Multiple cuts in the analytic center cutting plane method

Jean-Louis GOFFIN* Jean-Philippe VIAL**

HEC/Logilab Technical Report 98.10
June 4, 1998; revised April, 1999, and October, 1999.

Abstract

We analyze the multiple cut generation scheme in the analytic center cutting plane method. We propose an optimal primal and dual updating direction when the cuts are central. The direction is optimal in the sense that it maximizes the product of the new dual slacks and of the new primal variables within the trust regions defined by Dikin's primal and dual ellipsoids. The new primal and dual directions use the variance–covariance matrix of the normals to the new cuts in the metric given by Dikin's ellipsoid.

We prove that the recovery of a new analytic center from the optimal restoration direction can be done in $O(p\log(p+1))$ damped Newton steps, where p is the number of new cuts added by the oracle, which may vary with the iteration. The results and the proofs are independent of the specific scaling matrix —primal, dual or primal-dual— that is used in the computations.

The computation of the optimal direction uses Newton's method applied to a self-concordant function of p variables.

The convergence result of [19] holds here also: the algorithm stops after $O^*(\frac{\bar{p}^2 n^2}{\varepsilon^2})$ cutting planes have been generated, where $\bar{p}$ is the maximum number of cuts generated at any given iteration..

Keywords Primal Newton algorithm, Analytic center, Cutting Plane Method, Multiple cuts.

This work has been completed with the support from the Fonds National Suisse de la Recherche Scientifique, grant 12-42503.94, from the Natural Sciences and Engineering Research Council of Canada, grant number OPG0004152 and from the FCAR of Quebec.

* GERAD/Faculty of Management, McGill University, 1001, Sherbrooke West, Montreal, Que., H3A 1G5, Canada. E-mail: ma56@musica.mcgill.ca.

** LOGILAB/Management Studies, University of Geneva, 102, Bd Carl-Vogt, CH-1211 Genève 4, Switzerland. E-mail: jpvial@uni2a.unige.ch.

1 Introduction

The analytic center cutting plane (ACCPM) algorithm [5, 18] is an efficient algorithm in practice [2, 4]. The complexity of related algorithms was given in [1, 13], and subsequently in [6]. Extensions to deep cuts were given in [7] and to very deep cuts in [8]. The method studied in [8] corresponds to the practical implementation of ACCPM [11] with a single cut.

In practice, it often occurs that the oracle in the cutting plane scheme generates multiple cuts. The paper by Ye [19] show that it is possible to handle several cuts at a time provided they are central; the direction used is the primal direction suggested by Mitchell and Todd [12]. Although this analysis show how one can recover feasibility after introducing multiple cuts, there is no clear argument as to the choice of a feasibility restoration direction. Intuitive, but well justified, arguments about how to introduce multiple cuts were given in [2] in the context of a primal projective algorithm and two cuts (one shallow, one deep) and in [10] with an infeasible primal–dual approach to the introduction of several cuts in general position.

The case of two central cuts was analyzed in [9]. It was shown that there exist explicit primal and dual directions which allow a best move towards primal and dual feasibility. An argument using the primal, dual and primal–dual potentials at this new optimal primal and dual point proves that $O(1)$ damped Newton steps are enough to recover centrality. The updating direction depends on the cosine in the metric of Dikin's ellipsoid of the normals to the cuts.

In this paper, we analyze the multiple central cut generation scheme in the analytic center cutting plane method. An approach based upon weighted potentials applied to the primal direction proposed in [12] and studied in [19] leads to a number of recentering steps that depends upon the total number of cuts and on the data.

We propose an optimal updating direction when the cuts are central. The direction is optimal in the sense that it maximizes the product of the new slacks within the trust region defined by Dikin's ellipsoid. The new primal and dual directions use the variance–covariance matrix of the normals to the new cuts in the metric given by Dikin's ellipsoid.

We prove that the recovery of a new analytic center from the optimal restoration point can be done in $O(p \log(p+1))$ damped Newton steps, where p is the number of new cuts added by the oracle; the number of cuts may vary at each iteration. The number of damped Newton steps does not depend upon the data, i.e. it is strongly polynomial.

The results and the proofs are independent of the specific scaling matrix — primal, dual or primal-dual— that is used in the computations. The proof of a complexity that does not depend on the data of the problem relies on the use of the primal-dual potential function as a proximity measure.

The computation of the optimal direction uses Newton's method applied to a self-concordant function of p variables. The number of iterations needed is polynomial in the problem data; but this could be very advantageous in practice if the number of cuts p is a small multiple of n, the dimension of the space, as Newton's method takes place in a p-dimensional space.

The argument made here is very classical in nonlinear optimization, and involves computing an optimal direction within a trust region, here defined by Dikin's ellipsoid, and only then seaching for the new center.

The convergence result of [19] holds here also: the algorithm stops after $O^*(\frac{\bar{p}^2 n^2}{\varepsilon^2})$ cutting planes have been generated, where $\bar{p}$ is the maximum number of cuts generated by the oracle at any iteration. No improvement on this result can be offered here, as the worst case answer from the oracle is p copies of the same cutting plane, in which case the optimal direction proposed here is the same as the one studied in [19]. The long–step analysis given in [19] shows an average number of Newton steps of $O(\bar{p})$.

2 Analytic center cutting plane method

2.1 Cutting planes

The problem of interest is that of finding a point in a convex set $C \subset \mathbb{R}^n$. We make the following assumptions.

Assumption 2.1 *The set C is convex, contains a ball of radius $\varepsilon > 0$ and is contained in the cube $0 \leq y \leq e$.*

Assumption 2.2 *The set C is described by an oracle. That is, the oracle either confirms that $y \in C$, or answers at least one cutting plane that contains C and does not contain y in its interior.*

A cut at $\bar{y} \notin C$ takes the form

$$a^T y \leq a^T \bar{y} - \bar{\gamma}.$$

If $\bar{\gamma} > 0$, the cut is deep; if $\bar{\gamma} < 0$, the cut is shallow; if $\bar{\gamma} = 0$, the cut passes through $\bar{y}$, and we will refer to this as a central cut.

The algorithm may generate multiple cuts at a time. They take the form

$$a_j^T y \leq a_j^T \bar{y} - \bar{\gamma}_j,\ j = 1, \dots, p,\ \forall y \in C.$$

We define the matrix B by

$$B = (a_1, a_2, \dots, a_p);$$

p may vary at each iteration and, when necessary, this shall be denoted as p_k.

Assumption 2.3 *All the cutting planes generated have been scaled so that $\|a\| = 1$ (wlog). We also assume that $\bar{\gamma} = 0$, and thus that all cuts go through $\bar{y}$.*

A cutting plane algorithm constructs a sequence of query points $\{y^k\}$. The answers of the oracle to the queries, together with the cube $0 \leq y \leq e$, define a polyhedral outer approximation

$$\mathcal{F}_D = \{y \,:\, A^T y \leq c\}$$

of C. Since A contains the identity matrix associated with the cube, A has full row rank. Therefore there is a one-to-one correspondence between points $y \in \mathcal{F}_D$ and the slack $s = c - A^T y$, leading to the equivalent definition of $\mathcal{F}_D$

$$\mathcal{F}_D = \{s \geq 0 \,:\, A^T y + s = c\}.$$

The number of columns in A is denoted as m (or m_k) and is equal to $2n$ plus the number of cutting planes generated until the k^{th} iteration; i.e. $m_k = 2n + \sum_{j=0}^{k-1} p_j \leq 2n + k * \bar{p}$.

The analytic center cutting plane method chooses as a query point an approximate analytic center of $\mathcal{F}_D$.

2.2 Analytic center

The analytic center of $\mathcal{F}_D$ is the unique point maximizing the dual potential

$$\varphi_D(s) = \sum_{i=1}^{m} \log s_i$$

with $s = c - A^T y > 0$. We formally introduce the optimization problem

$$\max\left\{\varphi_D(s) \,:\, s = c - A^T y > 0\right\} \tag{1}$$

and the associated first order optimality conditions

$$\begin{aligned} xs &= e, & & \\ A^T y + s &= c, & s &> 0, \\ Ax &= 0, & x &> 0, \end{aligned}$$

where x is a vector in R^m. The notation xs indicates the Hadamard or componentwise product of the two vectors x and s.

The analytic center can alternatively be defined as the optimal solution of

$$\max\left\{\varphi_P(x) \,:\, Ax = 0,\, x > 0\right\}, \tag{2}$$

where

$$\varphi_P(x) = -c^T x + \sum_{i=1}^{m} \log x_i$$

denotes the primal potential. One easily checks that problem (2) shares with (1) the same first order optimality conditions.

At this stage it is convenient to introduce the primal-dual potential

$$\varphi_{PD}(x, s) = \varphi_P(x) + \varphi_D(s),$$

and an associated duality relationship.

Lemma 2.4 *Let $x \in \mathrm{int}\mathcal{F}_P$ and $s \in \mathrm{int}\mathcal{F}_D$. Then $\varphi_{PD}(x, s) \leq -m$, with equality if and only if $xs = e$.*

Proof

Consider the simple inequality

$$\log t \leq t - 1, \forall t > 0, \tag{3}$$

with equality if and only if $t = 1$. Let $x \in \mathrm{int}\mathcal{F}_P$ and $s \in \mathrm{int}\mathcal{F}_D$. Apply (3) with $t = x_i s_i$. By summing the resulting inequalities, one gets

$$\sum_{i=1}^{m} \log x_i + \sum_{i=1}^{m} \log s_i \leq x^T s - m = c^T x - m,$$

with equality if and only if $xs = e$. Therefore,

$$\varphi_P(x) + \varphi_D(s) \leq -m, \tag{4}$$

with equality if and only if $xs = e$. ■

Finally, we define approximate centers by relaxing the condition $xs = e$ in the first order optimality conditions. Formally, any solution (x, s) of

$$\begin{aligned} \|e - xs\| &\leq \theta < 1, && (5)\\ A^T y + s &= c, \quad s > 0, && (6)\\ Ax &= 0, \quad x > 0. && (7) \end{aligned}$$

defines a pair of θ-approximate centers, or θ-centers in short.

2.3 Analytic center cutting plane method

ACCPM can be shortly stated as follows.

Initialization Let $\mathcal{F}_D^0 = \{y \geq 0 \,:\, y \leq e\}$ be the unit cube and $y^0 = \frac{1}{2}e$ be its center. The centering parameter is $0 < \theta < 1$.

Basic Step y^k is a θ-center of $\mathcal{F}_D^k$; $m_k = 2n + \sum_{j=0}^{k-1} p_j$ the total number hyperplanes describing $\mathcal{F}_D^k$.

1. The oracle returns the cuts a_{m_k+j}, $j = 1, \ldots p_k$, at y^k.
2. Update
   $\mathcal{F}_D^{k+1} = \mathcal{F}_D^k \cap \{y \,:\, a_{m_k+j}^T(y - y^k) \leq 0,\, j = 1, \ldots p_k\}$.
3. Compute a θ-center of $\mathcal{F}_D^{k+1}$.

The computation of a new θ-center after adding new cuts will be discussed in a further section.

3 Some useful properties

The literature on interior point methods essentially proposes three approaches for computing analytic centers. All of them are based on Newton's method. The primal (resp. dual) Newton direction is initiated at an interior primal (resp. dual) feasible point; it involves the scaling matrix $D = X$ (resp. $D = S^{-1}$). (We recall the standard notation X which denotes the diagonal matrix diag(x).) The primal-dual direction is initiated at an interior primal-dual feasible pair, i.e., $(x, s) \in \text{int}\mathcal{F}_P \times \text{int}\mathcal{F}_D$; it involves the scaling matrix $D = (XS^{-1})^{1/2}$.

Let us shortly recall the formulas. The primal direction is given by $\Delta x = xp(x)$ with $p(x) = e - xs(x)$, $s(x) = c - A^T(AD^2A^T)^{-1}AX^2c$ and $D = X$. The dual direction is given by $\Delta s = sq(s)$ with $q(s) = e - sx(s)$, $x(s) = (I - DA^T(AD^2A^T)^{-1}AD)e$ and $D = S^{-1}$. Finally the primal dual direction is $\Delta s = A^T(AD^2A^T)^{-1}As^{-1}$, $\Delta x = -x + s^{-1} - D^2\Delta s$ and $D = (XS^{-1})^{1/2}$.

3.1 Properties of the Newton step

There are two basic properties, a local one in the vicinity of the analytic center, and a global one. Since the results are well-known we state them without proofs. Missing proofs can be found in the books [17] or [20].

Let us start with the local properties. Proximity to analytic center is measured with the quantity $\|e - sx\|$. In this definition, either *i)* $x \in \text{int}\mathcal{F}_P$ and $s = s(x)$, (primal case),

ii) $s \in \text{int}\mathcal{F}_D$ and $x = x(s)$, (dual case),

iii) $x \in \text{int}\mathcal{F}_P$ and $s \in \text{int}\mathcal{F}_D$, (primal-dual case).

Note that if $\|e - sx\| \le \theta < 1$ then $s(x) > 0$ and thus $s(x) \in \mathcal{F}_D$ (primal case), and $x(s) > 0$ and thus $x(s) \in \mathcal{F}_P$ (dual case). The Newton step defines a pair (x^+, s^+) as follows.

i) $x^+ = x + \Delta x$ and $s^+ = s(x^+)$ (primal case),

ii) $s^+ = s + \Delta s$ and $x^+ = x(s^+)$ (dual case),

iii) $x^+ = x + \Delta x$ and $s^+ = s + \Delta s$ (primal-dual case).

Theorem 3.1 *Assume $\|e - sx\| \le \theta < \frac{2}{3}$. Let (x^+, s^+) be the point resulting from a Newton step (primal, dual, or primal-dual). Then, $(s^+, x^+) \in int\mathcal{F}_D \times int\mathcal{F}_P$.*

In the primal and dual cases, the theorem holds with any $0 < \theta < 1$.

One can derive from the above theorem a useful corollary that yields lower bounds on the potentials near the analytic center. Let (x^c, y^c) be the pair of exact analytic centers. Denote $\varphi_P^c = \varphi_P(x^c)$ and $\varphi_D^c = \varphi_D(s^c)$.

Corollary 3.2 *Assume (5)–(7) at (x, s). Then*

1. $\varphi_P^c \ge \varphi_P(x) \ge \varphi_P^c - \frac{\theta^2}{1-\theta^2}$.
2. $\varphi_D^c \ge \varphi_D(s) \ge \varphi_D^c - \frac{\theta^2}{1-\theta^2}$.
3. $-m \ge \varphi_{PD}(x, s) \ge -m - 2\frac{\theta^2}{1-\theta^2}$.

Let us now consider the global properties of a damped Newton step. The properties are consequences of the well-known inequality on the logarithm function [17] p 439.

Lemma 3.3 *Let h be any point in R^m such that $\|h\| < 1$. Then,*

$$\sum_{i=1}^{m} \log(1 + h_i) \ge e^T h + \|h\| + \log(1 - \|h\|).$$

The main result bounds the variation of the potentials after a damped Newton step.

Theorem 3.4 *Assume $\|e - xs\| \ge \theta > 0$, where $0 < \theta < 1$. Define $x(\alpha) = x + \alpha\Delta x$ and $s(\alpha) = s + \alpha\Delta s$. ($\Delta x$ and Δs may be the primal, dual or primal-dual directions.) Then, there exists a step size $\alpha > 0$ and absolute constants σ_P, σ_D and σ_{PD} such that*

1. $\varphi_P(x(\alpha)) \geq \varphi_P(x) + \sigma_P$;
2. $\varphi_D(x(\alpha)) \geq \varphi_D(s) + \sigma_D$;
3. $\varphi_{PD}(x(\alpha), s(\alpha)) \geq \varphi_{PD}(x, s) + \sigma_{PD}$.

In the primal and dual cases the constants are $\sigma_P = \sigma_D = \theta - \log(1+\theta)$, while in the primal dual case $\sigma_{PD} = \frac{\theta}{2(1+\theta)} - \log(1 + \frac{\theta}{2(1+\theta)})$. The above result allows to design a potential increase algorithm based on damped Newton steps. The convergence estimate is given by the following theorem.

Theorem 3.5 *Let $x^0 \in int\mathcal{F}_P$ and $s^0 \in int\mathcal{F}_D$. Any potential increase algorithm (primal, dual, primal-dual) produces an interior feasible pair such that $\|e - xs\| \leq \theta < 1$ in a number of iterations not greater than*

$$\left\lceil \frac{\varphi_{PD}(x^0, s^0) + m}{\sigma} \right\rceil,$$

with $\sigma = \sigma_P$, σ_D or σ_{PD}, depending on which approach (primal, dual or primal-dual) is taken.

3.2 Dikin's ellipsoids

Let $x \in \text{int}\mathcal{F}_P$. From the observation that $x + \Delta x > 0$ for all Δx such that $\|x^{-1}\Delta x\| < 1$, we can define an ellipsoidal neighborhood of x that is entirely contained in $\mathcal{F}_P$. Formally,

$$\mathcal{E}_P = \left\{ \Delta x \, : \, A\Delta x = 0, \left\| X^{-1}\Delta x \right\| \leq 1 \right\}.$$

We shall be particularly concerned with ellipsoids around a θ-center.

We can extend the definition of Dikin ellipsoid to include a different scaling.

Lemma 3.6 *Let (x, s) be a pair of θ-centers.*

1. *If $D = S^{-1}$ (dual scaling),*

$$(1-\theta)\mathcal{E}_P \subset \left\{ \Delta x \, : \, A\Delta x = 0, \left\| D^{-1}\Delta x \right\| \leq 1 \right\} \subset (1+\theta)\mathcal{E}_P.$$

2. *If $D = X^{1/2}S^{-1/2}$ (primal-dual scaling),*

$$\sqrt{1-\theta}\mathcal{E}_P \subset \left\{ \Delta x \, : \, A\Delta x = 0, \left\| D^{-1}\Delta x \right\| \leq 1 \right\} \subset \sqrt{1+\theta}\mathcal{E}_P.$$

Proof

For the dual scaling the proof follows from

$$\left\| X^{-1}\Delta x \right\| = \left\| DX^{-1}D^{-1}\Delta x \right\| \leq \left\| (XS)^{-1} \right\|_\infty \left\| D^{-1}\Delta x \right\| \leq \frac{1}{1-\theta},$$

and

$$\left\|D^{-1}\Delta x\right\| = \left\|D^{-1}XX^{-1}\Delta x\right\| \leq \|XS\|_\infty \left\|X^{-1}\Delta x\right\| \leq 1+\theta.$$

For the primal-dual scaling the proof follows from

$$\left\|X^{-1}\Delta x\right\| = \left\|DX^{-1}D^{-1}\Delta x\right\| \leq \left\|(XS)^{-1/2}\right\|_\infty \left\|D^{-1}\Delta x\right\| \leq \frac{1}{\sqrt{1-\theta}},$$

and

$$\left\|D^{-1}\Delta x\right\| = \left\|D^{-1}XX^{-1}\Delta x\right\| \leq \left\|(XS)^{1/2}\right\|_\infty \left\|X^{-1}\Delta x\right\| \leq \sqrt{1+\theta}.$$

■

We can similarly define Dikin's ellipsoids in the dual. Let $s \in \mathrm{int}\mathcal{F}_D$. The dual ellipsoid is

$$\mathcal{E}_D = \left\{\Delta s \,:\, \Delta s = -A^T\Delta y, \left\|S^{-1}\Delta s\right\| \leq 1\right\}.$$

The extension of Dikin's ellipsoid to a different scaling at a θ-center is given by

Lemma 3.7 *1. If $D = X$ (primal scaling),*

$$(1-\theta)\mathcal{E}_D \subset \left\{\Delta s \,:\, \Delta s = -A^T\Delta y, \|D\Delta s\| \leq 1\right\} \subset (1+\theta)\mathcal{E}_D.$$

2. If $D = X^{1/2}S^{-1/2}$ (primal-dual scaling),

$$\sqrt{(1-\theta)}\mathcal{E}_D \subset \left\{\Delta s \,:\, \Delta s = -A^T\Delta y, \|D\Delta s\| \leq 1\right\} \subset \sqrt{(1+\theta)}\mathcal{E}_D.$$

The proof is the same as for Lemma 3.6.

It is well-known that an homothety of Dikin's ellipsoid contains the feasible set. We shall use this property in the restricted context of the set $\mathcal{F}_D$.

Lemma 3.8 *Let (x,s) be a θ-centered feasible pair. Then*

$$\mathcal{F}_D \subset \left\{\Delta s \,:\, \Delta s = -A^T\Delta y, \|D\Delta s\| \leq \frac{1+\theta}{1-\theta}(m+1)\right\}.$$

Proof

Let (x,s) be a θ-centered feasible pair and $\tilde{s} = c - A^T\tilde{y} > 0$ be any interior point of $\mathcal{F}_D$. Since $x(\tilde{s}-s)$ and e are orthogonal,

$$\begin{aligned} \|x(\tilde{s}-s)\|^2 + m &= \|x\tilde{s} + (e - xs)\|^2, \\ &\leq (\|x\tilde{s}\| + \|e - xs\|)^2. \end{aligned}$$

Since $\tilde{s} > 0$, then $\|x\tilde{s}\| \leq x^T\tilde{s} = x^T s$. From $\|e - xs\| \leq \theta$, one has $x^T s \leq (1+\theta)m$. We thus obtain the weak bound

$$\|x(\tilde{s} - s)\| \leq (1+\theta)m + \theta \leq (1+\theta)(m+1).$$

Finally, from $\|D(\tilde{s} - s)\| \leq \left\|Dx^{-1}\right\|_\infty \|x(\tilde{s} - s)\|$, one gets

$$\|D(\tilde{s} - s)\| \leq \frac{1+\theta}{1-\theta}(m+1).$$

Hence,

$$\mathcal{F}_D \subset \left\{\Delta s \,:\, \Delta s = -A^T \Delta y, \|D\Delta s\| \leq \frac{1+\theta}{1-\theta}(m+1)\right\}.$$

■

4 Multiple central cuts

We assume now that a θ-center $(x; s, y)$ has been computed, i.e.,

$$\begin{aligned} \|e - xs\| &\leq \theta < 1, && (8)\\ A^T y + s &= c, \quad s > 0, && (9)\\ Ax &= 0, \quad x > 0. && (10) \end{aligned}$$

The cuts are

$$a_{m+j}^T \tilde{y} \leq a_{m+j}^T y, \; j = 1, \ldots, p, \; \forall y \in C.$$

We define

$$B = (a_{m+1}, a_{m+2}, \ldots, a_{m+p}).$$

The new cuts lead to two new sets:

$$\widetilde{\mathcal{F}}_D = \left\{\tilde{y} \,:\, A^T\tilde{y} \leq c, B^T\tilde{y} \leq B^T y\right\}$$

or

$$\widetilde{\mathcal{F}}_D = \left\{\tilde{s} = (\hat{s}, \gamma) \geq 0 \,:\, A^T\tilde{y} + \hat{s} = c, B^T\tilde{y} + \gamma = B^T y\right\},$$

and

$$\widetilde{\mathcal{F}}_P = \left\{\tilde{x} = (\hat{x}, \beta) \geq 0 \,:\, A\hat{x} + B\beta = 0\right\}.$$

We shall use the notation

$$\Delta y = (\tilde{y} - y),$$

so

$$\gamma = -B^T \Delta y.$$

After adding the cuts, one has

$$\tilde{s} = \begin{pmatrix} s \\ \gamma = 0 \end{pmatrix} \in \widetilde{\mathcal{F}}_D$$

and

$$\tilde{x} = \begin{pmatrix} x \\ \beta = 0 \end{pmatrix} \in \widetilde{\mathcal{F}}_P.$$

Let us introduce the notation

$$\tilde{c} = \begin{pmatrix} c \\ B^T y \end{pmatrix}.$$

The primal and dual potentials at the new points $(\hat{x}, \beta)$ and $(\hat{s}, \gamma)$ are:

$$\tilde{\varphi}_D(\tilde{s}) = \sum_{i=1}^{m} \log \hat{s}_i + \sum_{i=1}^{p} \log \gamma_i = \varphi_D(\hat{s}) + \sum_{i=1}^{p} \log \gamma_i$$

and

$$\begin{aligned} \tilde{\varphi}_P(\tilde{x}) &= -c^T \hat{x} + \sum_{i=1}^{m} \log \hat{x}_i - y^T B\beta + \sum_{i=1}^{p} \log \beta_i, \\ &= \varphi_P(\hat{x}) - y^T B\beta + \sum_{i=1}^{p} \log \beta_i. \end{aligned}$$

The points $\tilde{x}$ and $\tilde{s}$ (or $\tilde{y}$) lie on the boundary of the new primal and dual sets respectively. To recover the new analytic center, one has to increase the components β and γ. Since the terms $\sum_{i=1}^{p} \log \beta_i$ and $\sum_{i=1}^{p} \log \gamma_i$ are dominant near $\beta = 0$ and $\gamma = 0$, maximizing those terms while limiting the variation on φ_P and φ_D is likely to produce a good step towards the solution.

This approach requires the knowledge of the level sets of the potential, something that we don't have, but that can be approximated by Dikin's ellipsoids. Therefore, we are interested in solving the following problems

$$\max \left\{ \sum_{i=1}^{p} \log \beta_i \,:\, \beta \geq 0,\ A\Delta x + B\beta = 0, \left\| D^{-1} \Delta x \right\| \leq 1 \right\} \tag{11}$$

and

$$\max \left\{ \sum_{i=1}^{p} \log \gamma_i \,:\, \gamma \geq 0,\ B^T \Delta y + \gamma = 0, \left\| DA^T \Delta y \right\| \leq 1 \right\}. \tag{12}$$

Here D is one of the scaling matrices X, S^{-1} or $(XS^{-1})^{\frac{1}{2}}$ depending whether the computations are done with the primal, the dual or the primal-dual algorithm.

Let show here that the above problems are well-defined and have a finite optimum.

Lemma 4.1 *Under Assumptions 2.1 and 2.2, Problems (11) and (12) are well defined and have a finite optimum that is uniquely defined by the first order optimality conditions.*

Proof Both problems have a strictly concave objective. Their optimum, if it exists, is unique in β (resp., γ).

By Assumptions 2.1 and 2.2, there exists a $\bar{\gamma} > 0$ and a $\bar{\Delta}y$ such that $B^T \bar{\Delta}y + \bar{\gamma} = 0$. Problem (12) is well-defined. Since Δy is bounded, γ is bounded and the feasible set is compact. Since the objective tends to $-\infty$ close to the boundary, the problem has a finite solution that is uniquely defined by the set of first order optimality conditions.

To show that Problem (11) is also well-defined, we note that the equation $A\Delta x + B\beta = 0$ has a solution for any $\beta > 0$ since A has full row rank. Let us show that the feasible set is bounded. Indeed, let $\beta \geq 0$ and $B\beta = 0$; then,

$$0 = \beta^T (B^T \bar{\Delta}y + \bar{\gamma}) = \beta^T \bar{\gamma}.$$

Since $\beta \geq 0$ and $\bar{\gamma} > 0$, then, $\beta = 0$. Recalling that A has full row rank, we conclude from $A\Delta x + B\beta = 0$ that $\Delta x \neq 0$ whenever $\beta \geq 0$, $\beta \neq 0$; thus β is bounded, since Δx is bounded by $\|D^{-1}\Delta x\| \leq 1$. Problem (11) is thus well-defined and has a finite optimum. ■

The solutions of Problems (11) and (12) define the primal dual pair of rays

$$\tilde{x}(\alpha) = \begin{pmatrix} x + \alpha\Delta x \\ \alpha\beta \end{pmatrix}$$

and

$$\tilde{s}(\alpha) = \begin{pmatrix} s + \alpha\Delta s \\ \alpha\gamma \end{pmatrix} = \begin{pmatrix} s - \alpha A^T \Delta y \\ \alpha\gamma \end{pmatrix}$$

for $\alpha > 0$.

If $\|e - xs\| \leq \theta < 1$, then for $\alpha < 1 - \theta$

$$\tilde{x}(\alpha) \in \text{int}\tilde{\mathcal{F}}_P \quad \text{and} \quad \tilde{s}(\alpha) \in \text{int}\tilde{\mathcal{F}}_D.$$

The following positive semidefinite matrix

$$V = B^T (AD^2 A^T)^{-1} B$$

plays a fundamental role in the analysis. V can be interpreted as variance-covariance matrix between the vectors (a_{m+j}), $j = 1, \ldots p$, in the metric induced by the matrix $(AD^2A^T)^{-1}$, i.e., Dikin's metric.

Theorem 4.2 *The solution of Problems (11) and (12) is given by*

$$\Delta x = -D^2 A^T (AD^2 A^T)^{-1} B\beta$$

and

$$\Delta y = -(AD^2 A^T)^{-1} B\beta,$$

with β *defined as the unique solution of*

$$\max\{-\frac{p}{2}\beta^T V\beta + \sum_{i=1}^{p} \log \beta_i\}, \tag{13}$$

and

$$\gamma = V\beta.$$

Proof Let $\lambda \in R^n$ and σ^2 be the multipliers associated with the constraints of Problem (11). The optimality conditions are

$$\begin{aligned} \beta^{-1} + B^T\lambda &= 0, \\ A^T\lambda - \sigma^2 D^{-2}\Delta x &= 0, \\ A\Delta x + B\beta &= 0, \\ \sigma^2(1 - \|D^{-1}\Delta x\|) &= 0. \end{aligned}$$

From the definition of Δx, one immediately sees that $A\Delta x + B\beta = 0$. Letting $\lambda = -p(AD^2A^T)^{-1}B\beta$ and $\sigma^2 = p$, we have

$$A^T\lambda = -pA^T(AD^2A^T)^{-1}B\beta = \sigma^2 D^{-2}\Delta x.$$

This proves the second relation. To prove the first relation, we shall use the optimality condition for Problem (13). However, we must check first that (13) has a bounded optimum. In Lemma 4.1 we proved that $B\beta = 0$ has no nonzero nonegative solution. Thus, for all $\beta \geq 0$, $\beta \neq 0$, one has

$$\beta^T V\beta = \beta^T B^T (AD^2A^T)^{-1} B\beta > 0.$$

This proves that the objective $-\frac{p}{2}\beta^T V\beta + \sum_{i=1}^{p} \log \beta_i$ is bounded above and Problem (13) has a unique optimum.

The optimality condition for Problem (13) is

$$-pV\beta + \beta^{-1} = 0.$$

Replacing β^{-1} by $pV\beta$ we get the identity

$$pV\beta + B^T\lambda = pB^T(AD^2A^T)^{-1}B\beta - pB^T(AD^2A^T)^{-1}B\beta \equiv 0.$$

It remains to check that $\|D^{-1}\Delta x\| = 1$. Indeed,

$$\|D^{-1}\Delta x\|^2 = \beta^T B^T (AD^2A^T)^{-1} B\beta = \beta^T V\beta$$

and

$$\beta^T V \beta = \frac{1}{p} \beta^T \beta^{-1} = 1.$$

Let us now consider Problem (12). The optimality conditions are

$$\begin{aligned} \gamma^{-1} - \mu &= 0, \\ \gamma + B^T \Delta y &= 0, \\ B\mu + \sigma^2 (AD^2A^T)\Delta y &= 0, \\ \sigma^2 (1 - \|DA^T \Delta y\|) &= 0, \end{aligned}$$

where $\mu \in R^p$ and $\sigma^2 \in R_+$ are the multipliers associated with the two constraints.

We want to show that $\mu = p\beta$ and $\sigma^2 = p$ are the optimal multipliers, where β is the optimal solution of Problem (13). Solving for Δy, one gets:

$$\Delta y = -\frac{1}{\sigma^2}(AD^2A^T)^{-1} B\mu = -(AD^2A^T)^{-1} B\beta.$$

Now

$$0 = \gamma + B^T \Delta y = \gamma - B^T (AD^2A^T)^{-1} B\beta = \gamma - V\beta.$$

Remembering the optimality condition for β, one may replace $\gamma = V\beta$ by $(p\beta)^{-1}$, and thus check that the first optimality condition $\gamma^{-1} = p\beta = \mu$ holds.

Finally,

$$1 = \Delta y^T (AD^2A^T) \Delta y = \beta^T V \beta = \|DA^T \Delta y\|$$

proves that with our choice of multipliers, the last optimality condition also holds. ■

Remark 4.1 *If V is nonsingular, γ is also the unique solution of*

$$\max\{-\frac{1}{2} p \gamma^T V^{-1} \gamma + \sum_{i=1}^{p} \log \gamma_i\}. \tag{14}$$

We can now give an explicit formula for the restoration direction. Noting that

$$\Delta s = -\frac{1}{p} A^T \Delta y = A^T (AD^2A^T)^{-1} B\beta,$$

we have the new primal-dual pair

$$\tilde{x}(\alpha) = \begin{pmatrix} x + \alpha \Delta x \\ \alpha\beta \end{pmatrix} = \begin{pmatrix} x - \alpha D^2 A^T (AD^2A^T)^{-1} B\beta \\ \alpha\beta \end{pmatrix}, \tag{15}$$

$$\tilde{s}(\alpha) = \begin{pmatrix} s + \alpha\Delta s \\ \alpha\gamma \end{pmatrix} = \begin{pmatrix} s + \alpha A^T(AD^2A^T)^{-1}B\beta \\ \alpha V\beta \end{pmatrix}, \tag{16}$$

and

$$\tilde{y}(\alpha) = \begin{pmatrix} y + \alpha\Delta y \end{pmatrix} = \begin{pmatrix} y - \alpha(AD^2A^T)^{-1}B\beta \end{pmatrix}. \tag{17}$$

Remark 4.2 *We note a significant dissymmetry between the primal and dual directions:*

1. *any positive value of β, say $\beta = e$ gives a primal feasible direction*
2. *but $\beta > 0$ does not guarantee $\gamma = V\beta > 0$; however if V is nonsingular then taking $\beta = pV^{-1}\widehat{\gamma}$, with $\widehat{\gamma} > 0$, gives a feasible dual direction.*

Different stepsizes (α_P, α_D) could be used in the primal and dual space.

Note that, by construction, $\|D^{-1}\Delta x\| = 1$ and $\|D\Delta s\| = 1$, and that if $\lambda = p\beta$, then $D^{-1}\Delta x = D\Delta s$. At the optimum direction, one has $p\gamma\beta = e$.

The computation of β requires solving the nonlinear optimization problem (13). Since the function $F(\beta) = -\sum_{i=1}^{p} \log\beta_i + \frac{p}{2}\beta^T V\beta$ is self-concordant, it can easily be minimized by classical Newton schemes. We postpone to a later section the discussion on the complexity estimate for getting approximate solutions.

For the sake of a simpler presentation we shall assume in our analysis of ACCPM that the minimizers are exact. However, this is not the case in practice and we must be concerned with the impact of errors on β and γ on the performance of ACCPM. This discussion is also postponed to a later section. Below, we sketch the result that enables an easy extension of our analysis of ACCPM with multiple cuts in the case of inexact computations of β and γ.

The convergence analysis of section 5 relies on the following properties:

i) $\|D^{-1}\Delta x\| = \beta^T V\beta = 1$,

ii) $\|D\Delta s\| = \frac{1}{p^2}\gamma^{-T}V\gamma^{-1} = 1$,

iii) $p\beta\gamma = e$.

If we can guarantee that the solutions satisfy $p\beta\gamma \approx e$ and $\frac{1}{p^2}\gamma^{-T}V\gamma^{-1} \approx 1 \approx \beta^T V\beta$, then the convergence result on ACCPM is essentially unaffected, while the proofs need only minor adjustments.

We give here a theorem that stipulates the condition that must be met by β and γ to carry the analysis with inexact minimizers. In a later section we shall show that classical interior point schemes make it possible to meet the condition.

Lemma 4.3 *Assume $\beta > 0$ and $\|p\beta(V\beta) - e\| \leq \eta$. Let $\gamma = V\beta$. Then,*

$$(1 - \eta)e \leq p\beta\gamma \leq (1 + \eta)e,$$

and

$$1 - \eta \leq \beta^T V\beta = \beta^T\gamma \leq 1 + \eta.$$

In particular, $\gamma = V\beta > 0$ if $\eta < 1$.

Proof

The first set of inequalities follows directly from the assumption and the definition of γ. These inequalities also imply that $\gamma = V\beta > 0$ if $\eta < 1$.

Multiplying these inequalities by e^T one gets

$$p(1 - \eta) \leq p\beta^T V\beta \leq p(1 + \eta).$$

■

5 Convergence analysis

We now assume that (x, s) is a pair of θ-centers and that Δx and Δs are computed as in Section 4 with β and γ being the exact minimizers of problems (13) and (14). We assume that the computations are done with either the primal, the dual of the primal-dual scaling.

Lemma 5.1 *Independently of the specific scaling matrix D (primal, dual or primal-dual), one has, for any $\alpha < 1 - \theta$, $\left\|\alpha X^{-1}\Delta x\right\| < 1$ and $\left\|\alpha S^{-1}\Delta s\right\| < 1$.*

Proof

By construction $\left\|D^{-1}\Delta x\right\| = \|D\Delta s\| = 1$. From Lemma 3.6, for any primal, dual or primal-dual scaling D, we have $\left\|X^{-1}\Delta x\right\| \leq \frac{1}{1-\theta}\left\|D^{-1}\Delta x\right\| = \frac{\alpha}{1-\theta} < 1$. The proof is the same in the dual case. ■

Remark 5.1 *The above result can be sharpened by considering separately the three different scaling matrices D. However, we prefer the weaker result since it allows a single formulation for the three cases.*

Lemma 5.2 *The following inequalities hold:*

$$|c^T \Delta x + y^T B\beta - e^T X^{-1}\Delta x| \leq \frac{\theta}{1-\theta},$$

and

$$|e^T S^{-1}\Delta s| \leq \frac{\theta}{1-\theta}.$$

Proof

From $B\beta = -A\Delta x$, one has

$$c^T \Delta x + y^T B\beta = c^T \Delta x - y^T A\Delta x = e^T(S\Delta x).$$

Thus,

$$\begin{aligned} |c^T \Delta x + y^T B\beta - e^T X^{-1}\Delta x| &= |e^T(S - X^{-1})\Delta x|, \\ &= |(sx-e)^T X^{-1}\Delta x|, \\ &\leq \|e - sx\| \left\|X^{-1}\Delta x\right\|, \\ &\leq \frac{\theta}{1-\theta}. \end{aligned}$$

To prove the second statement, we note that $x^T \Delta s = 0$ since $Ax = 0$. Thus

$$\begin{aligned} |e^T S^{-1}\Delta s| &= |e^T(S^{-1} - X)\Delta s|, \\ &= |(sx-e)^T S^{-1}\Delta s|, \\ &\leq \|e - sx\| \left\|S^{-1}\Delta s\right\|, \\ &\leq \frac{\theta}{1-\theta}. \end{aligned}$$

■

In view of the above lemmas, we can bound the potentials $\widetilde{\varphi}_P$ and $\widetilde{\varphi}_D$ at the new pair of points $(\tilde{x}(\alpha), \tilde{s}(\alpha))$.

Lemma 5.3 *For any $0 < \alpha < 1-\theta$, the new potentials satisfy*

$$\widetilde{\varphi}_P(\tilde{x}(\alpha)) \geq \varphi_P(x) + p\log\alpha + \alpha + \log(1 - \frac{\alpha}{1-\theta}) + \sum_{i=1}^{p} \log\beta_i, \tag{18}$$

$$\widetilde{\varphi}_D(\tilde{s}(\alpha)) \geq \varphi_D(s) + p\log\alpha + \alpha + \log(1 - \frac{\alpha}{1-\theta}) + \sum_{i=1}^{p} \log\gamma_i, \tag{19}$$

and

$$\widetilde{\varphi}_{PD}(\tilde{x}(\alpha), \tilde{s}(\alpha)) \geq \varphi_{PD}(x,s) + 2p\log\alpha + 2\alpha + 2\log(1 - \frac{\alpha}{1-\theta}) - p\log p. \tag{20}$$

Proof

Let us prove first the inequality on the primal potential. At the updated point $\tilde{x}(\alpha)$ the potential is

$$\begin{aligned}\widetilde{\varphi}_P(\tilde{x}(\alpha)) &= -\tilde{c}^T\tilde{x}(\alpha) + \sum_{i=1}^{m}\log x_i(\alpha) + \sum_{i=1}^{p}\log\alpha\beta_i,\\ &= -c^Tx - \alpha c^T\Delta x - \alpha y^T B\beta + \sum_{i=1}^{m}\log x_i(\alpha) + \sum_{i=1}^{p}\log\alpha\beta_i,\\ &= \varphi_P(x) - \alpha c^T\Delta x + \sum_{i=1}^{m}\log(1+\alpha x_i^{-1}(\Delta x)_i) - \alpha y^T B\beta + \sum_{i=1}^{p}\log\alpha\beta_i.\end{aligned}$$

Let $h_P = \alpha x^{-1}\Delta x$. By Lemma 5.1 $\|h_P\| < 1$. We can apply Lemma 3.3 to get

$$\sum_{i=1}^{m}\log(1+\alpha x_i^{-1}(\Delta x)_i) \geq \alpha e^T x^{-1}\Delta x + \|h_P\| + \log(1-\|h_P\|).$$

Then, by Lemma 5.2

$$\alpha e^T x^{-1}\Delta x - \alpha c^T\Delta x - \alpha y^T B\beta \geq -\frac{\alpha\theta}{1-\theta}.$$

Since $t+\log(1-t)$ is decreasing, we can bound $\|h_P\|+\log(1-\|h_P\|)$ by $\frac{\alpha}{1-\theta} + \log(1-\frac{\alpha}{1-\theta})$ and get

$$\widetilde{\varphi}_P(\tilde{x}(\alpha)) \geq \varphi_P(x) + \alpha + \log(1-\frac{\alpha}{1-\theta}) + \sum_{i=1}^{p}\log\alpha\beta_i.$$

Let us prove now the dual case. We have

$$\begin{aligned}\widetilde{\varphi}_D(\tilde{s}(\alpha)) &= \sum_{i=1}^{m}\log s_i(\alpha) + \sum_{i=1}^{p}\log\alpha\gamma_i,\\ &= \varphi_D(s) + \sum_{i=1}^{m}\log(1+\alpha s_i^{-1}(\Delta s)_i) + \sum_{i=1}^{p}\log\alpha\gamma_i.\end{aligned}$$

Let $h_D = \alpha S^{-1}\Delta s$. By Lemma 5.1 $\|h_D\| < 1$. We can apply Lemma 3.3 to get

$$\sum_{i=1}^{m}\log(1+\alpha s_i^{-1}(\Delta s)_i) \geq \alpha e^T s^{-1}\Delta s + \|h_D\| + \log(1-\|h_D\|).$$

Since by Lemma 5.2

$$\alpha e^T s^{-1}\Delta s \geq -\frac{\alpha\theta}{1-\theta}$$

we obtain, by putting the inequalities together, the same result as in the primal case

$$\tilde{\varphi}_D(\tilde{s}(\alpha)) \geq \varphi_D(s) + \alpha + \log(1 - \frac{\alpha}{1-\theta}) + \sum_{i=1}^{p} \log \alpha\gamma_i.$$

To conclude the proof of the theorem, we just sum the inequalities on $\tilde{\varphi}_P$ and $\tilde{\varphi}_D$ and use $\beta\gamma = \frac{1}{p}e$ to get

$$\tilde{\varphi}_{PD}(\tilde{x}(\alpha), \tilde{s}(\alpha)) \geq \varphi_{PD}(x, s) + 2p\log\alpha + 2\alpha + 2\log(1 - \frac{\alpha}{1-\theta}) - p\log p.$$

■

5.1 Recovering the new analytic center

Theorem 5.4 *The number of Newton steps to compute the updated θ-analytic center is bounded by*

$$\nu = \frac{-p-\rho}{\sigma} = O(p\log(p+1)),$$

where

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

and, depending on the Newton scheme,

$$\sigma = \sigma_P,\ \sigma_D \text{ or } \sigma_{PD}.$$

Proof

To bound the number of Newton steps, we compute the optimality gap

$$\Delta\tilde{\varphi}_{PD} = (\tilde{\varphi}_P^c + \tilde{\varphi}_D^c) - \tilde{\varphi}_{PD}(\tilde{x}(\alpha), \tilde{s}(\alpha))$$

for the sum of the primal and dual potentials. On the one hand,

$$\tilde{\varphi}_P^c + \tilde{\varphi}_D^c = -(m+p).$$

On the other hand, we can write

$$\tilde{\varphi}_{PD}(\tilde{x}(\alpha), \tilde{s}(\alpha)) \geq \varphi_{PD}(x, s) + 2\alpha + 2p\log\alpha + 2\log(1 - \frac{\alpha}{1-\theta}) - p\log p.$$

Finally,

$$\varphi_{PD}(x, s) \geq \varphi_P(x^c) + \varphi_D(s^c) - \frac{2\theta^2}{1-\theta^2} = -m - \frac{2\theta^2}{1-\theta^2}.$$

Hence

$$\begin{aligned}&\tilde{\varphi}_{PD}(\tilde{x}(\alpha),\tilde{s}(\alpha))\\ \geq\quad & -m+\frac{2\theta^2}{1-\theta^2}+2\alpha+2p\log\alpha+2\log(1-\frac{\alpha}{1-\theta})-p\log p.\end{aligned}$$

Thus

$$\Delta\tilde{\varphi}_{PD}\leq -p-\rho.$$

Using theorem 3.4 and the above bound on the potential variation we conclude the proof of the theorem. ■

5.2 Convergence of ACCPM with multiple cuts

The next lemma is a first step on bounding the number of calls to the oracle.

Theorem 5.5 *For all* $0<\alpha<1-\theta$

$$\tilde{\varphi}^c_D\leq\varphi^c_D+\sum_{i=1}^{p}\log\tau_i+\kappa(\theta,\alpha,p),$$

with

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

and τ *is the vector whose components are the square roots of the diagonal elements of* V.

Proof

The first inequality uses $\tilde{\varphi}^c_P\geq\tilde{\varphi}_P(\tilde{x}(\alpha))$, the duality on potential and Lemma 2.4 to yield

$$\begin{aligned}-\tilde{\varphi}^c_D &= (m+p)+\tilde{\varphi}^c_P,\\ &\geq (m+p)+\tilde{\varphi}_P(\tilde{x}(\alpha)),\\ &\geq m+p+\varphi_P(x)+p\log\alpha+\alpha+\log(1-\frac{\alpha}{1-\theta})+\sum_{i=1}^{p}\log\beta_i. \quad (21)\end{aligned}$$

We now need to deal with the contribution of the new variables

$$\sum_{i=1}^{p}\log\beta_i.$$

Since β solves (13), we have $\beta^T V\beta = 1$ and

$$\begin{aligned}\sum_{i=1}^{p} \log \beta_i - \frac{p}{2} &= \max_{\beta'}\{\sum_{i=1}^{p} \log \beta'_i - \frac{p}{2}\beta'^T V\beta'\}, \\ &\geq \sum_{i=1}^{p} \log \beta'_i - \frac{p}{2}\beta'^T V\beta',\end{aligned}$$

for any arbitrary β'.

Let us define the vector τ by

$$\tau_i = \sqrt{a_{m+i}^T (AX^2A^T)^{-1} a_{m+i}}.$$

Note that $\tau^2 = \text{diag}V$ while the off-diagonal terms of V are

$$\tau_{ij} = a_{m+i}^T (AX^2A^T)^{-1} a_{m+j}.$$

The off-diagonal elements satisfy

$$|\tau_{ij}| \leq \tau_i \tau_j.$$

Those properties are typical of a variance-covariance matrix. Let us choose

$$\beta' = \frac{\tau^{-1}}{\sqrt{\tau^{-T} V \tau^{-1}}}.$$

Then

$$\beta'^T V\beta' = 1.$$

The matrix $R = \text{diag}(\tau^{-1})V\text{diag}(\tau^{-1})$ is a correlation matrix: all its coefficient are bounded in absolute value by 1, and

$$\tau^{-T} V\tau^{-1} = e^T Re \leq p^2.$$

Thus

$$\begin{aligned}\sum_{i=1}^{p} \log \beta_i &\geq \sum_{i=1}^{p} \log \beta'_i \\ &= -\sum_{i=1}^{p} \log \tau_i - p \log \sqrt{\tau^{-T} V\tau^{-1}} \\ &\geq -\sum_{i=1}^{p} \log \tau_i - p \log p. \qquad (22)\end{aligned}$$

Using corollary 3.2 we have

$$\varphi_P(x) \geq \varphi_P^c - \frac{\theta^2}{1-\theta^2} = -\varphi_D^c - m - \frac{\theta^2}{1-\theta^2}. \qquad (23)$$

Putting together (21), (22) and (23) yields

$$\tilde{\varphi}_D^c \leq \varphi_D^c + \frac{\theta^2}{1-\theta^2} - p - \alpha - \log(1 - \frac{\alpha}{1-\theta}) + p \log \frac{p}{\alpha} + \sum_{i=1}^{p} \log \tau_i.$$

■

The bound

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

can be analyzed by selecting, somewhat arbitrarily, $\alpha = 1/\sqrt{2}$ and $\theta = .25$, guaranteeing $\alpha \leq 1 - \theta$ but also

$$\begin{aligned} \kappa(\alpha, \theta, p) &= \frac{\theta^2}{1-\theta^2} - p - \alpha - \log(1 - \frac{\alpha}{1-\theta}) + p \log \frac{p}{\alpha} \\ &\leq p \log(p+1); \end{aligned} \quad (24)$$

which is exactly the same result as in [19], but with a rather different derivation, as we show that this inequality is actually achieved at the iterate obtained by the restoration step.

Remark 5.2 *If the p cuts generated are identical, then the correlation matrix R is the rank-one matrix ee^T. Otherwise for the optimal β^**

$$\sum_{i=1}^{p} \log \beta_i^* + \sum_{i=1}^{p} \log \tau_i + p \log p \quad (25)$$

may be significantly greater than 0 and speed the convergence in practice, even though this does not appear to affect the worst case complexity bound.

5.3 Convergence of ACCPM

The convergence analysis uses the proof given in [19], for the case of multiple cuts.

Denote

$$P = \varphi_D(s^c) = \max\{\varphi_D(s) \,:\, s \in \mathcal{F}_D\}$$

and let P^k be the same value after k calls to the oracle, that is, after adding $m_k - 2n = \sum_{j=0}^{k-1} p_j$ cuts, where p_j denotes the number of cuts added at iteration j. By Theorem 5.5 and the observation (24) the following inequality holds

$$P^{k+1} \leq P^0 + \sum_{j=0}^{k} \sum_{i=1}^{p_j} \log \tau_i^j + \sum_{j=0}^{k} p_j \log(p_j + 1).$$

Theorem 10 of [19] can be used here, with $\bar{p} \leq n$ denotes the maximum number of cuts generated by any call to the oracle.

Theorem 5.6 *The algorithm stops with a solution as soon as k satisfies:*

$$\frac{\varepsilon^2}{(\bar{p}+1)^2} \geq \frac{\frac{n}{2} + \frac{18n^2}{15}\log(1+\frac{m_{k+1}}{8n^2})}{m_{k+1}}.$$

Furthermore the number of damped Newton steps per call to the oracle is $O(\bar{p}\log(\bar{p}+1))$. The number of cutting planes generated is at most $O^*(\frac{\bar{p}^2 n^2}{\varepsilon^2})$[1]*.*

The assumption that $\bar{p} \leq n$ is not required in the proof of [19], and in fact $\bar{p} = O(n)$ would still lead to $O^*(\frac{\bar{p}^2 n^2}{\varepsilon^2})$ cutting planes (this would only impact the constant).

6 Computing the optimal direction of restoration

The restoration direction requires the solution of the concave problem

$$\max\{F(\beta) = -\frac{1}{2}p\beta^T V\beta + \sum_{i=1}^{p} \log\beta_i\}.$$

We note that in the computation of the restoration direction a significant absence of symmetry occurs: it is easy to give a feasible value for β, say $\beta = \frac{e}{\sqrt{e^T V e}}$ or $\beta = \frac{\tau^{-1}}{\sqrt{\tau^{-T} V \tau^{-1}}}$ that gives a feasible solution to the problem of finding a feasible direction, but, in general, this is not the case for the dual side. τ is the vector whose components are the square roots of the diagonal elements of V.

If V is invertible, then the dual direction could also be computed by maximizing

$$G(\gamma) = -\frac{p}{2}\gamma^T V^{-1}\gamma + \sum_{i=1}^{p} \log\gamma_i.$$

A good starting value for γ could also be given, say $\gamma = \frac{e}{\sqrt{e^T V^{-1} e}}$ or $\gamma = \frac{\tau_D^{-1}}{\sqrt{\tau_D^{-T} V^{-1} \tau_D^{-1}}}$, with τ_D is the vector whose components are the square roots of the diagonal elements of V^{-1}.

The following bounds on $F(\beta)$ will be useful in the computation of complexity estimate of a Newton method to solve (13).

[1] The notation O^* indicates that lower order terms are ignored.

Theorem 6.1 *For*

$$\beta^0 = \frac{\tau^{-1}}{\sqrt{\tau^{-T} V \tau^{-1}}},$$

$$F(\beta^0) \geq -\sum_{i=1}^{p} \log \tau_i - p,$$

and

$$\max_{\beta>0} F(\beta) \quad \leq \quad p \log p - p/2 - p \log \left(\frac{\varepsilon(1-\theta)}{(m+1)(1+\theta)} \right).$$

Proof

The inequality on $F(\beta^0)$ was derived in the proof of theorem 5.5. (See (22).)

Let us construct an upper bound on $F(\beta^*)$ where β^* denotes the optimal solution of problem (13), and $\gamma^* = V\beta^*$. From the optimality condition

$$F(\beta^*) + \sum_{i=1}^{p} \log \gamma_i^* = \sum_{i=1}^{p} \log(\beta_i^* \gamma_i^*) - (p/2)\beta^{*T} V \beta^* \quad = \quad p \log p - p/2.$$

Hence,

$$F(\beta^*) = p \log p - p/2 - \max\{\sum_{i=1}^{p} \log \gamma_i : B^T \Delta y + \gamma = 0, \left\| DA^T \Delta y \leq 1 \right\|\}.$$

By Lemma 3.8, an homothety of Dikin's ellipsoid contains the current set of localization, i.e.,

$$\mathcal{F}_D \subset \left\{ \Delta s \; : \; \Delta s = -A^T \Delta y, \|D \Delta s\| \leq \frac{1+\theta}{1-\theta}(m+1) \right\}.$$

By assumption (2.1) and the fact that the algorithm has not terminated, a sphere of radius ε is contained in $\mathcal{F}_D$. Hence,

$$\left\{ \Delta s \; : \; \Delta s = -A^T \Delta y, \|D \Delta s\| \leq 1 \right\} \cap \mathcal{F}_D$$

contains a sphere of radius $\frac{\varepsilon(1-\theta)}{(m+1)(1+\theta)}$. Denoting by y_c the center of this sphere, and selecting $\gamma = -B^T(y_c - y)$ one has

$$-\sum_{i=1}^{p} \log \gamma_i \leq -p \log \left(\frac{\varepsilon(1-\theta)}{(m+1)(1+\theta)} \right)$$

with $\left\| DA^T(y_c - y) \right\| \leq 1$.

And thus

$$F(\beta^*) \quad \leq \quad p \log p - p/2 - p \log \left(\frac{\varepsilon(1-\theta)}{(m+1)(1+\theta)} \right).$$

■

If V is invertible, one can derive alternative upper bounds on $F(\beta^*)$ as follows. Using

$$F(\beta) + G(\gamma) \le F(\beta^*) + G(\gamma^*) = p \log p - p,$$

we have

$$\begin{aligned} F(\beta^*) &\le p\log p - p - G(\gamma), \quad (\forall \gamma > 0) \\ &= p \log p - p - \sum_{i=1}^{p} \log \gamma_i + (p/2)\gamma^T V^{-1} \gamma, \\ &\le p\log p - p/2 + (p/2)\log(e^T V^{-1} e), \quad (\text{setting } \gamma = e/\sqrt{e^T V^{-1} e}). \end{aligned}$$

If instead of $\gamma = e$ we set $\gamma = \frac{\tau_D^{-1}}{\sqrt{\tau_D^{-T} V^{-1} \tau_D^{-1}}}$, then,

$$F(\beta^*) \le p \log p - p/2 + \sum_{j=1}^{p} \log(\tau_D)_j + (p/2)\log(\tau_D^{-T} V^{-1} \tau_D^{-1}).$$

The bounds on $F(\beta^0)$ and $F(\beta^*)$ are used to derive a complexity estimate for the computation of an approximate optimal solution. Using the fact that the function F is self-concordant [15], we can resort to a potential increase scheme. The scheme uses the Newton direction

$$-[F''(\beta)]^{-1} F'(\beta).$$

Let us denote $\|u\|_H = \sqrt{u^T H u}$ the norm of an arbitrary vector u in the metric induced by the positive definite matrix H. The norm $\|F'(\beta)\|_{[F''(\beta)]^{-1}}$ plays a critical role in the analysis. The potential increase scheme is based on an extension of lemma 3.3. The proof can be found in the unpublished lecture notes [14]. (The proof is also made available in [16].)

Lemma 6.2 *Let $\Delta\beta$ be such that $\|\Delta\beta\|_{[F''(\beta)]^{-1}} < 1$. Then,*

$$F(\beta + \Delta\beta) \le F(\beta) + \Delta\beta^T F'(\beta) - t - \log(1-t),$$

with $t = \|\Delta\beta\|_{[F''(\beta)]^{-1}}$.

Assume now $\|F'(\beta)\|_{[F''(\beta)]^{-1}} \ge \eta$, for a fixed $0 < \eta < 1$. Let $\Delta\beta = -[F''(\beta)]^{-1} F'(\beta)$ and $\alpha = (1 + \|F'(\beta)\|_{[F''(\beta)]^{-1}})^{-1}$. Then, $\alpha\Delta\beta$ satisfies the condition of the above lemma. Thus,

$$F(\beta + \alpha\Delta\beta) \le F(\beta) + \alpha\Delta\beta^T F'(\beta) - \alpha \|F'(\beta)\|_{[F''(\beta)]^{-1}} + \log(1 + \|F'(\beta)\|_{[F''(\beta)]^{-1}}).$$

Since $\Delta\beta^T F'(\beta) = -\alpha \left\|F'(\beta)\right\|^2_{[F''(\beta)]^{-1}}$, we have

$$F(\beta + \alpha\Delta\beta) \leq F(\beta) - \sigma,$$

where $\sigma = \|F'(\beta)\|_{[F''(\beta)]^{-1}} - \log(1 + \|F'(\beta)\|_{[F''(\beta)]^{-1}})$. One easily shows that $\sigma \geq \eta - \log(1+\eta)$ is bounded from below by an absolute constant.

The complexity estimate for the potential increase scheme follows directly from the above analysis and a bound on the achievable potential increase $(F(\beta^*) - F(\beta^0))$.

Theorem 6.3 *Let $\beta^0 = \frac{\tau^{-1}}{\sqrt{\tau^{-T} V \tau^{-1}}}$. The potential increase algorithm applied to the maximization of F produces a point β such that $\|F'(\beta)\|_{[F''(\beta)]^{-1}} \leq \eta < 1$ in a number of iterations not greater than*

$$\left\lceil \frac{\sum_{i=1}^{p} \log \tau_i - p \log(\frac{\varepsilon(1-\theta)}{(m+1)(1+\theta)}) - p + p\log p}{\eta - \log(1+\eta)} \right\rceil.$$

Remark 6.1 *The total number of Newton steps involved in the computation of all the approximate optimum directions can easily bounded by $m_{k^*} \log(1/\varepsilon)$, using theorem (6.3), where k^* is the number of calls to the oracle at termination, and $m_{k^*} = O^*(\frac{p^2 n^2}{\varepsilon^2})$. A long step argument similar to the one given in [19] could most likely be used to reduce this bound.*

Remark 6.2 *Looking at every iteration individually, and using the fact that $A^T y \leq c$ contains the cutting planes $0 \leq y \leq e$, we can assert that*

$$(AS^{-2}A^T)^{-1} \prec (Y^{-2} + (I - Y)^{-2})^{-1} \prec (4I + 4I)^{-1} \prec \frac{1}{8} I,$$

and hence

$$\tau_j^2 = a_j^T (AS^{-2}A^T)^{-1} a_j \leq \frac{1}{8} \|a\|^2 = \frac{1}{8}.$$

This indicates that, in practice, the number of iterations needed at each iteration to compute the optimal β should not increase with the number of cutting planes.

Remark 6.3 *The number of iterations needed to compute this approximate optimal direction is polynomial in the data, as $\log m$ is polynomial in the data.*

It remains to prove that the potential increase scheme yields a solution β that meets the proximity condition $\|p\beta(V\beta) - e\| \leq \theta$ used in Theorem 4.3. In other words, we must show that for η small enough the condition $\|F'(\beta)\|_{[F''(\beta)]^{-1}} \leq \eta$ implies $\|p\beta(V\beta) - e\| \leq \theta$. To this end, we adapt some results and proofs of [3] developed for quadratic programming.

We then relate a few critical norms.

Lemma 6.4 *Let* $\Delta\beta = -[F''(\beta)]^{-1}F'(\beta) = (diag(\beta^{-2}) + pV)^{-1}(\beta^{-1} - pV\beta)$. *The following inequality holds*

$$\left\|\beta^{-1}\Delta\beta\right\| \leq \|F'(\beta)\|_{[F''(\beta)]^{-1}} \leq \|p\beta(V\beta) - e\|.$$

Proof

Since

$$\begin{aligned}
\|F'(\beta)\|^2_{[F''(\beta)]^{-1}} &= (pV\beta - \beta^{-1})^T(\mathrm{diag}(\beta^{-2}) + V)^{-1}(pV\beta - \beta^{-1}), \\
&= \Delta\beta^T \left(\mathrm{diag}(\beta^{-2}) + pV\right)\Delta\beta, \\
&\geq \Delta\beta^T \mathrm{diag}(\beta^{-2})\Delta\beta = \left\|\beta^{-1}\Delta\beta\right\|^2.
\end{aligned}$$

This proves the left-hand side inequality.

As V is positive semidefinite, one has

$$(\mathrm{diag}(\beta^{-2}) + pV)^{-1} \preceq (\mathrm{diag}(\beta^{-2}))^{-1} = \mathrm{diag}(\beta^2).$$

Therefore

$$\begin{aligned}
\|F'(\beta)\|^2_{[F''(\beta)]^{-1}} &\leq (pV\beta - \beta^{-1})^T\mathrm{diag}(\beta^2)(pV\beta - \beta^{-1}), \\
&= (p\beta(V\beta) - e)^T(p\beta(V\beta) - e) = \|(p\beta(V\beta) - e)\|^2.
\end{aligned}$$

■

We can now prove the main result of the section.

Lemma 6.5 *Assume* $\|F'(\beta)\|_{[F''(\beta)]^{-1}} \leq \eta < 1$ *and let*

$$\Delta\beta = -F''(\beta)]^{-1}F'(\beta) = (diag(\beta)^2 + pV)^{-1}(\beta^{-1} - pV\beta).$$

Then, $\beta^+ = \beta + \Delta\beta > 0$. *Besides,*

$$\left\|F'(\beta^+)\right\|_{[F''(\beta^+)]^{-1}} \leq \left\|p\beta^+(V\beta^+) - e\right\| \leq \|F'(\beta)\|^2_{[F''(\beta)]^{-1}}.$$

Proof

Since $\left\|\beta^{-1}\Delta\beta\right\| \leq \|F'(\beta)\|_{[F''(\beta)]^{-1}} < 1$, then $\beta^+ = \beta(e + \beta^{-1}\Delta\beta) > 0$. Moreover,

$$\left\|p\beta^+(V\beta^+) - e\right\| = \|p(\beta + \Delta\beta)(V(\beta + \Delta\beta)) - e\|.$$

From

$$\left(\mathrm{diag}(\beta^{-2}) + pV\right)\Delta\beta = \beta^{-1} - pV\beta$$

we get

$$pV(\beta + \Delta\beta) = -\mathrm{diag}(\beta^{-2})\Delta\beta + \beta^{-1}.$$

We conclude that

$$\begin{aligned}\left\|p\beta^+(V\beta^+) - e\right\| &= \left\|(\beta+\Delta\beta)(-\mathrm{diag}(\beta^{-2})\Delta\beta + \beta^{-1}) - e\right\|, \\ &= \left\|(\beta^{-1}\Delta\beta)^2\right\| \le \left\|\beta^{-1}\Delta\beta\right\|^2.\end{aligned}$$

■

The above lemma shows that once the condition $\|F'(\beta)\|_{[F''(\beta)]^{-1}} \le \eta < 1$ is met, one more Newton step is enough to generate point satisfying

$$\left\|p(\beta^+)(V\beta^+) - e\right\| \le \left\|\beta^{-1}\Delta\beta\right\|^2 \le \|F'(\beta)\|^2_{[F''(\beta)]^{-1}} \le \eta^2,$$

and thus, by Lemma 4.3, a point $\gamma^+ = V\beta^+ > 0$ such that

$$(1-\eta^2)e \le p\beta^+\gamma^+ \le (1+\eta^2)e.$$

7 Conclusion

In this paper, we defined an efficient direction to restore primal and dual feasibility and centrality after adding p new central cuts simultaneously. The direction is efficient in the sense that it maximizes the the product of the new variables brought into the primal or the dual potentials, under the constraints that the other variables remains within the Dikin ellipsoid. The computation of the optimal direction takes place in a space of dimension p equal to the number of cuts added at a given iteration. If p is sufficiently smaller than n, then significant gains in efficiency can be expected.

The analysis has been derived under the assumption that the cuts are central. If deep cuts are present, which is to be expected in practice, primal feasibility can always be recovered but dual feasibility appears difficult to achieve in general, except by the use of a primal Newton method. One could then extend the long step argument of [8] in the case of one deep cut to multiple deep cuts.

The implementation of ACCPM [11] uses $\beta = \frac{1}{p}e$. Other choices using the variance-covariance matrix V, if it is invertible, have been proposed in [10], and the analysis of this paper actually strengthens that line of thinking.

Both the heuristic and optimal choices for β and γ need to be tested in practice, and extensions to multiple deep cuts deserve a more thorough study.

References

[1] D. S. Atkinson and P. M. Vaidya (1995), "A cutting plane algorithm that uses analytic centers", *Mathematical Programming*, series B, **69**, 1–43.

[2] O. Bahn, O. du Merle, J.-L. Goffin and J.P. Vial (1995), "A Cutting Plane Method from Analytic Centers for Stochastic Programming", *Mathematical Programming*, Series B, **69**, 45–73.

[3] D. den Hertog (1994), *Interior Point Approach to Linear, Quadratic and Convex Programming*, Kluwer, London.

[4] J.-L. Goffin, J. Gondzio, R. Sarkissian and J.-P. Vial (1997), "Solving Nonlinear Multicommodity Flows Problems by the Analytic Center Cutting Plane Method", *Mathematical Programming*, Series B, vol 76 **1**, 131–154.

[5] J.-L. Goffin, A. Haurie and J.-P. Vial (1992), "Decomposition and nondifferentiable optimization with the projective algorithm", *Management Science* **38**, 284–302.

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

[7] J.-L. Goffin and F. Sharifi Mokhtarian (1999), "Using the Primal Dual Infeasible Newton Method in the Analytic Center Method for Problems Defined by Deep Cutting Planes", *Journal of Optimization Theory and Applications*, vol. 101, 35–58.

[8] J.-L. Goffin and J.-P. Vial (1999),"Shallow, deep and very deep cuts in the analytic center cutting plane method", *Mathematical Programming* **84** 1999, 89–103.

[9] J.-L. Goffin and J.-P. Vial (1999), "A two-cut approach in the analytic center cutting plane method", *Mathematical Methods of Operations Research* **49** 1, 149–169.

[10] J. Gondzio (1998), "Warm start of the Primal–Dual Method Applied in the Cutting Plane Scheme", *Mathematical Programming* 83 No 1, 125–143.

[11] J. Gondzio, O. du Merle, R. Sarkissian and J.-P. Vial (1996), "ACCPM - A Library for Convex Optimization Based on an Analytic Center Cutting Plane Method", *European Journal of Operational Research*, **94**, 206–211.

[12] J. E. Mitchell and M.J. Todd (1992), "Solving combinatorial optimization problems using Karmarkar's algorithm," *Mathematical Programming* **56**, 245–284.

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

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

[15] Yu. Nesterov and A. Nemirovsky. *Interior Point Polynomial Algorithms in Convex Programming : Theory and Applications,* SIAM, Philadelphia, 1994.

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

[17] C. Roos, T. Terlaky and J.-P. Vial (1997), *Theory and Algorithms for Linear Optimization: An Interior point Approach* . John Wiley & Sons, Chichester.

[18] Y. Ye (1992), "A potential reduction algorithm allowing column generation", *SIAM Journal on Optimization* **2**, 7–20.

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

[20] Y. Ye (1997), *Interior Point Algorithms: Theory and Analysis.* John Wiley & Sons, New York.