Selective Gram-Schmidt orthonormalization for conic cutting surface algorithms[1]

John E. Mitchell

Mathematical Sciences

Rensselaer Polytechnic Institute

Troy, NY 12180

mitchj@rpi.edu

http://www.rpi.edu/~mitchj

Vasile L Basescu

basesv@verizon.net

November 29, 2006

Abstract

It is not straightforward to find a new feasible solution when several conic constraints are added to a conic optimization problem. Examples of conic constraints include semidefinite constraints and second order cone constraints. In this paper, a method to slightly modify the constraints is proposed. Because of this modification, a simple procedure to generate strictly feasible points in both the primal and dual spaces can be defined. A second benefit of the modification is an improvement in the complexity analysis of conic cutting surface algorithms. Complexity results for conic cutting surface algorithms proved to date have depended on a condition number of the added constraints. The proposed modification of the constraints leads to a stronger result, with the convergence of the resulting algorithm not dependent on the condition number.

Keywords: Semidefinite programming, conic programming, column generation, cutting plane methods.

[1]Research supported in part by NSF grant number DMS–0317323

1 Introduction

Given an m-dimensional Hilbert space $(Y, \langle \cdot, \cdot \rangle_Y)$ and a nonempty convex set $C \subseteq Y$, the *convex feasibility problem* is to find a point in the set. Many algorithms for this problem generate a point $\bar{y} \in Y$ and determine whether the point is in C. If the point is not in C, extra information is provided which can be used to modify the point. We assume that if C is nonempty then it contains a ball of radius ε. In particular, there exists $\hat{y} \in C$ such that $\hat{y} + \varepsilon u \in C$ for any $u \in Y$ of norm one.

In this paper, we consider setting up a conic programming relaxation in order to determine the point $\bar{y}$. If $\bar{y}$ is not in C, we assume the extra information available takes the form of conic constraints that separate $\bar{y}$ from C. Thus, we generate a sequence of relaxations of the form

$$\text{Find } y \text{ satisfying } A^* y \preceq_K c \tag{1}$$

where K is a cone in an n-dimensional Hilbert space $(X, \langle \cdot, \cdot \rangle_X)$, the linear operator $A : X \to Y$ has adjoint A^*, the vector $c \in X$, and $u \preceq_K v$ if and only if $v - u \in K$. The point $\bar{y}$ is chosen to be an approximate analytic center of the system given in (1). The effect of the cut is illustrated in Figures 1 and 2. Note that K may be the cartesian product of a number of lower dimensional convex cones. Note that a linear programming relaxation is of this form, with $K = \mathbb{R}^n_+$, the cartesian product of n nonnegative half-lines requiring that the slack variable in the linear constraint be nonnegative.

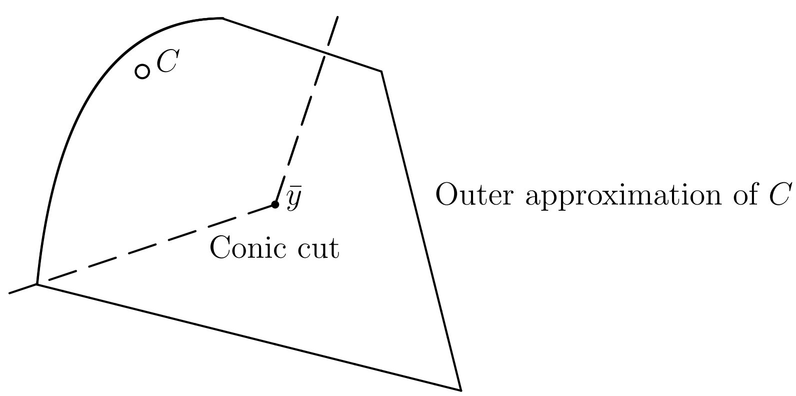

Figure 1: Adding a conic cut at the current approximate analytic center $\bar{y}$ of the current outer approximation to C.

For more information on cone programming, see [9] and [13]. Cones of particular interest are the cone of positive semidefinite matrices and the second order cone of

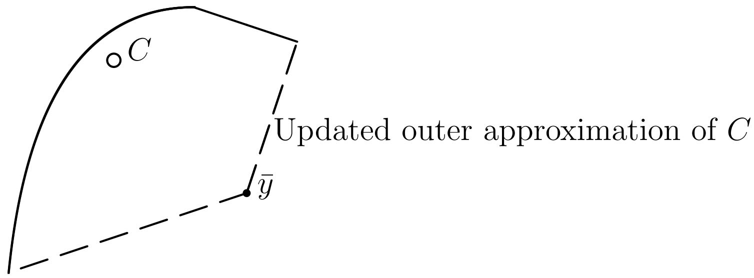

Figure 2: Updated outer approximation after adding a single conic cut

vectors $(x_0, x_1, \ldots, x_n)$ satisfying $x_0 \geq \sqrt{\sum_{i=1}^{n} x_i^2}$. These are examples of closed convex cones which possess strongly nondegenerate intrinsically self concordant barrier functionals.

Conic programming problems arise in many practical problems, with positive semidefiniteness and restrictions on norms of vectors being common requirements. For problems where the number of constraints is large compared to the number of variables, a cutting plane approach can be useful. Cutting plane methods can also be employed to replace a large conic constraint with several smaller constraints; for example, a semidefinite constraint can be replaced by linear and/or second order cone constraints. For more information on cutting plane methods for semidefinite programming and second order cone programming, see [1, 2, 3, 6, 10, 11, 12, 14, 15, 16].

We've previously described a selective orthonormalization procedure for polyhedral relaxations of the convex feasibility problem [7]. In this paper, we generalize the procedure to conic relaxations. The method for modifying the constraints is presented in §2, and the exploitation of this modification to regain strict feasibility is demonstrated in §3. Analysis of the resulting cutting surface algorithm requires potential functions, which are introduced in §4. Convergence to a new approximate analytic center is shown in §5 and convergence of the cutting plane algorithm to a point in C is shown in §6. The specialization of the algorithm to the cases of semidefinite programming and second order cone programming is the subject of §7.

Notation and scaling

Given two points u and v in a Hilbert space $(X, \langle .,. \rangle_X)$, and given a positive definite matrix M of the appropriate size, define $\langle u, v\rangle_M := \langle u, Mv\rangle_X$. Further, $||u||_M^2 := \langle u, Mu\rangle$. Each cone $K \subseteq X$ with barrier f has a fixed vector e in its interior with norm one, and we will use the local inner product between two points u and v in X, so $\langle u, v\rangle := \langle u, v\rangle_{H(e)} = \langle u, H(e)v\rangle_X$, where $H(e)$ is the Hessian of f evaluated at e. When we denote a cone as K_i, we denote the corresponding fixed vector in its interior as e_i. Let x be a point in the interior of K; then $\langle u, v\rangle_x := \langle u, v\rangle_{H(x)} = \langle u, H(x)v\rangle$. Define

$$\mathcal{M} := \{y \in Y : y + \varepsilon u \in C \ \forall u \in Y \text{ with } ||u|| = 1\}$$

and

$$\mathcal{M}_s := \{s : s = c - A^* y \text{ with } y \in \mathcal{M}\}.$$

2 A selective Gram-Schmidt orthonormalization procedure

We assume our current relaxation takes the form

$$A^* y \preceq_K c$$

and we have a point $\bar{y}$ that satisfies these constraints strictly. Define $\bar{s} := c - A^*\bar{y}$. We assume the operator A is surjective. The cone K has a barrier function f, with Hessian $H(s)$, well-defined for s in the interior of K. We define the positive definite matrix

$$\Gamma := (AH(\bar{s})A^*)^{-1}. \tag{2}$$

This matrix is used in the restart directions defined below to scale the directions. It is used because of its relationship with the Dikin ellipsoid, and hence it allows some control over the change in potential function values.

An oracle returns q conic constraints

$$A_i^* y \preceq_{K_i} c_i, \qquad i = 1, \ldots, q$$

which separate $\bar{y}$ from C. We want to modify the constraints so that we can find primal and dual strictly feasible iterates easily. In order to find a feasible solution, we will weaken the cuts if necessary to make them central (this is a standard assumption in the analysis of interior point and ellipsoid cutting plane algorithms). Thus, the cuts have the form

$$A_i^* y \preceq_{K_i} A_i^* \bar{y}, \qquad i = 1, \ldots, q.$$

In order to regain a strictly feasible solution, it is necessary to find a direction d satisfying

$$A_i^* d \prec_{K_i} 0, \qquad i = 1, \ldots, q.$$

The purpose of the selective orthonormalization procedure is to modify the added constraints in order to make it straightforward to find such a direction d.

Let's introduce the following notation. For all $a \in X, b \in Y$, we define the operator $ab^T : Y \to X$ by:

$$(ab^T)u = \langle b, u\rangle a \tag{3}$$

for all $u \in Y$. The conjugate (transpose) of this operator is given by:

$$(ab^T)^* = ba^T. \tag{4}$$

The proof is immediate; in particular for all $u \in Y, v \in X$:

$$\langle v, (ab^T)u\rangle = \langle v, \langle b, u\rangle a\rangle = \langle\langle v, a\rangle b, u\rangle = \langle (ba^T)v, u\rangle. \tag{5}$$

We first need the following assumption.

Assumption 1 *$A_i x_i \neq 0$, for all $x_i \in int(K_i)$.*

A justification for Assumption 1 is contained in the following theorem of the alternative. In particular, the assumption must hold if the set of feasible solutions to $A_i^* d \prec_{K_i} 0$ is to be nonempty. (Self-dual cones are defined formally in Appendix A.)

Theorem 1 *Let $\bar{K}$ be a self-dual cone with interior point x. Let A be an operator defining a constraint $A^* d \preceq_{\bar{K}} 0$. If $Ax = 0$ then there is no direction d satisfying $A^* d \preceq_{\bar{K}} 0$ with $A^* d \neq 0$.*

Proof: Let $v = A^* d \in \bar{K}$. Assume $v \neq 0$. Since x is in the interior of $\bar{K}$, the point $x - \alpha v$ is also in $\bar{K}$ for α sufficiently small and positive. The inner product $\langle v, x - \alpha v\rangle = -\alpha\langle v, v\rangle + \langle d, Ax\rangle = -\alpha\langle v, v\rangle < 0$. This contradicts the assumption that $\bar{K}$ is self-dual. Therefore, we must have $v = 0$. □

It should be noted that the assumption is only necessary for cuts that are added centrally. For example, Oskoorouchi and Goffin [11] initialize their cutting plane algorithm for second order cone programming (SOCP) by requiring that y satisfy a ball constraint $||y|| \leq M$ for some scalar M. This constraint can be written in the form $A^* y \preceq_K c$ with K the second order cone and $Ae = 0$ where $e = [1; 0, \ldots, 0]^T$. Assumption 1 holds for any SOCP cut that has the first column of A_i not equal to the zero vector.

If the added constraint is an SDP conic constraint, of the form $\sum_{j=1}^m A_j y_j \preceq \sum_{j=1}^m A_j \bar{y}_j$, Assumption 1 requires that the trace of at least one of the matrices A_j be nonzero: consider $x = I$. It is clear that this is a necessary requirement for the existence of a vector d with $\sum_{j=1}^m A_j d_j$ positive definite.

The following lemma shows how a cutting surface can be modified without cutting off any feasible points. The modification takes a nonnegative combination of the original constraint with a linearization of another valid constraint.

Lemma 1 *Let e_l be the selected interior points for K_l, $k = 1, 2$. Let A_l be linear operators from $\mathbb{R}^{p_l} \to \mathbb{R}^m$ for $l = 1, 2$. Let $\bar{A}_1 = A_1 + \lambda A_2 e_2 e_1^T$ for some $\lambda \geq 0$. If d satisfies $A_l^* d \preceq_{K_l} 0$ for $l = 1, 2$ then $\bar{A}_1^* d \preceq_{K_1} 0$.*

Proof: We have

$$\bar{A}_1^* d = A_1^* d + \lambda e_1 e_2^T A_2^* d \in K_1$$

since $A_1^* d \preceq_{K_1} 0$, $A_2^* d \preceq_{K_2} 0$ so $\langle e_2, \bar{A}_2^* d \rangle \leq 0$ from the self-duality of K_2, $\lambda \geq 0$, and $e_1 \succeq_{K_1} 0$. □

Note that this lemma applies even if $K_1 = K_2$ and $A_1 = A_2$; that is, the lemma can also be used to modify a single constraint so that it is easier to restart. The Selective Gram-Schmidt Orthonormalization process that we define in Figure 3 is designed to modify the constraints using a construction as in Lemma 1, with the purpose of creating a set of constraints that satisfies the criterion given in Theorem 2 below. We show in Section 3 how the algorithm can be restarted if such a criterion is satisfied.

The criterion we require of the constraints is that $\bar{A}_i^* \Gamma \bar{A}_j e_j$ be in cone K_i for each pair i and j. The following lemma shows that for a given ordered pair of constraints, a multiplier λ can be chosen using a line search so that the modified constraint matrices will satisfy the criterion, at least for this ordered pair.

Lemma 2 *Let e_l be the selected interior points for K_l, $l = 1, 2$. Let A_l be linear operators from $\mathbb{R}^{p_l} \to \mathbb{R}^m$ for $l = 1, 2$. Let M be a symmetric positive definite operator from $\mathbb{R}^m \to \mathbb{R}^m$. Let*

$$\bar{\lambda} := \max\{0, \min\{\lambda : \lambda e_1 + \frac{1}{||A_2 e_2||_M^2} A_1^* M A_2 e_2 \succeq_{K_i} 0\}\}.$$

Let $\bar{A}_1 = A_1 + \tilde{\lambda} A_2 e_2 e_1^T$ with $\tilde{\lambda} \geq \bar{\lambda}$. Then $\bar{A}_1^* M A_2 e_2 \succeq_{K_1} 0$.

Proof: We have

$$\begin{aligned} \bar{A}_1^* M A_2 e_2 &= (A_1^* + e_1 e_2^T \tilde{\lambda} A_2^*) M A_2 e_2 \\ &= ||A_2 e_2||_M^2 (\tilde{\lambda} e_1 + \frac{1}{||A_2 e_2||_M^2} A_1^* M A_2 e_2) \\ &\succeq_{K_1} 0. \end{aligned}$$

$\square$

Note that the conclusion of the lemma still holds even if $K_2 = K_1$ and $A_2 = A_1$, as we show explicitly in the next lemma.

Lemma 3 *Let e_l be the selected interior point for K_l. Let A_l be a linear operator from $\mathbb{R}^{p_l} \to \mathbb{R}^m$. Let M be a symmetric positive definite operator from $\mathbb{R}^m \to \mathbb{R}^m$. Let*

$$\bar{\lambda} := \max\{0, \min\{\lambda : \lambda e_l + \frac{1}{||A_l e_l||_M^2} A_l^* M A_l e_l \succeq_{K_i} 0\}\}.$$

Let $\bar{A}_l = A_l + \tilde{\lambda} A_l e_l e_l^T$ where $\tilde{\lambda} \geq \bar{\lambda}$. Then $\bar{A}_l^* M \bar{A}_l e_l \succeq_{K_l} 0$.

Proof: We have

$$\begin{aligned}
\bar{A}_l^* M \bar{A}_l e_l &= (A_l^* + e_l e_l^T \tilde{\lambda} A_l^*) M (A_l + \tilde{\lambda} A_l e_l e_l^T) e_l \\
&= (1 + \tilde{\lambda}\langle e_l, e_l\rangle)||A_l e_l||_M^2 (\tilde{\lambda} e_l + \frac{1}{||A_l e_l||_M^2} A_l^* M A_l e_l) \\
&\succeq_{K_l} 0
\end{aligned}$$

as required. $\square$

The Selective Gram-Schmidt Orthonormalization algorithm is contained in Figure 3. The algorithm modifies a linear operator based on its interaction with another operator. If all the added constraints are single linear constraints, then each e_i is just a positive scalar, and the process is a selective orthogonalization procedure where an operator is only modified if the normals of the corresponding constraints make negative inner product. In this case, the check at step 3 of the subroutine determines whether the constraints make nonnegative inner product, and if not then the standard Gram-Schmidt update is performed in steps 4 and 5. Mitchell and Ramaswamy [7] explored this case.

The proof of the correctness of the algorithm repeatedly uses the following technical lemma.

Lemma 4 *Let e_l be the selected interior points for K_l, $l = 1, 2, 3$. Let A_l be linear operators from $\mathbb{R}^{p_l} \to \mathbb{R}^m$ for $l = 1, 2, 3$. Let $\bar{\lambda} \geq 0$. Let $\bar{A}_1 = A_1 + \bar{\lambda} A_2 e_2 e_1^T$.*

1. *If $A_3^* \Gamma A_2 e_2 \succeq_{K_3} 0$ then $A_3^* \Gamma \bar{A}_1 e_1 \succeq_{K_3} A_3^* \Gamma A_1 e_1$.*

2. *If $A_2^* \Gamma A_3 e_2 \succeq_{K_2} 0$ then $\bar{A}_1^* \Gamma A_3 e_3 \succeq_{K_1} A_1^* \Gamma A_3 e_3$.*

3. *$\bar{A}_1^* \Gamma A_2 e_2 \succeq_{K_1} A_1^* \Gamma A_2 e_2$.*

Figure 3: Selective Gram-Schmidt Orthonormalization

Main algorithm:

1. Take ν and ω to be scalars between 0 and 1.
2. Set $\bar{A}_i = A_i$ for $i = 1, \ldots, q$.
3. Let *first=1, last=q*. Call *SelOrth(first,last)*.
4. Let *first=q, last=1*. Call *SelOrth(first,last)*.
5. For $i = 1, \ldots, q$:
   (a) If $\bar{A}_i^* \bar{A}_i e_i \not\succeq_{K_i} 0$, set $\bar{\lambda} := \min\{\lambda : \lambda e_i + \frac{1}{||\bar{A}_i e_i||} \bar{A}_i^* \bar{A}_i e_i \succeq_{K_i} 0\}$. Update $\bar{A}_i \leftarrow \bar{A}_i + \bar{\lambda} \bar{A}_i e_i e_i^T$ and renormalize so that $||\bar{A}_i|| = 1$.
   (b) If $\bar{A}_i^* \bar{A}_i e_i \not\succeq_{K_i} \omega e_i$, update $\bar{A}_i \leftarrow (1 - \sqrt{\omega}) \bar{A}_i + \frac{\sqrt{\omega}}{||\bar{A}_i e_i||} \bar{A}_i e_i e_i^T$. Renormalize so that $||\bar{A}_i|| = 1$.
   (c) Let $\eta_i = \sqrt{e_i^* \bar{A}_i^* \Gamma \bar{A}_i e_i}$. If $\bar{A}_i^* \Gamma \bar{A}_i e_i \not\succeq_{K_i} \nu \eta_i^2 e_i$, update $\bar{A}_i \leftarrow (1 - \nu) \bar{A}_i + \nu \bar{A}_i e_i e_i^T$. Renormalize so that $||\bar{A}_i|| = 1$.
6. End.

Subroutine *SelOrth(first,last)*:

For $i =$*first*$,\ldots,$*last*:
    For $j =$*first*$,\ldots,i$:
        If $\bar{A}_i^* \Gamma \bar{A}_j e_j \not\succeq_{K_i} 0$ then
            $\lambda_{ij} := \min\{\lambda : \lambda e_i + \frac{1}{||\bar{A}_j e_j||_\Gamma^2} \bar{A}_i^* \Gamma \bar{A}_j e_j \succeq_{K_i} 0\}$.
            $\bar{A}_i \leftarrow \bar{A}_i + \lambda_{ij} \bar{A}_j e_j e_i^T$
        End if.
    End for.
    Normalize $\bar{A}_i$, so $||\bar{A}_i|| = 1$.
End for, and Return.

4. *If* $\bar{A}_1^* \Gamma A_2 e_2 \succeq_{K_1} 0$ *then* $\bar{A}_1^* \Gamma \bar{A}_1 e_1 \succeq_{K_1} A_1^* \Gamma A_1 e_1$.

5. *If* $A_2^* \Gamma A_2 e_2 \succeq_{K_2} 0$ *then* $A_2^* \Gamma \bar{A}_1 e_1 \succeq_{K_2} A_2^* \Gamma A_1 e_1$.

Proof: We prove the first two parts.

1. We have
$$A_3^* \Gamma \bar{A}_1 e_1 = A_3^* \Gamma A_1 e_1 + \lambda A_3^* \Gamma A_2 e_2 e_1^T e_1$$
and the result follows from the conditions of the lemma.

2. Similarly,
$$\bar{A}_1^* \Gamma A_3 e_3 = A_1^* \Gamma A_3 e_3 + \lambda e_1 e_2^T A_2^* \Gamma A_3 e_3$$
leading to the required result.

The proofs of the remaining parts are similar. □

The matrix $\bar{A}_i$ is updated using $\bar{A}_j$, which affects the value of $\bar{A}_i^* \Gamma \bar{A}_j e_j$. It also affects the value of $\bar{A}_i^* \Gamma \bar{A}_k e_k$ for all other values of k, and the proof that this term is in K_i will require induction.

Proposition 1 *After the first call to the subroutine* SelOrth(.,.), *we have* $\bar{A}_i^* \Gamma \bar{A}_j e_j \succeq_{K_i} 0$, *for* $1 \le j \le i \le q$.

Proof: We use induction.

Base cases: From Lemmas 2 and 4, we have $\bar{A}_1^* \Gamma \bar{A}_1 e_1 \in K_1$, $\bar{A}_2^* \Gamma \bar{A}_1 e_1 \in K_2$, $\bar{A}_2^* \Gamma \bar{A}_2 e_2 \in K_2$, and $\bar{A}_3^* \Gamma \bar{A}_1 e_1 \in K_3$ prior to the update given in step 5 of subroutine *SelOrth* with $i = 3$ and $j = 2$.

Inductive step: Given i and j with $i \ge j$. Assume $\bar{A}_k^* \Gamma \bar{A}_l e_l \succeq_{K_k} 0$ for $k \le i$ and $l < j$ before the the update given in step 5 of subroutine *SelOrth*. After the update, we have $\bar{A}_i^* \Gamma \bar{A}_j e_j \succeq_{K_i} 0$ from Lemma 2 and $\bar{A}_i^* \Gamma \bar{A}_p e_p \succeq_{K_i} 0$ for $p < j$ from Lemma 4.

□

Now we can analyze the effect of the second call to the subroutine *SelOrth*.

Theorem 2 *After the second call to the subroutine* SelOrth, *we have* $\bar{A}_i^* \Gamma \bar{A}_j e_j \succeq_{K_i} 0, \forall i, j = 1, \ldots, q$.

Proof: From Proposition 1 and Lemma 4, we must have $\bar{A}_i^* \Gamma \bar{A}_j e_j \succeq_{K_i} 0$ for $i \geq j$ when the algorithm terminates. The proof that $\bar{A}_i^* \Gamma \bar{A}_j e_j \succeq_{K_i} 0$ for $1 \leq i < j \leq q$ is then very similar to the proof of Proposition 1. □

The updates in Step 5 leave $\bar{A}_j^* \Gamma \bar{A}_k^* e_k \succeq_{K_j} 0$, $\forall j, k = 1, \ldots, q$ from Lemma 4. The update in step 5c of the algorithm enables the construction of a strictly feasible dual iterate, as shown in §3. The update in step 5b is needed in §6 to prove global convergence of the cutting plane algorithm. Note first that it follows from Lemma 3 that the update in Step 5a results in $\bar{A}_i^* \bar{A}_i e_i \succeq_{K_i} 0$. We now show that Steps 5b and 5c result in $\bar{A}_i^* \bar{A}_i e_i \succeq \omega e_i$ and $\bar{A}_i^* \Gamma \bar{A}_i e_i \succeq_{K_i} \nu \eta_i^2 e_i$ for $i = 1, \ldots, q$.

Lemma 5 *Let* $\bar{A}_i = (1 - \sqrt{\omega}) A_i + \frac{\sqrt{\omega}}{||A_i e_i||} A_i e_i e_i^T$. *Assume* $0 \leq \sqrt{\omega} \leq 1$, $A_i^* A_i e_i \succeq_{K_i} 0$, $||e_i|| = 1$, *and* $||A_i|| = 1$. *Then* $||\bar{A}_i|| \leq 1$ *and* $\bar{A}_i^* \bar{A}_i e_i \succeq_{K_i} \omega e_i$.

Proof: The proof is by direct calculation. Note that

$$\bar{A}_i^* \bar{A}_i e_i = ((1 - \sqrt{\omega})^2 + \frac{\sqrt{\omega}(1 - \sqrt{\omega})}{||A_i e_i||}) A_i^* A_i e_i + (1 - \sqrt{\omega})\sqrt{\omega} ||A_i e_i|| e_i + \omega e_i \succeq_{K_i} \omega e_i$$

where we used the fact that $\langle e_i, e_i \rangle = 1$.

Further, we have $||\bar{A}_i|| \leq (1 - \sqrt{\omega}) ||A_i|| + \frac{\sqrt{\omega}}{||A_i e_i||} ||A_i e_i|| ||e_i|| = 1$. □

Lemma 6 *Let* $\bar{A}_i = (1 - \nu) A_i + \nu A_i e_i e_i^T$. *Assume* $0 \leq \nu \leq 1$, $A_i^* \Gamma A_i e_i \succeq_{K_i} 0$, $A_i^* A_i e_i \succeq_{K_i} \omega e_i$, *and* $||A_i|| = 1$. *Then* $||\bar{A}_i|| \leq 1$, $\bar{A}_i^* \Gamma \bar{A}_i e_i \succeq_{K_i} \nu \eta_i^2 e_i$ *and* $\bar{A}_i^* \bar{A}_i e_i \succeq_{K_i} \omega e_i$.

Proof: The proof is by direct calculation. First, $||\bar{A}_i|| \leq (1-\nu)||A_i|| + \nu ||A_i|| ||e_i||^2 = 1$. Note that $\bar{A}_i e_i = A_i e_i$ and so

$$\bar{A}_i^* \Gamma \bar{A}_i e_i = (1 - \nu) A_i^* \Gamma A_i e_i + \nu \eta_i^2 e_i \succeq_{K_i} \nu \eta_i^2 e_i$$

and

$$\bar{A}_i^* \bar{A}_i e_i = (1 - \nu) A_i^* A_i e_i + \nu ||A_i e_i||^2 e_i \succeq (1 - \nu) \omega e_i + \nu \omega e_i = \omega e_i$$

where we used the facts that $\langle e_i, e_i \rangle = 1$ and $||A_i e_i||^2 = e_i^T A_i^* A_i e_i \geq e_i^T (\omega e_i)$ since e_i and $A_i^* A_i e_i$ are both in K_i. □

If it is necessary to modify $\bar{A}_i$ in Step 5c of the algorithm, the subsequent rescaling of $\bar{A}_i$ will also change η_i, but it will keep $\bar{A}_i^* \Gamma \bar{A}_i e_i \succeq_{K_i} \nu \eta_i^2 e_i$.

We summarize various properties of the constraints after the Selective Orthonormalization procedure is completed in the following theorem. These results follow from Lemmas 1, 5, and 6, and Theorem 2.

Theorem 3 *When the algorithm terminates, we have*

$$\{y : A_i^* y \preceq_{K_i} A_i^* \bar{y}\} \subseteq \{y : \bar{A}_i^* y \preceq_{K_i} \bar{A}_i^* \bar{y}\}. \tag{6}$$
$$||\bar{A}_i|| = 1 \text{ for } i = 1, \ldots, q. \tag{7}$$
$$\bar{A}_i^* \Gamma \bar{A}_i e_i \succeq_{K_i} \nu \eta_i^2 e_i \text{ for } i = 1, \ldots, q, \text{ with } \eta_i = \sqrt{e_i^* \bar{A}_i^* \Gamma \bar{A}_i e_i}. \tag{8}$$
$$\bar{A}_i^* \bar{A}_i e_i \succeq_{K_i} \omega e_i \text{ for } i = 1, \ldots, q. \tag{9}$$
$$\bar{A}_i^* \Gamma \bar{A}_j e_j \succeq_{K_i} 0 \text{ for } i = 1, \ldots, q \text{ and } j = 1, \ldots, q. \tag{10}$$

3 Finding a new feasible point

After modifying the constraints as detailed in §2, we have the conic feasibility problem

$$\begin{array}{llll} \max & 0 & & \\ \text{subject to} & A^* y & \preceq_K & c \\ & \bar{A}_i^* y & \preceq_{K_i} & \bar{c}_i \quad i = 1, \ldots, q \end{array} \qquad (Dq)$$

where $\bar{c}_i = \bar{A}_i^* \bar{y}$, $i = 1, \ldots, q$. The corresponding dual problem is

$$\begin{array}{llllll} \min & \langle c, x \rangle & + & \sum_{i=1}^q \langle \bar{c}_i, x_i \rangle & & \\ \text{subject to} & Ax & + & \sum_{i=1}^q \bar{A}_i x_i & = & 0 \\ & x & & & \succeq_K & 0 \\ & & & x_i & \succeq_{K_i} & 0, \quad i = 1, \ldots, q. \end{array} \qquad (Pq)$$

We assume we have a point $\bar{y}$ with $c - A^* y \succ_K 0$, and a corresponding primal solution $\bar{x} \succ_K 0$ with $Ax = 0$. Taking $y = \bar{y}$, $x = \bar{x}$, and $x_i = 0$ for $i = 1, \ldots, q$ is feasible in (Dq) and (Pq), but it is not interior. So we use the direction

$$d_y = -\sum_{i=1}^q \frac{1}{\eta_i} \Gamma \bar{A}_i e_i \tag{11}$$
$$d_x = H(\bar{s}) A^* d_y \tag{12}$$
$$d_{x_i} = \frac{1}{\eta_i} e_i \tag{13}$$

This gives interior feasible solutions to (Pq) and (Dq) for small positive step lengths. The direction d_y is illustrated in Figure 4.

Theorem 4 *The directions defined in (11)–(13) result in strictly feasible primal and dual iterates for small positive steplengths.*

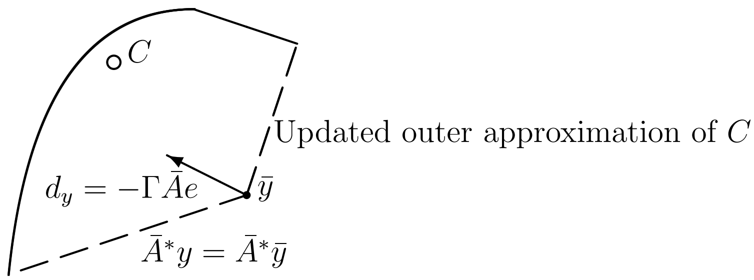

Figure 4: Direction d_y for regaining a strictly feasible dual solution when adding a single conic cut

Proof: Note that $Ad_x = \Gamma^{-1}d_y = -\sum_{i=1}^q \frac{1}{\eta_i}\bar{A}_i e_i = -\sum_{i=1}^q \bar{A}_i d_{x_i}$ so the equality constraints in (Pq) hold for any steplength. Further, $d_{x_i} \succ_{K_i} 0$ from the choice of e_i, so the primal solution is strictly feasible for any sufficiently small positive steplength.

Further, $\bar{A}_j^* d_y = -\sum_{i=1}^q \frac{1}{\eta_i}\bar{A}_j\Gamma\bar{A}_i e_i \preceq_{K_i} -\nu\eta_j e_j$ from Theorem 3, showing strict feasibility in (Dq) for any sufficiently small positive steplength. □

The steplengths can be chosen to ensure the iterates remain within appropriately defined Dikin ellipsoids. Given a cone K with barrier function f, the Dikin ellipsoid around a point $\hat{s}$ in the interior of K is

$$\mathcal{E} := \{s : ||s - \hat{s}||_{H(\hat{s})} \leq 1\}$$

and is contained in K.

Lemma 7 *Let $y(\alpha) = \bar{y} + \alpha d_y$. If $0 < \alpha < 1/q$ then $y(\alpha)$ is strictly feasible in (Dq).*

Proof: Note that $-A_j^* d_y \succ_{K_j} 0$ for $j = 1, \ldots, q$, from Theorem 3. The change in the dual slack variable $s = c - A^*y$ is $d_s = -A^*d_y$. We have

$$||d_s||_{H(\bar{s})} = ||\sum_{i=1}^q \frac{1}{\eta_i} A^*\Gamma\bar{A}_i e_i||_{H(\bar{s})} \leq \sum_{i=1}^q \frac{1}{\eta_i}||A^*\Gamma\bar{A}_i e_i||_{H(\bar{s})} = q$$

from (2) and (8). Feasibility then follows from the observation about the Dikin ellipsoid centered at $\bar{s}$. □

It is also useful to obtain a valid lower bound on the maximum possible primal steplength, using the Dikin ellipsoid centered at $\bar{x}$. Rather than looking at a bound based on $||d_x||_{H(\bar{x})}$, we construct one based on a slightly different norm, namely $||d_x||_{H(\bar{s})^{-1}}$. The two norms can be related using the fact that the current iterate is an approximate analytic center; for details see [9] or [13].

Lemma 8 *We have* $||d_x||_{H(\bar{s})^{-1}} \leq q$.

Proof: Note that

$$||d_x||_{H(\bar{s})^{-1}} = ||\sum_{i=1}^{q} \frac{1}{\eta_i} H(\bar{s}) A^* \Gamma \bar{A}_i e_i||_{H(\bar{s})^{-1}} \leq \sum_{i=1}^{q} \frac{1}{\eta_i} ||H(\bar{s}) A^* \Gamma \bar{A}_i e_i||_{H(\bar{s})^{-1}} = q$$

from (2). $\square$

For simplicity, we use the same steplength α in both the primal and the dual. The restart point is then

$$y(\alpha) = \bar{y} + \alpha d_y \tag{14}$$

$$s(\alpha) = \bar{s} - \alpha A^* d_y \tag{15}$$

$$s_i(\alpha) = -\alpha A_i^* d_y \tag{16}$$

$$x(\alpha) = \bar{x} + \alpha d_x \tag{17}$$

$$x_i(\alpha) = \frac{\alpha}{\eta_i} e_i \tag{18}$$

for some $\alpha > 0$.

If the constraints do not satisfy the conditions of Theorem 3 then a nonlinear program has to be solved to find a restart point. The variables of this problem are the primal directions d_{x_i} and it is necessary to include constraints based on Dikin ellipsoids in order to ensure that the restart point has an appropriate potential function value. For details, see [10, 11, 12, 2].

4 Potential function preliminaries

In this section we give some properties of potential functions. Most of the results on potential functions are taken from Renegar [13], and the remainder can be found in Basescu and Mitchell [2]. These results are used in §5 and §6. In §5, we look at the change in the potential function when moving in the direction given in §3, and hence show that the algorithm is able to regain an approximate analytic center quickly. Global convergence of an interior point cutting plane method using selective orthonormalization is the subject of §6.

The definitions of self-concordant barrier functionals, logarithmically homogeneous functionals, conjugate functions, and self-dual cones can be found in Appendix A.

Lemma 9 *Let f be a self-concordant logarithmically homogeneous barrier functional with domain D_f. Let $\bar{D}_f$ denote the closure of D_f. If $x \in D_f$, $y \in \bar{D}_f$ and for all $t \geq 0$ then*

$$f(x + ty) \leq f(x). \tag{19}$$

If the domain of f is a cone K then the geometrical interpretation of Lemma 9 is that x maximizes f over the cone $x + K$.

Adding a scalar to an intrinsically self-conjugate barrier functional results in another such function. Therefore, we can make the following assumption without loss of generality.

Assumption 2 *Given a cone K with fixed vector e in its interior, the conjugate barrier function satisfies $f^*(e) = 0$.*

Let ϑ_f denote the parameter of the intrinsically self-concordant barrier functional $f : \text{int}(K) \to I\!R$, in the terminology of Nesterov and Nemirovskii [9]. This parameter is called the complexity value of f by Renegar [13].

Lemma 10 *(Proposition 3.5.1. [13].) If f : int$(K) \to I\!R$ is an intrinsically self-conjugate barrier functional, then*

$$f^*(s) = f(s) + \vartheta_f.$$

Note that a consequence of Assumption 2 and Lemma 10 is that $f(e) = -\vartheta_f$. The standard dual barriers for linear programming, semidefinite programming, and second order cone programming satisfy Assumption 2. For example, for semidefinite programming the dual slack matrix $S = I$ satisfies $f^*(S) = -\ln\det(S) = 0$. The assumption is stated using the conjugate function rather than the original barrier function because the dual barrier function $f^*(.)$ is used in §6 to prove global convergence.

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

Definition 1 *For an instance of the algorithm described by the functional f, the vector c and the linear operator A, we define the* primal-dual potential *to be:*

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

It is customary to call $\langle c, x\rangle + f(x)$ the primal potential *and $f^*(s)$ the* dual potential. *Note that if $Ax = 0$ and $s = c - A^*y$ for some y then $\langle c, x\rangle = \langle s, x\rangle$.*

We conclude this section with an upper bound on the potential function value of a θ - analytic center.

Theorem 5 *Let (x, y, s) be a θ - analytic center corresponding to an instance of the algorithm described by the functional f, the linear operator A and the vector c. Then,*

$$\Phi_{PD}(x,s) \leq \frac{\theta^3}{3(1-\theta)} + \frac{\theta^2}{2}. \tag{20}$$

5 Local convergence

In this section, we show that the algorithm can quickly recover an approximate analytic center if the added cuts are those returned by the selective orthonormalization procedure. Let $(\bar{x}, \bar{y}, \bar{s})$ be the current θ - analytic center for (P) and (D) with the corresponding primal-dual potential:

$$\bar{\phi} := \langle c, \bar{x}\rangle_X + f(\bar{x}) + f^*(\bar{s}).$$

Note that Theorem 5 gives an upper bound on the value of $\bar{\phi}$. After adding the cuts described by f_i, A_i and c_i for $i = 1, \ldots, q$, we take a scaled step to get back into the feasible region. At this new point, the primal-dual potential is:

$$\begin{aligned} \phi_{new} &:= \langle c, x(\alpha)\rangle + \sum_{i=1}^{q}\langle c_i, x_i(\alpha)\rangle + f(x(\alpha)) + \sum_{i=1}^{q} f_i(x_i(\alpha)) + f(s(\alpha)) + \sum_{i=1}^{q} f_i^*(s_i(\alpha)) \\ &= \bar{\phi} + \alpha(\langle c, d_x\rangle + \sum_{i=1}^{q}\langle c_i, d_{x_i}\rangle) + \sum_{i=1}^{q} f_i(x_i(\alpha)) + \sum_{i=1}^{q} f_i^*(s_i(\alpha)) + F \end{aligned}$$

with

$$F := f(x(\alpha)) - f(\bar{x}) + f^*(s(\alpha)) - f^*(\bar{s}). \tag{21}$$

Because the cuts are central, $A_i^*\bar{y} = c_i$ for $i = 1, \ldots, q$. Hence

$$\sum_{i=1}^{q}\langle c_i, d_{x_i}\rangle = \sum_{i=1}^{q}\langle A_i^*\bar{y}, d_{x_i}\rangle = \langle \bar{y}, \sum_{i=1}^{q} A_i d_{x_i}\rangle = -\langle \bar{y}, Ad_x\rangle = -\langle A^*\bar{y}, d_x\rangle$$

and so

$$\langle c, d_x \rangle + \sum_{i=1}^{q} \langle c_i, d_{x_i} \rangle = \langle \bar{s}, d_x \rangle.$$

So, finally:

$$\phi_{new} = \bar{\phi} + \sum_{i=1}^{q} f_i(x_i(\alpha)) + \sum_{i=1}^{q} f_i^*(s_i(\alpha)) + F + \alpha \langle \bar{s}, d_x \rangle. \tag{22}$$

Now let's evaluate $F + \alpha\langle \bar{s}, d_x \rangle$. From [2], we have

$$\alpha \langle \bar{s}, d_x \rangle + f(x(\alpha)) - f(\bar{x}) \le \theta\zeta + \frac{1}{2}\zeta^2 + \frac{\zeta^3}{3(1-\zeta)}. \tag{23}$$

and

$$f^*(s(\alpha)) - f^*(\bar{s}) \le \theta\zeta + \frac{1}{2}\zeta^2 + \frac{\zeta^3}{3(1-\zeta)} \tag{24}$$

where $0 < \zeta < 1$ and

$$\alpha < (1 - \frac{\theta}{\sqrt{\vartheta_f}})\frac{\zeta}{q} \tag{25}$$

The contributions to the new potential function value in (22) from the new primal variables and dual slacks can be bounded above.

Lemma 11 *The sum of the primal and dual barrier function values $f_i(x_i(\alpha))$ and $f_i^*(s_i(\alpha))$ is bounded above for $i = 1, \ldots, q$. In particular, $f_i(x_i(\alpha)) + f_i^*(s_i(\alpha)) \le -\vartheta_{f_i} - 2\vartheta_{f_i} \ln \alpha - \vartheta_{f_i} \ln \nu$.*

Proof: Note from the proof of Theorem 4 that $s_i(\alpha) \succeq_{K_i} \alpha\nu\eta_i e_i$. The result then follows immediately from Lemmas 9 and 10, and the logarithmic homogeneity of f_i and f_i^*. □

This bound is the final piece that enables us to show that the algorithm can regain a new approximate analytic center efficiently.

Theorem 6 *After the addition of q new cuts with barrier functionals with complexity values ϑ_{f_i}, a new θ - analytic center can be obtained in $O(\sum_{i=1}^{q} \vartheta_{f_i} \ln q)$ Newton steps.*

Proof: Equations (22), (23), and (24), Theorem 5, and Lemma 11 enable us to place a bound on ϕ_{new}. The factor of $\ln(q)$ arises from the upper bound on α in (25). As shown in [2], this allows us to find a new approximate analytic center in a proportional number of steps. □

6 Global convergence

The proofs of global convergence in the literature have examined upper and lower bounds on the dual potential function value (see, for example, [2, 4, 5, 10, 11, 12]). The lower bound increases more quickly than the upper bound, and the algorithm must terminate before the two bounds meet. In all of these references, the lower bound depends only on the barrier parameters of the added constraints, and not on any other property of these constraints. Hence, these lower bounds are still valid when the constraints are modified using selective orthonormalization.

The analysis of the upper bound on the dual potential function has to consider certain cases that can lead to significant increases in the bound. The use of selective orthonormalization prevents these cases and hence the upper bound does not grow as fast, so it is possible to prove a stronger convergence result. We first give an upper bound on the size of the contribution to the dual barrier function of the additional terms due to the cuts.

Lemma 12 *Assume $\hat{y}+\varepsilon u$ is feasible in (Dq) for any u with norm no greater than 1. Let $\hat{s}_i = \bar{c}_i - \bar{A}_i^* \hat{y}$ for $i = 1, \ldots, q$. Then $f_i^*(\hat{s}_i) \leq -\vartheta_{f_i} \ln \varepsilon\omega$ for $i = 1, \ldots, q$.*

Proof: For any such u we have $\bar{c}_i \succeq_{K_i} \bar{A}_i^*(\hat{y} + \varepsilon u)$, so $\hat{s}_i \succeq_{K_i} \varepsilon \bar{A}_i^* u$. Let $u_i = \bar{A}_i e_i$. Then $\hat{s}_i \succeq_{K_i} \varepsilon \bar{A}_i^* \bar{A}_i e_i \succeq_{K_i} \varepsilon\omega e_i$ from (9). The result follows from Lemma 9, the logarithmic homogeneity of f_i^*, and Assumption 2. □

In order to extend this lemma to the complete dual barrier function, the algorithm must be described in more detail. First, we discuss initialization.

Assumption 3 *The algorithm is initialized with either a box $-2^L e \leq y \leq 2^L e$ or with a ball $||y|| \leq R$, for some positive constants L and R. If a box is used, the initial dual barrier consists of $2m$ linear barriers, each with barrier function parameter $\vartheta = 1$. The ball constraint is a second order cone constraint, with barrier function parameter $\vartheta = 2$. In either case, the initial dual analytic center is $y = 0$. The initial cone is denoted K_0. The dual slack for the initial dual barrier function is denoted s_0 and the initial dual barrier function is denoted $f_0^*(s_0)$.*

Lemma 13 *Assume $\hat{y}+\varepsilon u$ is feasible in (Dq) for any u with norm no greater than 1. Let $\hat{s}_0$ denote the dual slack corresponding to the initial set of constraints. Then $f_0^*(\hat{s}_0) \leq -\vartheta_{f_0} \ln \varepsilon$.*

Proof: For either initialization, we have $\hat{s}_0 \geq \varepsilon e$. The result follows. □

Conic constraints are added to the current formulation. The current cone K in (Dq) and (Pq) is the cartesian product of several cones. The algorithm finds an approximate analytic center, adds several violated conic constraints if the current approximate center is not in C, and repeats the process. At stage j the algorithm adds p_j conic constraints; we denote their barrier functions by $f_i^j(x_i^j)$ for $i = 1, \ldots, p_j$. After t stages, the dual barrier function is

$$f^*(s) = f_0(s_0) + \sum_{j=1}^{t} \sum_{i=1}^{p_j} f_i^j(s_i^j).$$

An upper bound on this function follows from Lemmas 12 and 13.

Theorem 7 *Assume $\hat{y} + \varepsilon u$ is dual feasible for any u with norm no greater than 1. Then the dual barrier function is no larger than*

$$-\vartheta_{f_0} \ln \varepsilon - \ln(\varepsilon\omega) \sum_{j=1}^{t} \sum_{i=1}^{p_j} \vartheta_{f_i^j}$$

Note that this upper bound is valid for the analytic center, from our assumption that C contains a ball of radius ε as long as it is nonempty.

A lower bound can be constructed by establishing an upper bound on the primal potential function, since the optimal value of $\Phi_{PD}(x, s)$ is zero. This upper bound comes from the restart point in (17) and (18). Before adding the cuts, we have an approximate analytic center so (20) holds. Lemma 11 and (23) give an upper bound in the change in the primal potential function, so a valid lower bound on the increase in the barrier function value of the dual analytic center is contained in the following lemma.

Lemma 14 *In moving from an approximate analytic center for (D) to one for (Dq), the dual potential function increases by at least*

$$\Lambda + \sum_{i=1}^{q} \vartheta_{f_i}(1 + \ln(\alpha/\eta_i))$$

where

$$\Lambda = -\frac{\theta^3}{3(1-\theta)} - \frac{\theta^2}{2} - \theta\zeta - \frac{1}{2}\zeta^2 - \frac{\zeta^3}{3(1-\zeta)}.$$

Proof: Let $(x^{AC}, \{x_i^{AC} : i = 1, \ldots, q\})$ and $(y^{AC}, (s^{AC}, \{s_i^{AC} : i = 1, \ldots, q\}))$ denote the analytic center after the addition of the q cuts. The new analytic center after the addition of the cuts has dual barrier function value

$$\begin{aligned} f^*(s^{AC}) + \sum_{i=1}^{q} f_i^*(s_i^{AC}) &= -\langle c, x^{AC}\rangle - \sum_{i=1}^{q}\langle c_i, x_i^{AC}\rangle - f(x^{AC}) - \sum_{i=1}^{q} f_i(x_i^{AC}) \\ &\geq -\langle c, x(\alpha)\rangle - \sum_{i=1}^{q}\langle c_i, x_i(\alpha)\rangle - f(x(\alpha)) - \sum_{i=1}^{q} f_i(x_i(\alpha)) \\ &\geq -\langle c, \bar{x}\rangle - f(\bar{x}) - \theta\zeta - \frac{1}{2}\zeta^2 - \frac{\zeta^3}{3(1-\zeta)} - \sum_{i=1}^{q} f_i(\frac{\alpha}{\eta_i}e_i) \\ &\geq f^*(\bar{s}) + \Lambda - \sum_{i=1}^{q} f_i(\frac{\alpha}{\eta_i}e_i). \end{aligned}$$

The result follows from Lemmas 9 and 10, and the logarithmic homogeneity of f_i. □

Similar bounds have been derived without exploiting the restart point given by the SGSO construction. Nesterov [8] provided such a bound for linear programming, and this was used in the work of Goffin *et al.* [4, 5]. This result was generalized by Oskoorouchi and Goffin to the case of a single semidefinite cut [10] and a single SOCP cut [11], and then to the case of multiple SOCP cuts by Oskoorouchi and Mitchell [12]. Basescu and Mitchell [2] extended the result further to general conic cuts. For larger values of t, the lower bound result from Basescu and Mitchell [2] can be simplified to:

$$\bar{f}_t^*(\bar{s}_t) \geq \bar{f}_0^*(\bar{s}_0) + 0.5\sum_{j=1}^{t}\sum_{i=1}^{p_j}\vartheta_{f_i^j}(\ln(\sum_{j=1}^{t} p_j) - \ln\frac{2C_0\Theta P^3}{\alpha^2} - \ln(m\ln\frac{\sum_{j=1}^{t} p_j}{m})) \tag{26}$$

where Θ is an upper bound on $\arg\max\{\vartheta_{f_j^i}\}$, $C_0 = 16||c_0||^2$ with $||c_0||$ an upper bound on the norm of y derived from Assumption 3, and $P = \arg\max\{p_j\}$. Comparing this lower bound with the upper bound in Theorem 7 leads to the following complexity result.

Theorem 8 *The Selective Gram-Schmidt Orthonormalization cutting plane algorithm terminates as soon as*

$$(\sum_{j=1}^{t}\sum_{i=1}^{p_j}\vartheta_{f_i^j})(\ln(Hm\ln\frac{\sum_{j=1}^{t} p_j}{m}) - \ln\sum_{j=1}^{t} p_j) \leq 2\bar{f}_0^*(\bar{s}_0) + 2\vartheta_{f_0}\ln\varepsilon$$

where $H = \frac{2C_0\Theta P^3}{\alpha^2\varepsilon^2\omega^2}$. *The number of cuts added is at most* $O^*(\frac{mC_0\Theta P^3}{\varepsilon^2})$ *(here* O^* *means that low orders are ignored).*

The upper bound in Theorem 7 only uses the fact that (9) holds after the SGSO procedure. The lower bound result in (26) does not exploit the SGSO procedure. Therefore, the global convergence result in Theorem 8 holds even if only steps 5(a) and 5(b) of the SGSO procedure are performed.

The complexity of the algorithms of Oskoorouchi and Goffin for semidefinite programming [10] and second order cone programming [11] and the algorithm of Basescu and Mitchell [2] depend on a condition number. If the selective Gram-Schmidt orthonormalization procedure is used to modify the cuts, a fixed positive lower bound (namely a simple function of ω) can be placed on these condition numbers, so the dependence on the condition number can be removed. The condition number is needed in the appropriate analogues of Theorem 7, and plays exactly the same role as our parameter ω. Without the SGSO procedure, the condition number has to appear in the complexity bound, since it is not under control. The condition numbers for SDP and SOCP are considered in more detail in the next section.

7 SGSO applied to SDP and SOCP

In this section, we specialize the SGSO algorithm to semidefinite programming and second order cone programming. In particular, we look at the construction of the selectively orthonormalized constraints and examine the complexity of the resulting algorithms.

7.1 Semidefinite programming

We consider the case of adding a single semidefinite programming cut of the form

$$\mathcal{A}^* y := \sum_{j=1}^{m} A^j y_j \preceq \sum_{j=1}^{m} A^j \bar{y}_j \tag{27}$$

where each A^j is a $q \times q$ symmetric matrix, where q may vary from iteration to iteration. An element in the interior of the semidefinite cone is the $q \times q$ identity matrix I. The quantity $\bar{A}^* \bar{A} I$ constructed in the selective orthogonalization procedure is equal to

$$\mathcal{A}^* \mathcal{A}(I) = \sum_{j=1}^{m} A^j \text{trace}(A^j). \tag{28}$$

Let σ be the smallest eigenvalue of this matrix. If $\sigma < 0$ the update given in step 5a of the SGSO procedure gives

$$A^j \leftarrow A^j + \lambda \text{trace}(A^j) I, \tag{29}$$

where

$$\lambda := \frac{-\sigma}{\sum_{k=1}^{m} \text{trace}(A^k)^2}. \tag{30}$$

Note that if $\text{trace}(A^j)$ was negative then the update will serve to make it more negative, and if it was zero then A^j is not changed. With this update, we obtain:

$$\begin{aligned} \bar{\mathcal{A}}(X) &= \mathcal{A}(X) + \lambda \text{trace}(X) \begin{pmatrix} \text{trace}(A^1) \\ \vdots \\ \text{trace}(A^m) \end{pmatrix} & (31) \\ \bar{\mathcal{A}}^*(y) &= \mathcal{A}^*(y) + \lambda \text{trace}(\mathcal{A}^* y) I & (32) \\ \bar{\mathcal{A}}^*(\bar{\mathcal{A}}(I)) &= (1 + \lambda p)(\sum_{j=1}^{m} A^j \text{trace}(A^j) - \sigma I). & (33) \end{aligned}$$

Thus, from (33) the matrix $\bar{\mathcal{A}}^*(\bar{\mathcal{A}}(I))$ is positive semidefinite (as must be the case from Lemma 2). Any y satisfying (27) will also satisfy the linear constraint

$$\text{trace}(\mathcal{A}^*(y)) \leq \text{trace}(\mathcal{A}^*(\bar{y})). \tag{34}$$

Therefore, it follows from (32) that it will also satisfy the modified constraint $\bar{\mathcal{A}}^*(y) \preceq \bar{\mathcal{A}}^*(\bar{y})$ (as must be the case from Lemma 1).

Let β be a fixed positive scalar. If we choose

$$\bar{\lambda} = \frac{\max\{0, \beta - \sigma\}}{\sum_{k=1}^{m} \text{trace}(A^k)^2} \tag{35}$$

and update $A^j \leftarrow A^j + \bar{\lambda}\text{trace}(A^j) I$, then the smallest eigenvalue of $\bar{\mathcal{A}}^*(\bar{\mathcal{A}}(I))$ is at least β.

The condition number defined in [10] is

$$\mu := \max\{\det(\bar{\mathcal{A}}^* u) : ||u|| = 1, \bar{\mathcal{A}}^* u \succeq 0\} \tag{36}$$

where the operator $\bar{\mathcal{A}}$ is normalized so that $\text{trace}(A^j) \leq 1$ and $||a_{lq}|| \leq 1$, with $a_{lq} = vec({A^1}_{lq}, {A^2}_{lq}, \ldots, {A^m}_{lq})$, and at least one of these inequalities holds at equality. The algorithm in [10] requires the addition of no more than $O(\frac{p^2 m^3}{\mu^2 \epsilon^2})$ cuts to get within ϵ of optimality, where p and μ are upper bounds on their values for each cut.

The SGSO procedure allows the condition number μ to be controlled. As long as $||\bar{\mathcal{A}}|| \leq 1$, the condition number can be underestimated by choosing $u = \bar{\mathcal{A}}(I)$. The update in Step 5b allows us to choose any value for $\omega < 1$, and then $\mu \geq \det(\bar{\mathcal{A}}^* \bar{\mathcal{A}}(I)) \geq \omega^p$. For example, choosing $\omega = (1 - 1/p)$ for $p \geq 4$ and $\omega = 0.67$ for $p = 1, 2, 3$ gives $\mu \geq 0.3$ for any cut. Thus, the SGSO cutting plane algorithm requires the addition of no more than $O(\frac{p^2 m^3}{\epsilon^2})$ cuts to get within ϵ of optimality.

7.2 SOCP

In a second order cone program, the additional cut takes the form

$$A^T y \preceq_K A^T \bar{y} \tag{37}$$

where A is an $m \times (1+q)$ matrix and q may vary from cut to cut. We write $A = [a;\ \hat{A}]$. The usual element in the interior of the second order cone of dimension $1+q$ is the vector $e = (1; 0, \ldots, 0)^T$. The vectors calculated in Step 5a of the SGSO procedure are $A^* A e = [a^T a; a^T \hat{A}]^T$ and $Aee^T = [a; 0]$, so the effect of the procedure is to rescale the first column of A. Thus, $\bar{A} = \xi[\bar{a}; \hat{A}]$ where $\bar{a} = ra$ for some $r \geq 1$ and $\xi = 1/||[\bar{a}; \hat{A}]||$.

The condition number defined in [11] is

$$\mu := \max\{\det(A^* u) := (a^T u)^2 - ||\hat{A}^T u||^2 : ||u|| = 1, \bar{A}^T u \succeq 0\} \tag{38}$$

and the algorithm requires the addition of no more than $O(\frac{m}{\varepsilon^2 \mu})$ cuts. With an SGSO approach, note that $\mu \geq \det(\bar{A}^* \bar{A} e) \geq \omega \det(e) = \omega$ provided (9) is satisfied. Hence the parameter ω can be chosen to be any fixed positive number between 0 and 1 in Step 5b and it will follow that the cutting plane algorithm with the SGSO procedure requires no more than $O(\frac{m}{\varepsilon^2})$ cuts.

8 Conclusions

A cutting surface algorithm using SGSO results in the addition of slightly weaker constraints. There are two benefits: first it is easy to restart the algorithm, and second the overall complexity of the algorithm can be shown to be independent of any condition number of the cuts returned by the oracle.

The complexity of the original algorithms should be at least as good as these weakened versions, at least in terms of the number of outer iterations. However, the current analysis framework can't be used to get rid of the condition number. The upper bound on the dual potential function is smaller when SGSO cuts are added because these cuts better relate s and y: if y is at least some distance from the boundary then the corresponding slack must have a reasonable value, leading to the bound on the potential function. The lower bound doesn't depend on the condition number explicitly, it just uses the fact that some constraint was added, with no exploitation of the structure of the constraint. If the lower bound could be improved to reflect the condition number then it may be possible to prove a complexity result for unmodified cuts that does not depend on a condition number.

A Appendix: Definitions

In this appendix, we define various properties of barrier functions, self-dual cones, conjugate functions, and approximate analytic centers. These definitions are taken from Renegar [13] unless otherwise stated.

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

$$B_x(y,r) = \{z : \|z-y\|_x \leq r\}. \tag{39}$$

Definition 2 *A functional f is said to be* (strongly nondegenerate) self-concordant *if for all $x \in D_f$ we have $B_x(x,1) \subseteq D_f$, and if whenever $y \in B_x(x,1)$ we have:*

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

Let SC be the family of such functionals.

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

$$g_x(y) \;:=\; H(x)^{-1} g(y), \tag{40}$$

$$H_x(y) \;:=\; H(x)^{-1} H(y). \tag{41}$$

Definition 3 *A functional is a* (strongly nondegenerate self-concordant) barrier functional *if $f \in SC$ and*

$$\vartheta_f := \sup_{x \in D_f} \|g_x(x)\|_x^2 < \infty. \tag{42}$$

Let SCB be the family of such functionals.

Definition 4 *Let K be a closed convex cone and $f \in SCB$, $f : int(K) \to I\!R$. f is* logarithmically homogeneous *if for all $x \in int(K)$ and $t > 0$:*

$$f(tx) = f(x) - \vartheta_f \ln(t). \tag{43}$$

Definition 5 *Let K be a cone and $z \in int(K)$. The* dual cone *of K is*

$$K^* = \{s \in X : \langle x, s \rangle_X \geq 0 \text{ for all } x \in K\}. \tag{44}$$

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

$$K_z^* := \{s \in X : \langle x, s \rangle_z \geq 0, \text{ for all } x \in K\}. \tag{45}$$

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

Definition 6 *The* conjugate *of $f \in SCB$ with respect to $\langle \cdot, \cdot \rangle$ is*

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

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

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

Definition 7 *A functional $f \in SCB$ is* intrinsically self-conjugate *if f is logarithmically homogeneous, if K is intrinsically self-dual, and for each $z \in int(K)$ there exists a constant C_z such that $f_z^*(s) = f(s) + C_z$ for all $s \in int(K)$. A cone K is* self-scaled *or* symmetric *if $int(K)$ is the domain of an intrinsically self-conjugate barrier functional.*

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

Definition 8 *The* analytic center *(the AC) of the domain $\mathcal{F}_P := \{x \in int(K) : Ax = 0\}$ with respect to $f(x) + \langle c, x \rangle_X$ is the exact solution to the problem:*

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

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

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

For simplicity we will say that x or y or s is an analytic center if they are the components of an analytic center. We can introduce the notion of θ - analytic center by relaxing some of the previous equalities. First we will define this notion, then the following lemma will give an insight for this definition.

Definition 9 *[2] (x, y, s) is a θ -* analytic center *for $\mathcal{F}_P$, $\mathcal{F}_D$ iff $x \in \mathcal{F}_P$, $s \in \mathcal{F}_D$ and*

$$\|I - H(x)^{-\frac{1}{2}} H(s)^{-\frac{1}{2}}\| \leq \frac{\theta}{\sqrt{\vartheta_f}}. \tag{46}$$

Lemma 15 *[2] Let (x, y, s) be a θ - analytic center. Then:*

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

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

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

with e being the vector of all ones and xs the Hadamard product of the vectors x and s. Our definition for a θ analytic center is equivalent to requiring that the ∞-norm of $(e - xs)$ be no larger than $\frac{\theta}{\sqrt{\vartheta_f}}$. Note that if (x, y, s) is the analytic center for the intrinsically self-conjugate barrier functional f then $H(s)H(x) = I$.

References

[1] V. L. Basescu. *An analytic center cutting plane method in conic programming.* PhD thesis, Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, August 2003.

[2] V. L. Basescu and J. E. Mitchell. An analytic center cutting plane approach for conic programming. Technical report, Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, June 2005.

[3] S. K. Chua, K. C. Toh, and G. Y. Zhao. An analytic center cutting plane method with deep cuts for semidefinite feasibility problems. *Journal of Optimization Theory and Applications*, 123:291–318, 2004.

[4] J.-L. Goffin, Z.-Q. Luo, and Y. Ye. Complexity analysis of an interior cutting plane method for convex feasibility problems. *SIAM Journal on Optimization*, 6:638–652, 1996.

[5] J.-L. Goffin and J.-P. Vial. Multiple cuts in the analytic center cutting plane method. *SIAM Journal on Optimization*, 11(1):266–288, 2001.

[6] K. Krishnan and J. E. Mitchell. A unifying framework for several cutting plane methods for semidefinite programming. *Optimization Methods and Software*, 21(1):57–74, 2006.

[7] J. E. Mitchell and S. Ramaswamy. Using selective orthonormalization to update the analytic center after the addition of multiple cuts. *Journal of Optimization Theory and Applications*, 125(2):431–451, 2005.

[8] Y. Nesterov. Cutting plane algorithms from analytic centers: efficiency estimates. *Mathematical Programming*, 69(1):149–176, 1995.

[9] Y. E. Nesterov and A. S. Nemirovsky. *Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms*. SIAM Publications. SIAM, Philadelphia, USA, 1993.

[10] M. R. Oskoorouchi and J.-L. Goffin. The analytic center cutting plane method with semidefinite cuts. *SIAM Journal on Optimization*, 13(4):1029–1053, 2003.

[11] M. R. Oskoorouchi and J.-L. Goffin. An interior point cutting plane method for the convex feasibility problem with second-order cone inequalities. *Mathematics of Operations Research*, 30(1):127–149, 2005.

[12] M. R. Oskoorouchi and J. E. Mitchell. A second-order cone cutting surface method: complexity and application. Technical report, College of Business Administration, California State University San Marcos, California, May 2005.

[13] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. Number 1 in MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2001.

[14] K. Sivaramakrishnan, G. Plaza, and T. Terlaky. A conic interior point decomposition approach for large scale semidefinite programming. Technical report, Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, December 2005.

[15] J. Sun, K. C. Toh, and G. Y. Zhao. An analytic center cutting plane method for semidefinite feasibility problems. *Mathematics of Operations Research*, 27:332–346, 2002.

[16] K. C. Toh, G. Zhao, and J. Sun. A multiple-cut analytic center cutting plane method for semidefinite feasibility problems. *SIAM Journal on Optimization*, 12(4):1126–1146, 2002.