Multiple Cuts with a Homogeneous Analytic Center Cutting Plane Method

Olivier Péton* Jean–Philippe Vial*

May 17, 2001

Abstract

This paper analyzes the introduction of multiple central cuts in a conic formulation of the analytic center cutting plane method (in short ACCPM). This work extends earlier work on the homogeneous ACCPM, and parallels the analysis of the multiple cut process in the standard ACCPM. The main issue is the calculation of a direction that restores feasibility after introducing p new cutting planes at the query point. We prove that the new analytic center can be recovered in $O(p \log \omega p)$ damped Newton iterations, where ω is a parameter depending of the data. We also present two special cases where the complexity can be decreased to $O(p \log p)$. Finally, we show that the number of calls to the oracle is the same as in the single cut case, up to a factor $\sqrt{p}$.

keywords: Cutting planes, analytic center, conic formulation, multiple cuts

1 Introduction

Cutting plane methods aim to solve nondifferentiable convex optimization problems. Loosely speaking, the methods consist in approximating the solution set by a localization set made of cutting planes. This approximation is refined by iteratively adding new cutting planes. The procedure that generates the cutting planes is usually called oracle. In practice, the oracle often produces multiple cuts at once. It is well known [1, 3] that this feature is likely to bring considerable speed-up. However, managing efficiently the multiple cuts raises important implementation and theoretical issues [2, 4, 11].

In this paper we propose to extend the analysis of [2] to the conic formulation of Nesterov and Vial [9]. The supporting or separating hyperplanes provided by the oracle define a polyhedral model of the feasible set and the objective function of the problem of interest. In [9], the original problem and its polyhedral model are embedded into a projective space. The original problem is thus approximated by a polyhedral cone, which is described by a ν-normal function [7]. The addition of a quadratic term to the barrier function guarantees the existence of a minimum for the barrier. By extension, this minimum is named "analytic center". The oracle is called at the analytic center or in its vicinity [8].

*HEC/LOGILAB, University of Geneva, 40 bd du Pont d'Arve, CH-1211-Geneva 4, Switzerland, peton@hec.unige.ch, vial@hec.unige.ch

Computing a new analytic center after adding new cuts can be viewed as a special interior point form of a warm start problem. In the standard non-conic formulation the warm start procedure has been issued in the context of central cuts [11]. In [2], Goffin and Vial focus the analysis on the choice of a good search direction to restore the feasibility. This local problem reduces to the minimization of an unconstrained self-concordant function, which can be done in $O(p \ln p)$ iterations, where p is the number of new cuts. The computation of the new analytic center from a well chosen point in that direction is shown to be $O(1)$ and the overall complexity of the analytic center computation is $O(p \ln p)$.

The main idea in [2] for defining an optimal restoration direction is to replace the polyhedral model that was used to define the current analytic center by the Dikin's ellipsoid at the analytic center. The restoration direction is chosen as the point within Dikin's ellipsoid that maximizes the product of the slacks to the new constraints. The key argument in proving that the auxiliary problem is $O(p \ln p)$ uses the fact that the barrier function for the polyhedral model is a ν-normal barrier, i.e., logarithmically homogeneous.

Due to the normalizing quadratic term in the conic approach, this argument does not hold anymore; the compound barrier function is no longer logarithmically homogeneous. Using a recent result of Nesterov and Vial [10], we show that in two special cases - where the new cuts form two by two acute (resp. obtuse) angles [1] - the auxiliary problem is $O(p \ln p)$, independently of any other data. In the general case, computing the re-entering direction requires $O(p \ln \omega p)$ iterations, where $\omega \in R$ is a characteristic factor of the direction finding problem.

2 The Homogeneous Scheme with Multiple Cuts

2.1 The canonical problem

The cutting plane scheme applies to the canonical problem:

$$\text{Find } x \in K \bigcap X^*, \quad x \neq 0, \tag{2.1}$$

where $K \subset R^n$ is a closed convex cone with nonempty interior and $X^* \subset R^n$ is a closed convex cone. K is a set in which we can find an initial point. X^* plays the role of a target set. For example, from an optimization point of view, X^* would be the set of points satisfying the optimality conditions. Many problems can be rewritten into formulation (2.1). The examples of convex feasibility problems, convex constrained minimization and monotone variational inequalities are detailed in [9].

Before describing the algorithm to solve (2.1), we introduce a few definitions and assumptions. Some of them are based on the theory of self-concordant functions (see Appendix A).

Definition 2.1 *A separation oracle is a mapping $a(x)$ such that $\langle a(x), x - x^* \rangle \geq 0$ for any $x^* \in X^*$. The oracle is homogeneous if $a(tx) = a(x)$ for any $x \in \operatorname{int} K$ and $t > 0$, and $\langle a(x), x \rangle = 0$.*

Assumption 2.1 *Problem (2.1) is endowed with a homogeneous separation oracle $a(x)$, $x \in \operatorname{int} K$. That is, the oracle either confirms that $x \in X^*$, or returns p central cutting planes that contain X^* and do not contain x in its interior.*

[1] in the metric defined by the Dikin's ellipsoid.

At iteration i, the j^{th} cutting plane at query point x_i takes the following form:

$$\langle a_j(x_i), x_i - x \rangle \geq 0, \qquad \forall x \in \operatorname{int} K.$$

From now on, we introduce the notation $a_{ij} = a_j(x_i)$. Note that Definition 2.1 implies $\langle a_{ij}, x_i - x \rangle = -\langle a_{ij}, x \rangle$.

Assumption 2.2 *Without loss of generality, we suppose that all cutting planes have been scaled to* $\| a_{ij} \| = 1$.

Assumption 2.3 K *is equipped with a* ν*-normal barrier* F*. (See the definition of* ν*-normal barriers in Appendix A.)*

2.2 The homogeneous scheme with multiple cuts

Let $\mathcal{F}_D$ be an outer approximation of X^*. We name it "localization set". The homogeneous analytic center cutting plane method consists in an iterative refinement of $\mathcal{F}_D$. Starting with $\mathcal{F}_D^0 = K$, we compute at iteration k an η-approximate analytic center of set $\mathcal{F}_D^k$, generate the p cutting planes and update the localization set by adding the p cutting planes in its definition.

More formally, the scheme can be described as follows:

Initialization

Let $\mathcal{F}_D^0 = K$ and $x_0 \in \operatorname{int} K$.

Define a ν-normal barrier $F(x)$ for K.

Set $F_0(x) = \frac{1}{2} \| x \|^2 + F(x)$.

Basic Step (2.2)

1. Compute an approximate center x_k such that $\| F_k'(x_k) \|_{[F_k''(x_k)]^{-1}} \leq \eta$.
2. The oracle returns p cuts $a_{kj}, \quad j = 1 \dots p$.
3. Update $\mathcal{F}_D^{k+1} = \mathcal{F}_D^k \bigcap \{x \, , \, \langle a_{kj}, x_k - x \rangle \geq 0, \; j = 1 \dots p\}$.
4. Set $F_{k+1}(x) = F_k(x) - \sum_{j=1}^{p} \log\left(\langle a_{kj}, x_k - x \rangle\right)$.
5. If $x_k \in X^*$, STOP, otherwise replace k by $k+1$ and go to 1.

We recall that $\| u \|_B = \langle Bu, u \rangle^{\frac{1}{2}}$ denotes the local norm of the vector u relatively to the symmetric semidefinite matrix B. The centering parameter η is chosen in $]0, \frac{1}{3}[$.

2.3 Convergence of the homogeneous scheme

The main convergence theorem in [9] states that a certain function $\mu_k(x)$ can be made small enough with respect to $\| x \|$ as k increases. To fit the multiple cut case, we extend the definition of μ_k using the parameters

$$\begin{aligned} \lambda_{ijk} &= \frac{1}{\langle a_{ij}, x_i - x_k \rangle} > 0, \qquad i = 0 \dots k-1, \quad j = 1 \dots p, \\ S_k &= \sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk}. \end{aligned}$$

Then $\mu_k(x) = \frac{1}{S_k} \sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk} \langle a_{ij}, x_i - x \rangle$.

This value can be interpreted as a weighted average of the distance of x to the existing cutting planes. It is shown in [9] that small values of $\mu_k(x)$ imply near optimality in a number of applications.

The next theorem and its proof involve certain constants that depend on the iteration number and on the number on cutting planes. These constants are denoted $\sigma_1(p)$ to $\sigma_6(p)$, $\psi_1(k,p)$ and $\psi_2(k,p)$. Appendix B provides the extensive convergence proof of the homogeneous scheme with multiple cuts, as well as the analytic expression of the constants.

Theorem 2.1 *There exist constants $\sigma_1(p)$, $\sigma_2(p)$, $\psi_1(k,p)$ and $\psi_2(k,p)$ such that*

$$\begin{aligned} &\sigma_1(p) \textit{ is bounded from below by an absolute constant,} \\ &\sigma_2(p) = O(p), \\ &\sqrt{\nu + kp} \leq \psi_1(k,p) \leq \psi_2(k,p) \leq \left(\tfrac{1+3\eta+\eta^2}{1-\eta^2} \right) \sqrt{\nu + kp}. \end{aligned} \tag{2.3}$$

The following statements hold for any $k > 0$

1. *For any $x \in K$ we have*

$$\mu_k(x) \leq \frac{\psi_1(k,p)}{k\sigma_1(p)} \exp\left[\frac{F(x_k) - F(x_0)}{kp} \right] \| x \| + \eta \, \frac{1 + S_k}{S_k} \| x \| . \tag{2.4}$$

2. *For any $x \in \mathcal{F}_D$ we have*

$$\mu_k(x) \leq \frac{\psi_2(k,p)}{k\sigma_1(p)} \exp\left[\frac{F(x_k) - F(x_0)}{kp} \right] \| x \| . \tag{2.5}$$

Furthermore, we have in both cases $F(x_k) - F(x_0) \leq k\sqrt{\nu}\, \sigma_2(p)$.

Theorem 2.3 is used to compute the number of outer iterations for Scheme (2.2) for varied applications (see [9]). Setting $p = 1$, we retrieve the same expressions as in Theorem 1 of [8]. Setting also $\eta = 0$, we retrieve the same expression as in [9].

3 The re-entering direction

In what follows, we focus on the internal mechanism of the analytic center computation (Step 1 of Scheme (2.2)).

To evaluate the overall complexity of the general scheme, we need to give a complexity estimate of the computation of the analytic center after adding p new cuts. At iteration k, the j^{th} cutting plane given by the oracle at the query point x_k takes the form $\langle a_{kj}, x_k - x\rangle > 0$, and the ν-normal barrier is updated as follows

$$F_{k+1}(x) = F_k(x) - \sum_{j=1}^{p} \log\langle a_{kj}, x_k - x\rangle.$$

This function is not defined at x_k. The general procedure consists in a standard damped Newton scheme from an admissible starting point, as done in the single cut case [8]. We must look for a search direction of the form $d = x - x_k$ along which to look for a "good" starting point for the computation of the next approximate center x_{k+1}. This re-entering direction must be compatible with the whole set of new cutting planes. In [8], the starting point is chosen on a re-entering direction that is given by a closed form formula. In the multiple cut case, finding a good starting point is an issue in itself. We resort to an auxiliary problem to define the re-entering direction (subsection 3.1). The complexity of this computation is analyzed in subsection 3.2. Since the auxiliary problem cannot be solved exactly, we show that a suboptimal solution still allows to construct a good re-entering direction (subsection 3.3). Finally, the choice of a point along this direction is made by taking an appropriate step length (subsection 3.4).

3.1 Recovering Feasibility: the direction finding subproblem

Following the ideas of [2], we look for the direction d^* which maximizes the product of the p slacks $\langle a_{kj}, -d\rangle$. Since we want the direction to yield an admissible point, we restrict this maximization to Dikin's ellipsoid around x_k, which is guaranteed to be included in the localization set.

The direction d^* is defined as follows

$$\arg\max_{d\in R^n} \left\{ \sum_{j=1}^{p} \log\langle a_{kj}, -d\rangle \;\mid\; \langle F_k''(x_k)d, d\rangle \le 1 \right\}.$$

Let us denote $A = (a_{k1}, \ldots, a_{kp})$. The direction finding problem can be written as

$$\max \left\{ \sum_{j=1}^{p} \log\gamma_j \;\mid\; \gamma + A^T d = 0, \| d \|_{F_k''(x_k)} \le 1 \right\}. \tag{3.1}$$

This problem is equivalent to problem (12) of [2].

The positive semidefinite matrix $V = A^T F_k''(x_k)^{-1} A$ plays a fundamental role in the analysis. V can be interpreted as the variance-covariance matrix between the vectors a_{kj} with respect to Dikin's metric.

It turns out that the solution of Problem (3.1) can be obtained by solving the unconstrained auxiliary problem

$$\min_{\beta>0} \left\{ G(\beta) = \frac{p}{2}\langle \beta, V\beta\rangle - \sum_{j=1}^{p} \log\beta_j \right\}, \tag{3.2}$$

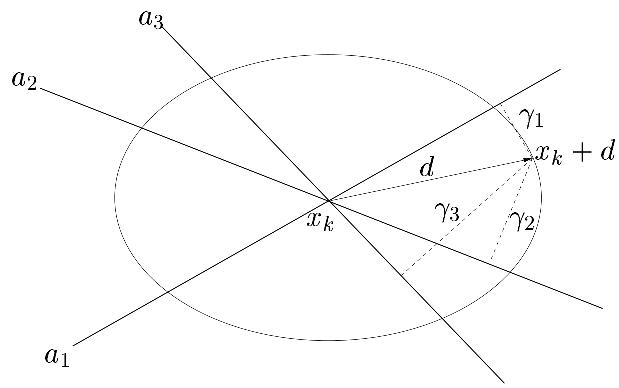

Figure 1: Re-entering direction

where β belongs to the positive orthant of R^p. The equivalence between (3.1) and (3.2) is stated in the following theorem

Theorem 3.1 *Let (d^*, γ^*) and β^* be the optimal solutions of (3.1) and (3.2) respectively. The following equalities hold:*

i) $(\beta^*)^{-1} = pV\beta^*$,

ii) $d^* = -F_k''(x_k)^{-1}A\beta^*$ *and* $\gamma^* = -A^T d^* = V\beta^*$,

iii) $\| d^* \|^2_{F_k''(x_k)} = \langle \beta^*, V\beta^* \rangle = 1$,

iv) $\sum_{j=1}^{p} \log \beta_j^* + \sum_{j=1}^{p} \log \gamma_j^* + p \log p = 0.$

The proof of Theorem 3.1 is given in [2] (Theorem 4.2): it is a straightforward application of the necessary and sufficient first order optimality conditions for (3.1). It can be shown that the existence of a direction d such that $Ad < 0$ guarantees that (3.2) has a finite optimum. The computation of the optimal direction d^* requires solving the unconstrained optimization problem (3.2). Since the function $G(\beta)$ is self-concordant, we apply the following Newton scheme to compute a close approximation of the optimal solution:

Initialization	Set $\beta^0$and choose a parameter $0 < \theta < 1$.	
Damped steps	While $\| G'(\beta) \|_{[G''(\beta)]^{-1}} > \sqrt{\theta}$, perform damped Newton steps.	(3.3)
Full step	Perform one additional pure Newton step.	

3.2 Complexity of the direction finding subproblem

From its definition (3.2), it is clear that the self-concordant function G has parameter p. The complexity of the process depends on p, θ, and of the starting point β^0. First, a complexity estimate is given when the p cuts are in a general position. Then, we study two special cases in which the complexity is independent of the data.

The following result ensures that the last pure Newton steps yields a θ-approximate solution:

Lemma 3.1 (Lemma 6.5 of [2]) *For any β such that*

$$\| G'(\beta) \|_{[G''(\beta)]^{-1}} \leq \sqrt{\theta} < 1,$$

the point $\beta^+ = \beta - G''(\beta)^{-1} G'(\beta)$ satisfies the inequality

$$\| G'(\beta^+) \|_{G''(\beta^+)^{-1}} \leq \| p\beta^+(V\beta^+) - e \| \leq \theta.$$

General Case

In the context of standard (non homogeneous) analytic center cutting plane method, Goffin and Vial [2] prove that a θ-approximate solution can be found in $O(p \log p)$. Their argument, based on the property that the localization set is included in a homothety of Dikin's ellipsoid, is no more valid in the present conic formulation.

Recently, Nesterov and Vial [10] studied the class of so-called augmented self-concordant barriers. Such problems are of the form

$$\min_{x \in C} \left\{ \frac{1}{2} \langle Qx, x \rangle + F(x) \right\},$$

where Q is a positive semi-definite matrix and F and p-normal barrier for a cone C.

The matrix V is positive semi-definite and the expression $-\sum_{j=1}^{p} \log \beta_j$ is a p-normal barrier for the positive orthant. Thus, G is an augmented self-concordant barrier. Following the ideas of [10], we define for any given point $\hat{\beta}$ in the positive orthant the two characteristics

$$\gamma_u \geq \max \left\{ \langle V\beta, \beta \rangle \;\mid\; \sum_{j=1}^{p} \frac{\beta_j}{\hat{\beta}_j} = 1, \quad \beta > 0 \right\}, \tag{3.4}$$

$$\gamma_l \leq \min \left\{ \langle V\beta, \beta \rangle \;\mid\; \sum_{j=1}^{p} \left(\frac{\beta_j}{\hat{\beta}_j} \right)^2 = 1, \quad \beta > 0 \right\}. \tag{3.5}$$

The complexity of Scheme (3.3) is given by the following theorem

Theorem 3.2 *(Theorem 3 of [10]) Taking $\beta_0 = \frac{\hat{\beta}}{\sqrt{p \, \gamma_u}}$ as an initial point, an upper bound for the number of iterations of a damped Newton scheme to minimize an augmented self-concordant barrier is as follows:*

$$O\left(p \log \frac{p \, \gamma_u}{\gamma_l} \right).$$

This complexity can be decreased to $O\left(\sqrt{p} \log p\right) + O\left(\ln \frac{\sqrt{p} \, \gamma_u}{\gamma_l} \right)$ by a path-following scheme.

Corollary 3.1 *No more than*

$$1 + \left\lceil \frac{\frac{p}{2} \log \frac{p \, \gamma_u}{\gamma_l}}{\sqrt{\theta} - \log(1 + \sqrt{\theta})} \right\rceil$$

iterations are necessary to get a θ-approximate solution for problem (3.2).

The proof of this corollary follows directly from the fact that each damped Newton step reduces the self-concordant function G by constant $\sqrt{\theta} - \log(1+\sqrt{\theta})$.

Special case #1 : acute angles

Let us analyze the case where all vectors a_{kj} in Problem (3.1) form pairwise acute angles with respect to the metric induced by F''. In other words, V is a symmetric positive semidefinite matrix with non-negative entries.

Let us mention first that G is bounded from below on the positive orthant, so that Problem (3.2) is well defined. This result follows from the assumption that the solution set of our original problem is a cone with nonempty interior, and is proved in Lemma 4.1 of [2]. In the present case, we shall make the stronger assumption that the a_{kj} are linearly independent. The following lemma gives the complexity estimate in this case.

Lemma 3.2 *Assume that V has non-negative entries and let $\beta^0 = \frac{1}{p}\, e$. Then the Newton scheme (3.3) converges in at most $O(p\log p)$.*

Proof:

We assume without loss of generality that V diagonal entries are scaled to 1. Since V (after scaling) is a correlation matrix, then V is componentwise less than the matrix E of all ones. Hence,

$$G(\beta^0) \le \frac{1}{2p}\langle Ee, e\rangle + p\log p \le \frac{1}{2}p + p\log p.$$

On the other hand, using the fact that the off-diagonal elements of V are non-negative, we have

$$G(\beta^*) \ge \sum_{i=1}^{p}\left(\frac{p}{2}\beta_i^2 - \log\beta_i\right).$$

The right-hand side of the last inequality is separable and achieves its minimum at $\beta_i = \frac{1}{\sqrt{p}}, \quad i = 1\ldots p$. Thus

$$G(\beta^*) \ge \frac{1}{2}(p + p\log p)$$

and

$$G(\beta^0) - G(\beta^*) \le \frac{1}{2}p\log p.$$

Since the Newton process decreases G at each iteration by quantity larger than an absolute constant, we conclude that the process converges in $O(p\log p)$. □

Remark 3.1 *Note that the complexity estimate in Lemma 3.2 is independent of the data a_{kj}.*

Corollary 3.2 *If the matrix V has only non-negative entries, no more than*

$$1 + \left\lceil \frac{\frac{p}{2}\log p}{\sqrt{\theta} - \log(1+\sqrt{\theta})} \right\rceil$$

iterations are enough to get a θ-approximate solution for problem (3.2).

Special case #2: obtuse angles

The case where all cuts form pairwise acute angles is too optimistic. Indeed, the limiting case would be that the cuts are all the same, which clearly is the less informative answer that a multicut oracle can deliver. Let us consider the case where the cuts form pairwise obtuse angles with respect to the metric induced by F''. In other words, V is a symmetric positive semi-definite matrix with non-positive off-diagonal entries. We shall further assume that the cuts are linearly independent, so that V^{-1} exists.

We introduce a new problem that is closely related to (3.1) and (3.2):

$$\min\{H(\gamma) = \frac{p}{2}\langle V^{-1}\gamma, \gamma\rangle - \sum_{i=1}^{p} \log \gamma_i\}. \tag{3.6}$$

Using the first order optimality condition, one easily check that $\gamma^* = V\beta^*$ is the optimal solution of (3.6). Since γ^* is also an optimal solution of (3.1), solving problem (3.6) yields the solution of (3.1). It remains to show that (3.6) can be solved in $O(p \log p)$ iterations. This is a direct consequence of the following lemma

Lemma 3.3 *Assume V is a full rank matrix with non-positive off-diagonal elements. Then V^{-1} exists and has non-negative elements.*

Proof:

Assume V is scaled to have a main diagonal of ones. Then, $V = I - K$, where K is a matrix with non-negative elements (and a zero main diagonal). Since V is positive definite,

$$\langle Kx, x\rangle < \| x \|^2 .$$

This proves that the maximal eigenvalue of matrix K is less than 1. By the Perron-Frobenius theorem, the spectral radius of a positive matrix is bounded by its maximal eigenvalue. Thus K has its eigenvalues bounded by 1 in absolute value, and $K^n \to 0$ as $n \to \infty$. We may write

$$V^{-1} = I + K + K^2 + \dots$$

Since K is positive, V^{-1} is clearly positive. □

Thus, problem (3.6) has the same form as (3.2). From Lemma 3.2, Scheme (3.3) can be used to yield a θ-approximate solution in $O(p \log p)$. Note that Corollary 3.2 also applies in the present case.

3.3 Using a near optimal reentering direction

From the first order optimality conditions for problems (3.1) and (3.2), one can check that β^* and γ^* satisfy $p\beta^*\gamma^* = e$.

The main point is that Scheme (3.3) only returns an approximate value of minimizer β^*, and the re-entering direction may be affected by such an approximation. The solution β^+ meets the proximity condition $\| p\beta^+(V\beta^+) - e \| < \theta$. Following Theorem 3.1, we define the restoration direction as follows

$$d^+ = -F_k''(x_k)^{-1}A\beta^+,$$

and

$$\gamma^+ = V\beta^+.$$

There is no guarantee that $x_k + d^+ \in \text{dom } F_k$, *i.e.* is feasible for the older cutting planes, since we may well have that $\| d^+ \|_{F_k''(x_k)} > 1$.

Hopefully the following result gives useful bounds characterizing d^+ and γ^+, and ensures that the analysis can be carried with these approximate values.

Lemma 3.4 (Theorem 4.3 of [2]) *The following inequalities hold*

i) $(1-\theta)e \le p\beta^+\gamma^+ \le (1+\theta)e$,

ii) $1-\theta \le \langle \beta^+, V\beta^+ \rangle \le 1+\theta$,

iii) $\gamma^+ > 0$ *and* $\sqrt{1-\theta} \le \| d^+ \|_{F_k''(x_k)} \le \sqrt{1+\theta}$.

Hence, an admissible step can be defined by the scaled vectors

$$\hat{d} = \frac{1}{\sqrt{1+\theta}}\, d^+,$$

and

$$\hat{\gamma} = \frac{1}{\sqrt{1+\theta}}\, \gamma^+.$$

It remains to show that this approximation of the optimal re-entering direction has only a limited impact on the objective value of problem (3.1).

Lemma 3.5

$$0 \le \sum_{j=1}^{p} \left(\log \gamma_j^* - \log \hat{\gamma}_j \right) \le \frac{\theta p}{2} + p \log \frac{\sqrt{1+\theta}}{1-\theta}.$$

Proof:

From Theorem 3.1, we have

$$\begin{aligned} G(\beta^*) &= \frac{p}{2} - \sum_{j=1}^{p} \log \beta_j^*, \\ &= \frac{p}{2} + p \log p + \sum_{j=1}^{p} \log \gamma_j^* \le G(\beta^+). \end{aligned}$$

Using definition of $G(\beta^+)$ and Lemma 3.4, we can provide an upper bound for $\sum_{j=1}^{p} \log \gamma_j^*$:

$$\sum_{j=1}^{p} \log \gamma_j^* \le \frac{p}{2} \langle \beta^+, V\beta^+ \rangle - \sum_{j=1}^{p} \log \beta_j^+ - \frac{p}{2} - p \log p,$$

$$\sum_{j=1}^{p} \log \gamma_j^* \leq \frac{\theta p}{2} - p \log p - \sum_{j=1}^{p} \log \beta_j^+. \tag{3.7}$$

Let us now compute a lower bound for $\sum_{j=1}^{p} \log \hat{\gamma}_j$. From Lemma 3.4 and the definition of $\hat{\gamma}$, we have $\frac{1-\theta}{\sqrt{1+\theta}} e \leq p\beta^+ \hat{\gamma}$. Taking the logarithm and summing, we get

$$p \log \frac{1-\theta}{\sqrt{1+\theta}} \leq p \log p + \sum_{j=1}^{p} \log \beta_j^+ + \sum_{j=1}^{p} \log \hat{\gamma}_j.$$

$$-\sum_{j=1}^{p} \log \hat{\gamma}_j \leq -p \log \frac{1-\theta}{\sqrt{1+\theta}} + p \log p + \sum_{j=1}^{p} \log \beta_j^+. \tag{3.8}$$

Summing (3.7) and (3.8) concludes the proof. □

3.4 Choosing a step length along the reentering direction

Once we have computed the optimal restoration direction, we have to set a step length α in this direction. Points such as $x^+ = x + \alpha \hat{d}$, with $0 < \alpha \leq 1$, belong to Dikin's ellipsoid and make appropriate candidates for the restoration step. Keeping in mind that the objective at iteration $k+1$ is to minimize function F_{k+1}, we first give an upper bound for the difference $F_{k+1}(x_k + \alpha \hat{d}) - F_{k+1}(x_{k+1}^*)$, and then determine the value of α which minimizes this gap.

Lemma 3.6 *For any point $x = x_k + \alpha \hat{d}$ $(0 < \alpha \leq 1)$ along Dikin's direction, the following inequality holds*

$$F_{k+1}(x_k + \alpha \hat{d}) - F_{k+1}(x_{k+1}^*) \leq \Delta_k(\alpha), \tag{3.9}$$

where

$$\begin{aligned}
\Delta_k(\alpha) &= -(1-\eta)\alpha - \log(1-\alpha) - p \log \alpha + \sigma_6(p), \\
\sigma_3(p) &= \frac{\sqrt{4p + (p-\eta)^2} - (p+\eta)}{2\,(1-\eta)}, \\
\sigma_4(p) &= \frac{\sqrt{4p + (p-\eta)^2} + (p+\eta)}{2\,(1-\eta)}, \\
\sigma_6(p) &= \eta \sigma_4(p) + p \log \sigma_4(p) - \sigma_3(p) + \log(1 + \sigma_3(p)) + \\
&\quad \frac{\theta p}{2} + p \log \frac{\sqrt{1+\theta}}{1-\theta} = O\left(p \log(p)\right).
\end{aligned}$$

Proof:

Using the definition of F_{k+1},

$$F_{k+1}(x_{k+1}^*) = F_k(x_{k+1}^*) - \sum_{j=1}^{p} \log \langle a_{kj}, x_k - x_{k+1}^* \rangle.$$

From inequality A.3, we get

$$F_k(x_{k+1}^*) \geq F_k(x_k) + \langle F_k'(x_k), x_{k+1}^* - x_k \rangle + \omega(\| x_{k+1}^* - x_k \|_{F_k''(x_k)}).$$

where $\omega(t) = t - \log(1+t)$. Recall that $\| F'(x_k) \|_{F_k''(x_k)^{-1}} \leq \eta$. Using Lemma B.3, and since ω is an increasing function, we get

$$F_k(x_{k+1}^*) \geq F_k(x_k) - \eta\sigma_4(p) + \sigma_3(p) - \log\left(1 + \sigma_3(p)\right).$$

Hence,

$$\begin{aligned} -F_{k+1}(x_{k+1}^*) \leq -F_k(x_k) + \eta\sigma_4(p) - \sigma_3(p) + \log\left(1 + \sigma_3(p)\right) \\ + \sum_{j=1}^{p} \log\langle a_{kj}, x_k - x_{k+1}^* \rangle. \end{aligned} \tag{3.10}$$

On the other hand, we still have by definition of F_{k+1}

$$F_{k+1}(x_k + \alpha\hat{d}) = F_k(x_k + \alpha\hat{d}) - \sum_{j=1}^{p} \log\langle a_{kj}, -\alpha\hat{d} \rangle. \tag{3.11}$$

Let us bound separately the two terms in the right-hand side of (3.11).

Using inequality A.3, we have

$$F_k(x_k + \alpha\hat{d}) \leq F_k(x_k) + \langle F'(x_k), \alpha\hat{d} \rangle - \| \alpha\hat{d} \|_{F_k''(x_k)} - \log(1 - \| \alpha\hat{d} \|_{F_k''(x_k)})$$

Since $\| \alpha\hat{d} \|_{F_k''(x_k)} \leq \alpha$, we get

$$F_k(x_k + \alpha\hat{d}) \leq F_k(x_k) - (1-\eta)\alpha - \log(1-\alpha).$$

By Lemma 3.5, we have

$$-\sum_{j=1}^{p} \log\langle a_{kj}, -\alpha\hat{d} \rangle = -\sum_{j=1}^{p} \log\hat{\gamma}_j \leq \frac{\theta p}{2} + p\log\frac{\sqrt{1+\theta}}{1-\theta} - \sum_{j=1}^{p} \log\gamma_j^*. \tag{3.12}$$

Since $\sum_{j=1}^{p} \log\gamma_j^*$ is the optimal solution of problem (3.1), it is larger than any feasible solution. Since x_{k+1}^* is feasible relatively to the new cuts, $\gamma = \frac{x_{k+1}^* - x_k}{\sigma_4(p)}$ is feasible to (3.1). Moreover $\| \gamma \|_{F_k''(x_k)} \leq 1$. Thus,

$$-\sum_{j=1}^{p} \log\gamma_j^* \leq p\log\sigma_4(p) - \sum_{j=1}^{p} \log\langle a_{kj}, x_{k+1}^* - x_k \rangle,$$

and by (3.12),

$$\begin{aligned} -\sum_{j=1}^{p} \log\langle a_{kj}, -\alpha\hat{d} \rangle \leq \frac{\theta p}{2} + p\log\left(\frac{\sqrt{1+\theta}}{1-\theta}\,\sigma_4(p)\right) \\ - \sum_{j=1}^{p} \log\langle a_{kj}, x_k - x_{k+1}^* \rangle. \end{aligned} \tag{3.13}$$

An upper bound for $F_{k+1}(x_k + \alpha\hat{d})$ is then given by

$$F_{k+1}(x_k + \alpha\hat{d}) \le F_k(x_k) - (1-\eta)\alpha - \log(1-\alpha) + \frac{\theta p}{2} + p\log\left(\frac{\sqrt{1+\theta}}{1-\theta}\,\frac{\sigma_4(p)}{\alpha}\right) - \sum_{j=1}^{p} \log\langle a_{kj}, x_k - x_{k+1}^*\rangle. \tag{3.14}$$

Summing (3.10) and (3.14) concludes the proof. □

Lemma 3.6 yields an upper-bound for the gap $F_{k+1}(x_k + \alpha\hat{d}) - F_{k+1}(x_{k+1}^*)$. Minimizing this bound in α provides the best step length along direction $\hat{d}$. Hence we set

$$\alpha^* = \arg\min_{0<\alpha\le 1} \left\{-(1-\eta)\alpha - \log(1-\alpha) - p\log\alpha\right\}.$$

We get $\alpha^* = \frac{\sqrt{4p+(p-\eta)^2}-(p+\eta)}{2\,(1-\eta)} = \sigma_3(p)$ (See Lemma B.3).

4 Computation of an approximate analytic center

With $x = x_k + \alpha\hat{d}$ as an initial point, we resort to a Newton scheme to minimize F_{k+1}. In a first stage, we perform damped Newton steps until an acceptable approximation is attained. In a second stage, starting from this approximate center, the standard Newton method with full steps converges quadratically to an η-approximate analytic center.

Recall that Newton's direction d_N is given by

$$d_N = -[F''_{k+1}(x)]^{-1}\, F'_{k+1}(x).$$

The scheme is designed as follows:

$$\begin{array}{ll} 0. & x \leftarrow x_k + \alpha\hat{d}, \\ 1. & \text{Define } d_N = -[F''_{k+1}(x)]^{-1}F'_{k+1}(x) \quad \text{and} \quad \lambda(x) = \| d_N \|_{F''_{k+1}(x)}, \\ 2. & \text{a) If } \lambda(x) \ge \frac{1}{3}, \text{ then } x \leftarrow x + \frac{1}{1+\lambda(x)}\, d_N, \\ & \text{b) If } \lambda(x) < \frac{1}{3}, \text{ then } x \leftarrow x + d_N, \\ & \text{c) If } \lambda(x) \le \eta, \text{ then STOP}, \\ 3. & \text{Goto 1.} \end{array} \tag{4.1}$$

Theorem 4.1 *Scheme (4.1) converges to an η-approximate center after at most $K_1 + K_2$ iterations, with*

$$K_1 = \left\lceil \frac{\Delta_k(\alpha)}{\frac{1}{3} - \log\frac{4}{3}} \right\rceil \quad \textit{and} \quad K_2 = \left\lceil \log_2\left(\frac{\log\frac{4}{9\eta}}{\log\frac{4}{3}}\right) \right\rceil.$$

Proof:

Each damped Newton step reduces the potential F_{k+1} by the value

$$\omega(\lambda(x)) \ge \omega(\frac{1}{3}) = \frac{1}{3} - \log\frac{4}{3}.$$

Lemma 3.6 provides an upper bound for the gap $\Delta_k(\alpha)$ between $F_{k+1}(x_k+\alpha\hat{d})$ and $F_{k+1}(x^*_{k+1})$. Dividing this gap by the guaranteed reduction $\frac{1}{3}-\log\frac{4}{3}$ at each damped Newton step, we obtain the desired bound for K_1.

Once a $\frac{1}{3}$-approximate center is attained, the scheme performs full Newton steps. Theorem 2.2.2 of [7] states that:

$$\lambda(x+d_N) \leq \left(\frac{\lambda(x)}{1-\lambda(x)}\right)^2 \leq \frac{1}{2}\lambda(x) < \frac{1}{3}.$$

We can see that $\hat{\lambda}(x)=\frac{9}{4}\lambda(x)$ satisfies

$$\hat{\lambda}(x+d_N) \leq (\hat{\lambda}(x))^2 \leq \frac{9}{16}.$$

This quadratic convergence property yields the following stopping criterion for the number k of full steps:

$$\hat{\lambda}(x) \leq \left(\frac{3}{4}\right)^{2^k} \leq \frac{9}{4}\,\eta.$$

Solving in k, we easily prove that only K_2 Newton steps are enough to reach an η-approximate center. □

Corollary 4.1 *For any $k \geq 1$, $O\left(p+\log(p+1)\right)$ damped Newton steps, followed by $O\left(\log\log\frac{1}{\eta}\right)$ full Newton steps generate an η-approximate analytic center.*

Note that the bound K_2 for the number of full Newton steps does not depend on p.

5 Conclusion

The present paper extends the homogeneous scheme proposed in [9] to multiple cuts by adapting the approach followed by Goffin and Vial [2]. Finding the re-entering direction after introducing p cuts is performed in $O(p\ln\omega p)$ in the general case, where ω is a problem dependent characteristic. In two cases of interest this complexity reduces to $O(p\ln p)$, which is independent of the problem data. After this initial step, finding a new analytic center through a damped Newton scheme is $O(p\ln p)$.

In this paper, the stopping criterion is defined w.r.t. $\mu_k(x)$ as in [9]. Indeed, it is shown in [9] that for an appropriate $\varepsilon>0$, the criterion $\mu_k(x)\leq\varepsilon$ yields a solution to the original problem (feasibility, convex optimization, variational inequality). From Theorem 2.3, the number of iterations to meet $\mu_k(x)\leq\varepsilon$ is easily shown to be $O(\frac{p\exp(\sqrt{\nu})}{\varepsilon^2})$. The number of cutting planes generated is $O(\frac{p^2\exp(\sqrt{\nu})}{\varepsilon^2})$. This bound is independent of the dimension n of the underlying space. Thus, for large n, the present method is optimal (see [5]).

Compared to the single cut case, the complexity is multiplied by a factor p. This negative result is compensated by the fact that multiple cuts contribute to restrict the localization set to a much smaller set from one iteration to another; an important speed-up can be expected.

Our paper assumes that all cuts are central, though in practice most cuts are deep. An obvious extension of the present research would be to deal with deep cuts.

References

[1] O. du Merle, J.-L. Goffin, and J.-Ph. Vial. "On Improvements to the Analytic Center Cutting Plane Method", *Computational Optimization and Applications*, **11**, pp. 37–52, (1998).

[2] J.-L. Goffin and J.-Ph. Vial. "Multiple cuts in the analytic center cutting plane method", *SIAM Journal on Optimization* **11**, pp. 266-288, (2000).

[3] K.L. Jones, I.J. Lustig, J.M. Farvolden, and W.B. Powell, "Multicommodity network flows: the impact of formulation on decomposition", *Mathematical Programming*, **62**, pp. 95–117, (1993).

[4] Z.-Q. Luo. "Analysis of a Cutting Plane Method That Uses Weighted Analytic Center and Multiple Cuts", *SIAM Journal on Optimization*, **4**, pp. 697–716, (1994).

[5] A. Nemirovsky and D. Yudin. "Informational Complexity and Efficient Methods for Solution of Convex Extremal Problems". J. Wiley & Sons, New York, (1983).

[6] Yu. Nesterov. "Introductory lectures on Convex Optimization". Unpublished manuscript, Louvain, Belgium, (1996).

[7] Yu. Nesterov and A. Nemirovsky. "Interior Point Polynomial Algorithms in Convex Programming : Theory and Applications", SIAM, Philadelphia, (1994).

[8] Yu. Nesterov, O. Péton and J.-Ph. Vial. "Homogeneous Analytic Center Cutting Plane Methods with Approximate Centers", *Optimization methods and software*, **11/12**, pp. 243–73, (1999).

[9] Yu. Nesterov and J.-Ph. Vial. "Homogeneous Analytic Center Cutting Plane Methods for Convex Problems and Variational Inequalities", *SIAM Journal on Optimization* **9**, pp. 707–728 (1999).

[10] Yu. Nesterov and J.-Ph. Vial. "Augmented self-concordant barriers and nonlinear optimization problems with finite complexity", Logilab Technical Report, University of Geneva, Switzerland, (2000).

[11] Y. Ye. "Complexity Analysis of the Analytic Center Cutting Plane Method That Uses Multiple Cuts", *Mathematical Programming*, **78**, pp. 85–104, (1997).

A Useful Properties of Self-Concordant Functions

This section introduces general results on self-concordant functions, that are used in the paper and in Appendix B for the convergence proof of scheme (2.2). Most of these results can be found in the notes [6] and the book [7]. Since the notes have not yet appeared in the open literature, we include the following results here for the sake of completeness.

Properties of self-concordant functions

Let f be a self-concordant function defined on some open domain. A point $x \in \operatorname{dom} f$ is called an η-approximate analytic center associated to f if

$$\| f'(x) \|_{[f''(x)]^{-1}} = \langle [f''(x)]^{-1} f'(x), f'(x) \rangle^{1/2} \leq \eta.$$

This definition also applies if f is a self-concordant or a ν-normal barrier. The analytic center is an η-approximate center with $\eta = 0$. We denote it x^*.

Let $t = \| y - x \|_{f''(x)}$.

Theorem A.1 (Theorem 4.1.5 of [6]) *If $t < 1$,*

$$\frac{t}{1+t} \leq \| x - y \|_{f''(y)} \leq \frac{t}{1-t}. \tag{A.1}$$

The left-hand side holds for any $x, y \in \dim f$.

Theorem A.2 (Theorem 4.1.6 of [6]) *Let $x, y \in \operatorname{dom} f$ such that $t < 1$, we have*

$$(1-t)^2 f''(x) \leq f''(y) \leq \frac{1}{(1-t)^2} f''(x).$$

Theorem A.3 (Theorems 4.1.7 and 4.1.8 of [6]) *If $t < 1$ then*

$$\frac{t^2}{1+t} \leq \langle f'(y) - f'(x), y - x \rangle \leq \frac{t^2}{1-t}, \tag{A.2}$$

$$f(x) + \langle f'(x), y - x \rangle + \omega(t) \leq f(y) \leq f(x) + \langle f'(x), y - x \rangle + \omega_*(t), \tag{A.3}$$

where

$$\omega(t) = t - \log(1+t), \text{ and } \omega_*(t) = -t - \log(1-t).$$

Besides, the left-hand sides hold for all t.

Theorem A.4 (Theorem 4.1.11 of [6]) *For any η-approximate center $x \in \operatorname{dom} f$, we have*

$$\frac{\eta}{1+\eta} \leq \| x - x^* \|_{f''(x)} \leq \frac{\eta}{1-\eta}. \tag{A.4}$$

Corollary A.1 *Let x^* and x be respectively the exact and an η-approximate analytic center of* $\operatorname{dom} f$*, with $0 \leq \eta < \frac{1}{3}$, and $y \in \operatorname{dom} f$. Then,*

$$\| x^* - y \|_{f''(x)} \leq \frac{1-2\,\eta}{1-3\,\eta} \; \| x^* - y \|_{f''(x^*)} \, .$$

Proof:

In view of Theorem A.2 we have

$$f''(x) \preceq \frac{1}{(1 - \| x - x^* \|_{f''(x^*)})^2} \; f''(x^*).$$

Using (A.1) and (A.4), we get

$$f''(x) \preceq \frac{1}{\left(1 - \frac{\frac{\eta}{1-\eta}}{1-\frac{\eta}{1-\eta}}\right)^2} f''(x^*) = \left(\frac{1-2\eta}{1-3\eta}\right)^2 f''(x^*).$$

Thus,

$$\langle f''(x)(x^* - y), x^* - y \rangle \leq \left(\frac{1-2\eta}{1-3\eta}\right)^2 \langle f''(x^*)(x^* - y), x^* - y \rangle.$$

□

Properties of ν-normal barriers

Definition A.1 *A function F is a ν-normal barrier for K if it is self-concordant and logarithmically homogeneous, that is*

$$F(\tau x) = F(x) - \nu \ln \tau, \qquad \forall x \in \operatorname{int} K, \; \tau > 0.$$

Corollary 2.3.2 of [6] states that a ν-normal barrier is a ν-self-concordant-barrier for its domain. Finally, since f is logarithmically homogeneous, we have:

$$\langle f''(x)x, x \rangle = \nu. \tag{A.5}$$

This result follows from the definition of the logarithmically homogeneous barrier by differentiating twice this identity in τ at $\tau = 1$.

B Convergence of the Homogeneous Scheme

This appendix contains the convergence proof of scheme (2.2). It is a straighforward extension of the convergence proofs of [8] and [9] to the multiple cuts case.

Let us write down the expression of function F_k:

$$F_k(x) = \frac{1}{2} \| x \|^2 + F(x) - \sum_{i=0}^{k-1} \sum_{j=1}^{p} \log \langle a_{ij}, x_i - x \rangle,$$

where a_{ij} stands for $a_j(x_i)$. We also have

$$F_k'(x) = x + F'(x) + \sum_{i=0}^{k-1}\sum_{j=1}^{p} \frac{a_{ij}}{\langle a_{ij}, x_i - x\rangle}, \tag{B.1}$$

and

$$F_k''(x) = I + F''(x) + \sum_{i=0}^{k-1}\sum_{j=1}^{p} \frac{a_{ij}a_{ij}^T}{\langle a_{ij}, x_i - x\rangle^2}. \tag{B.2}$$

Lemma B.1 *The inequality*

$$|\parallel x_k \parallel^2 - (\nu + kp)| \leq \sqrt{\frac{2\eta^2}{1-\eta}}\sqrt{\nu + kp}$$

holds for all $k \geq 0$.

For $p = 1$, Lemma B.1 is the same as Lemma 1 of [8]. For any $p > 1$, the proof follows exactly the same pattern. The interested reader will refer to [8] to read the proof.

Lemma B.2 *1. For any* $x \in K$, *we have*

$$\mu_k(x) \leq \frac{1}{S_k}\left(\parallel x_k \parallel + \eta[S_k + 1] \right) \parallel x \parallel . \tag{B.3}$$

2. If $x \in K$ *satisfies the inequalities*

$$\langle a_{ij}, x_i - x\rangle \geq 0, \quad i = 1, \ldots, k-1, \quad j = 1, \ldots, p,$$

then

$$\mu_k(x) \leq \frac{1}{(1-\eta^2)S_k}\left(\parallel x_k \parallel + \eta\sqrt{1 - \eta^2 + \parallel x_k \parallel^2}\right) \parallel x \parallel . \tag{B.4}$$

Proof:

For any $x \in K$, we have

$$\begin{aligned} S_k\mu_k(x) &= \sum_{i=0}^{k-1}\sum_{j=1}^{p} \lambda_{ijk}\langle a_{ij}, x_i - x\rangle = -\sum_{i=0}^{k-1}\sum_{j=1}^{p} \lambda_{ijk}\langle a_{ij}, x\rangle, \\ &= \langle x_k, x\rangle + \langle F'(x_k), x\rangle - \langle F_k'(x_k), x\rangle, \\ &\leq \parallel x_k \parallel \cdot \parallel x \parallel + \langle F'(x_k), x\rangle + \eta \parallel x \parallel_{F_k''(x_k)} . \end{aligned}$$

Using (B.2), we may write

$$\parallel x \parallel^2_{F_k''(x_k)} = \parallel x \parallel^2 + \langle F''(x_k)x, x\rangle + \sum_{i=0}^{k-1}\sum_{j=1}^{p} \lambda_{ijk}^2 \langle a_{ij}, x\rangle^2.$$

Since x is a recession direction[2] of the cone K,

$$\| x \|^2_{F''_k(x_k)} \leq \| x \|^2 + \langle F'(x_k), x\rangle^2 + \sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk}^2 \langle a_{ij}, x\rangle^2.$$

Thus,

$$S_k \mu_k(x) \leq \| x_k \| \cdot \| x \| + \langle F'(x_k), x\rangle + \eta \left[\| x \|^2 + \langle F'(x_k), x\rangle^2 + \sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk}^2 \langle a_{ij}, x\rangle^2 \right]^{\frac{1}{2}}. \tag{B.5}$$

In view of equation (2.3.13) of [7] and Theorem 4.2.4 of [6], we have respectively $\langle F'(x_k), x_k\rangle = -\nu$ and $\langle F'(x_k), x - x_k\rangle \leq \nu$. Thus, $\langle F'(x_k), x\rangle \leq 0$. Since $\eta < 1$, the right-hand side of (B.5) is an increasing function of $\langle F'(x_k), x\rangle$. Hence, we can exhibit a simpler upper bound for $S_k \mu_k(x)$ by replacing this term by 0 in inequality (B.5). Therefore,

$$S_k \mu_k(x) \leq \| x_k \| \cdot \| x \| + \eta \left[\| x \|^2 + \sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk}^2 \langle a_{ij}, x\rangle^2 \right]^{\frac{1}{2}}. \tag{B.6}$$

By Assumption 2.2, $\| a_{ij} \| = 1$ and $\langle a_i, x\rangle^2 \leq \| x \|^2$. Therefore

$$\sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk}^2 \langle a_{ij}, x\rangle^2 \leq \| x \|^2 \ \sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk}^2 \leq \| x \|^2 \ \ S_k^2.$$

Putting $\| x \|$ as a common factor in (B.6), and noting that $\sqrt{1 + S_k^2} \leq 1 + S_k$, we prove (B.3).

Assume further that $\langle a_{ij}, x_i - x\rangle \geq 0$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, p$. Then, we can prove that

$$\sum_{i=0}^{k-1} \sum_{j=1}^{p} \lambda_{ijk}^2 \langle a_{ij}, x\rangle^2 \leq S_k^2 \mu_k^2(x).$$

In view of (B.6) we have

$$S_k \mu_k(x) \leq \|x_k\| \cdot \|x\| + \eta \left[\| x \|^2 + S_k^2 \mu_k^2(x) \right]^{\frac{1}{2}}.$$

Solving this inequality in $\mu_k(x)$ we get (B.4). □

Recall that x_k^* and x_k denote respectively the exact analytic center and an η-approximate analytic center at iteration k.

[2]If $F(x)$ is a self-concordant barrier for a set Q and $x \in \operatorname{int} Q$ then $\langle F''(x)h, h\rangle \leq \langle F'(x), h\rangle^2$ for any recession direction h of the set Q. This result follows from Corollary 2.3.1 of [7].

Lemma B.3 *For any* $k \geq 0$,

$$\sigma_3(p) \leq \| x^*_{k+1} - x_k \|_{F''_k(x_k)} \leq \sigma_4(p),$$

where

$$\sigma_3(p) = \frac{\sqrt{4p + (p-\eta)^2} - (p+\eta)}{2\,(1-\eta)}$$

and

$$\sigma_4(p) = \frac{\sqrt{4p + (p-\eta)^2} + (p+\eta)}{2\,(1-\eta)}.$$

Proof:

$x^*_{k+1} = \arg\min F_{k+1}(x)$ is the exact analytic center at step $k+1$. The first order optimality condition at this point is

$$F'_{k+1}(x^*_{k+1}) = F'_k(x^*_{k+1}) + \sum_{j=1}^{p} \frac{a_{kj}}{\langle a_{kj}, x_k - x^*_{k+1} \rangle} = 0. \tag{B.7}$$

Multiplying (B.7) by $x^*_{k+1} - x_k$, and adding $\langle F'_k(x_k), x_k - x^*_{k+1} \rangle$ on both sides, we obtain

$$\langle F'_k(x^*_{k+1}) - F'_k(x_k), x^*_{k+1} - x_k \rangle = p + \langle F'_k(x_k), x_k - x^*_{k+1} \rangle. \tag{B.8}$$

Denote $s = \| x_k - x^*_{k+1} \|_{F''_k(x_k)}$. In view of (A.2), we have a lower bound for the left-hand side

$$\frac{s^2}{1+s} \leq \langle F'_k(x^*_{k+1}) - F'_k(x_k), x^*_{k+1} - x_k \rangle.$$

In view of the centering condition, we can bound the right-hand side of equation (B.8) by

$$p + \langle F'_k(x_k), x_k - x^*_{k+1} \rangle \leq\ p + \eta\, s.$$

Therefore, the following inequality is valid for s

$$\frac{s^2}{1+s} \leq p + \eta\, s.$$

Solving for s, we have

$$s \leq \frac{\sqrt{4p + (p-\eta)^2} + p + \eta}{2\,(1-\eta)} = \sigma_4(p).$$

On the other hand, assume that $s < 1$. Using inequality (A.2) and the Cauchy-Schwartz inequality in equation (B.8), we obtain

$$p \leq \frac{s^2}{1-s} + \eta s.$$

Solving for s, we get

$$s \geq \frac{\sqrt{4p + (p-\eta)^2} - (p+\eta)}{2(1-\eta)} = \sigma_3(p).$$

Note that $\sigma_3(p) < 1$. Hence if we assume $s \geq 1$, the bound $s \geq \sigma_3(p)$ remains valid. □

Lemma B.4 *For any $k \geq 0$,*

$$F_{k+1}(x_{k+1}) \geq F_{k+1}(x^*_{k+1}) \geq F_k(x_k) + \sigma_5(p),$$

with $\sigma_5(p) = \sigma_3(p) - \eta\sigma_4(p) - \log(1+\sigma_3(p)) - p\log\sigma_4(p)$.

Proof:

The inequality on the left holds trivially. From inequality (A.3), we get

$$F_k(x^*_{k+1}) \geq F_k(x_k) + \langle F'_k(x_k), x^*_{k+1} - x_k\rangle + \omega(\| x^*_{k+1} - x_k \|_{F''_k(x_k)}).$$

where $\omega(t) = t - \log(1+t)$. Using Lemma B.3, and since ω is an increasing function, we get

$$F_k(x^*_{k+1}) \geq F_k(x_k) - \eta\sigma_4(p) + \sigma_3(p) - \log(1+\sigma_3(p)). \tag{B.9}$$

Since $\| a(x) \|= 1$, we get by Cauchy-Schwartz inequality

$$\begin{aligned} F_{k+1}(x^*_{k+1}) &= F_k(x^*_{k+1}) - \sum_{j=1}^{p} \log\langle a_{kj}, x_k - x^*_{k+1}\rangle \\ &\geq F_k(x_k) + \sigma_3(p) - \log(1+\sigma_3(p)) - \eta\sigma_4(p) - p\log\| x_k - x^*_{k+1} \|. \end{aligned}$$

Since $I \preceq F''_k$, we have

$$\| x_k - x^*_{k+1} \| \leq \| x_k - x^*_{k+1} \|_{F''_k(x_k)} \leq \sigma_4(p).$$

Hence

$$F_{k+1}(x^*_{k+1}) \geq F_k(x_k) + \sigma_3(p) - \log(1+\sigma_3(p)) - \eta\sigma_4(p) - p\log\sigma_4(p).$$

□

Lemma B.5 *For any $k \geq 0$ we have*

$$S_k \geq k\,\sigma_1(p)\exp\left[\frac{F(x_0) - F(x_k)}{kp}\right],$$

where $\sigma_1(p) = p\exp\left\{-\frac{1}{2} + \frac{\sigma_5(p)}{p} - \frac{1}{kp}\sqrt{\nu + kp}\right\}$.

Proof:

From the definition of F_k,

$$\sum_{i=0}^{k-1}\sum_{j=1}^{p} \log\frac{1}{\langle a_{ij}, x_i - x_k\rangle} = -\frac{1}{2}\| x_k \|^2 - F(x_k) + F_k(x_k). \tag{B.10}$$

By Lemma B.1

$$\frac{1}{2}\| x_k \|^2 \leq \frac{\nu + kp}{2} + \frac{\eta\sqrt{\nu + kp}}{\sqrt{2(1-\eta)}}, \tag{B.11}$$

while by repetitive use of Lemma B.4 and Lemma B.1 at x_0,

$$\begin{aligned} F_k(x_k) &\geq F_0(x_0) + k\sigma_5(p), \\ &\geq F(x_0) + k\sigma_5(p) + \frac{\nu}{2} - \frac{\eta\sqrt{\nu}}{\sqrt{2(1-\eta)}}. \end{aligned} \tag{B.12}$$

Using (B.11) and (B.12) in (B.10), we get

$$\sum_{i=0}^{k-1}\sum_{j=1}^{p} \log\frac{1}{\langle a_{ij}, x_i - x_k\rangle} \geq F(x_0) - F(x_k) + k\left(\sigma_5(p) - \frac{p}{2}\right) - \frac{\eta\left(\sqrt{\nu} + \sqrt{\nu + kp}\,\right)}{\sqrt{2(1-\eta)}}.$$

Since $\eta < \frac{1}{2}$ and $\sqrt{\nu} \leq \sqrt{\nu + kp}$, we get

$$\sum_{i=0}^{k-1}\sum_{j=1}^{p} \log\frac{1}{\langle a_{ij}, x_i - x_k\rangle} \geq F(x_0) - F(x_k) + k\left(\sigma_5(p) - \frac{p}{2}\right) - \sqrt{\nu + kp}.$$

Using the inequality between the arithmetic and the geometric means, we have

$$S_k = \sum_{i=0}^{k-1}\sum_{j=1}^{p} \frac{1}{\langle a_{ij}, x_i - x_k\rangle} \geq kp\exp\left\{\frac{1}{kp}\sum_{i=0}^{k-1}\sum_{j=1}^{p} \log\frac{1}{\langle a_{ij}, x_i - x_k\rangle}\right\}.$$

Thus, we obtain the desired bound on S_k

$$S_k \geq kp\exp\left\{-\tfrac{1}{2} + \frac{\sigma_5(p)}{p} - \frac{1}{kp}\sqrt{\nu + kp}\right\}\exp\left\{\frac{F(x_0) - F(x_k)}{kp}\right\}.$$

□

Remark B.1 *Note that* $\sigma_3(p) - \log(1 + \sigma_3(p)) \geq 0$, *then* $\sigma_5(p) \geq -\eta\sigma_4(p) - p\log\sigma_4(p)$. *We also have* $\frac{\sqrt{\nu+kp}}{kp} \leq \sqrt{\nu+1}$. *Using these bounds in the definition of* $\sigma_1(p)$, *we get*

$$\sigma_1(p) \geq \frac{p}{\sigma_4(p)}\exp\left\{-\tfrac{1}{2} - \sqrt{\nu+1} - \frac{\eta\sigma_4(p)}{p}\right\}.$$

One easily checks that $\sigma_4(p)/p$ *is a decreasing function of* p. *Therefore,*

$$\sigma_1(p) \geq \frac{1}{\sigma_4(1)}\exp\left\{-\tfrac{1}{2} - \sqrt{\nu+1} - \eta\sigma_4(1)\right\}.$$

Therefore, $\sigma_1(p)$ *is bounded from below by an absolute constant.*

Lemma B.6 *For any* $k \geq 0$,

$$\| x_{k+1} - x_k \|_{F_k''(x_{k+1})} \leq \sigma_2(p)$$

where

$$\sigma_2(p) = \frac{1-2\eta}{2(1-3\eta)}\,(p + \eta\sigma_4(p))\left(1 + \sqrt{1 + \frac{4}{p + \eta\sigma_4(p)}}\right) + \frac{\eta}{1-\eta}.$$

Proof:

To prove the lemma, we resort to the triangular inequality

$$\| x_{k+1} - x_k \|_{F''_k(x_{k+1})} \leq \| x_{k+1} - x^*_{k+1} \|_{F''_k(x_{k+1})} + \| x^*_{k+1} - x_k \|_{F''_k(x_{k+1})} \,. \tag{B.13}$$

We shall bound separately the two terms in the right-hand side. Since $F''_k \preceq F''_{k+1}$, and since x_{k+1} is an approximate center at iteration $k+1$, we get by Theorem A.4

$$\| x_{k+1} - x^*_{k+1} \|_{F''_k(x_{k+1})} \leq \| x_{k+1} - x^*_{k+1} \|_{F''_{k+1}(x_{k+1})} \leq \frac{\eta}{1-\eta}. \tag{B.14}$$

From Corollary A.1, we may write

$$\| x^*_{k+1} - x_k \|_{F''_k(x_{k+1})} \leq \frac{1-2\eta}{1-3\eta} \| x^*_{k+1} - x_k \|_{F''_k(x^*_{k+1})} \,. \tag{B.15}$$

In view of (B.8) and (A.2), we have

$$\frac{\| x^*_{k+1} - x_k \|^2_{F''_k(x^*_{k+1})}}{1 + \| x^*_{k+1} - x_k \|_{F''_k(x^*_{k+1})}} \leq p + \langle F'_k(x_k), x_k - x^*_{k+1} \rangle.$$

Using Lemma B.3, we obtain

$$\frac{\| x^*_{k+1} - x_k \|^2_{F''_k(x^*_{k+1})}}{1 + \| x^*_{k+1} - x_k \|_{F''_k(x^*_{k+1})}} \leq p + \eta\sigma_4(p).$$

$\| x^*_{k+1} - x_k \|_{F''_k(x^*_{k+1})}$ must be smaller than the larger root of

$$\| x^*_{k+1} - x_k \|^2_{F''_k(x^*_{k+1})} - \left[1 + \| x^*_{k+1} - x_k \|_{F''_k(x^*_{k+1})} \right] (p + \eta\sigma_4(p)) = 0.$$

Thus

$$\| x^*_{k+1} - x_k \|_{F''_k(x^*_{k+1})} \leq \tfrac{1}{2} \left[p + \eta\sigma_4(p) + \sqrt{(p+\eta\sigma_4(p))^2 + 4(p+\eta\sigma_4(p))} \right].$$

Using (B.15), we have

$$\| x^*_{k+1} - x_k \|_{F''_k(x_{k+1})} \leq \frac{1-2\eta}{2(1-3\eta)} \left[\sqrt{(p+\eta\sigma_4(p))^2 + 4(p+\eta\sigma_4(p))} + p + \eta\sigma_4(p) \right]. \tag{B.16}$$

Inserting the bounds (B.14) and (B.16) into the decomposition (B.13) proves the lemma. □

Lemma B.7 *For any $k > 0$,* $F(x_k) - F(x_0) \leq k \sqrt{\nu}\, \sigma_2(p)$.

Proof:

This lemma is a direct extension of Lemma 7 of [8] to the multicut case [replace θ_2 in [8] by $\sigma_2(p)$]. □

In view of Lemmas B.1 and B.2, we introduce the following constants:

$$\begin{aligned}
\psi_1(\eta, k, p) &= \sqrt{\nu + kp + \sqrt{\frac{2\eta^2}{1-\eta}(\nu + kp)}}, \\
\psi_2(\eta, k, p) &= \frac{1}{1-\eta}\psi_1(\eta, k, p) + \frac{\eta}{\sqrt{1-\eta^2}}.
\end{aligned}$$

Since $\eta \leq \frac{1}{3}$, it can be easily shown that $\psi_1(\eta, k, p)$ and $\psi_2(\eta, k, p)$ are roughly proportional to $\sqrt{\nu + kp}$. The following bounds hold :

$$\sqrt{\nu + kp} \leq \psi_1(\eta, k, p) \leq \psi_2(\eta, k, p) \leq \left(\frac{1 + 3\eta + \eta^2}{1 - \eta^2}\right)\sqrt{\nu + kp}. \tag{B.17}$$

From Lemma B.1, we have $\| x_k \| \leq \psi_1(\eta, k, p)$. Combining Lemmas B.2, B.5 and B.7 proves the main convergence result (Theorem 2.3).