi^2 (\bar{A}_P \bar{A}_P^T)^{-1} b\tilde{x} e_{m+1}.$$

5.5 Dual Algorithm for Weighted Analytic Center

Let

$$\Omega_D = \left\{ y_z \in R^{m+1} : \mathcal{A}^T y_z \preceq C,\ A^T y_z \leq c,\ z \leq \theta \right\},$$

where $z \leq \theta$ is repeated ρ times, be the current set of localization in dual space and

$$\phi_D(S, s, \sigma) = \log\det S^{-1} + \sum_{j=1}^{n_l} \log s_j^{-1} + \rho \log \sigma^{-1},$$

where $S = C - \mathcal{A}^T y_z$, $s = c - A^T y_z$ and $\sigma = \theta - z$, be the dual potential function for Ω_D as defined in Section 5.3.

Let y_z be a strictly feasible point of Ω_D and consider the quadratic approximation of ϕ_D around $dS = (dS)^T$, ds and $d\sigma$:

$$\begin{aligned}
&\phi_D(S + dS, s + ds, \sigma + d\sigma) = \\
&\qquad \log\det(S + dS)^{-1} + \sum \log(s_j + ds_j)^{-1} + \rho \log(\sigma + d\sigma)^{-1} \\
&= \phi_D(S, s, \sigma) - \mathbf{tr} S^{-1} dS + \frac{1}{2} \mathbf{tr} S^{-1}(dS) S^{-1}(dS) \\
&\quad -(s^{-1})^T ds + \frac{1}{2}(ds)^T S_{lp}^{-2} ds - \rho\sigma^{-1}(d\sigma) + \frac{\rho}{2}\sigma^{-2}(d\sigma)^2,
\end{aligned}$$

where S_{lp} is a diagonal matrix made up of s.

On the other hand, dS, ds and $d\sigma$ are feasible directions if

$$\mathcal{A}^T dy_z + dS = 0,$$

$$A^T dy_z + ds = 0,$$

and

$$dz + d\sigma = 0.$$

Thus

$$\begin{aligned}&\phi_D(S + dS, s + ds, \sigma + d\sigma) = \\ &\quad \phi_D(S, s, \sigma) + (\mathcal{A}S^{-1})^T dy_z + (As^{-1})^T dy_z + \rho\sigma^{-1}(dz) \\ &\quad + \tfrac{1}{2}(dy_z)^T(\mathcal{A}_D\mathcal{A}_D^T)dy_z + \tfrac{1}{2}(dy_z)^T(AS_{lp}^{-2}A^T)dy_z + \tfrac{\rho}{2}\sigma^{-2}(dz)^2,\end{aligned} \tag{5.33}$$

where $\mathcal{A}_D\mathcal{A}_D^T \in \mathcal{S}^{m+1}$ is the symmetric matrix defined in Section 3.4, i.e., $(\mathcal{A}_D\mathcal{A}_D^T)_{ij} = \mathbf{tr}A_iS^{-1}A_jS^{-1}$.

Minimizing (5.33) over $dy_z \in R^{m+1}$ gives

$$dy_z = -(G^D)^{-1}g^D,$$

where

$$G^D = \mathcal{A}_D\mathcal{A}_D^T + A(S_{lp}^{-2})A^T + \rho\sigma^{-2}(e_{m+1}e_{m+1}^T),$$

and

$$g^D = \mathcal{A}S^{-1} + As^{-1} + \rho\sigma^{-1}e_{m+1}.$$

The dual direction dy_z with a step size α, obtained by a line search method, is used to updated the dual iteration $y_z^+ = y_z + \alpha dy_z$. We refer to G^D as the dual Gram matrix.

5.5.1 Adding Cuts in Dual Algorithm

The dual algorithm for computing an approximate analytic center as we mentioned before is very efficient. The dual potential function is reduced at each iteration by at least a constant amount. The rate of convergence, however, is quadratic when the iteration is close to the analytic center, i.e., within the Dikin ellipsoid. After adding new cuts, we need to recover feasibility. In practice, the new cuts could be deep, shallow or central. In the dual setting, recovering feasibility in case of the deep cuts is not trivial. In this section, we provide a procedure based on a path-following algorithm to recover feasibility and obtain a warm start for recentering procedure in case of semidefinite shallow and deep cuts. For the sake of simplicity, we consider only semidefinite cuts. The extension of our algorithm to the case in which we have mixed linear and semidefinite cuts is trivial. Let us first define semidefinite deep and shallow cuts:

Definition 28 *Let Ω_D be the current set of localization and $\bar{y}_z$ be an approximate center of Ω_D. Let the oracle returns a p-dimensional semidefinite cut*

$$\mathcal{B}^T y_z \preceq D,$$

then

1. *If $D \succ \mathcal{B}^T \bar{y}_z$, then the semidefinite cut is called "completely shallow"*
2. *If $D \prec \mathcal{B}^T \bar{y}_z$, then the semidefinite cuts is called "completely deep"*
3. *If $D = \mathcal{B}^T \bar{y}_z$, then the semidefinite cut is central.*

In any other situation, the cut is called "partially deep".

Now let

$$\Omega_D^+ = \Omega_D \cap \{y_z \in R^{m+1} : \mathcal{B}^T y_z \preceq D\}.$$

If $\mathcal{B}^T y_z \preceq D$ is a p dimensional shallow cut, the current center $\bar{y}_z$ is clearly a strictly feasible point of Ω_D^+ and can be used as an initial point for the Newton algorithm to compute an approximate center of Ω_D^+. In any other situation, we need to compute a feasible interior point of Ω_D^+. In Section 4.1, we presented a procedure to obtain an interior point of Ω_D^+ as a "warm start" for the Newton algorithm if the semidefinite cut is central. We use this procedure with a path-following algorithm to recover feasibility when the semidefinite cut is deep (partially or completely). The idea is as follows:

Let $\mathcal{B}^T y_z \preceq D$ be a p-dimensional semidefinite cut which is completely or partially deep. We first place the cut at the center and update the set of localization by adding the central cut $\mathcal{B}^T y_z \preceq \mathcal{B}^T \bar{y}_z$. We then use the procedure described in Section 4.1 to obtain a "warm start" and use this point to compute $\bar{y}_z^+$, an approximate center of the updated set. If $D \succ \mathcal{B}^T \bar{y}_z^+$, then update Ω_D by adding the original cut and use the current center as an initial point for the Newton algorithm. Otherwise, move the cut to the new center and repeat the procedure. The path-following algorithm is used after each call to the oracle where the return cut is deep.

For the optimization problem (5.3), the oracle returns a subgradient cut and updates the upper bound cut at the same time. The localization set should therefore be updated accordingly. In this case we add two types of cut at the same time; the subgradient cut(semidefinite) and the upper bound cut(linear). We prove that these two cuts cannot be both deep. More precisely, when one cut is deep the other one is shallow and vice versa. Let

$$\bar{\Omega} = \{y_z \in R^{m+1} : \bar{\mathcal{A}}^T y_z \preceq \bar{C}, z \leq \theta\},$$

be the set of localization arising from Problem (5.3), and $(\bar{y}, \bar{z})$ be an approximate center of $\bar{\Omega}$. Let $\mathcal{B}^T y_z \preceq D$ be the new subgradient cut returned by the oracle, where

$$B_i = -\tau \bar{Q}^T A_i \bar{Q} + b_i I, \quad i = 1, ..., m, \tag{5.34}$$

and

$$B_{m+1} = -I, \tag{5.35}$$

and

$$D = -\tau \bar{Q}^T C \bar{Q}, \tag{5.36}$$

where $\bar{Q}$ is the orthonormal matrix defined in (5.5). That is

$$(C - \mathcal{A}^T \bar{y})\bar{Q} = \lambda_{\max}(C - \mathcal{A}^T \bar{y})\bar{Q}. \tag{5.37}$$

In the next lemma we show that the subgradient cut is completely shallow when the upper bound cut is deep and vice versa.

Lemma 29 *Let $\bar{\Omega}$ and $\mathcal{B}^T$ be defined as above and let $(\bar{y}, \bar{z})$ be an approximate center of $\bar{\Omega}$. Let θ^+ be the new upper bound updated by $f(\bar{y})$. Then, the subgradient cut $\mathcal{B}^T y_z \preceq D$ is*

1. *"completely shallow" if the upper bound cut $z \leq \theta^+$ is deep, or $\bar{z} > \theta^+$,*
2. *"completely deep", if the upper bound cut $z \leq \theta^+$ is shallow, or $\bar{z} < \theta^+$,*
3. *"central", if the upper bound cut $z \leq \theta^+$ is central, or if $\bar{z} = \theta^+$*

Proof. Let the upper bound is updated by $\theta^+ = f(\bar{y})$. From (5.34) through (5.36)

$$\begin{aligned} D - \mathcal{B}^T \bar{y}_z &= -\tau \bar{Q}^T C \bar{Q} + \tau(\bar{Q}^T \mathcal{A}^T \bar{Q})\bar{y} - (b^T \bar{y})I + \bar{z}I \\ &= -\tau \bar{Q}^T (C - \mathcal{A}^T \bar{y})\bar{Q} - (b^T \bar{y})I + \bar{z}I. \end{aligned}$$

In view of (5.37)

$$\begin{aligned} D - \mathcal{B}^T \bar{y}_z &= -\tau \lambda_{\max}(C - \mathcal{A}^T \bar{y})\bar{Q}^T \bar{Q} - (b^T \bar{y})I + \bar{z}I \\ &= -f(\bar{y})I + \bar{z}I. \end{aligned}$$

Claims 1, 2, and 3 are immediate now. ■

The same lemma is valid when the subgradient cut is linear. For the analysis of linear cuts see Goffin and Vial [17].

The disadvantage of the dual algorithm is that the recentering procedure employs a path-following algorithm when the subgradient cut is deep. This drastically increases the number of Newton steps and therefore the computational time. Recovering feasibility in the case of deep cuts without using a path-following algorithm at this point remains an open problem.

5.6 A Lower Bound

Consider the semidefinite relaxations arising from the Max-Cut problem derived in Section 2.2:

$$\begin{array}{ll} \max & C \bullet X \\ s.t. & \\ & diag(X) = e \\ & X \succeq 0, \end{array} \tag{5.38}$$

where $C = \frac{1}{4}L$ is the coefficient matrix and $X \in \mathcal{S}_+^n$ is the matrix variable. Problem (5.38) is a semidefinite program in the primal form of (5.1), where $b = e$ and $A_i = e_i e_i^T$, for $i = 1, ..., m$. Observe that $\mathtt{tr}X = n$ for any primal feasible matrix X, and therefore it satisfies our assumption of this chapter. The dual Problem is

$$\begin{array}{ll} \min & e^T y \\ s.t. & \\ & Diag(y) - S = C \\ & S \succeq 0, \end{array} \tag{5.39}$$

which is equal to the eigenvalue optimization

$$\min \quad n\lambda_{\max}(C - Diag(y)) + e^T y. \tag{5.40}$$

We apply our primal and dual algorithms, developed in Sections 3.3 and 3.4 with the implementation issues derived in Sections 5.4 and 5.5 to find y^*, the minimizer of (5.40) for randomly generated graphs with different sizes and edge densities. y^* serves as the optimal solution of Problem (5.39). One advantage of the ACCPM is that the algorithm does not need to update the primal matrix at each iteration, yet a primal matrix very close to the optimal can be generated from y^* at the end. We describe this procedure now.

Let y_z^* be the optimal solution of Problem (5.40) computed by primal or dual ACCPM algorithm. Then y^* is an approximate analytic center of

$$\Omega_D = \left\{ y_z \in R^{m+1} : \mathcal{A}^T y_z \preceq C,\ A^T y_z \leq c,\ z \leq \theta \right\},$$

where $\mathcal{A}^T y_z = \sum_{i=1}^{m} y_i A_i + z A_{m+1}$ and $A = [A^0, a^1, \ldots, a^{n_{lp}}]$, where A_is are the semidefinite cut matrices, a_is are the linear cuts, and $A^0 = [I_{m+1}, -I_{m+1}]$ is the box constraint. Note that

$$a_i^j = -\tau (q^j)^T e_i e_i^T q^j + 1, \text{ for } i = 1, \ldots, m, , j = 1, \ldots, n_{lp}, \tag{5.41}$$

and $a_{m+1}^j = -1$, and

$$A_i = Diag\left(A_i^1, \ldots, A_i^{n_{sd}}\right), \text{ for } i = 1, \ldots, m+1, \tag{5.42}$$

where

$$A_i^j = -\tau (Q^j)^T e_i e_i^T Q^j + I_{p_j}, \text{ for } i = 1, \ldots, m,\ j = 1, \ldots, n_{sd}, \tag{5.43}$$

and

$$A_{m+1}^j = -I_{p_j}, \text{ for } j = 1, \ldots, n_{sd}. \tag{5.44}$$

Since y_z^* is an approximate weighted analytic center of Ω_D, it satisfies the first order conditions (5.8). Thus there exist $X \succ 0$, $x > 0$, and $\xi > 0$ such that

$$\mathcal{A}X + Ax + \xi e_{m+1} = 0, \tag{5.45}$$

where $X = Diag\,(X^1, \ldots, X^{n_{sd}})$, with $X^j \in \mathcal{S}^{p_j}_{++}$ and

$$x^T = \left[[(x^1_{box})^T, (x^2_{box})^T], x_1, \ldots, x_{n_{lp}}\right],$$

with $x^k_{box} \in R^{m+1}_+$, $k = 1, 2$.

First observe that for $i = 1, \ldots, m$,

$$A_i \bullet X + (Ax)_i = 0. \tag{5.46}$$

From (5.42) and (5.43)

$$\begin{aligned} A_i \bullet X &= \sum_{j=1}^{n_{sd}} A^j_i \bullet X^j \\ &= \sum_{j=1}^{n_{sd}} \left(-\tau (Q^j)^T e_i e_i^T Q^j + I_{p_j}\right) \bullet X^j \\ &= \sum_{j=1}^{n_{sd}} \left(-\tau (e_i e_i^T) \bullet Q^j X^j (Q^j)^T\right) + \sum_{j=1}^{n_{sd}} \mathbf{t}r X^j \\ &= -(e_i e_i^T) \bullet \hat{X}^1 + \alpha^1, \end{aligned} \tag{5.47}$$

where

$$\hat{X}^1 = \tau \sum_{j=1}^{n_{sd}} Q^j X^j (Q^j)^T,$$

and $\alpha^1 = \mathbf{t}r X$.

On the other hand

$$\begin{aligned} (Ax)_i &= (x^1_{box} - x^2_{box})_i + \sum_{j=1}^{n_{lp}} x_j a^j_i \\ &= (e_i e_i^T) \bullet (X^1_{box} - X^2_{box}) + \sum_{j=1}^{n_{lp}} x_j a^j_i, \end{aligned} \tag{5.48}$$

where X^k_{box} is a diagonal matrix made up of x^k_{box}. Now from (5.41)

$$\begin{aligned} \sum_{j=1}^{n_{lp}} x_j a^j_i &= \sum_{j=1}^{n_{lp}} \left(-\tau (q^j)^T e_i e_i^T q^j x_j\right) + \sum_{j=1}^{n_{lp}} x_j \\ &= -(e_i e_i^T) \bullet \hat{X}^2 + \alpha^2, \end{aligned} \tag{5.49}$$

where $\hat{X}^2 = \tau \sum_{j=1}^{n_{lp}} q^j (q^j)^T x_j$ and $\alpha^2 = \sum_{j=1}^{n_{lp}} x_j$.

In view of (5.46) and with regard to (5.47) through (5.49), one has

$$(e_i e_i^T) \bullet \left(\hat{X}^1 + \hat{X}^2 - X^1_{box} + X^2_{box} \right) = \alpha^1 + \alpha^2,$$

or $diag(\tilde{X}) = e$, where

$$\tilde{X} = \frac{1}{\alpha^1 + \alpha^2} \left(\hat{X}^1 + \hat{X}^2 - X^1_{box} + X^2_{box} \right).$$

Note that $\hat{X}^k \succ 0$ and $x^k_{box} > 0$, $k = 1, 2$. In general, $\tilde{X}$ may not be a positive semidefinite matrix because of the negative sign of x^1_{box}. However in practice, at the optimal solution y^*_z, it seems that the primal matrices $\hat{X}^k$ corresponding to the semidefinite and linear cuts, dominate the primal vectors x^k_{box} of the box constraint. In most cases as we will see in the numerical results, $\tilde{X}$ is the optimal solution of Problem (5.38). That is $\tilde{X} \succeq 0$ and the relative error between the primal and dual objective values is small.

Computing $\tilde{X}$ is a very expensive task because X^j and A^j_i are dense matrices. However, since we compute it only once it does not make a significant increase in the over all computational time of our algorithm. Moreover, α^1 and α^2 can be computed in a less expensive way as follows. In view of (5.45)

$$A_{m+1} \bullet X + (Ax)_{m+1} + \xi = 0.$$

From (5.44)

$$\begin{aligned} A_{m+1} \bullet X &= \sum_{j=1}^{n_{sd}} A^j_{m+1} \bullet X^j \\ &= \sum_{j=1}^{n_{sd}} (-I_{p_j} \bullet X^j) \\ &= -\mathbf{t}rX \\ &= -\alpha^1. \end{aligned}$$

On the other hand since $a^j_{m+1} = -1$ for all j,

$$(Ax)_{m+1} = (x^1_{box} - x^2_{box})_{m+1} - \sum_{j=1}^{n_{lp}} x_j,$$

or $(Ax)_{m+1} = (x^1_{box} - x^2_{box})_{m+1} - \alpha^2$. Thus

$$\alpha^1 + \alpha^2 = (x^1_{box})_{m+1} - (x^2_{box})_{m+1} + \xi.$$

In the rare cases, where $\tilde{X}$ does not satisfy the positive definiteness constraint, we can construct a positive definite matrix from $\tilde{X}$ that is a primal feasible and it is close to the optimal solution:

Let

$$\beta = \min_j \left\{ (x^2_{box} - x^1_{box})_j \right\}.$$

If $\beta \geq 0$, then $\tilde{X} \succ 0$. Assume that $\beta < 0$, then clearly

$$\hat{X}^1 + \hat{X}^2 - X^1_{box} + X^2_{box} - \beta I \succ 0.$$

Let

$$X^f = \frac{\hat{X}^1 + \hat{X}^2 - X^1_{box} + X^2_{box} - \beta I}{\alpha^1 + \alpha^2 - \beta}. \tag{5.50}$$

Then X^f is positive definite and $diag(X^f) = 1$. Thus X^f is a feasible matrix for Problem (5.38).

5.7 Numerical Results

We coded the primal and dual algorithms by Matlab 6.0 and tested them with a number of random graphs generated by the graph generator "Rudy"[1] on a Pentium(R) 4 CPU 1.60 GHz, with 128 MB of RAM. In this section, we present our numerical results.

There are some advantages and some disadvantages with both algorithms. However, the primal algorithm seems to behave better than the dual algorithm in general. The main difficulty which first appeared in both algorithms is initiating the "box constraint". Recall that, in order to make sure that

[1]The graph generator Rudy can be downloaded from the following site:
http://www.zib.de/helmberg/sdp_software.html

the initial set of localization Ω^0 is bounded, we bound y_z by a box constrain. That is $l_i \leq y_i \leq u_i$ and $\bar{z} \leq z \leq \underline{z}$. On the one hand, the parameters l and u should be chosen such that Ω^0 contains the optimal solution y^*. On the other hand, having a big initial set requires too many cuts to satisfy the stopping criterion and consequently slows down the algorithm. To over pass this problem we initiate $l_i = -2$ and $u_i = 2$ and start with the small box constraint $-2 \leq y_i \leq 2$. At each iteration if y_i^k, the i^{th} component of the analytic center is very close to its boundary, we loosen up the constraint by multiplying the bound by 2. We initiate $\bar{z} = b^T y^0$, and $\underline{z} = trC$, since I is a feasible solution of Problem (5.1).

Computing the Gram matrices G^P and G^D is another difficulty in both algorithms. As the dimension of the semidefinite cut matrices increases the computation time of the Gram matrix and its Cholesky factorization also increases. In fact, Matlab is extremely slow when it comes to a computation task which involves "for" loops. To overcome this problem we wrote Mex files in "C", to compute $\mathcal{A}_P\mathcal{A}_P^T$ and $\mathcal{A}_D\mathcal{A}_D^T$, and imported them into Matlab. This made a significant improvement in the computation time when there are too many semidefinite cuts. Nevertheless, the fact that we have to compute the Gram matrix several times at each iteration is the main weakness of our algorithm from the implementation point of view.

As we mentioned earlier the primal algorithm works better than the dual algorithm in practice. The advantage of the primal algorithm is that there is basically no need to identify the deep cuts and take special procedures to update the analytic center after adding a cut. In primal algorithm shallow, central and deep cuts are treated in the same way and the updating procedure is quite efficient. The dual algorithm, however, cannot recover feasibility in an efficient way after adding a deep cut. As explained in Section 5.5, a path-following algorithm is applied in such cases. The Gram matrix must

be computed several times at each recentering process of the path-following algorithm and it drastically increases the computational time.

Note that the numerical results reported in this section are not competitive with the Spectral Bundle method. There are a number of reasons for that.

First of all, our algorithm for computing the analytic center is based on interior point methods. Consequently, the Gram matrix and its Cholesky factorization has to be computed at each iteration for both primal and dual algorithms. This is a very expensive task. Secondly, the numerical results in [23] are obtained using a computer code by C^{++}, whereas we used Matlab for coding our algorithm. Matlab, as mentioned, is more than 20 times slower than C^{++}, especially for the computational algorithms involved in many loops. Finally, because of the difficulties with the dual algorithm, we have to work in primal setting and therefore we lose sparsity of the original problem. This is also the reason that the ACCPM works better on the dense problems than it does on the sparse ones.

The Stopping Criterion

The ACCPM creates a sequence of points y^k to reduce the objective function $f(y)$. As y^k gets closer to the minimizer of $f(y)$, the reduction $f(y^{k-1}) - f(y^k)$ becomes smaller and smaller. This is a common property of almost all nonsmooth optimization techniques. We measure this reduction at each iteration. When we cannot expect a significant improvement between two iterations, i.e., when

$$f(y^{k-1}) - f(y^k) < \varepsilon |f(y^{k-1})|, \tag{5.51}$$

we compute a feasible primal matrix X^f from 5.50 and the relative error between the upper bound $b^T y^k$ and the lower bound $C \bullet X^f$. We terminate

the algorithm if

$$\bar{e} = \frac{b^T y^* - C \bullet X^f}{b^T y^*},$$

is small. After running our algorithm with many data (Table 5.2), we realized that when ε is small enough ($\varepsilon = 10^{-5}$), the relative error $\bar{e}$ is reasonably small ($< 5 \times 10^{-3}$) and therefore in most cases the algorithm stops when (5.51) is satisfied. Note that X^f is computed only when the reduction in f is small; in most cases, only once. This matrix then can be used as a primal feasible solution very close to the optimal.

In Table 5.1 we illustrate the result of the primal algorithm when applied to the randomly generated graphs with 100, 200, 300, 400 and 500 nodes. We tested the algorithm for each class of graph with different edge densities vary between 6% to 90%. The number of edges of a graph with n nodes and density d is given by the integer closest to $(n(n-1) \times d)/200$.

The first two columns of the table demonstrate the number of nodes (n) and the edge density of the graph. the third and the forth columns show the number of linear and semidefinite cuts at the optimal solution; dim(cut) indicates the dimension of the optimal cut matrix, which includes linear and semidefinite cuts. The last column of the table gives the multiplicity of the maximum eigenvalue at the optimal solution. The key property of Table 5.1 is that as the edge density of a graph increases the dimension of the optimal cut matrix decreases. In other words, the denser the graph, the less cuts (linear and semidefinite) is needed to find the optimal solution. This means that ACCPM is more successful when applied to the more complicated graphs with many edges as compared to sparse graphs.

Table 5.2 provides additional information on the performance of the primal algorithm. The third column of the table shows $b^T y^*$, the optimal objective value of Problem (5.2). The forth column of the table illustrates a lower bound on the dual objective function by the primal objective value $C \bullet X^f$.

node n	density (%)	Linear cuts (n_{lp})	Semidefinite cuts (n_{sd})	dim(cut)	p
100	6	140	37	235	4
100	20	89	21	135	2
100	40	75	29	149	4
100	80	28	33	118	5
100	90	22	16	57	3
200	6	144	55	298	5
200	20	102	46	234	4
200	40	102	35	200	5
200	80	37	41	181	4
200	90	17	25	107	5
300	6	167	69	338	5
300	20	117	54	275	4
300	40	100	72	301	4
300	80	40	46	209	5
300	90	22	33	163	5
400	6	168	68	375	5
400	20	149	81	394	5
400	40	107	100	445	4
400	80	40	65	306	5
400	90	23	33	177	5
500	6	379	9	404	2
500	20	407	12	431	1
500	40	367	26	421	2
500	80	204	32	270	4
500	90	125	37	205	3

Table 5.1: The primal algorithm performance on a randomly generated graphs with 100, 200, 300, 400, and 500 nodes

The relative error between the upper bound and the lower bound is shown in the next column. The data in this column shows that the objective function value of the primal problem is relatively close to the optimal objective value of the dual problem. Therefore, as it was expected, X^f is very close to the optimal solution of Problem (5.1). The CPU time in the last column is in the format *minutes:seconds*.

Next, we illustrate the recentering process after adding a cut. Recall that after adding a p-dimensional cut we compute an updating direction by solving the optimization Problem 4.1. In Section 5.4.1 we mentioned that computing an exact Newton direction for this problem is impossible when the cut is semidefinite. Table 5.3 illustrates the number of iterations and the the CPU time for solving Problem 4.1. Although, the convergence rate is not quadratic but the CPU time is reasonably low. The reason is that the computational time per iteration of this algorithm is not expensive since we do not need to compute the Gram matrix and its Cholesky factorization.

In Theorem 14 we proved that, starting from the "warm start" the number of Newton steps after adding a p-dimensional cut is bounded by $O(p\log(p+1))$. We mentioned that the recentering procedure is much slower if we make the cut shallow and use the current analytic center as the initial point for the Newton algorithm. Table 5.4 shows one step of the recentering procedure, with and without computing the optimal updating direction for a random graph with 300 nodes and 90% edge density.

The numerical results demonstrated in this table further confirm the mathematical result established in Theorem 14. Only 2 Newton iterations are needed to update the analytic center after updating the upper bound starting from the "warm start" and the CPU time is 0.20 seconds, whereas these number for the same cut when starting from the current analytic center are 30 and 3.76. The situation is more or less the same for the linear ($p = 1$)

node n	density (%)	upper bound $(b^T y^*)$	lower bound $(C \bullet X^f)$	relative error $(\bar{e})$	cpu time (mm:ss)
100	6	246.46	245.57	3.5e-3	1:20
100	20	666.22	658.01	9.3e-3	0:42
100	40	1210.06	1202.75	6.0e-3	0:46
100	80	2171.74	2165.61	2.8e-3	0:48
100	90	2378.93	2359.49	8.1e-3	0:16
200	6	892.50	887.03	6.1e-3	336
200	20	2500.82	2486.64	5.6e-3	4:15
200	40	4623.48	4589.48	7.4e-3	3:57
200	80	8500.30	8480.57	2.3e-3	3:49
200	90	9374.95	9334.52	4.3e-3	1:30
300	6	1905.71	1895.29	2.7e-3	13:37
300	20	5442.62	5403.05	7.2e-3	12:16
300	40	10146.02	10063.68	8.1e-3	23:21
300	80	18941.84	18752.76	9.9e-3	10:31
300	90	20956.18	20901.29	2.6e-3	5:16
400	6	3265.83	3243.41	6.8e-3	25:35
400	20	9444.34	9398.43	4.8e-3	44:32
400	40	17783.01	17611.14	9.6e-3	58:13
400	80	33480.24	33228.52	7.5e-3	28:26
400	90	37113.40	36902.81	5.6e-3	9:58
500	6	4239.61	4198.10	9.7e-3	46:28
500	20	14589.08	14443.99	9.9e-3	33:35
500	40	27543.51	27309.76	8.4e-3	53:00
500	80	52081.13	51664.28	8.0e-3	36:51
500	90	57810.31	57503.22	5.3e-3	25:18

Table 5.2: The primal algorithm performance on a randomly generated graphs

$n = 300$ density: 90%	upper bound cut	linear cut p=1	SD cut p=2	SD cut p=3	SD cut p=4
No. Newton steps for computing warm start	1	1	9	8	11
CPU time(sec)	0.0000	0.0160	0.5748	0.6132	0.7921

Table 5.3: Number of Newton steps and the CPU time for computing the optimal updating direction and the warm start

$n = 300$ density: 90%	upper bound cut	linear cut p=1	SD cut p=2	SD cut p=3	SD cut p=4
No. steps starting from warm start	2	2	3	3	3
CPU time(sec)	0.20	0.30	0.35	0.38	0.43
No. steps starting from previous center	30	25	25	27	27
CPU time(sec)	3.76	3.84	4.07	4.29	4.69

Table 5.4: Number of Newton steps and the CPU time of the recentering procedure after adding a p-dimensional cut

and semidefinite ($p = 2, 3, 4$) subgradient cuts. Note that the computation of the Gram matrix is needed for each step of the Newton algorithm. This is why the algorithm is about 20 times faster when we compute the optimal updating direction. As the dimension of the cut matrix increases, the Gram matrix becomes more and more complex and the computational time increases exponentially. Therefore bringing the optimal updating direction into play is essential in practice.

Chapter 6

Conclusion

We proposed an analytic center cutting plane method with semidefinite cuts for convex feasibility problems. Convex feasibility problem is a problem of finding a point in a convex set, which contains a full dimensional ball with ε radius and is contained in a compact convex set described by matrix inequalities. Although, the convex feasibility problem is an abstract problem by nature, however, there are many applications of this problem in nonsmooth optimization such as minmax eigenvalue optimization.

The ACCPM is an efficient technique for nondifferentiable optimization. The method has been studied in the past few years in the Euclidean space $\Re^m$ with single and multiple linear cuts and with quadratic cuts. We employed for the first time, a nonpolyhedral model into the ACCPM by means of the semidefinite cuts. As a result, the ACCPM can now be applied to a wide variety of optimization problems such as linear programming, quadratic programming and semidefinite programming.

From the theoretical point of view we studied the convergence analysis of the method in the context of convex feasibility problem. At each iteration of our algorithm, an oracle returns a p-dimensional semidefinite cut. We up-

date the set of localization by placing the cut at the analytic center of the set. Then a special updating procedure is applied to find a strictly feasible point in the updated set of localization as a "warm start" for computing the next analytic center. In Section 4.2 we proved that starting from the "warm start" the Newton algorithm finds the analytic center of the updated set of localization in $O(p\log(p+1))$. We showed in Section 4.5 that the updating direction itself can be obtained very efficiently. Applying the Newton algorithm, the problem comes to optimality in at most $O\left(\frac{p\log\frac{1}{\varepsilon^*}+\log\frac{1}{\mu}}{\beta-\log(1+\beta)}\right)$ iterations, where β is the Newton decrement, $\mu > 0$ is a condition number on the field of cuts, $\varepsilon^* = \frac{(1-\theta)\varepsilon}{(1+\theta)(1+n)}$ and θ is a positive constant less than 1. We established the main theoretical result of our work in Section 4.4. We proved that the ACCPM with semidefinite cuts is a fully polynomial approximation scheme. That is, the algorithm stops when the dimension of the accumulated block diagonal cut matrix reaches to the bound $O^*(\frac{p_{\max}^2 m^3}{\mu^2\varepsilon^2})$, where $p_{\max}$ is the maximum dimension of the semidefinite cuts returned by the oracle.

From the practical point of view we applied our algorithm to the semidefinite programs arising from combinatorial optimization. In particular, we considered the semidefinite relaxation on the max-cut problem. Reformulation of the semidefinite programming as an eigenvalue optimization is well known. We applied the ACCPM, as a nonsmooth optimization technique, to minimize the maximum eigenvalue of an affine combination of symmetric matrices. We transformed the optimization problem into a feasibility problem such that the set of localization contains the minimizer of the max-eigenvalue function. The set of localization in this case is bounded from below by a set of subgradients of the max-eigenvalue function, and from above by an upper bound cut.

The weighted analytic center of the set of localization when a weight equal to the dimension of the current cut matrix is set on the upper bound

cut was defined. We introduced the weighted analytic center because in practice, as the number of subgradient cuts increases the analytic center approaches to the upper bound and therefore we lose the centrality. The weight on the upper bound cut, which is somewhat arbitrary, pushes back the analytic center to the center of the set. We derived the first order optimality conditions for the weighted analytic center and presented its computational algorithms in primal and dual settings.

In primal algorithm, we decomposed the Newton direction into three categories; (a) the semidefinite direction, for the primal matrix variables corresponding to the semidefinite cuts, (b) the linear direction, for the primal vector variables corresponding to the linear cuts, and (c) the direction corresponding to the upper bound cut with weight. This devision is done for computational reasons. In theory, the three groups are combined in the semidefinite cut. However, it is practically very expensive and unnecessary to treat a linear cut as a semidefinite cut.

At each iteration if the max-eigenvalue function is differentiable at the current analytic center the oracle returns the gradient of the function as a single linear cut. Otherwise the oracle returns a set of subgradients, which form a p-dimensional semidefinite cut. We updated the set of localization by adding the cut to the current working set. In practice, a single or semidefinite cut is either shallow or deep. These issues are crucial for recovering feasibility in implementation of our algorithm. We made a distinction between linear and semidefinite cuts in both primal and dual cases and between shallow and deep cuts in the dual algorithm.

The attractiveness of the primal algorithm is in recovering the feasibility after adding a cut. There is basically no needs to determine whether a cut is deep and to take any special procedure for such a case. We discussed two cases for recovering feasibility; (1) adding a p-dimensional semidefinite cuts,

(2) adding a single linear cut. In each case we derived a strictly feasible point in the updated primal set as a "warm start". We illustrated by the numerical results that starting from such a point the Newton algorithm finds the next analytic center in a very efficient way. Such a result was expected as it was established by a theoretical bound in Theorem 14.

Although, the dual algorithm is capable of exploiting sparsity and it is the preferred algorithm when dealing with the sparse problems, however, is not as efficient as the primal one in recovering feasibility. At this point there is basically no efficient way to recover feasibility after adding a deep cut in dual algorithm. We dealt with this issue by applying a path-following algorithm: when the oracle returns a deep linear or a deep semidefinite cut, we place it at the center and use the procedure described in Section 4.1 to obtain a strictly feasible point and recover the centrality. Then we move the cut to the new center and repeat this procedure until we obtain a strictly feasible point for the updated dual set. At each recentering procedure the Gram matrix is computed several times, which is the main cause of slowing down the method. This makes the dual algorithm not a favored one at this point. The problem of restoration and make use of the full capacity of the dual algorithm is a potential future research problem.

An advantage of the ACCPM is that after solving the minmax optimization problem, which is equal to the dual problem of semidefinite relaxation of the max-cut problem, a primal feasible matrix very close to the optimal can be computed. We provided this course of action in Section 5.6.

Introducing the semidefinite cuts into the analytic center cutting plane method in this thesis, opened a door to the new applications for this method. The theoretical results and the complexity bounds established in Chapter 4 demonstrate that the ACCPM is as efficient in nonpolyhedral cone as it is in linear programming. The ACCPM can now handle semidefinite cuts as good

as it does the linear or the quadratic cuts. The numerical results however, presented in Section 5.7, illustrate that the ACCPM can efficiently be applied only to moderate-size semidefinite programs. The reason that the ACCPM is not efficient when dealing with large-scale problems is basically due to the computational difficulties of semidefinite programming.

As we mentioned in Chapter 1, the main purpose of this thesis was to integrate the ACCPM with the semidefinite cuts and not to design an algorithm for large-scale semidefinite programs. However, with the aptitude of the ACCPM in recovering centrality after adding cuts, we strongly believe that a combination of the algorithms presented in this thesis with a branch and cut or a branch and price algorithm will efficiently solve the moderate-size max-cut problem to optimality with an integer solution. This will be the line of our future research.

Bibliography

[1] F. ALIZADEH, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM Journal on Optimization, 5 (1995), pp. 13–51.

[2] F. ALIZADEH, J-P. A. HAEBERLY, AND M. L. OVERTON, *A new primal-dual interior point methods for semidefinite programming*, In J. G. Lewis, editor, Proceedings of the fifth SIAM Conference on Applied Linear Algebra, (1994), pp. 113–117.

[3] F. ALIZADEH, J-P. A. HAEBERLY, AND M. L. OVERTON, *Complementarity and nondegeneracy in semidefinite programming*, Mathematical Programming, 77 (1997), pp. 111–128.

[4] F. ALIZADEH, J-P. A. HAEBERLY, AND M. L. OVERTON, *Primal-dual interior point methods for semidefinite programming: Convergence rate, stability and numerical results*, SIAM Journal on Optimization, 8 (1998), pp. 746–768.

[5] J. ALLWRIGHT, *Positive semidefinite matrices: characterization via conial hulls and least-squares solution of a matrix equation*, SIAM Journal on Control and Optimization, 26 (1988), pp. 537–556.

[6] D. S. Atkinson and P. M. Vaidya, *A cutting plane algorithm for convex programming that uses analytic centers*, Mathematical Programming, series B, 69 (1995), pp. 1–43.

[7] R. Bellman and K. Fan, *On systems of linear inequalities in hermitian matrix variables in convexity*, V. L. Klee, ed., Proceedings of Symposia in Pure Mathematics, American Mathematical Society, 7 (1963), pp. 1–11.

[8] J. S. Benson, Y. Ye, and X. Zhang, *Solving large scale sparse semidefinite programming for combinatorial optimization*, SIAM Journal on Optimization, 10(2) (2000), pp. 443–461

[9] S. K. Chua, K. C. Toh, and G. Y. Zhao, *An analytic center cutting plane method with deep cuts for semidefinite feasibility problems*, Working paper, Department of Mathematics, National University of Singapore, (2002).

[10] B. Craven and B. Mond, *Linear programming with matrix variable*, Linear Algebra and Appl., 38 (1981), pp. 73–80.

[11] C. Delorme and S. Poljak, *Laplacian eigenvalues and the maximum cut problem*, Mathematical Programming, 62 (1993) pp. 557–574.

[12] R. Fletcher, *Semidefinite matrix constraint in optimization*, SIAM Journal on Control and Optimization, 23 (1985), pp. 493–513.

[13] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for Maximum Cut and satisfiability problems using semidefinite programming*, Journal of ACM, 42 (1995) pp. 1115–1145.

[14] J.-L. Goffin, J. Gondzio, R. Sarkissian, and J.-P. Vial, *Solving nonlinear multicommodity flow problems by the analytic center cutting plane method*, Mathematical programming, 76(1) (1997), pp. 131–154.

[15] J.-L. GOFFIN, A. HAURIE, AND J.-P. VIAL, *Decomposition and non-differentiable optimization with the projective algorithm*, Management Science, 38 (1992), pp. 284–302.

[16] J.-L. GOFFIN, Z.-Q. LUO, AND Y. YE, *Complexity analysis of an interior cutting plane for convex feasibility problems*, SIAM Journal on Optimization, 6 (1996), pp. 638–652.

[17] J.-L. GOFFIN AND J.-P. VIAL, *Shallow, deep and very deep cuts in the analytic center cutting plane methods*, Mathematical programming, 84 (1999), pp. 89–103.

[18] J.-L. GOFFIN AND J.-P. VIAL, *Multiple cuts in the analytic center cutting plane methods*, SIAM Journal on Optimization, 11 (2000), pp. 266–288.

[19] J. GONDZIO, *Warm start of the primal–dual method applied in the cutting plane scheme*, Mathematical programming, 83 (1998), pp. 125–143.

[20] J. GONDZIO AND R. KOUWENBERG, *High-performance computing for asset-liability management*, Operations Research, 49(6) (2001), pp. 879–891.

[21] C. HELMBERG, *Semidefinite Programming for Combinatorial Optimization*, Konrad-Zuse-Zentrum, fur Informationstechnik, Berlin (October 2000).

[22] C. HELMBERG, *A Cutting Plane Algorithm for Large Scale Semidefinite Relaxations*, ZIB-Report 01-26, Konrad-Zuse-Zentrum Berlin, Takustrasse 7, D-14195 Berlin, Germany, October 2001

[23] C. HELMBERG AND F. RENDL, *A spectral bundle method for semidefinite programming*, SIAM Journal on Optimization, 10(3) (2000) pp. 673–696.

[24] C. HELMBERG AND F. RENDL, *An interior-point method for semidefinite programming*, SIAM Journal on Optimization, 6(2) (1996) pp. 342–361.

[25] F. JARRE, *An interior point method for minimizing the maximum eigenvalue of a linear combination of matrices*, SIAM Journal on Control and Optimization, 31 (1993), pp. 1360–1377.

[26] K. C. KIWIEL, *Proximity control in bundle methods for convex nondifferentiable minimization*, Mathematical Programming, 46 (1990), pp. 105–122.

[27] M. KOJIMA, S. KOJIMA, AND S. HARA, *Linear algebra for semidefinite programming*, Technical report 1004, Dept. of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan, 1997. Linear matrix inequalities and positive semidefinite programming (Japanese) (Kyoto, 1996)

[28] M. KOJIMA, S. SHINDOH, AND S. HARA, *Interior point methods for the monotone semidefinite linear complementarity problem in symmetric matrices*, SIAM Journal on Optimization, 7(1) (1997), pp. 86–125.

[29] K. KRISHNAN, *Linear programming approaches to semidefinite programming problems*, PhD thesis, Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, July 2002.

[30] K. KRISHNAN AND J. MITCHELL, *A linear programming approach to semidefinite programming*, Journal of Combinatorial Optimization, to appear.

[31] M. LAURENT AND S. POLJAK, *On a positive semidefinite relaxation of the cut polytope*, Linear Algebra and its Applications, 223/224 (1995) pp. 439–461.

[32] L. Lovász, *On the shannon capacity of a graph*, IEEE Transactions on Information Theory, IT-25(1) (1979) pp. 1–7.

[33] Z.-Q. Luo and J. Sun, *An analytic center based on column generation algorithm for convex quadratic feasibility problems*, SIAM Journal on Optimization, 9(1) (1998) pp. 217–235.

[34] H.J. Lüthi and B Büeler, *The Analytic Center Quadratic Cut Method (ACQCM) for Strongly Monotone Variational Inequality Problems*, SIAM Journal on Optimization, 10(2) (2000), pp. 415–426.

[35] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, San Diego, California 92101-4495, 1979.

[36] J. E. Mitchell and M.J. Todd, *Solving combinatorial optimization problems using Karmarkar's algorithm*, Mathematical Programming, 56 (1992), pp. 245–284.

[37] J. E. Mitchell, *Computational experience with an interior point cutting plane algorithm*, SIAM Journal on Optimization, 10(4) (2000), pp. 1212–1227.

[38] J. E. Mitchell and B. Borchers, *Solving linear ordering problemswith a combined interior point/simplex cutting plane algorithm*, in H. L. Frenk et al., editor, High-Performance Optimization, chapter 14, pages 348–366. Kluwer Academic Publishers, The Netherland 2000.

[39] J. E. Mitchell and S. Ramaswamy, *A long-step cutting plane algorithm for linear and convex programming*, Annals of Operations Research, 99 (2000), pp. 95–122.

[40] R. D. C. MONTEIRO, *Primal-dual path following algorithms for semidefinite programming*, SIAM Journal on Optimization, 7(3) (1997), pp. 663–678.

[41] Y. NESTEROV, *Cutting plane algorithms from analytic centers: efficiency estimates*, Mathematical Programming, series B, 69 (1995), pp. 149–176.

[42] Y. NESTEROV AND A. NEMIROVSKI, *Polynomial barrier methods in convex programming*, Ékonom. i Mat. Metody, 24(6) (1988), pp. 1084–1091.

[43] Y. NESTEROV AND A. NEMIROVSKI, *Optimization over positive semidefinite matrices: Mathematical background and user's manual*, USSR Acad. Sci. Centr. Econ. & Math. Inst., 32 Krasikova St., Moscow 117418 USSR, 1990.

[44] Y. NESTEROV AND A. NEMIROVSKI, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.

[45] Y. NESTEROV AND M. J. TODD, *Self-scaled barriers and interior point methods for convex programming*, Mathematics of Operations Research, 22(1) (1997), pp. 1–42.

[46] Y. NESTEROV AND M. J. TODD, *Primal-dual interior point methods for self-scaled cones*, SIAM Journal on Optimization, 8 (1998), pp. 324–364.

[47] M. L. OVERTON, *Large-Scale Optimization of Eigenvalues*, SIAM Journal on Optimization, 2(1) (1992), pp. 88–120.

[48] I. S. PRESSMAN AND S. JIBRIN, *A weighted analytic center for linear matrix inequalities*, Journal of Inequalities in Pure and Applied Mathematics, 2(3), (2001), Article 29.

[49] F. RENDL, R. J. VANDERBEI, AND H. WOLKOWICZ, *Max-min eigenvalue problems, primal-dual interior point algorithms, and trust region subproblems*, Optim. Methods Softw., 5 (1995), pp. 1–16.

[50] C. ROOS, T. TERLAKY, AND J.-P. VIAL, *Theory and Algorithms for Linear Optimization*, John Wiley and Sons Ltd., Baffins Lane, Chichester, England, (1997).

[51] A. SHAPIRO, *Extermal problems on the set of nonnegative definite matrices*, Linear Algebra and Appl. 67 (1985), pp. 7–18.

[52] F. SHARIFI MOKHTARIAN AND J.-L. GOFFIN, *An analytic center quadratic cut method for the convex quadratic feasibility problem*, GERAD research report, G00 April 2000.

[53] G. SONNEVEND, *New algorithms in convex programming based on a notation of center and on rational extrapolations*, International Series of Numerical Mathematics, Birkhauser Verlag, Basel, Switzerland, 84 (1988), pp. 311–327.

[54] J. SUN, K. C. TOH, AND G. Y. ZHAO, *An analytic center cutting plane method for semidefinite feasibility problems*, Mathematics of Operations Research, to appear.

[55] C. M. THEOBALD , *An inequality for the trace of the product of two symmetric matrices*, Proc. Cambridge Philos. Soc., 77 (1975), pp. 265–267.

[56] K. C. TOH, *Primal-dual path-following algorithms for determinant maximization problems with linear matrix inequalities*, Computational Optimization and Applications, 14 (1999), pp. 309–330.

[57] K. C. Toh, G. Y. Zhao, and J. Sun, *A multiple-cut analytic center cutting plane method for semidefinite feasibility problems*, SIAM Journal on Optimization, to appear.

[58] L. Vandenberghe and S. Boyd, *Semidefinite Programming*, SIAM Review, 38(1) (1996), pp. 49–95.

[59] L. Vandenberghe, S. Boyd and S. P. Wu, *Determinant maximization with linear matrix inequality constraint*, SIAM Journal on Matrix Analysis and Applications, 19(2) (1998), pp. 499–533.

[60] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications* Kluwer Academic Publishers, Norwell, Massachusetts, (2000).

[61] Y. Ye, *A potential reduction algorithm allowing column generation*, SIAM Journal on Optimization, 2 (1992), pp. 7–20.

[62] Y. Ye, *Interior Point Algorithms, Theory and Analysis* John Wiley Inc., New York, NY, (1997).

[63] Y. Ye, *Complexity analysis of the analytic center cutting plane method that uses multiple cuts*, Mathematical Programming, 78 (1997), pp. 85–104.